Aquila: a spatial omics database and analysis platform

Yimin Zheng [1,2,3,4], Yitian Chen [1,2,3,4], Xianting Ding [5], Koon Ho Wong [1,2,3,4] and Edwin Cheung [1,2,3,4,*]

[1]Cancer Centre, University of Macau, Taipa 999078, Macau SAR, [2]Centre for Precision Medicine Research and Training, University of Macau, Taipa 999078, Macau SAR, [3]MoE Frontiers Science Center for Precision Oncology, University of Macau, Taipa 999078, Macau SAR, [4]Faculty of Health Sciences, University of Macau, Taipa 999078, Macau SAR and [5]Institute for Personalized Medicine, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

ABSTRACT

Spatial omics is a rapidly evolving approach for exploring tissue microenvironment and cellular networks by integrating spatial knowledge with transcript or protein expression information. However, there is a lack of databases for users to access and analyze spatial omics data. To address this limitation, we developed Aquila, a comprehensive platform for managing and analyzing spatial omics data. Aquila contains 107 datasets from 30 diseases, including 6500+ regions of interest, and 15.7 million cells. The database covers studies from spatial transcriptome and proteome analyses, 2D and 3D experiments, and different technologies. Aquila provides visualization of spatial omics data in multiple formats such as spatial cell distribution, spatial expression and co-localization of markers. Aquila also lets users perform many basic and advanced spatial analyses on any dataset. In addition, users can submit their own spatial omics data for visualization and analysis in a safe and secure environment. Finally, Aquila can be installed as an individual app on a desktop and offers the RESTful API service for power users to access the database. Overall, Aquila provides a detailed insight into transcript and protein expression in tissues from a spatial perspective. Aquila is available at https://aquila.cheunglab.org.

INTRODUCTION

Highly complex biological processes, such as the development of the neurological system or the progression of diseases like diabetes and cancer, are extremely challenging to study due to the heterogeneous cellular nature of tissues and the complex interactions between cells. However, the spatial distribution of cells within tissues is positively related to their function and lineage trajectories (1,2). Thus, understanding the spatial organization of tissues and the evolutionary relationship among cells is vital for capturing the different steps of developmental processes and the complete portrait of diseases. The rapid development and adoption of spatial omics technologies recently have accelerated our understanding of these essential processes.

Many spatial omics technologies have been developed so far, with the majority geared toward interrogating the spatial transcriptome of tissues (3). Spatial transcriptome methods can be broadly divided into either sequencing- or imaging-based techniques. Sequencing-based techniques capture and preserve spatial gene expression information of tissues by *in situ* hybridizing RNA transcripts onto spatially immobilized arrays of indexed sequences. Visium is currently the most popular sequencing-based tool, although, at present, this technology and similar technologies like Slide-seq (4) and DBiT-seq (5) are not yet at single-cell resolution (6). In contrast, newer techniques such as spatial enhanced resolution omics sequencing (Stereo-seq) (7), Seq-Scope (8) and sci-Space (9) are at single-cell or subcellular resolution. Imaging-based spatial transcriptome technologies include sequential fluorescence *in situ* hybridization (seqFISH) (10), multiplexed error-robust fluorescence *in situ* hybridization (MERFISH) (11) and spatially resolved transcript amplicon readout mapping (STARmap) (12). These technologies require designing fluorescent-labeled probes to hybridize with known RNA targets and a super-resolution microscope to capture the fluorescent signal.

Besides measuring transcript information, spatial omics technologies can also detect the proteome of tissues (13). Compared to spatial transcriptome technologies, the throughput of spatial proteomic technologies is currently limited due to the availability of fluorescent or metal-conjugated antibodies required for detecting target proteins. As with spatial transcriptome technologies, spatial proteomic technologies can also be categorized into two main groups. For example, cyclic immunofluorescence (14)

and co-detection by indexing (15) are imaging-based techniques that measure protein abundance by repeatedly staining tissues with three to five fluorescent-labeled antibodies at a time. In comparison, imaging mass cytometry (IMC) (16) and mass ion beam imaging (MIBI) (17) are mass cytometry-based technologies that stain and quantify tissues using a maximum of 40 metal-conjugated antibodies at once.

Since spatial omics technologies produce a rich amount of data, they have been widely adopted in the scientific field, resulting in many exciting new findings. For example, Stereo-seq has been applied extensively for the spatial–temporal gene expression profiling of early embryonic stages at unprecedented resolution in mice (7), flies (18) and zebrafish (19). In a multimodal study integrating scRNA-seq, Visium and MIBI, Khavari and colleagues examined the spatial architecture of squamous cell carcinoma, revealing a multifaceted immunosuppression mechanism, including the exhaustion of T cells (20). Recently, IMC was used to map the spatial and temporal landscape of the lung at subcellular resolution during the progression of coronavirus disease 2019 (21). This study showed that the abundance, hyperactivities, the interplay between different inflammatory cells and structural changes contribute to the poor immune response in the late stage of the disease (21).

The recent precipitous expansion of spatial omics has also produced many datasets, and the increase will continue as the technology matures and becomes more widely accessible. A database that users can easily access, visualize and use for analyzing different types of spatial omics datasets is immensely beneficial for the scientific community. Unfortunately, no databases currently provide users with all these features in one complete package. At present, SpatialDB (22) is the only spatial omics database available, but it hosts only 24 transcriptome datasets, lacks proteomics data and does not support 3D datasets. SpatialDB also does not offer users any type of spatial analysis. While users can turn to software such as Giotto (23) and Squidpy (24) to analyze their data, these programs are targeted toward bioinformaticians with programming skills, which are unsuitable for most biologists.

Thus, we developed Aquila, a combined spatial omics database and analysis platform that supports all types of spatial omics data, including transcriptome and proteome, 2D and 3D, and from different technologies (Figure 1A). Aquila currently hosts 107 datasets from 15 technologies with >15.7 million cells. Aquila provides users the visualization of spatial omics data in multiple formats, such as cell maps, expression maps and co-localization of markers (Figure 1B). Users can also easily retrieve any dataset and run a repertoire of integrated analyses directly in the same browser, including advanced spatial analyses that typically require coding skills (Figure 1C). Furthermore, users can submit spatial omics data of their own to Aquila and perform the same analyses as the datasets found in the database. We address data privacy issues by storing all user datasets locally without uploading them to the remote server. Finally, Aquila can also be installed as an individual application on a desktop through the progressive web application standard.

MATERIALS AND METHODS

Data collection and preprocessing

We manually searched for spatial omics-related keywords using PubMed and Google Scholar to acquire publications with relevant datasets. Only publicly available data were collected. To handle the different types of spatial omics data, they were stored in three parts: expression matrix (matrix market format), spatial location and region of interest (ROI) annotation information. For the expression matrix, we filtered out genes or proteins expressed in <100 cells (or spots) and cells (or spots) expressing <1 gene or protein. Any negative expression values were replaced with a minimum value of 0 to ensure all the expression values were nonnegative. Afterward, the expression matrix was log transformed as $\log(1 + X)$. The spatial and annotation information was left intact without any transformation. The processing scripts were written in Python using Numpy (25), Pandas (26) and Scanpy (27).

Determining the dimension of ROIs

Aquila uses the envelope of an ROI as its dimension, which is the maximum boundary of the ROI. For a series of 2D points, the envelope is calculated as $[\text{Min}(x)\text{Min}(y)\text{Max}(x)\text{Max}(y)]$. This equation is also extended to 3D data.

Mixing colors for co-localization of markers

Aquila uses the following algorithm to determine mixed colors by combining two or more colors. Colors are represented in RGB color space, with color A as (R_1, G_1, B_1) and color B as (R_2, G_2, B_2), and the mixed color result as $(R_{\text{mix}}, G_{\text{mix}}, B_{\text{mix}})$. For each channel, the mixing value is calculated as $C_{\text{mix}} = \sqrt{(C_1^2 + C_2^2)/2}$.

Co-expression analysis

Aquila uses Pearson correlation and Spearman correlation to determine the relationship between the expression of any genes or markers. Users can select 2–50 genes to perform co-expression analysis. This analysis is implemented in Rust for optimized computing pair correlation using $(x_i - \bar{x})$ and $(y_i - \bar{y})$ as the kernel for acceleration. Co-expression is calculated as
$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}.$$
Spearman correlation has an additional process in which input vectors are ranked by Spearman ranking, and the ranked vectors are then used for subsequent calculations. Users do not need to provide any parameters for this analysis.

Cell distribution pattern analysis

The cell distribution pattern analysis examines whether cells are distributed randomly, evenly or in a cluster. Aquila offers users three methods to choose from, including the index of dispersion (28), the Morisita index (29) and the Clarks–Evans index (30). For all three methods, the null hypothesis
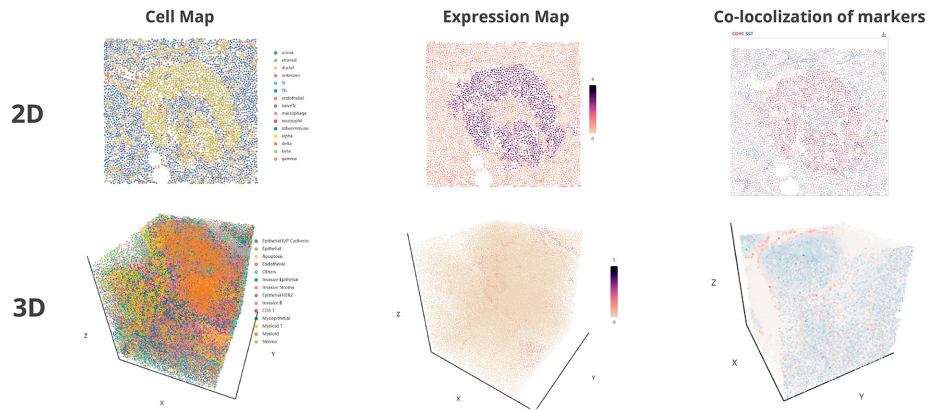
Figure 1. Schematic design of Aquila. (**A**) Various types of spatial omics datasets accessible in Aquila. (**B**) Examples of ROI visualization (i.e. cell map, expression map and co-localization of markers) in 2D and 3D. (**C**) A comprehensive list of spatial analyses with examples of results.

is that points are in a completely random state (Poisson random point process). Each method is tested against the null hypothesis to determine which pattern cells are distributed in an ROI.

The index of dispersion method uses a sampling window to sample an ROI multiple times randomly. Each time, we obtain the number of cells in a window. S^2 is defined as the variance and $\bar{x}$ is the mean. The χ^2-test is used to determine significance. The index value is calculated as

$$I = \frac{s^2}{\bar{x}}.$$

Since Aquila uses circles as sampling windows, users will need to define the radius for the circle and the number of sampling times. By default, Aquila uses one-tenth of the minimum side of the ROI envelope as the radius and samples 1000 times. For each run, the results will be different due to the difference between samplings.

The Morisita index method is based on quadratic statistics that rasterize an ROI into a grid. The number of points in a square is denoted as x. The χ^2-test is used to determine significance. The index value is calculated as

$$I = n \left[\frac{\sum x^2 - \sum x}{(\sum x)^2 - \sum x} \right].$$

Users can define the dimension of the grid to control the granularity of the rasterization process. The default setting uses a 10×10 grid to rasterize the ROI.

The Clark–Evans index is a nonparametric statistical method in which the distribution of cells is based on the distance between neighboring points. Here, ρ is defined as the density and n as the number of individuals.

$$\bar{D} = \frac{\sum_{i=1}^{n} d_i}{n},$$

$$E(D) = \frac{1}{2\sqrt{\rho}},$$

$$SE(D) = \frac{4 - \pi}{4\rho\pi},$$

$$R = \frac{\bar{D}}{E(D)},$$

$$Z = \frac{\bar{D} - E(D)}{SE(D)}.$$

The index of dispersion, the Morisita index and the Clarks–Evans index are all implemented in Rust.

Ripley's cell distribution analysis

Ripley's K function (31) is a cell distribution pattern analysis that works by calculating the average number of points from different distance intervals. Users need to define the number of intervals for the function. In Aquila, the K function and its variants G, F and L are made available by Pointpats (v2.2.0) in the PySAL ecosystem (32).

Spatial entropy analysis

Aquila uses two entropy metrics to measure tissue heterogeneity, the Shannon entropy (33) and Leibovici entropy (34). Shannon entropy is a classic method for evaluating the chaos of a system. In contrast, Leibovici entropy considers the contribution of spatial information to a system's entropy. It requires a distance parameter to determine the range of co-occurring events. By default, the distance is one-tenth of the minimum side of the ROI envelope. Shannon entropy is calculated as

$$H(X) = -\sum P_i \log_2 (P_i),$$

where P_i is defined as the proportion of cell type i in the ROI. Leibovici entropy is determined as

$$H(Z|d) = \sum_{r=1}^{I^m} p(z_r|d) \log \left(\frac{1}{p(z_r|d)} \right),$$

where $p(z_r|d)$ is equal to the proportion of different co-occurring events within the distance condition d. Spatial entropy analysis is implemented in Rust.

Spatial variable gene analysis

The spatial expression pattern of a gene may or may not depend on the spatial context. Aquila identifies spatial variable genes using a GPU accelerated version of SpatialDE (35,36). The current method for detecting spatial variable genes in a regular ROI on a web server is inefficient and can take more than a few minutes. Due to this limitation, users can select no more than 50 genes per run.

Constructing neighboring networks

The results from the neighboring network analysis lay the foundation for the following analyses: spatial community, cell centrality, cell–cell interaction, spatial co-expression and spatial autocorrelation. Aquila provides two methods, KD-Tree and Delaunay triangulation, to determine the neighbors of a cell. KD-Tree employs a binary tree structure approach to analyze spatial data, making it an efficient way for spatial queries. Users set two parameters to use this method, the search radius for each cell and the number of nearest neighbors for a cell. Delaunay triangulation is a nonparametric method that can be thought of as filling a discrete set of points (i.e. cells) in a 2D plane with triangles. Each point is used as the vertex of a triangle such that no points lie in the circumcircle of any triangle. Cells are considered neighbors if they are connected to the edge of the triangle. Both methods construct a spatial network and display the results graphically with nodes representing cells and edges as the Euclidean distance between cells. With the network graph, users can examine the quality of the network and determine which cells have neighbors. They can also check the Euclidean distance between cells by hovering over the edges. Both methods are implemented in Rust using Kiddo (v0.2.3) and Delaunator (v1.0.1).

Spatial community analysis

Tissue phenotyping can be conducted by grouping cells within a local environment as a community (37). Thus, it is essential to identify these groups of cells that are spatially distinguished. Aquila includes three community detection methods, Leiden (38), Louvain (39) and Infomap (40), for clustering cells based on the above-mentioned neighboring network results. For the Louvain method, a multilevel version was implemented, which has a better computation efficiency than the original (41). The Leiden method requires users to input a resolution parameter that affects the size of each community. No input parameters are needed for either Louvain or Infomap. These three methods are executed using Python-igraph (v0.9.10).

Cell centrality analysis

Cell centrality analysis measures the importance of a cell in a network. Aquila provides three ways to determine cell centrality based on 'degree', 'closeness' and 'betweenness' (42). 'Degree' centrality works by counting the number of links connected to a node, while 'closeness' centrality is based on calculating the number of times a node acts as a bridge along the shortest path between two other nodes. Finally, 'betweenness' centrality measures the average length of the shortest path between a node and all other nodes in the network. Cell centrality analysis is implemented in Python using Python-igraph (v0.9.10).

Cell–cell interaction analysis

The spatial relationship between each cell type is determined by bootstrapping, a previously described approach used to analyze spatial single-cell data (43). For each ROI, the cell type identity is randomly reassigned, while the physical location of each cell type remains unchanged. First, a null distribution is generated for the number of cells A in the neighborhood of cell B. This result is derived from several times of bootstrapping. The null distribution is then compared to the actual distribution to determine the significance. A pseudo *P*-value (two-tailed) is calculated by

$$P_{\text{association}} = \frac{\text{number of } \left(\overline{\text{simulation}} \geq \overline{\text{real}}\right)}{\text{resample times} + 1},$$

$$P_{\text{avoidance}} = \frac{\text{number of } \left(\overline{\text{simulation}} \leq \overline{\text{real}}\right)}{\text{resample times} + 1}.$$

If a *Z*-score is used, the significance is calculated as

$$z = \frac{\text{real} - \overline{\text{simulation}}}{\sigma}.$$

The *Z*-score will then be converted into a *P*-value. An interaction result with a *P*-value <0.05 is considered significant. If two cells are likely to occur in each other's neighborhood, their relationship is considered an association (+1); otherwise, their relationship is classified as avoidance (−1). Users need to define the number of times to perform the bootstrapping (500–2000). The cell–cell interaction analysis is implemented in Rust.

Spatial co-expression analysis

The spatial co-expression analysis is similar to the co-expression analysis described above, except that here Aquila uses the adjacency list from the neighbor network result to construct an expression vector. For example, the expression of marker A is $\{A_1, A_2, A_3, \ldots, A_x\}$ and that of marker B is $\{B_1, B_2, B_3, \ldots, B_x\}$.

Spatial autocorrelation analysis

Spatial autocorrelation analysis calculates the degree of random spatial distribution of cells. If cells tend to locate closely together, they are considered positively spatially correlated. Moran's I (44) measures global spatial autocorrelation, while Geary's C (45) measures local spatial autocorrelation. Moran's I is defined as

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2},$$

while Geary's C is defined as

$$C = \frac{(N-1) \sum_i \sum_j w_{ij} (x_i - x_j)^2}{2W \sum_i (x_i - \bar{x})^2}.$$

N is defined as the number of spatial units indexed by i and j, while w_{ij} is the spatial weight matrix derived from the neighbor network analysis and W is the sum of the matrix. The spatial autocorrelation analysis is implemented in Rust.

Structure and implementations

The Aquila web user interface (UI) was generated using Next.js, a framework based on React. The web UI was designed using MUI, following the Material Design Guidelines. The visualization was created using Echarts.js with WebGL support to enable GPU acceleration. The codes for processing user datasets were implemented using pure JavaScript and executed directly in the browser using Web Workers. All user data are stored locally using IndexedDB. The central API server was constructed using the Actix framework and written in Rust. Analyses including cell neighbor network, cell distribution, co-expression analysis, spatial autocorrelation analysis, spatial entropy and cell–cell interactions were also implemented in Rust. The secondary API server was constructed using FastAPI, written in Python and implemented the following analyses: Ripley statistics, spatial community, cell centrality and spatial variable gene. SpatialDE powered the spatial variable gene analysis with GPU support. Currently, we are running Aquila on a server with an NVIDIA GTX750 GPU. Nginx was used as the reverse proxy server to register the two API servers under the same API gateway.

DATABASE CONTENT AND USAGE

Querying and selecting data

To query and select datasets, users must first click on the 'Explore' button on the home page, which will direct them to the 'Dataset' page. This page displays all the datasets as a list of cards. Each card highlights the essential information
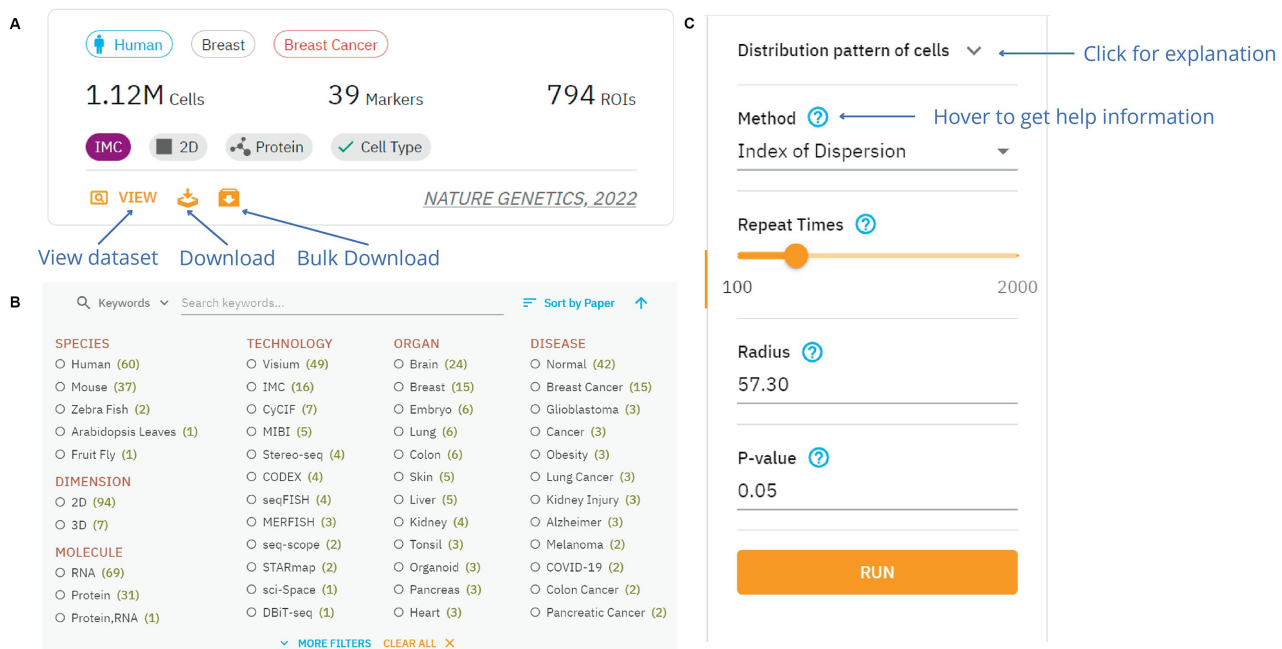
Figure 2. Navigating and using Aquila. (**A**) An example of a dataset card displaying different essential information. (**B**) A panel for filtering and searching datasets with a sort button. (**C**) An example of an analysis panel for selecting methods and inserting parameters to run an analysis.

of a dataset, including species, tissue, disease, technology, molecule, the statistics of datasets and associated publications. The card also contains buttons that users can choose to view the dataset, download the dataset or add the dataset to the download list to download multiple datasets at once (Figure 2A). Aquila has a powerful filtering panel on the same page that helps users look for the dataset(s) they are interested in faster. Users can select dataset(s) based on multiple conditions such as species, technology, tissue or disease (Figure 2B). They can also find datasets by running a fuzzy search with any keywords. Searching datasets by gene names (case-insensitive) is also supported (Figure 2B). After narrowing down their search, users can sort the filtered datasets by publication name, year or the number of cells/ROIs/genes in either an ascending or a descending order. Aquila also supports viewing datasets from the 'Publications' page since each publication may contain multiple datasets. Once users find their dataset of interest, they can explore it in detail by selecting the 'View' button. After entering the details page of data, users can also select to view different ROIs in each dataset through a feature-rich data table. The default will present the first ROI in the dataset.

Visualizing spatial maps

Spatial maps are shown in 2D or 3D, depending on the dataset (Figure 1B). Users can begin by obtaining a general overview of the dataset by choosing one or more ROIs they would like to view in the ROI preview panel. They can then pick from the ROI preview panel one ROI at a time that they would like to acquire additional details. Once users have selected an ROI, they can view different contents of the ROI by toggling between 'Cell Map', 'Expression Map' and 'Co-localization of Markers'. 'Cell Map' shows users the spatial

distribution of the different cell types in the ROI. 'Cell Map' also contains general information about the ROI, including the dimension (minimum bounding box) and the number of cells. As for 'Expression Map', this option lets users look at the spatial expression of individual markers. Users can select or search for their markers of interest using the dropdown menu. Multiple expression maps can be viewed at once, so users can compare the spatial expression patterns of the different markers. Users can click on an ROI to obtain a larger view. Quantification of the expression distribution for the selected marker is also provided under 'Expression Map'. Finally, by selecting 'Co-localization of Markers', users can view the expression of multiple markers in a cell, similar to immunofluorescent and fluorescent *in situ* hybridization analyses. Users can select up to two to five markers, assigning each marker a different color to see whether their expression is co-localized. All spatial maps allow users to adjust the point size and canvas size and zoom in and out to check tissue structures at different resolutions. All visualizations displayed in Aquila can be downloaded as PNGs.

Performing spatial analysis

Users can perform a comprehensive spatial analysis of any selected ROI within Aquila. Currently, 14 types of analyses are available. A concise explanation of each analysis can be found by clicking on the down arrowhead icon to the right of each analysis title. Users can choose to use the general default parameters supplied for every analysis, or they can input their own values. Users can also hover over the question mark at the end of each parameter to learn its functionalities in the popup card (Figure 2C). Many analyses come with more than one method, which allows users to try out different methods to achieve optimal results. Anal-

yses will start immediately after selecting the 'Run' button, and users can visualize the results as soon as they are completed. Some spatial analyses, as described above, rely on the results from the neighboring network analysis to work and, by default, are not accessible to users. To unlock and run these other spatial analyses, users will need to perform the 'neighboring network' analysis, which will then embed the results into these analyses. Other analyses require cell type information to run, and thus they are not available for datasets that do not provide this information. The details of each analysis and how to choose parameters are found in the 'Materials and Methods' section.

Analyzing user data

Users can use Aquila to analyze their data by uploading them on the 'Analysis' page. Users must upload three files containing ROI annotation, cell/spot coordination and gene expression values. After these files are processed, users can then view and analyze their data in Aquila with the exact layout and functionalities as the public datasets. To protect sensitive private information from patient samples or unpublished research data from possible leaks, Aquila processes the data and stores the results locally.

CONCLUSION AND FUTURE DEVELOPMENT

In this work, we introduce Aquila, a database and analysis package for spatial omics data. Aquila provides a 2D and 3D spatial omics data visualization platform and allows users to analyze existing datasets in the database. Moreover, Aquila also supports visualizing and analyzing user data. Overall, Aquila provides users with results in a detailed and comprehensive profile from a spatial perspective.

Many spatial omics datasets provide tissue images such as H&E staining and expression information. Combining expression information with images helps users better understand the datasets and gain deeper insights into their findings. However, Aquila currently does not support tissue images, mainly due to the vast storage size needed to archive them and the resulting performance degradation in the backend database. We will seek better technical solutions in the future to incorporate tissue images. Finally, we will continuously update Aquila with new datasets and integrate additional analyses to provide a richer understanding of the data.

DATA AVAILABILITY

Aquila is freely accessible to anyone with no registration requirements. The database can be accessed and all the datasets can be downloaded via https://aquila.cheunglab.org. For usage and API documentation, please refer to https://aquila.cheunglab.org/about. All preprocessing and source codes for deploying the database are found at https://github.com/Mr-Milk/Aquila-next.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. HuBMAP Consortium (2019) The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature*, **574**, 187.
2. Chen,X., Teichmann,S.A. and Meyer,K.B. (2018) From tissues to cell types and back: single-cell gene expression analysis of tissue architecture. *Annu. Rev. Biomed. Data Sci.*, **1**, 29–51.
3. Moses,L. and Pachter,L. (2022) Museum of spatial transcriptomics. *Nat. Methods*, **19**, 534–546.
4. Stickels,R.R., Murray,E., Kumar,P., Li,J., Marshall,J.L., Di Bella,D.J., Arlotta,P., Macosko,E.Z. and Chen,F. (2021) Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.*, **39**, 313–319.
5. Liu,Y., Yang,M., Deng,Y., Su,G., Enninful,A., Guo,C.C., Tebaldi,T., Zhang,D., Kim,D. and Bai,Z. (2020) High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell*, **183**, 1665–1681.
6. Ståhl,P.L., Salmén,F., Vickovic,S., Lundmark,A., Navarro,J.F., Magnusson,J., Giacomello,S., Asp,M., Westholm,J.O. and Huss,M. (2016) Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, **353**, 78–82.
7. Chen,A., Liao,S., Cheng,M., Ma,K., Wu,L., Lai,Y., Qiu,X., Yang,J., Xu,J. and Hao,S. (2022) Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell*, **185**, 1777–1792.
8. Cho,C.-S., Xi,J., Si,Y., Park,S.-R., Hsu,J.-E., Kim,M., Jun,G., Kang,H.M. and Lee,J.H. (2021) Microscopic examination of spatial transcriptome using Seq-Scope. *Cell*, **184**, 3559–3572.
9. Srivatsan,S.R., Regier,M.C., Barkan,E., Franks,J.M., Packer,J.S., Grosjean,P., Duran,M., Saxton,S., Ladd,J.J. and Spielmann,M. (2021) Embryo-scale, single-cell spatial transcriptomics. *Science*, **373**, 111–117.
10. Eng,C.-H.L., Lawson,M., Zhu,Q., Dries,R., Koulena,N., Takei,Y., Yun,J., Cronin,C., Karp,C. and Yuan,G.-C. (2019) Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*, **568**, 235–239.
11. Xia,C., Fan,J., Emanuel,G., Hao,J. and Zhuang,X. (2019) Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl Acad. Sci. U.S.A.*, **116**, 19490–19499.
12. Wang,X., Allen,W.E., Wright,M.A., Sylwestrak,E.L., Samusik,N., Vesuna,S., Evans,K., Liu,C., Ramakrishnan,C. and Liu,J. (2018) Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*, **361**, eaat5691.
13. Lewis,S.M., Asselin-Labat,M.-L., Nguyen,Q., Berthelet,J., Tan,X., Wimmer,V.C., Merino,D., Rogers,K.L. and Naik,S.H. (2021) Spatial omics and multiplexed imaging to explore cancer biology. *Nat. Methods*, **18**, 997–1012.
14. Lin,J.-R., Izar,B., Wang,S., Yapp,C., Mei,S., Shah,P.M., Santagata,S. and Sorger,P.K. (2018) Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife*, **7**, e31657.
15. Goltsev,Y., Samusik,N., Kennedy-Darling,J., Bhate,S., Hale,M., Vazquez,G., Black,S. and Nolan,G.P. (2018) Deep profiling of mouse

splenic architecture with CODEX multiplexed imaging. *Cell*, **174**, 968–981.

16. Giesen,C., Wang,H.A., Schapiro,D., Zivanovic,N., Jacobs,A., Hattendorf,B., Schüffler,P.J., Grolimund,D., Buhmann,J.M. and Brandt,S. (2014) Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods*, **11**, 417–422.

17. Keren,L., Bosse,M., Thompson,S., Risom,T., Vijayaragavan,K., McCaffrey,E., Marquez,D., Angoshtari,R., Greenwald,N.F. and Fienberg,H. (2019) MIBI-TOF: a multiplexed imaging platform relates cellular phenotypes and tissue structure. *Sci. Adv.*, **5**, eaax5851.

18. Wang,M., Hu,Q., Lv,T., Wang,Y., Lan,Q., Xiang,R., Tu,Z., Wei,Y., Han,K. and Shi,C. (2022) High-resolution 3D spatiotemporal transcriptomic maps of developing *Drosophila* embryos and larvae. *Dev. Cell*, **57**, 1271–1283.

19. Liu,C., Li,R., Li,Y., Lin,X., Zhao,K., Liu,Q., Wang,S., Yang,X., Shi,X. and Ma,Y. (2022) Spatiotemporal mapping of gene expression landscapes and developmental trajectories during zebrafish embryogenesis. *Dev. Cell*, **57**, 1284–1298.

20. Ji,A.L., Rubin,A.J., Thrane,K., Jiang,S., Reynolds,D.L., Meyers,R.M., Guo,M.G., George,B.M., Mollbrink,A., Bergenstråhle,J. *et al.* (2020) Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell*, **182**, 497–514.

21. Rendeiro,A.F., Ravichandran,H., Bram,Y., Chandar,V., Kim,J., Meydan,C., Park,J., Foox,J., Hether,T. and Warren,S. (2021) The spatial landscape of lung pathology during COVID-19 progression. *Nature*, **593**, 564–569.

22. Fan,Z., Chen,R. and Chen,X. (2020) SpatialDB: a database for spatially resolved transcriptomes. *Nucleic Acids Res.*, **48**, D233–D237.

23. Dries,R., Zhu,Q., Dong,R., Eng,C.-H.L., Li,H., Liu,K., Fu,Y., Zhao,T., Sarkar,A. and Bao,F. (2021) Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.*, **22**, 1–31.

24. Palla,G., Spitzer,H., Klein,M., Fischer,D., Schaar,A.C., Kuemmerle,L.B., Rybakov,S., Ibarra,I.L., Holmberg,O. and Virshup,I. (2022) Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods*, **19**, 171–178.

25. Harris,C.R., Millman,K.J., van der Walt,S.J., Gommers,R., Virtanen,P., Cournapeau,D., Wieser,E., Taylor,J., Berg,S., Smith,N.J. *et al.* (2020) Array programming with NumPy. *Nature*, **585**, 357–362.

26. McKinney,W. (2011) pandas: a foundational Python library for data analysis and statistics. In: *Proceedings of the Workshop onPython for High Performance and Scientific Computing*. Vol. **14**, pp. 1–9.

27. Wolf,F.A., Angerer,P. and Theis,F.J. (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.*, **19**, 15.

28. Fath,B.D. (2018) *Encyclopedia of Ecology*. Elsevier, Amsterdam, The Netherlands.

29. Morisita,M. (1962) I_σ-index, a measure of dispersion of individuals. *Res. Popul. Ecol.*, **4**, 1–7.

30. Clark,P.J. and Evans,F.C. (1954) Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, **35**, 445–453.

31. Ripley,B.D. (2005) *Spatial Statistics*. John Wiley & Sons, Hoboken, NJ.

32. Rey,S.J. and Anselin,L. (2010) *Handbook of Applied Spatial Analysis*. Springer, Berlin, pp. 175–193.

33. Shannon,C.E. (1948) A mathematical theory of communication. *Bell Syst. Tech. J.*, **27**, 379–423.

34. Leibovici,D.G., Claramunt,C., Le Guyader,D. and Brosset,D. (2014) Local and global spatio-temporal entropy indices based on distance-ratios and co-occurrences distributions. *Int. J. Geogr. Inf. Sci.*, **28**, 1061–1084.

35. Svensson,V., Teichmann,S.A. and Stegle,O. (2018) SpatialDE: identification of spatially variable genes. *Nat. Methods*, **15**, 343–346.

36. Kats,I., Vento-Tormo,R. and Stegle,O. (2021) SpatialDE2: fast and localized variance component analysis of spatial transcriptomics. bioRxiv doi: https://doi.org/10.1101/2021.10.27.466045, 27 October 2021, preprint: not peer reviewed.

37. Jackson,H.W., Fischer,J.R., Zanotelli,V.R., Ali,H.R., Mechera,R., Soysal,S.D., Moch,H., Muenst,S., Varga,Z. and Weber,W.P. (2020) The single-cell pathology landscape of breast cancer. *Nature*, **578**, 615–620.

38. Traag,V.A., Waltman,L. and van Eck,N.J. (2019) From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.*, **9**, 1–12.

39. Rotta,R. and Noack,A. (2011) Multilevel local search algorithms for modularity clustering. *J. Exp. Algorithmics*, **16**, 2.1–2.27.

40. Rosvall,M. and Bergstrom,C.T. (2007) Maps of information flow reveal community structure in complex networks. *PNAS*, **105**, 1118–1123.

41. Blondel,V., Guillaume,J., Lambiotte,R. and Lefebvre,E. (2008) Fast unfolding of community hierarchies in large networks. *J. Stat. Mech.*, **2008**, P10008.

42. West,D.B. (2001) *Introduction to Graph Theory*. Prentice Hall, Upper Saddle River, NJ.

43. Schapiro,D., Jackson,H.W., Raghuraman,S., Fischer,J.R., Zanotelli,V.R., Schulz,D., Giesen,C., Catena,R., Varga,Z. and Bodenmiller,B. (2017) histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat. Methods*, **14**, 873.

44. Moran,P.A. (1950) Notes on continuous stochastic phenomena. *Biometrika*, **37**, 17–23.

45. Geary,R.C. (1954) The contiguity ratio and statistical mapping. *Inc. Stat.*, **5**, 115–146.