

1 **Hiplot: a comprehensive and easy-to-use web service boosting** 2 **publication-ready biomedical data visualization**

3

4 Jianfeng Li^{1,2,18,*}, Benben Miao^{3,18}, Shixiang Wang^{4,18}, Wei Dong^{5,18}, Houshi Xu^{6,18},
5 Chenchen Si^{7,18}, Wei Wang⁸, Songqi Duan⁹, Jiacheng Lou¹⁰, Zhiwei Bao¹¹, Hailuan
6 Zeng¹², Zengzeng Yang¹³, Wenyan Cheng¹, Fei Zhao^{14,15}, Jianming Zeng¹⁶, Xue-Song
7 Liu^{4,*}, Renxie Wu^{3,*}, Yang Shen^{1,*}, Zhu Chen^{1,*}, Saijuan Chen^{1,*}, Mingjie Wang^{17,*},
8 and Hiplot Consortium

9

10 **Affiliations of authors:**

11 ¹ Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics,
12 National Research Center for Translational Medicine at Shanghai, Ruijin Hospital
13 Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China; ²
14 School of Life Sciences and Biotechnology, Shanghai Jiao Tong University,
15 Shanghai, China; ³ College of Fisheries, Guangdong Ocean University, Zhanjiang,
16 Guangdong, China; ⁴ School of Life Science and Technology, ShanghaiTech
17 University, Shanghai, China; ⁵ Zhongshan School of Medicine, Sun Yat-sen
18 University, Guangzhou, China; ⁶ Shanghai General Hospital, Shanghai Jiao Tong
19 University School of Medicine, Shanghai, China; ⁷ Department of Histology,
20 Embryology, Genetics and Developmental Biology, Shanghai Key Laboratory for
21 Reproductive Medicine, Shanghai Jiao Tong University School of Medicine,
22 Shanghai, China; ⁸ Department of Urology, First Hospital of Shanxi Medical
23 University, Taiyuan, China; ⁹ College of Food Science, Sichuan Agricultural
24 University, Yaan, China; ¹⁰ Department of Neurosurgery, The Second Affiliated
25 Hospital of Dalian Medical University, Dalian, China; ¹¹ State Key Laboratory for
26 Diagnosis and Treatment of Infectious Diseases, Collaborative Innovation Center for
27 Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital,

28 Zhejiang University School of Medicine, Hangzhou, China; ¹² Department of
 29 Endocrinology and Metabolism, Zhongshan Hospital, Fudan Institute for Metabolic
 30 Diseases, and Human Phenome Institute, Fudan University, Shanghai, China; ¹³
 31 Qinghai Academy of Animal and Veterinary Science, State Key Laboratory of Plateau
 32 Ecology and Agriculture in the Three River Head Waters Region, Qinghai Provincial
 33 Key Laboratory of Adaptive Management on Alpine Grassland Qinghai University,
 34 Xining, China; ¹⁴ National Key Laboratory of Plant Molecular Genetics, CAS Center
 35 for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology
 36 and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of
 37 Sciences, Shanghai, China; ¹⁵ University of the Chinese Academy of Sciences,
 38 Beijing, China; ¹⁶ Faculty of Health Sciences, University of Macau, Taipa, Macau,
 39 China; ¹⁷ Department of Gastroenterology, Ruijin Hospital, Shanghai Jiao Tong
 40 University School of Medicine, Shanghai, China; ¹⁸ These authors contributed
 41 equally: Jianfeng Li, Benben Miao, Shixiang Wang, Wei Dong, Houshi Xu, Chenchen
 42 Si;

43

44 *** Correspondence authors:**

45 Jianfeng Li (lee_jianfeng@sjtu.edu.cn), Xue-Song Liu (liuxs@shanghaitech.edu.cn),
 46 Renxie Wu (wurenxie@163.com), Yang Shen (yang_shen@sjtu.edu.cn), Zhu Chen
 47 (zchen@stn.sh.cn), Saijuan Chen (sjchen@stn.sh.cn), Mingjie Wang
 48 (huzai920621@126.com)

49 State Key Laboratory of Medical Genomics, Rui-Jin Hospital, 197 Rui-Jin Er Road,
 50 Shanghai 200025, China

51 Tel: +86-21-33562658

52 Fax: +86-21-64743206

53

54 **Abstract**

55 Modern web techniques provide an unprecedented opportunity for leveraging
56 complex biomedical data generating in clinical, omics, and mechanism experiments.
57 Currently, the functions for carrying out publication-ready biomedical data
58 visualization represent primary technical hurdles in the state-of-art omics-based web
59 services, whereas the demand for visualization-based interactive data mining is ever-
60 growing. Here, we propose an easy-to-use web service, Hiplot (<https://hiplot.com.cn>),
61 equipping with comprehensive and interactive biomedical data visualization functions
62 (230+) including basic statistics, multi-omics, regression, clustering, dimensional
63 reduction, meta-analysis, survival analysis, risk modeling, etc. We used the demo and
64 real datasets to demonstrate the usage workflow and the core functions of Hiplot. It
65 permits users to conveniently and interactively complete a few specialized
66 visualization tasks that previously could only be done by senior bioinformatics or
67 biostatistics researchers. A modern web client with efficient user interfaces and
68 interaction methods has been implemented based on the custom components library
69 and the extensible plugin system. The versatile output can also be produced in
70 different environments via using the cross-platform portable command-line interface
71 (CLI) program, Hctl. A switchable view between the editable data table and the file
72 uploader/path selection could facilitate data importing, previewing, and exporting,
73 while the plumber-based response strategy significantly reduced the time costs for
74 generating basic scientific graphics. Diversified layouts, themes/styles, and color
75 palettes in this website allow users to create high-quality and publication-ready
76 graphics. Researchers devoted to both life and data science may benefit from the
77 emerging web service.

78 **Keywords:** web service; data visualization; data mining; risk models; translational
79 medicine

80

81 **Introduction**

82 Exploration and mining of multidimensional biomedical data originating from
 83 experimental assays, omics studies, and clinical observations largely rely on modern
 84 graphics and statistics [1], such as statistical description/inference and disease
 85 diagnosis [2]. In addition, visualization-based data mining techniques play an
 86 increasingly critical role in further enhancing interpretability, reproducibility, and
 87 effectiveness of both hypothesis- and data-driven scientific research [3-6]. Over a
 88 decade ago, users could only use desktop applications with limited functions and
 89 scalability to perform daily visualization analysis of scientific data. Recently, web-
 90 based cloud applications with better scalability have therefore become one of the ideal
 91 options leveraging complex biomedical data for biologists and clinicians who lack
 92 programming skills [7-13]. Since the establishment of well-known bioinformatics
 93 cloud services, such as Galaxy and DNAnexus [14], a number of upstream data
 94 analysis tasks have been moderately simplified, such as sequence alignments, variant
 95 calling, epigenetic profiling, and other workflow-based pipelines.

96
 97 However, the common downstream functions including publication-ready scientific
 98 graphics and interactive data mining, visualization analysis in particular, based on the
 99 tabular data are still quite lacking in these websites [11, 15, 16]. The well-known
 100 bioinformatics cloud platform, Galaxy, only provides limited dozens of visualization-
 101 based plugins and still lacks adequate optimization for those lightweight biomedical
 102 visualization tasks. The visualization module of St. Jude Children's Research Hospital
 103 cloud portal offers 20 JavaScript-based plugins for the interactive cancer genomics
 104 visualizations, while the basic scientific graphics is blank [12, 17]. The imageGP
 105 merely developed 16 subfunctions for scientific graphics and analysis since 2017 [18].
 106 A huge amount of work still needs to be done for diverse visualization demands,
 107 which requires joint efforts from the entire scientific community. In addition,
 108 complicated user interfaces and inefficient interactions have become the major

negative factors for users skipping web-based tools with functions of biomedical data visualization. For example, it has been rarely supported in the existing web-based bioinformatics tools that could preview and edit the data in the online spreadsheet editor, like classic desktop commercial graphics software. The automatic arrangements of multiple graphics in publication layout, e.g., 4, 6, 9 items per page, also often were overlooked. Other explicit issues, such as untimely tasks output, inconvenient reproduction of parameters/results, and lacking the cross-platform and easy-to-use command-line program, may further prevent the web-based tools from being more widely used for conducting biomedical data visualization tasks.

To meet these challenges, we propose an emerging easy-to-use and scalable web service, Hiplot (<https://hiplot.com.cn>), and an interdisciplinary community focusing on creating interactive applications related to biomedical data visualization. Since October 2019, hundreds of interactive web plugins related to visualization-based data mining have been developed by the Hiplot collaborative group. The core subfunctions of Hiplot have been built based on the open-source and published methods, e.g. R base graphics, ggplot2, pheatmap, ComplexHeatmap [2], circlize [19], ggstatsplot, cola [20], Broad Gene Set Enrichment Analysis (GSEA) [21], clusterProfiler [22], DISCOVER [23], etc. It has covered the most common demands of biologists and clinicians for daily biomedical data visualization using concise user interfaces and efficient interaction methods, and tens of thousands of researchers have already been using our web services. We expect this toolkit could become a useful infrastructure of data visualization for a broad of researchers in biomedicine, life sciences, and data science.

133 **Implementation**

134 **Concise and easy-to-use web interfaces and interactions**

135 The modern UI development framework Vuetify.js (v2) using Vue.js (v2) syntax and
 136 in-house components library were integrated to construct the website. One of the
 137 major advantages of the Hiplot web service is its user-friendly and efficient web
 138 interfaces and interaction methods (**Fig. S1, S2, S3**). Based on the web development
 139 framework of Hiplot, we provide a comparatively uniform user experience in high-
 140 frequency modern scientific graphics. The plugins of this website share a similar web
 141 layout following two columns design (**Fig. S3**). The left column shows the thumbnail,
 142 data, parameters, task list, and output of plugins, while the right column contains all
 143 static content including the description and the documentation of plugins. Thus, the
 144 vast majority of user interactions can be done on a single working page.

145
 146 On the other hand, the online editable spreadsheet in data importing steps has been
 147 first introduced into this large-scale biomedical visualization cloud service with
 148 hundreds of plugins, which has better readability and editability to process small
 149 chunks of tabular data (**Fig. S4**). Meanwhile, the combination strategy of spreadsheet
 150 tabular input and switchable file uploader could improve the importing efficiency for
 151 different sizes of data files (**Fig. S4**). We implemented all operations of the core
 152 functional buttons and the switchable file uploader mode in the “hiplot-table-editor”
 153 web component. It supports preview, import, and edit the data in the spreadsheet
 154 view, which was developed based on the x-spreadsheet project [24] and Vuetify.js
 155 framework. Besides, it has been wrapped as a Vue component for more easily
 156 integrating into the web plugins of Hiplot. Through binding the options using table
 157 headers, the unique values of rows, and columns, the interactivity of visualization
 158 plugins has been further enhanced.

159

160 The pdf-collage plugin allows users to freely combine multiple visualization graphics
 161 of Hiplot in publication-ready layouts in batches. Previously, users needed to
 162 complete the routine task in the Adobe illustrator program with massive mouse clicks
 163 and drag-and-drop (**Fig. S5**). A pool of themes and hundreds of color palettes
 164 provided in the web plugins permit users to get high-quality and publication-ready
 165 outputs. The object of R data (.Rdata) can be used to reproduce or change the output
 166 style of graphics in the local R programming environment.

167

168 In terms of reproducibility of data, parameters, and issues, users can conveniently
 169 reproduce the history data/parameters and result in the plugins of Hiplot based on the
 170 standard JSON data objects from the local file or the remote file manager (**Fig. S6,**
 171 **S7**). To view the logging and input/output for debugging the possible errors via a
 172 confidential task index, Hiplot would generate a random temporary code in the
 173 browser cache for the newly submitted task, which has better task anonymity and
 174 security compared with the permanent cloud-based storage. Several interactive web
 175 applications based on the R Shiny and Python Streamlite development framework
 176 also have been developed for resolving different user demands (**Fig. S8**) [13]. The
 177 reproduction of errors in the Shiny- and Streamlite-based plugins are still challenging.
 178 Through the feedback system built in the Hiplot website, all submitted issues would
 179 be automatically synchronized to GitHub for unified management of issues.

180 **Cross-platform command-line interface program**

181 The cross-platform command-line interface program of Hiplot (Hctl) was developed
 182 based on the Golang programming language, which conferred great advantages in
 183 helping users to quickly visualize multiple datasets in different environments at the
 184 same time. We adopted the minimal design principles, providing two core
 185 subcommands *config* and *plot* in the Hctl. Users could use the subcommand *config* of
 186 Hctl to query the demo data and parameters stored in a single JSON file. To submit a

demo task of the heatmap plugin, users just need to execute the command “hctl plot -p <user_custom_input_path>/params.json -t heatmap -o <user_custom_output_path>”. The result files will be automatically retrieved when the task is finished or be manually requested via a random task key.

JSON-based plugin system for visualization tasks

To increase the scalability of the website, we developed a JavaScript Object Notation (JSON)-based plugin system dynamically deploying new web plugins and simultaneously supporting the functionality of the web client and command-line program. All standard native plugins of Hiplot comprise multiple components including documentation, JSON files, and one or more core scripts. The documentation shown on the web page is directly parsed from individual multilingual Markdown files, which also could be reused in the VuePress documentation system. The JSON files rendering the plugins of Hiplot include *Meta JSON*, *UI JSON*, and *Data JSON* (**Fig. S9A**). The application name, thumbnail, entry, version, short description, maintainer, contact information, citation, release date, updated date, and quality score were stored in the *Meta JSON*. In contrast, *UI JSON* was designed to organize available web components of main data tables and parameters using the standard layout of plugins. *Data JSON* contains the default parameters and demo data that are indispensable for the web client, command-line program, and backend service.

Apart from the custom “hiplot-table-editor” and “cloud-file” web components, other seven core components include combobox, autocomplete, text-field, slider, range-slider, switch, and color-picker components of Vuetify.js framework have been supported in the *UI JSON*. The structured description of web interfaces has been significantly simplified the development of web clients for most data visualization analysis tasks. For example, the AGFusion program for visualizing chemical fusion

genes has been wrapped in the AGFusionWeb project requiring massive development work in the front-end building and backend service [25], while we only need to use less than 200 lines of codes for realizing the web interface and backend function if as the plugin of the Hiplot website.

To reduce redundant parameters setting and programming handling, task parameters of *Data JSON* were divided into two parts. The general parameters were used in the web components and backend process functions that are shared in multiple plugins (**Fig. S9B**). The extra parameters only were used to control the specific task steps of plugins. Besides, we have been developed a standalone program HiSub [26] to render the JSON-based web plugin of Hiplot from a structured R script storing meta description, front-end fields, and backend utility (**Fig. S9C**).

Backend service and hardware layer improve tasks efficiency and reproducibility

It is noted that we used personalized task response methods for different visualization tasks with varied time costs, one of them is suitable for time-sensitive tasks. For example, under most circumstances, the users would like to obtain results of most basic visualization tasks in one to a few seconds. To reduce the time costs of R-based web plugins, we adopted a plumber-based multi-users R task response strategy in this work, allowing basic statistics graphics with multiple time-intensive dependencies to be completed in seconds. After starting a plumber session, the dependencies packages and environments will be pre-loaded as the resource representational state transfer (RESTful) application programming interfaces (APIs), which can efficiently process the new user task request and execute the backend R functions of task plugins. In contrast, the computation of gene enrichment analysis based on the clusterProfiler R program [22], and other command-line programs usually takes a few minutes.

240 Because the loading of dependencies only occupies a small portion of the program's
241 runtime, there is no significant change in the time cost.

242

243 The control of the runtime environment of Hiplot is provided by Conda, Singularity
244 image, and renv. The virtualization based on the Singularity image was used to install
245 the dependence software with complex requirements. The renv-based runtime allows
246 the Hiplot cloud service to be independent of the system R package environment and
247 provides a stable and consistent package runtime environment for analysis
248 reproducibility. The website is daily rolling updates in the development cycle. When
249 the stable version is released, a snapshot of front-end *UI JSON*, background scripts,
250 and third-party dependencies would be created. So, users can switch the version of
251 Hiplot by using the selection button at the left top of the navigation.

252

253 In the hardware layer, two high-performance computing nodes with a high-speed
254 internal network were included: 10 core CPUs with 40 threads, 90 TB storage, 128
255 GB memory; 20 core CPUs with 80 threads, 512 GB memory. Two computational
256 nodes have been introduced for balancing task load and reducing network blockage. It
257 is adequate to handle most lightweight tasks of biomedical data visualization based on
258 tabular data.

259 **Results**

260 **Overview of the comprehensive functions in the Hiplot**

261 Since 2019, massive interactive web-based visualization applications (230+) have
262 been developed by the Hiplot Consortium for different biomedical data mining tasks
263 (**Fig. 1A, Table S1**). To our knowledge, this is one of the largest community-driven
264 efforts to establish a free web service for interactively and comprehensively
265 conducting publication-ready biomedical data visualization. In fact, the most of
266 known modern statistics graphics have been implemented in this web service, which

is comparable to the GraphPad, a well-known commercial desktop-based software for scientific data visualization (**Fig. 1B, C**). Hence, users can use these open accessed visualization tools of Hplot to handle daily data analysis without limitations of the operating system or software environment for displaying and/or inferencing the data correlation, distribution, percentage, evolution, flow, ranking, and spatial features.

Excepting basic scientific graphics, users can freely and interactively explore cancer multi-omics datasets and comprehensively conduct multi-omics data visualizations, such as genome structure, chromosome distribution, genetic variations, population genetics, gene expression profiles, gene pathways enrichment, and tumor microenvironment (TME) (**Fig. 1B, C**). Besides, several machine learning-based visualization methods, such as unsupervised clustering, dimensional reduction algorithm (DRA), linear/non-linear regression, meta-analysis, survival analysis, and risk models, permit users to correlate multidimensional clinical features and conduct translational research (**Fig. 1B, C**).

Comparison between Hplot and similar visualization services

To demonstrate the advantages of Hplot, we compared the Hplot with representative websites (**Table 1**). The number and diversity of interactive visualization applications of Hplot (native Hplot, R Shiny, and Python Stremlit) are superior to other similar websites, the basic scientific graphics in particular. The implementation of many basic visualizations in Hplot has attracted users from the whole scientific community, while the omics and clinical data visualizations may further assist the researchers of biomedicine and biology.

In addition to the versatility of the task plugins, the efficient user interfaces and interaction methods in Hplot, significantly save the learning cost of users. Users can invoke the plentiful functions of Hplot via using the main web client or the

294 command-line program (**Fig. 2A**). The web plugins in different modules, such as
 295 basic, advanced, clinical-tools, and mini-tools, can be quickly located via fuzzy
 296 matching of keyword and tag searching or switch and jump plugins via the top path.
 297 Thumbnails on plugins of cards allow users to find which web tools are needed. It is
 298 noted that in the traditional bioinformatics website, users only be allowed to upload
 299 the data using the file path selection or text area. The spreadsheet data editor with
 300 better readability and editability in the plugins of Hiplot may become an optional
 301 replacement method for other similar websites. Standard JSON data objects from
 302 local or remote storage can be used to reproduce the input and output of plugins in
 303 seconds. The multi-user plumber-based response method further improved the task
 304 efficiency in the lightweight visualization tasks (**Fig. 2B**) compared with the
 305 workflow-based response method requiring *de novo* loading dependencies. Though
 306 the setting up tasks using workflow editor has not been supported, it is feasible to
 307 integrate the functions of Hiplot into the existing data analysis pipeline via the
 308 command-line program. In the future, if there is a strong users demand, we can build a
 309 Galaxy-like interface using the existing plugins of Hiplot for better workflow-based
 310 integration.

311 **Usage statistics represent popularity and potential impact**

312 The project of Hiplot was launched in 2019, and the first version was released in
 313 March 2021. Here, we summarized the visits statistics of the Hiplot website from 09
 314 July 2020 to 31 December 2021. It is encouraging that the Hiplot website has been
 315 visited more than 2,500,000 times from 100 countries worldwide in the web browser.
 316 Meanwhile, the website has reached averages of over 5,000 visits and 3,000 task
 317 submissions per day (**Fig. 3A**). More than 22,000 user accounts have been registered
 318 (**Fig. 3B**) and more than 160 plugins of Hiplot were visited more than 1,000 times
 319 (**Fig. 3C**). The basic heatmap plugin alone has been visited more than 70,000 times,
 320 and the correlation heatmap, bubble, boxplot, line-regression, and volcano more than

20,000 times. The UCSCXenaShiny Shiny application in the advanced module was visited 6,307 times at least [13], while the clusterProfiler-based and gene set enrichment analysis (GSEA)-based pathway analyses have been accessed more than 21,000 and 5,000 times respectively. The pdf-collage plugin in the mini-tools module has been visited more than 7,000 times. An increase in user traffic may reflect the potential impact of the web plugins that have been deployed and served in the web service.

Use case 1: basic scientific graphics

One of the common use cases of Hiplot is to draw basic scientific graphics based on tabular data. Users can investigate the correlation of the variables in scatter, chord, line, heatmap plugins, etc. (**Fig. S10A**). At least ten plugins including upset, venn, pie, parliament, waffle, donut, fan, moon charts, tree-map, and flower plot are capable to exhibit the data ratio or topology (**Fig. 1B**), while data distributions can be shown in the histogram, boxplot, violin, ridge, density, area, and bean plots (**Fig. 1B**). Other visualization functions involving data evolution, network relationships, and spatial characteristics also have been developed in the Hiplot, e.g., igraph-based network analysis (**Fig. S10B, C, D**), slopegraph, barplot-line-multiple, and waterfalls plots for viewing the trend of data changes (**Fig. S10E, F**).

Heatmap, the highest high-frequency scientific graphics in the Hiplot website, was chosen to exhibit the common usage of basic visualization tasks (**Fig. 4A**). After entering the plugin page, users can load the demo data via clicking the top/bottom demo button for checking the demo input. In the heatmap plugin, users need to input a numeric data table (one row represents a feature, and one column represents a sample) and optional row/column annotations. The above data tables can be imported from the clipboard, local file, or remote file server. The gene expression matrix of more than 2 megabytes (MB) is recommended to be uploaded in the mode of file path selection

with the file uploader. The general parameters of the heatmap plugin can be used to control the width, height, font, theme, color palettes of annotations rows/columns, font size of row/column text, and title. The extra parameters involve the heatmap colors, scale, top variance, the clustering method, the distance measure, and whether display numeric values. Different clustering methods and distance measures could be tried for finding a reasonable result. Default, the heatmap plugin used the Euclidean distance and ward.D2 method in the clustering analysis. If the discrimination of gradient color is not obvious enough, users could try to adjust the extra parameter “scale” for scaling data by row or column. The selection of top variance features is useful for conducting unsupervised hierarchical clustering if the number of inputted features is too large. After submitting the task, the data stream was encoded based on Base64 and then was transferred to the backend services (**Fig. 4B**). Submitted heatmap tasks will commonly be processed by the available plumber workers. Finally, the heatmap output, e.g., JPG and PDF, would be previewed and downloaded in the bottom preview window. To facilitate the next loading of submitted parameters, users can export the parameters as the local JSON file. In addition, the logged-in users can click the sync button in the preview window to permanently save the results to the cloud file manager.

Use case 2: omics-based data visualizations in cancer

Dozens of multiple multi-omics data visualizations plugins have been developed in the Hiplot, especially cancer genomics and transcriptomics, which mainly consists of the Shiny applications and the native plugins of Hiplot. In 2019, the open-source project UCSCXenaShiny (**Fig. 5A**) was launched by OpenbioX community. In this project, we implement a set of R functions and the Shiny-based web interface allowing users to quickly search, download, explore, analyze, and visualize the dataset from UCSC Xena data hubs [13, 27]. Here, we use the pan-cancer module of UCSCXenaShiny to exhibit that the high expression of *TRH* (thyrotropin releasing

hormone) is significantly associated with favorable prognosis and poor prognosis in the acute myeloid leukemia (AML) and glioma patients from The Cancer Genome Atlas (TCGA) database, respectively (**Fig. S11**). In fact, the prognostic significance of any other genes with de-regulated gene expression or sequence mutations in major types of cancer can be explored in this interactive application. Other Shiny-based applications involving cancer omics visualizations could be found via clicking the “Shiny” and other related tags in the advanced module of the Hiplot website, such as genome-wide association study (GWAS)-related Shiny plugins (**Fig. S12, S13**).

Apart from Shiny-based applications, several JSON-based native plugins of Hiplot can be used to interactively visualize large-scale cancer omics data. The fusion-circize plugin displays the demo gene fusions in B-cell precursor acute lymphocytic leukemia (BCP-ALL) at the chromosomal level [28], such as *BCR-ABL1*, *ETV6-RUNX1*, *DUX4* fusions, *ZNF384* fusions, *MEF2D* fusions, *KMT2A* fusions, and *NUTM1* fusions (**Fig. 5B**). In this plugin, the colors and ribbon width of fusion genes can be customized. The chromosomes-scatter plugin visualizes the numeric value using scatters of chromosomes (**Fig. 5C**), which may help to display the level of gene copy numbers and gene expressions on a large genome-scale. The oncoplot plugin was developed based on the ComplexHeatmap [2], which was used to display the selected mutant genes and gene pathways of patients from the TCGA LAML cohort (**Fig. 5D, E**). Based on the published variants data of patients with BCP-ALL [28], we invoked multiple plugins of Hiplot and validated that the major fusion genes and chromosomal abnormalities are mostly mutually exclusive in BCP-ALL, while *CRLF2* fusions, *DUX4* fusions, and *BCR-ABL1* significantly coexist with the *JAK2*, *MYC*, and *RUNX1* sequence variants respectively (**Fig. S14**). The coexistence mutual exclusion analysis is suggested to be done in the discover-mut-test plugin with better performance in recognition of exclusive events compared with the classical Fisher's exact test [23].

403
 404 The transcriptomic data can be visualized in several basic graphics, such as heatmap,
 405 boxplot, volcano, and pseudo-enhanced-ma, for displaying the gene expression level,
 406 characteristic genes, and conducting unsupervised clustering (**Fig. 1B**). For example,
 407 in the volcano plugin, users can interactively add the selected gene labels of
 408 differentially expressed genes (DEGs). The demo of consensus clustering based on
 409 gene expression data was completed in the cola plugin (**Fig. S15**). It provides the
 410 interface to get stable clusters via data sampling and integrating different top-value
 411 and clustering methods [20]. In addition, we used the complex-heatmap plugin to
 412 correlate multi-omics features using gene expression, gene mutations, and clinical
 413 features (**Fig. S16**) [2].

414
 415 In the clusterprofiler-go-kegg plugin, we performed a multi-group pathway
 416 enrichment analysis of DEGs in a single run based on the demo multi-columns data
 417 (**Fig. 5F**). It supports inputting the gene symbol, ensemble id, and gene id format data
 418 and will merge the multiple graphics and table output in a single PDF and Excel file.
 419 The gsea plugin is a complete web implementation of the original Broad GSEA
 420 command-line program. Compared with the desktop version, this web plugin allows
 421 users to simultaneously compare multiple subgroups and multiple gene sets (**Fig. 5G**).
 422 Besides, we constructed the first web interface of immunedeconv R package, which
 423 allows users to perform the TME analysis and to calculate the immune cell fraction
 424 based on multiple algorithms including quantiseq, TIMER, CIBERSORT,
 425 MCPCounter, xCell, and EPIC (**Fig. S17**) [29-35].

426 **Use case 3: dimensional reductions and clinical data visualizations**

427 Dimension reduction analysis (DRA), such as the principal component analysis
 428 (PCA), t-distributed stochastic neighbor embedding (tSNE), and uniform manifold
 429 approximation and projection (UMAP), etc., selects the most important dimensions

430 from multidimensional data according to the sorting method of features or distances
 431 as representative data for subsequent analysis, thus helping to identify potential new
 432 types of cells/patients with biological/clinical significance. Here, we used multiple
 433 DRA methods to visualize the well-known iris dataset and showed the known classes
 434 in different data spaces after dimensionality reduction (**Fig. S18**). In the clinical data
 435 visualizations tasks, we used the ezcox plugin to conduct a batch of Cox modeling
 436 using a multivariate survival risk model with control variable [36] (**Fig. 6A, B**). The
 437 metawho plugin is a simple web implementation of the “Meta-analytical method to
 438 Identify Who Benefits Most from Treatments” for conducting the meta-analysis tasks
 439 (**Fig. 6C**) [37]. Meanwhile, the risk-plot (**Fig. 6D**) and survival (**Fig. 6E**) plugins were
 440 used to visualize the disease risk models and survival data of patients. The risk-plot
 441 plugin can display the correlation between survival status and risk factors. The
 442 patients were sorted by the risk scores from low to high. In the nomogram plugin, we
 443 established the demo predictive risk score based on the lung dataset in the survival
 444 package (**Fig. 6F**).

445 **Conclusion**

446 In summary, a comprehensive and easy-to-use cloud service, Hiplot, is proposed in
 447 this work for interactively conducting biomedical data visualization. The concise user
 448 interfaces and efficient interaction methods of Hiplot have minimized the learning and
 449 usage costs of lightweight visualization tasks for users without programming skills.
 450 The demo and real datasets show the web-based visualization functions of Hiplot. Our
 451 work provides an important and useful online resource for biomedical researchers and
 452 other data scientists.

453 **Availability and requirements**

454 **Project name:** Hiplot

455 **Project home page:** <https://hiplot.com.cn>

456 **Any restrictions to use by non-academics:** license required

457 **List of abbreviations**

458 **CLI**, command-line interface

459 **JSON**, JavaScript Object Notation

460 **RESTful**, resource representational state transfer

461 **APIs**, application programming interfaces

462 **GO**, gene ontology

463 **KEGG**, Kyoto Encyclopedia of Genes and Genomes

464 **UCSC**, University of California, Santa Cruz

465 **TCGA**, The Cancer Genome Atlas

466 **GSEA**, gene sets enrichment analysis

467 **GWAS**, genome-wide association study

468 **HR**, hazard ratio

469 **CI**, confidence interval

470 **DEGs**, differentially expressed genes

471 **TME**, tumor microenvironment

472 **DRA**, dimensional reduction algorithm

473 **BCP-ALL**, B-cell precursor acute lymphocytic leukemia

474 **PCA**, principal component analysis

475 **tSNE**, t-distributed stochastic neighbor embedding

476 **UMAP**, uniform manifold approximation and projection

477 **Declarations**

478 **Ethics approval and consent to participate**

479 Not applicable.

480 **Consent for publication**

481 Not applicable.

482 **Availability of data and materials**

483 The website can be freely accessed via <http://hiplot.com.cn>. The available open-
484 source code of the website are located at <http://github.com/hiplot>.

485 **Competing interests**

486 The authors declare that they have no competing interests.

487 **Funding**

488 Not applicable.

489 **Authors' contributions**

490 L.J.F, W.M.J, C.S.J, C.Z, S.Y, W.R.X, L.X.S designed and supervised the study.
491 L.J.F, W.M.J, M.B.B, and W.S.X designed and implemented the native Hiplot
492 framework and infrastructure of the web service. L.J.F and M.B.B contributed the
493 command-line tool HCTL. L.J.F, W.M.J, M.B.B, W.S.X, D.W, and X.H.S developed
494 the major applications of Hiplot. D.S.Q, L.J.C, and B.Z.W developed the minor
495 applications of Hiplot. L.J.F, M.B.B, and W.S.X wrote the manuscript and other co-
496 authors critically reviewed and modified the manuscript.

497 **Acknowledgements**

498 The authors are grateful and indebted to the members of the Galaxy, UCSC Xena,
499 UCSCXenaShiny project, OpenbioX community, Shanghai Tengyun Biological

Technology Co., Ltd., and other developers of open-source scientific applications.
Thanks are due to Yang Xu of the National Earth System Science Data Center of
China for providing help in developing geoinformatics applications. Useful
suggestions given by Yutong Ding and other members of the China National
GeneBank are also acknowledged.

References

1. O'Donoghue SI, Baldi BF, Clark SJ, Darling AE, Hogan JM, Kaur S, Maier-Hein L, McCarthy DJ, Moore WJ, Stenau E, et al: **Visualization of Biomedical Data**. *Annual Review of Biomedical Data Science* 2018, **1**:275-304.
2. Gu Z, Eils R, Schlesner M: **Complex heatmaps reveal patterns and correlations in multidimensional genomic data**. *Bioinformatics* 2016, **32**:2847-2849.
3. Wainberg M, Merico D, Delong A, Frey BJ: **Deep learning in biomedicine**. *Nat Biotechnol* 2018, **36**:829-838.
4. Nielsen CB, Cantor M, Dubchak I, Gordon D, Wang T: **Visualizing genomes: techniques and challenges**. *Nat Methods* 2010, **7**:S5-S15.
5. Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, Kohlbacher O, Neuweger H, Schneider R, Tenenbaum D, Gavin AC: **Visualization of omics data for systems biology**. *Nat Methods* 2010, **7**:S56-68.
6. Aebersold R, Mann M: **Mass-spectrometric exploration of proteome structure and function**. *Nature* 2016, **537**:347-355.
7. Li J, Chen H, Wang Y, Chen MM, Liang H: **Next-Generation Analytics for Omics Data**. *Cancer Cell* 2021, **39**:3-6.
8. Lee CM, Barber GP, Casper J, Clawson H, Diekhans M, Gonzalez JN, Hinrichs AS, Lee BT, Nassar LR, Powell CC, et al: **UCSC Genome Browser enters 20th year**. *Nucleic Acids Res* 2020, **48**:D756-D761.
9. Edmonson MN, Patel AN, Hedges DJ, Wang Z, Rampersaud E, Kesserwan CA, Zhou X, Liu Y, Newman S, Rusch MC, et al: **Pediatric Cancer Variant Pathogenicity Information Exchange (PeCanPIE): a cloud-based platform for curating and classifying germline variants**. *Genome Res* 2019, **29**:1555-1565.
10. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C: **Nextflow enables reproducible computational workflows**. *Nat Biotechnol* 2017, **35**:316-319.
11. Jalili V, Afgan E, Gu Q, Clements D, Blankenberg D, Goecks J, Taylor J, Nekrutenko A: **The Galaxy platform for accessible, reproducible and**

- collaborative biomedical analyses: 2020 update. *Nucleic Acids Res* 2020, **48**:W395-W402.
12. Zhou X, Edmonson MN, Wilkinson MR, Patel A, Wu G, Liu Y, Li Y, Zhang Z, Rusch MC, Parker M, et al: **Exploring genomic alteration in pediatric cancer using ProteinPaint**. *Nature Genetics* 2016, **48**:4-6.
13. Wang S, Xiong Y, Zhao L, Gu K, Li Y, Zhao F, Li J, Wang M, Wang H, Tao Z, et al: **UCSCXenaShiny: An R/CRAN Package for Interactive Analysis of UCSC Xena Data**. *Bioinformatics* 2021.
14. Blankenberg D, Kuster GV, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J: **Galaxy: A Web-Based Genome Analysis Tool for Experimentalists**. *Current Protocols in Molecular Biology* 2010, **89**.
15. Yousif A, Drou N, Rowe J, Khalfan M, Gunsalus KC: **NASQAR: a web-based platform for high-throughput sequencing data analysis and visualization**. *BMC Bioinformatics* 2020, **21**:267.
16. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, et al: **The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data**. *Cancer Discov* 2012, **2**:401-404.
17. Zhou X, Wang J, Patel J, Valentine M, Shao Y, Newman S, Sioson E, Tian L, Liu Y, Brady SW, et al: **Exploration of Coding and Non-coding Variants in Cancer Using GenomePaint**. *Cancer Cell* 2021, **39**:83-95 e84.
18. Chen T, Liu YX, Huang L: **ImageGP: An easy-to-use data visualization web server for scientific researchers**. *iMeta* 2022.
19. Gu Z, Gu L, Eils R, Schlesner M, Brors B: **circlize Implements and enhances circular visualization in R**. *Bioinformatics* 2014, **30**:2811-2812.
20. Gu Z, Schlesner M, Hubschmann D: **cola: an R/Bioconductor package for consensus partitioning through a general framework**. *Nucleic Acids Res* 2021, **49**:e15.
21. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**:15545-15550.
22. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al: **clusterProfiler 4.0: A universal enrichment tool for interpreting omics data**. *Innovation (N Y)* 2021, **2**:100141.
23. Canisius S, Martens JW, Wessels LF: **A novel independence test for somatic alterations in cancer shows that biology drives mutual exclusivity but chance explains most co-occurrence**. *Genome Biol* 2016, **17**:261.
24. **x-spreadsheet: a web-based JavaScript (canvas) spreadsheet**. <https://github.com/myliang/x-spreadsheet>. Accessed 4 March 2021

- 579 25. **AGFusion: annotate and visualize gene fusions.**
580 <https://www.biorxiv.org/content/10.1101/080903v1>. Accessed 4 February
581 2022
- 582 26. **HiSub: speed up converting plotting function to a Hiplot plugin.**
583 <https://github.com/hiplot/hiplot-plugin-generator>. Accessed 21 May 2021
- 584 27. Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, Banerjee
585 A, Luo Y, Rogers D, Brooks AN, et al: **Visualizing and interpreting cancer**
586 **genomics data via the Xena platform.** *Nat Biotechnol* 2020, **38**:675-678.
- 587 28. Li JF, Dai YT, Lilljebjorn H, Shen SH, Cui BW, Bai L, Liu YF, Qian MX,
588 Kubota Y, Kiyoi H, et al: **Transcriptional landscape of B cell precursor**
589 **acute lymphoblastic leukemia based on an international study of 1,223**
590 **cases.** *Proc Natl Acad Sci U S A* 2018, **115**:E11711-E11720.
- 591 29. Sturm G, Finotello F, Petitprez F, Zhang JD, Baumbach J, Fridman WH, List
592 M, Aneichyk T: **Comprehensive evaluation of transcriptome-based cell-**
593 **type quantification methods for immuno-oncology.** *Bioinformatics* 2019,
594 **35**:i436-i445.
- 595 30. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD,
596 Diehn M, Alizadeh AA: **Robust enumeration of cell subsets from tissue**
597 **expression profiles.** *Nat Methods* 2015, **12**:453-457.
- 598 31. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, Krogsdam
599 A, Loncova Z, Posch W, Wilflingseder D, et al: **Molecular and**
600 **pharmacological modulators of the tumor immune contexture revealed by**
601 **deconvolution of RNA-seq data.** *Genome Med* 2019, **11**:34.
- 602 32. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS: **TIMER: A**
603 **Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune**
604 **Cells.** *Cancer Res* 2017, **77**:e108-e110.
- 605 33. Aran D, Hu Z, Butte AJ: **xCell: digitally portraying the tissue cellular**
606 **heterogeneity landscape.** *Genome Biol* 2017, **18**:220.
- 607 34. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J,
608 Laurent-Puig P, Sautes-Fridman C, Fridman WH, de Reynies A: **Estimating**
609 **the population abundance of tissue-infiltrating immune and stromal cell**
610 **populations using gene expression.** *Genome Biol* 2016, **17**:218.
- 611 35. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D: **Simultaneous**
612 **enumeration of cancer and immune cell types from bulk tumor gene**
613 **expression data.** *Elife* 2017, **6**.
- 614 36. **ezcox: An R/CRAN Package for Cox Model Batch Processing and**
615 **Visualization.** <https://arxiv.org/abs/2110.14232>. Accessed 09 January 2022
- 616 37. Fisher DJ, Carpenter JR, Morris TP, Freeman SC, Tierney JF: **Meta-**
617 **analytical methods to identify who benefits most from treatments: daft,**
618 **deluded, or deft approach?** *Bmj* 2017, **356**:j573.

619

620 Tables

621 **Table 1. Comparison between Hplot and similar web services**

Features	Hplot	ImageGP	Galaxy
Applications and functions			
Number of visualizations tasks	230+	16	60+
Basic graphics	Full	Small	Small
Statistical analyses	Large	Small	Intermediate
Omics visualization	Large	Small	Large
Clinical visualization	Large	No	Small
HTML/JavaScript applications	Yes	No	Yes
Shiny applications	Yes	No	Yes
Stremlit applications	Yes	No	No
User interfaces & Interactions			
Front-end framework	Vuetify.js+Vue.js	Bootstrap	Vue.js
Difficulty of getting started	Low	Intermediate	High
Thumbnails of applications	Yes	Yes	No
Spreadsheet mode	Yes	No	No
Easy demos (optional multiple)	Yes	Yes	No
Upload file before submit	Optional	No	Yes
Search of tools	Yes	No	Yes
Tags of tools	Yes	No	No
Plumber-based tasks response	Yes	No	No
Storage of history tasks	Local+JSON	No	Remote
Filtration of history tasks	Yes	No	No
Import history from local	Yes	No	No
Import history from remote	Yes	No	Yes
Permanent history	Optional	No	Yes
Documentation	Yes	Yes	Yes
File manager	Yes	Yes	Yes
Command-line program	Yes	No	No
Workflow usage	No	No	Yes
Environment and version			
Runtime manager	conda/renv/singularity	/	conda/docker/singularity
Switch version	Yes (website)	No	Yes (plugin)
Development			
Code repository	GitHub	GitHub	GitHub
Plugin system	JSON-based (Meta/UI/Data)	/	XML-based
Multi-language	Chinese+English	No	Yes
Structured R script	HiSub	No	No
Cycle of updates and online	Nightly	Nightly	Monthly
Other			
Feedback	Built-in/Email/GitHub	Email	Email/GitHub
Registration required	Partially	No	No
Community-driven project	OpenbioX	No	Galaxy
Release year	2019	2017	2005
Open source	Partially	Partially	Full

	License	Academic Free	Free	Academic Free
622	AutoML, automatic machine learning. UI, user interface.			
623				

624 **Figure legends**

625 **Fig. 1. Overview of core features and functions in the Hplot web service.** (A)

626 Three columns indicate the key features and advantages of the Hplot cloud service.
 627 Comprehensive biomedical visualization functions related to modern statistical
 628 graphics, omics, and clinical data visualization have been established in the Hplot
 629 website with a user-friendly web and command-line user interface. The spreadsheet
 630 and switchable file uploader simplified data importing for web-based lightweight
 631 visualization tasks. Basic graphics are completed in seconds based on the plumber
 632 workers. Users can directly use the Hplot to generate publication-ready visualization
 633 graphics. (B) Graph gally shows selected demo outputs of partial plugins of Hplot
 634 related to basic graphics, omics, and clinical data visualization. More demo output can
 635 be directly viewed in the cards list of applications on the website. (C) Classification of
 636 plugins and the four functions classes including basic graphics, omics, clinical, and
 637 other plugins. The left shows the core classes of basic graphics, and the right list the
 638 entry name of web plugins that were contributed by the Hplot Consortium.

639

640 **Fig. 2. Website infrastructure and components from client to backend services.**

641 (A) Two types of user interfaces are provided including the web client and command-
 642 line program, Hctl. The top window shows the screenshot of the Hplot web client on
 643 the basic module page and several cards of partial basic applications are listed. On the
 644 left of the page is the main menu to navigate the website in different modules. The
 645 notification window is also shown with the history task records. The bottom window
 646 shows the subcommands of Hctl including login, config, and submit. Login is
 647 required to use the Hctl program. The *config* and *submit* subcommands respectively to
 648 query the demo data/parameters and to submit tasks. (B) The infrastructure diagram
 649 illustrates the core backend services and hardware resources of the Hplot web
 650 service. The web and command-line clients of Hplot are communicated with the
 651 Nginx proxy/Node.js Express APIs/Web Socket services. The task plugins of Hplot

are distributed in four core modules including basic, advanced, mini-tools, and clinical tools. Apart from the JSON-based Vue.js plugin, the R Shiny and Python Streamlit frameworks are also introduced for the construction of interactive applications. The runtime environment of Hiplot was controlled by the renv, conda, and singularity. The plugins that are deployed in the Github of Hiplot or OpenbioX organization will be automatically synced to the development environment or production environment if they pushed a new commit.

Fig. 3. User visit statistics of Hiplot website from July 2020 to December 2021.

(A) The line plot shows the PV and UV statistics of Hiplot from July 2020 to December 2021. The website has been visited more than 2,500,000 times, and now reaches 5000 times visits per day. (B) The line plot demonstrates the growth trends of registered users of Hiplot, and more than 22,000 user accounts have been registered. (C) The bar plot indicates the PV of the top-visited plugins in the basic, advanced, and mini-tools module of Hiplot. The order of web plugins is sorted by the module names and PV. The visits of web plugins in the basic module are overall greater than other modules reflecting the common daily demands in the whole community. The heatmap, clusterprofiler-go-kegg, and pdf-collage are the top-visited applications in the basic, advanced, and mini-tools modules respectively. PV, page view. UV, unique visitor.

Fig. 4. Usage flow and task processing steps through the heatmap plugin.

(A) The web interface of the heatmap plugin and the major operation steps for submission of a new task. Tabular data tables are imported and can be previewed and edited in the spreadsheet web component. The general and extra parameters of the heatmap are listed on the right panel. These parameters can control the data pre-processing, clustering steps, and the output style of the heatmap. The parameter “top variance” can be used to select partial genes for conducting the clustering. The parameters of the

680 heatmap are stored in a JSON format data structure and can be directly exported for
 681 reproducing the input. The history results in the cloud file manager can be used to
 682 reproduce the input and output. (B) The backend response steps and the relevant
 683 services processing the submitted task. Submitted data will be encoded using the
 684 Base64 algorithm. The available plumber worker processes the heatmap task and the
 685 core codes of the heatmap plugin. When the task is finished, the backend returns a
 686 history JSON file storing task information and path of output for retrieving the image
 687 and logging output. (C) Demo output of heatmap plugin. One column represents a
 688 sample and one row represents a gene. Two clusters can be defined in the demo input
 689 (cancer vs. control). The row annotations and column annotations respectively show
 690 the classification of genes (cancer or metabolism). and sample types (cancer or
 691 control).

692

693 **Fig. 5 Representative use cases of omics-based visualization functions in Hiplot.**

694 (A) Screenshot of UCSCXenaShiny application, which is an R shiny-based
 695 application for interactively mining the published datasets from UCSC Xena Hub. (B)
 696 Demo output of the fusion-circlize plugin. It shows a part of known gene fusions
 697 including *ETV6-RUNX1*, *BCR-ABL1*, *DUX4* fusions, *ZNF384* fusions, *MEF2D*
 698 fusions, and *KMT2A* fusions in B-cell acute precursor lymphoblastic leukemia (BCP-
 699 ALL). Different chromosomes are ordered clockwise. Different chromosomes are
 700 ordered clockwise. The gene fusions are linked between chromosomes using ribbons.
 701 The color and the width of the link/text indicate the classes and frequency of the
 702 specific fusion genes. (C) Demo output of chromosomes-scatter plugin, which shows
 703 the simulated numeric values in different chromosomes with colors (D) Oncoplot
 704 shows selected mutant genes in the acute myoid leukemia (AML) cohort from The
 705 Cancer Genome Atlas (TCGA) database. The patients are split into two parts
 706 according to their survival status. Different mutation types are labeled in different
 707 colors. The bottom box draws the meta-annotation of patients. (E) Another Oncoplot

with extra the gene pathway rows and the patients are sorted by mutation types. (F) The demo outputs of gsea and clusterprofile-go-kegg plugins. The NES and p-value are shown in the GSEA output, which is regenerated to the PDF version. The bubble plot and network plot show the enriched GO pathways based on the demo data of clusterprofile-go-kegg plugin. KEGG, Kyoto Encyclopedia of Genes and Genomes. UCSC, University of California, Santa Cruz. TCGA, The Cancer Genome Atlas. GO, gene ontology. GSEA, gene sets enrichment analysis. NES, normalized enrichment score.

Fig. 6. Representative use cases of clinical data visualization in Hplot. (A) and (B) respectively shows the demo forest output using the lung dataset in the survival package and the web interface of the ezcox plugin. The COX models show that the sex and ph.ecog variables are respectively associated with high risk (HR:1.56, CI: 1.24-1.96) and low risk (HR:0.60, CI: 0.43-0.83) in the lung when age is used as the control variable. (C) Demo output of the metawho plugin. The HR of study subgroups and studies are displayed in the table. (D) Demo output of the risk-plot plugin. One point represents one patient, which is divided into high risk (red) and low risk (blue) subgroups according to the median value of risk scores. Dead patients are shown in red points in the middlebox. The expression of feature genes including *EMP3*, *TIMP1*, *PDPN*, and *TAGLN2*, displayed in the bottom box. (E) Demo output of the survival plugin. It indicates the three years survival curve in G1 (blue) and G2 (red) subgroups of simulated data. (F) Nomogram plot that can be used for predicting the prognosis in the lung dataset of the survival package. The p-value is calculated based on the log-rank test. HR, hazard ratio. CI, confidence interval.

A

Features of Hiplot platform

Easy and User-friendly

- ✓ Modern UI and well-designed user client
- ✓ Reproducible parameters and results
- ✓ Concise parameters and operations
- ✓ Switchable file uploader and online table
- ✓ Error tracking and feedback
- ✓ Command-line interface
- ✓ Demo, documentation, and tutorial

Comprehensive and fast

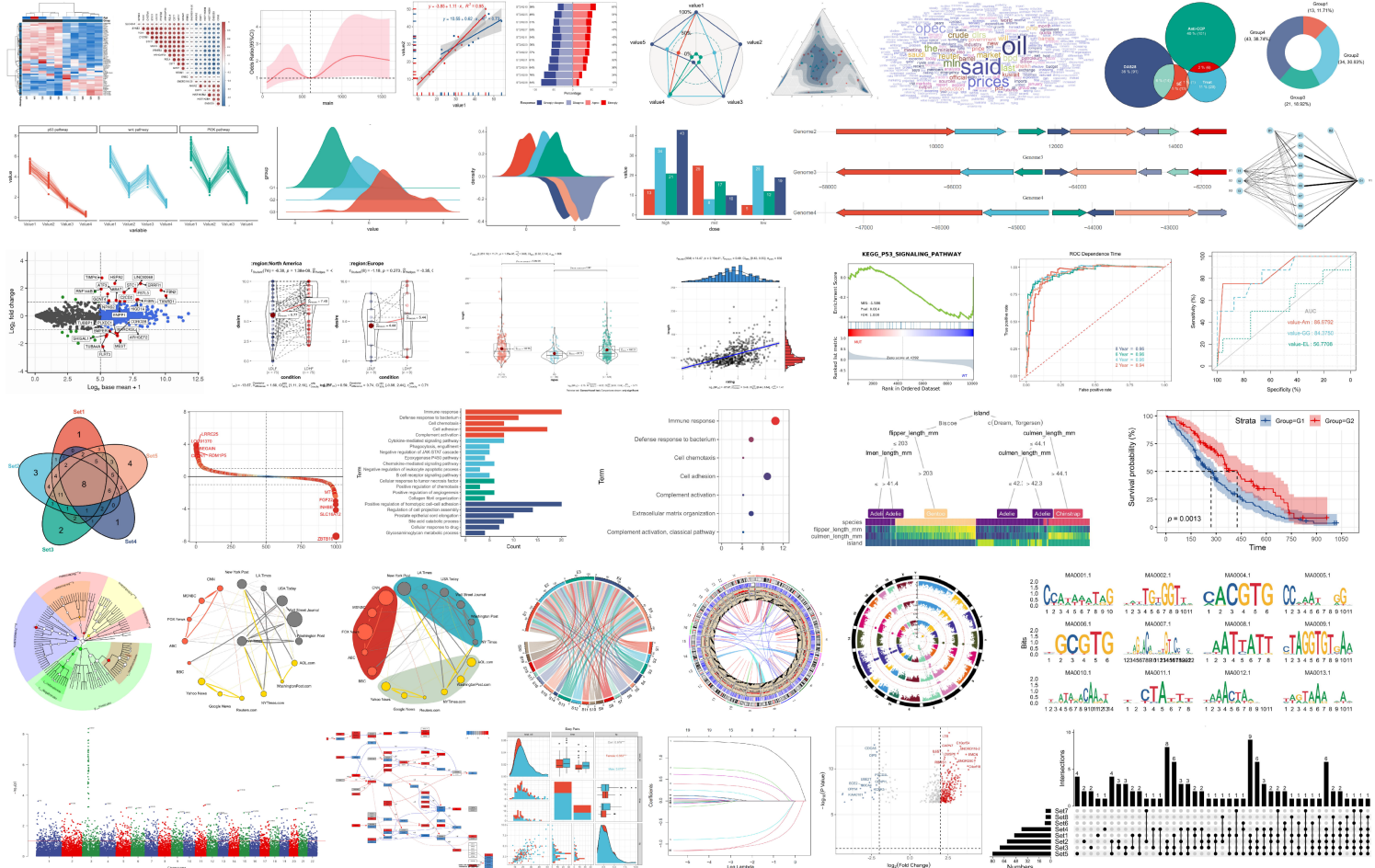
- ✓ More than 200+ functional plugins
- ✓ Statistical description and inference
- ✓ Multi-omics applications
- ✓ Machine learning-based visualization
- ✓ Multi-user plumber-based tasks
- ✓ Basic visualization in seconds
- ✓ Scalable plugins

Publication-ready output

- ✓ Validated algorithms and tools
- ✓ Rich parameters for complex output
- ✓ Predefined themes and layout
- ✓ Hundreds of color palettes
- ✓ Diversified output formats (PDF/SVG/...)
- ✓ Local modification based on R data object
- ✓ Grid arrangement of outputs in batch

B

Graph gallery



C

Correlation

Distribution

Flow

Ranking

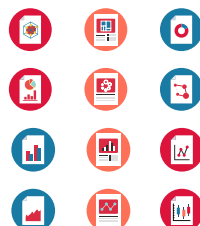
Part of whole

Evolution

Map

Other graphics

Scientific graphics



Omics

Biomedicine

Other

fishplot, gene-density, ggseqlogo, oncoplot, fusion-circize, discover-mut-test, ggtree-msa, hic-heatmap, doabsolute, agfusion, sigflow, gistic2, maf-summary, maf-oncoplots, maf-rainfall, maf-titv, genome-circos, gggenes, cmplot, manhattan-shiny, ucsc-xena-shiny, complex-heatmap, shinyseurat, heatmap, cola, corplot, cor-heatmap, clusterprofiler-go-kegg, gsea, pseudo-enhanced-ma, volcano, gocircle, gobar, gobubble, gene-rank, grdotplot, diy-gsea, ideogram-heat, chromosomes-scatter, pathview, immunedeconv, enhancedvolcano, anatogram-shiny, cemitool, cluster, go-plot, kegg-enrichplot, go-enrichplot, otu-tree, community-shiny, network-shiny.....

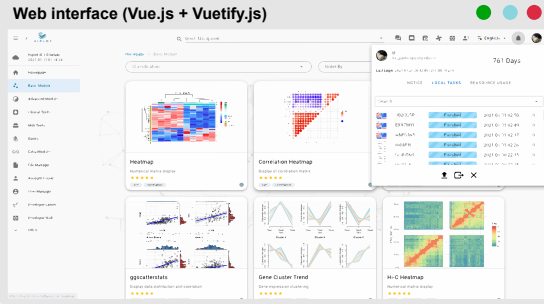
pca, pcatools, tsne, tsne-ellipse, pca-tsne, umap, umap-ellipse, forestplot, neural-network, treeheat, roc, time-roc, nomogram, nomogram-logistic, ezcox, survival, rcs-lrm, rcs-cox, risk-plot, metawho, meta-bin, meta-cont, calibration-curve, rda-visual, psm-matchit, lasso-ridge, pls-da, opds-da, smd-visual, nmds-ellipse, pcoa-ellipse, pca-ellipse, cluster, strml-tf-linear-regression, strml-svm-classifier, clinaml-gep.....

pdf-collage, bioextr-pdf, bioextr-pubmed, gene-ids, na-deal,

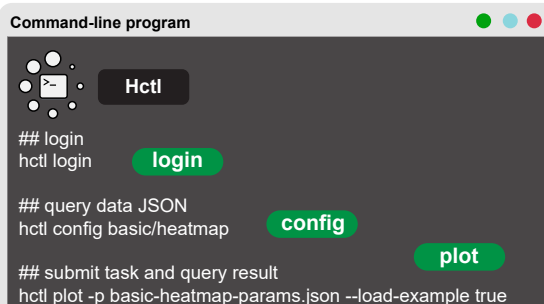
A

User client

Web interface (Vue.js + Vuetify.js)



Command-line program



```

Hctl

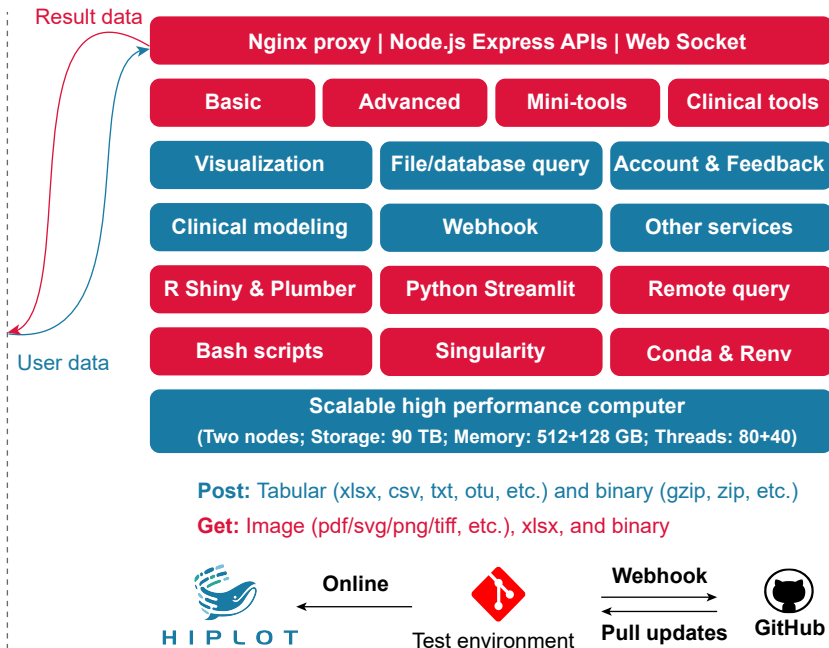
## login
hctl login

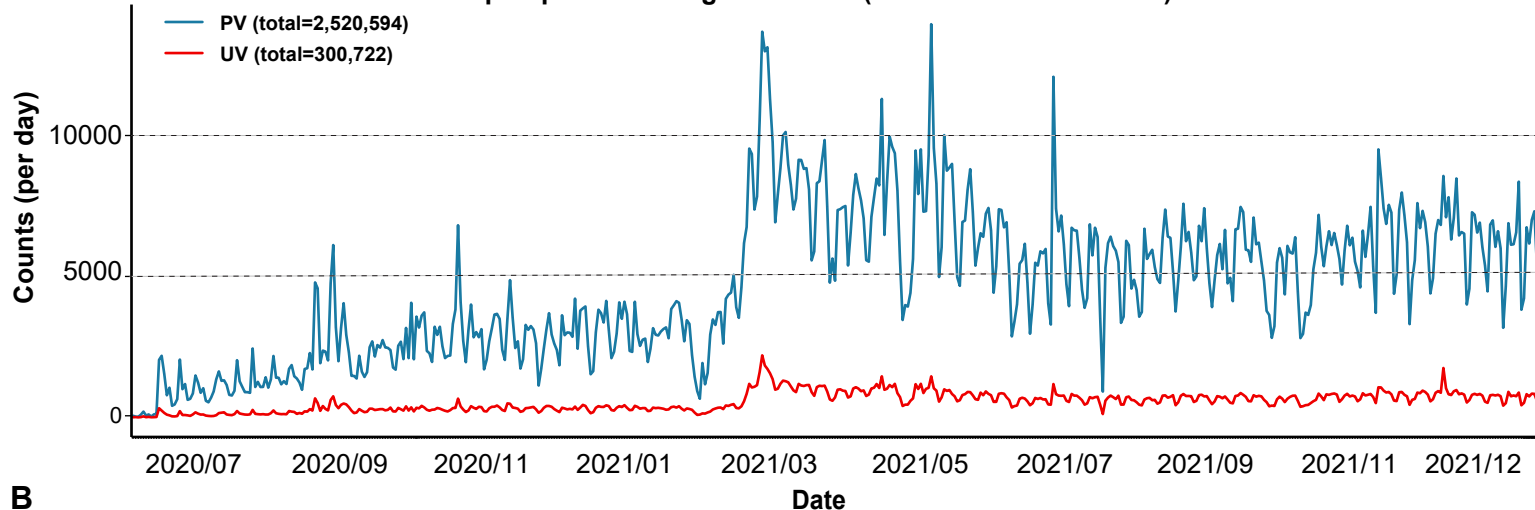
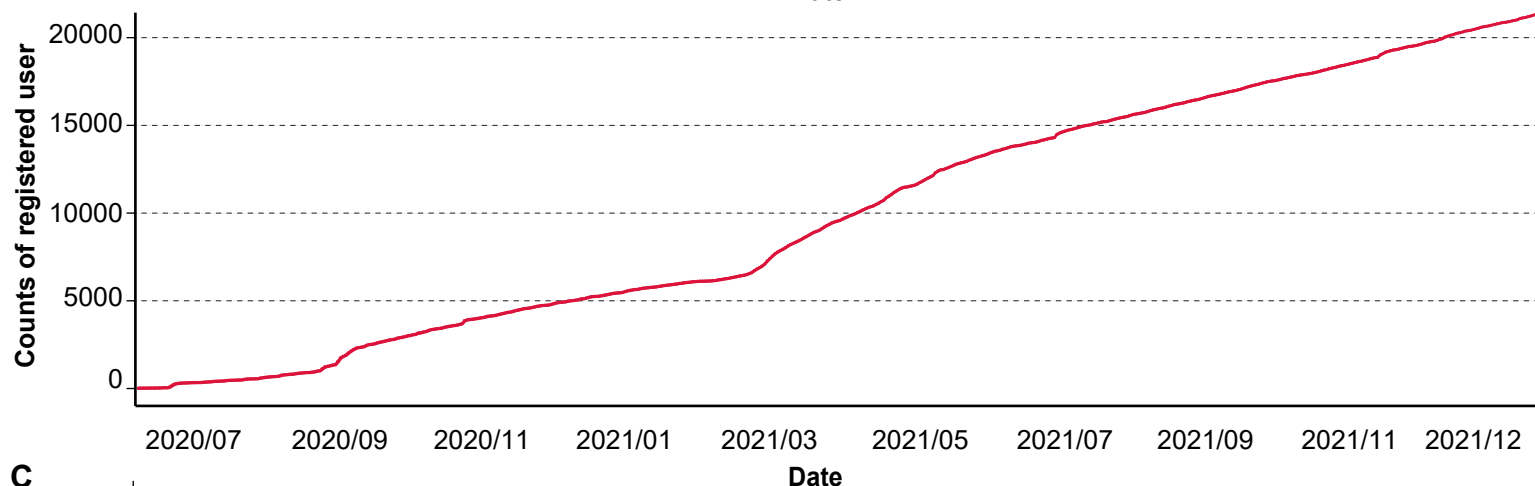
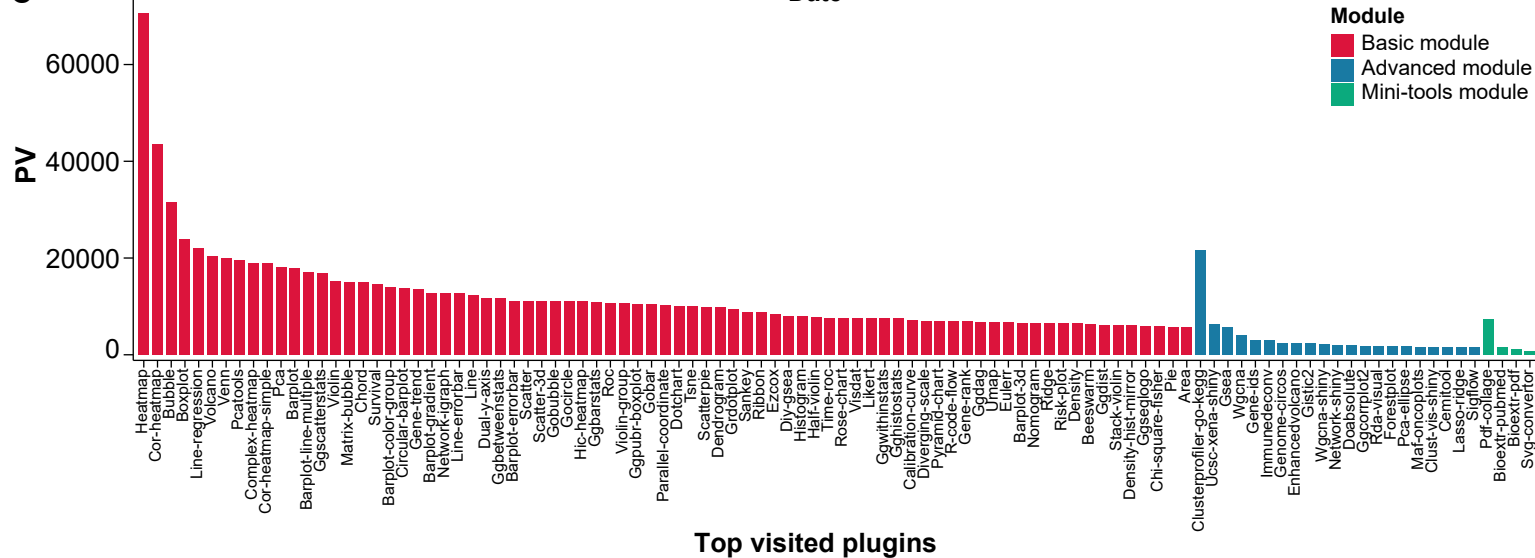
## query data JSON
hctl config basic/heatmap

## submit task and query result
hctl plot -p basic-heatmap-params.json --load-example true
  
```


B

Remote service



A**Hiplot platform usage statistics (2020/07/09 - 2021-12/31)****B****C**

A

Import Data  **Notion 1.1**

Counts

	A	B	C	D	E	F	G
1	Gene	M1	M2	M3	M4	M5	M6
2	GBPI4	6.5993441095	5.2262960248	3.6932875381	3.9385009528	4.5271930886	9.3081190324
3	BCAT1	5.7603803773	4.8927830209	5.4486281723	3.4854133857	3.8556691177	8.8620810404
4	CMKP2	5.96119551154	4.5491681565	3.9986540224	5.6143839805	3.9047928726	9.7907700044
5	STOX2	8.3964093157	8.7170552201	8.0390641101	7.6430595565	9.2746409091	4.4170130073
6	PADI2	8.4197658905	8.2684298023	8.4511808436	9.2007322844	8.5982072017	4.5900330127
7	SCARNAS	7.6530473097	5.7603926297	10.63350472	5.9136843629	8.8056047501	5.8901204765
8	ALOX12P2	9.0014211521	7.7535422310	4.9209595817	7.6363543709	8.8986047890	6.2119435438
9	SNORA74B	5.5387087152	7.1690533938	11.4416464368	5.9825656857	8.3944025246	4.7515592532
10	HIST1H2BL	6.1812704666	4.4484188901	5.6049610208	6.4851683738	2.5095743866	7.6301766056
11	MINDA	8.7026352495	7.3223519563	9.2718214201	9.1345154397	6.3391138436	4.5719022848
12	CLF4	9.345383049	6.4386096367	8.7167612840	9.685451685	7.6882818432	5.9698038656
13	FLT3	5.1643063209	4.4193554432	5.0635513633	5.4127835356	3.3778631484	8.5851643929
14	CHD7	8.3734749638	6.7280322425	8.4088176083	8.7164755127	8.3737706673	4.6131990181
15	NFPL3	8.0554164753	8.3650998813	9.2391936769	8.5368250774	8.4523508188	4.4785271277
16	HSPA1B	5.0081989999	9.1946502771	6.59886499179	7.1295082512	3.8852697172	5.7121787252
17							

Notion 2.1

SampleInfo

	A	B	C	D	E	F	G
1	Sample	Group	Age				
2	M1	Control	23				
3	M2	Control	34				
4	M3	Control	25				
5	M4	Control	52				
6	M5	Control	36				
7	M6	Cancer	28				
8	M7	Cancer	31				
9	M8	Cancer	22				
10	M9	Cancer	29				
11	M10	Cancer	34				
12							
13							
14							
15							
16							
17							

Set Parameters **Notion 3.1**

General Parameters

Task Name: 27109da0-50f6-11ec-866d-79be10e2505e Title: Heatmap Plot

Color Palette: starry2 Color Palette 2: starry2 Font Family: Arial

Image Export: jpeg pdf

Width: 8 Height: 6

FontSize (Row): 6 Fontsize (Column): 10

Extra Parameters **Notion 3.2**

Heatmap Color: custom Scale: none Cluster Method: ward.D2

Row Distance Measure: euclidean Column Distance Measure: euclidean

☒ Cluster Rows ☒ Cluster Column ☐ Display Value

Top Variance: 100

☒ Color Low ☐ Color Middle ☐ Color High

SUBMIT **RESET** **DEMO** **Notion 4.1**

Notion:

1.1 Load demo data/parameters

2.1 Table operations

Switch to file uploader

Load table demo data

Import data from cloud

Import data from local file

Export table data

Clear table data

3.1 Set general parameters

3.2 Set extra parameters

4.1 Task operations:

Submit task

Reset to default

Load demo data/parameters

Load history result

Upload local data/parameters

Export data/parameters

5.1 Demo output (Heatmap plugin)

B

User client
(Web browser or command-line interface)

Submit task and create local record
Base64 encoding object

Check task via sending socket message
(refresh page or polling)

Sync task to cloud library
Retrieve results

bash run.sh
Rscript run.R
python run.py
julia run.jl
.....

Command-line

Data import

Preprocess

Result export

Pre-defined utility

Pre-loaded Plumber

Create working directory
Write task data/parameters JSON file
Create task history file (JSON file)
Check task types (Plumber or command-line mode)
Check available Plumber APIs (If it is Plumber-based)
Invoke APIs (PID:host:port) or programs
Update history file after task completed or failed
Send socket message of task to client

Query static files
Feedback
Signup
Authorization

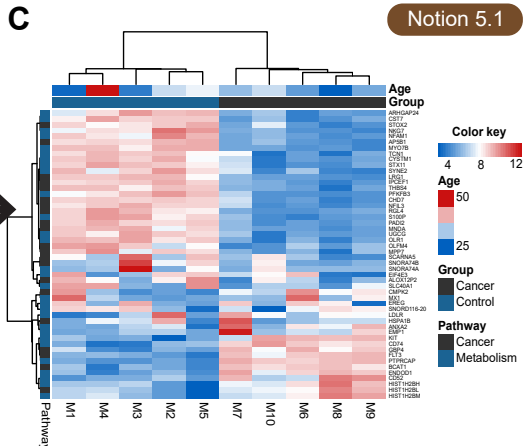
Running process

Idle process

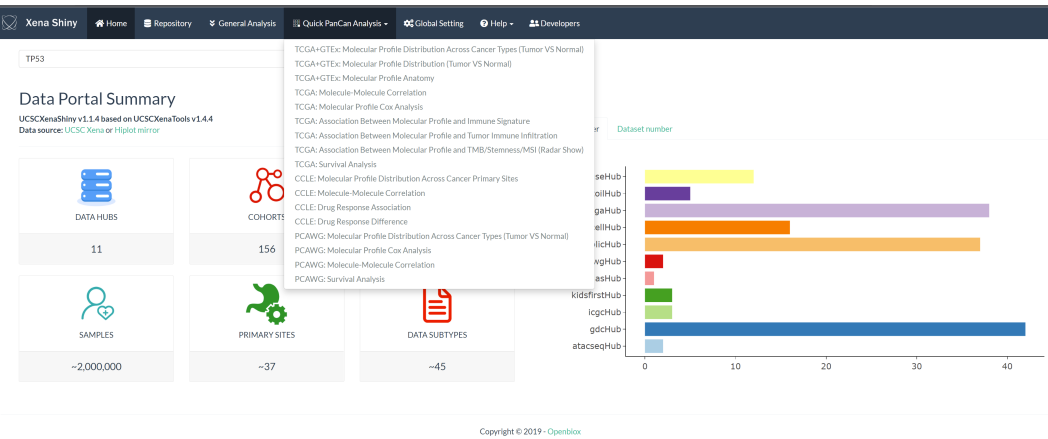
CONDA

Reproducible environment
(Conda/Spack + Singularity + Renv)

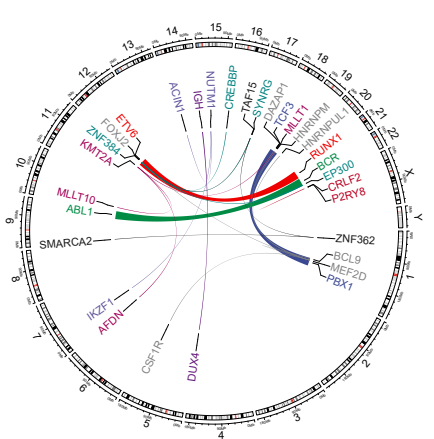
Service for native Hiplot application



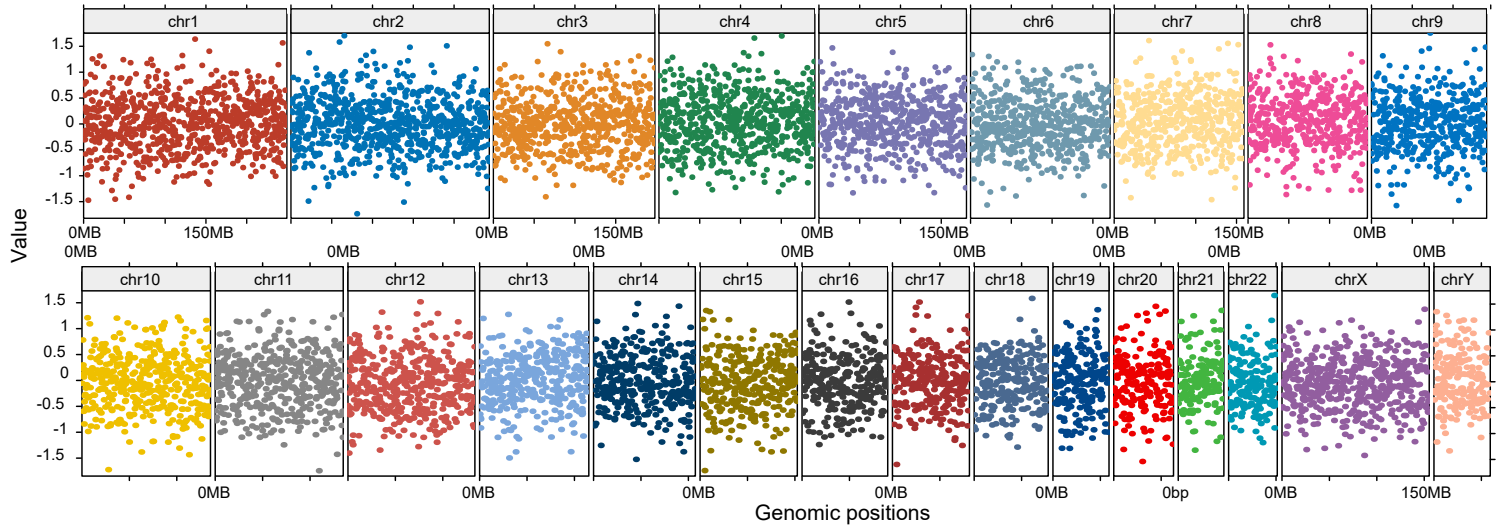
A UCSCXenaShiny



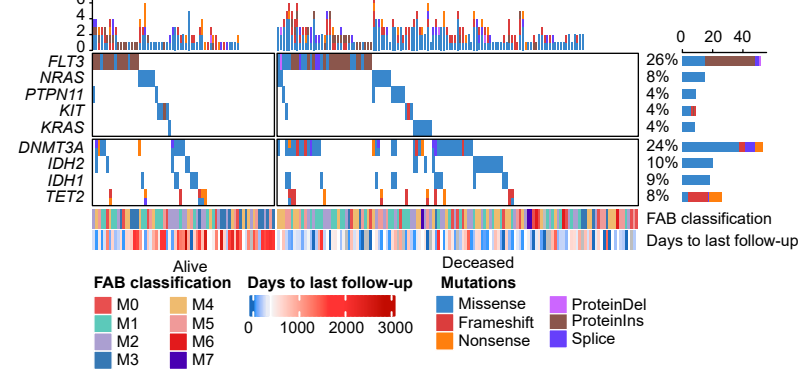
B Fusion-circlize



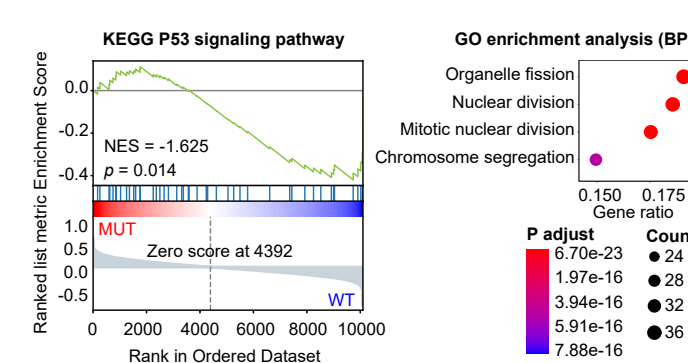
C Chromosomes-scatter



D Oncoplot



F Gene pathway analysis (GO/KEGG and GSEA)



E Oncoplot (Pathway with type sort)

