

SingPro: a knowledge base providing single-cell proteomic data

Xichen Lian^{1,2,3,†}, Yintao Zhang^{1,†}, Ying Zhou^{1,4,†}, Xiuna Sun¹, Shijie Huang¹, Haibin Dai^{1,*}, Lianyi Han^{2,*} and Feng Zhu^{1,3,*}

¹College of Pharmaceutical Sciences, The Second Affiliated Hospital, Zhejiang University School of Medicine, Zhejiang University, Hangzhou 310058, China

²Greater Bay Area Institute of Precision Medicine (Guangzhou), School of Life Sciences, Fudan University, Shanghai 315211, China

³Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Hangzhou 330110, China

⁴State Key Laboratory for Diagnosis and Treatment of Infectious Disease, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Zhejiang Provincial Key Laboratory for Drug Clinical Research and Evaluation, The First Affiliated Hospital, Zhejiang University, Hangzhou 310000, China

*To whom correspondence should be addressed. Tel: +86 18989466518; Fax: +86 57188208444; Email: zhufeng@zju.edu.cn

Correspondence may also be addressed to Lianyi Han. Email: hanlianyi@ipm-gba.org.cn

Correspondence may also be addressed to Haibin Dai. Email: haibindai@zju.edu.cn

[†]The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

Abstract

Single-cell proteomics (SCP) has emerged as a powerful tool for detecting cellular heterogeneity, offering unprecedented insights into biological mechanisms that are masked in bulk cell populations. With the rapid advancements in AI-based time trajectory analysis and cell subpopulation identification, there exists a pressing need for a database that not only provides SCP raw data but also explicitly describes experimental details and protein expression profiles. However, no such database has been available yet. In this study, a database, entitled ‘SingPro’, specializing in single-cell proteomics was thus developed. It was unique in (a) systematically providing the SCP raw data for both *mass spectrometry*-based and *flow cytometry*-based studies and (b) explicitly describing experimental detail for SCP study and expression profile of any studied protein. Anticipating a robust interest from the research community, this database is poised to become an invaluable repository for OMICS-based biomedical studies. Access to SingPro is unrestricted and does not mandate a login at: <http://idrblab.org/singpro/>.

Graphical abstract



Introduction

Single-cell proteomics (SCP) has emerged as a powerful tool for detecting cellular heterogeneity, offering unprecedented insights into biological mechanisms that are masked in bulk

cell populations (1,2). As shown in Figure 1, two techniques are widely adopted in current SCP study (3): *flow cytometry*-based one (FC-SCP) measuring up to 50 proteins per cell based on the antibodies, which demonstrates remarkable ability to

Received: July 30, 2023. Revised: September 3, 2023. Editorial Decision: September 19, 2023. Accepted: September 25, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

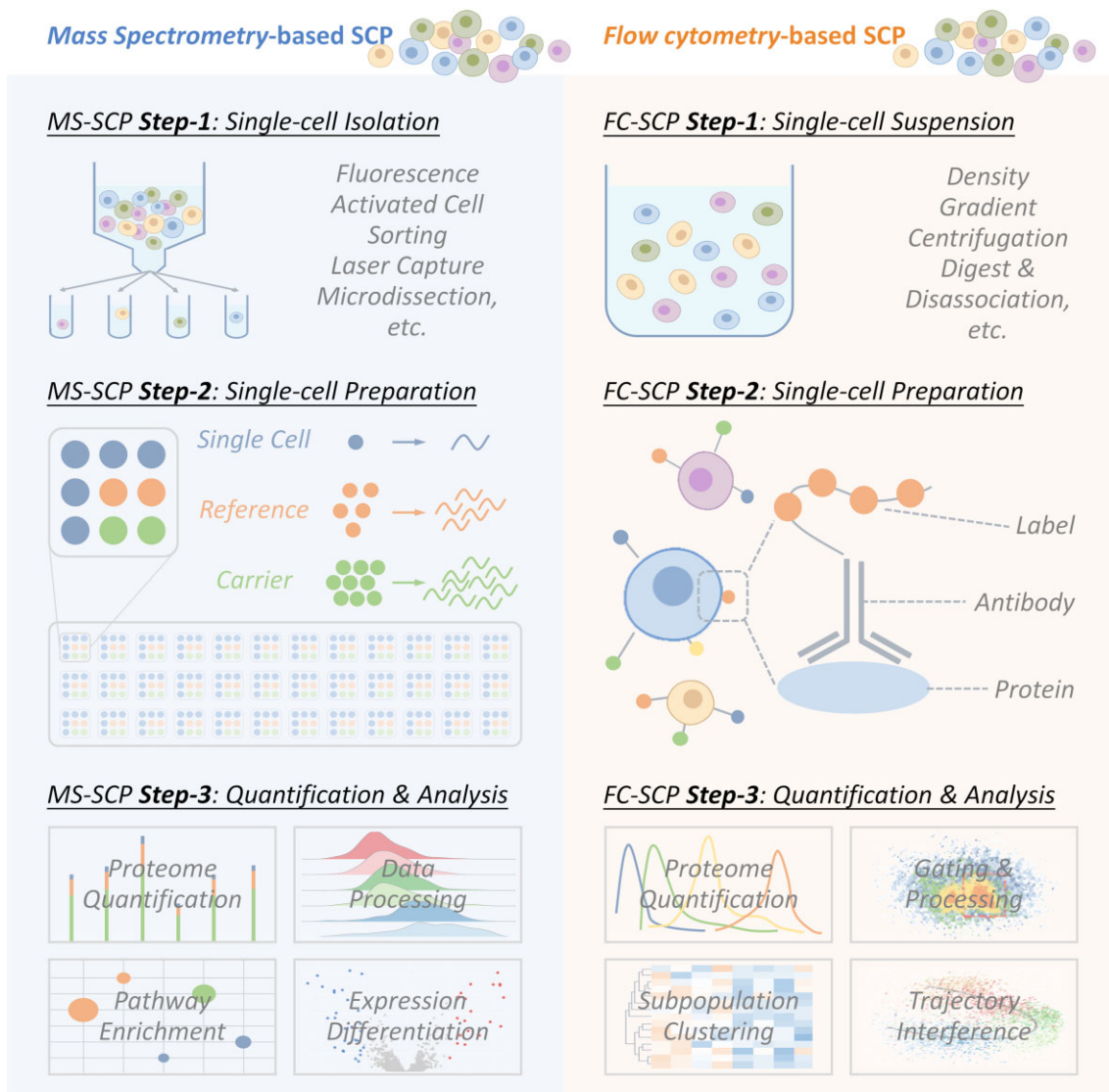


Figure 1. The flowchart of two typical experimental procedures adopted in single-cell proteomic (SCP) analysis, including *mass spectrometry*-based and *flow cytometry*-based SCP analyses. For *mass spectrometry*-based SCP, single cell is (a1) first isolated, (a2) then lysed, digested & labeled and (a3) finally quantified based on MS data & analyzed using pathway enrichments, expression differentiation, and so on. For *flow cytometry*-based SCP, all cells are (b1) first treated into single-cell suspension, (b2) then stained with antibody, and (b3) finally quantified based on FC data & analyzed using cell subpopulation identification, time trajectory interference, and so on.

identify disease-specific cell subpopulation and monitor signal transduction (4–7); *mass spectrometry*-based one (MS-SCP) quantifying over 600 proteins per cell but with limited throughput and relatively lower sensitivity when comparing to FC-SCP, which makes it suitable for identifying new markers and tracking rare cell populations (8–11). Both techniques are powerful and have been frequently adopted to measure the time of delivery (12), uncover the heterogeneity among cells (13,14), realize the high-content drug screening (15), and so on.

However, the extremely-high experimental cost and time-consuming analytical process limit the availability of the publicly accessible SCP data (16–18). An SCP study asks for sophisticated data processing and analysis procedure, and the raw data should be provided to select suitable process (19–21). Meanwhile, it is extremely difficult to conduct SCP-based meta- and multiomic-analysis if the corresponding raw SCP data are unavailable (22–25). For example, the integration of SCP and single-cell transcriptomics (SCT) data is re-

garded as revolutionary for the understandings of biological characteristics/dynamics (26,27), but it is greatly hampered by the unequal amount of raw data between SCP and SCT (28). With the rapid advancements in AI-based time trajectory analysis and cell subpopulation identification, there exists a pressing need for a database that not only provides SCP raw data but also explicitly describes experimental details and protein expression profiles (29–32).

So far, several proteomics-related databases have been developed (33–42). Some of them provide storage and download of *mass spectrometry*-based bulk proteomic data, such as ProteomeXchange (33), PRIDE (34), and iProX (36); some others are public repositories providing the experimental data generated using *cytometry* technique to facilitate cell sorting and immunophenotyping, such as FlowRepository (39) and ImmPort (40). There are also several R packages that can be used to obtain SCP data, such as scpdata (43). However, none of them are dedicated to provide SCP raw data for either MS-based or FC-based technique. Moreover, the existing databases are

General Information

Project ID

SCP57021

Project Title

Single-cell profiling of myasthenia gravis identifies a pathogenic T cell signature

Description

This project employed high-dimensional single-cell mass and spectral cytometry of blood and thymus samples from MG patients in combination with supervised and unsupervised machine-learning tools to gain insight into the immune dysregulation underlying MG. By creating a comprehensive immune map, two dysregulated ...

Click to Show/Hide

Research Type

Cell subpopulation identification

Comparative study: Cell population differences of peripheral blood mononuclear cells from healthy and myasthenia gravis patients.

Publication

Single-cell profiling of myasthenia gravis identifies a pathogenic T cell signature. Acta Neuropathol. 2021 Jun;141(6):901-915

Data Processing & Analysis Tools

ANPELA

Anpela is an online tool that focusing on cytometry-based single-cell protein quantification. It contain over 1,000 available data processing workflows and well established data analysis ...

Click to Show/Hide

Cytobank

Cytobank is a cloud-based platform that offers machine-learning assisted analysis of high dimensional single-cell data. It allows researchers to annotate, analyze, and share results ...

Click to Show/Hide

FLOWJO

FlowJo is a software package for single-cell flow cytometry analysis which has been wildly applied in immunophenotyping, cell cycle analysis, proliferation, kinetics studies and ...

Click to Show/Hide

Project Files

Data Download Tool for Batch Download Project Files:

SingPro provide the *Data Download Tool* to batch download project files.

You can click [here](#) to download the IDs of this project files and click [here](#) to view the user's guide of the tool.

In Total 4 Files as Follow:

Batch Download

Please Click and Select the File(s) for Batch Download

File Name	File Type	Download File	Download ID
<input checked="" type="checkbox"/> Spectral_flow_thymus_live_ICS_untrans_merged.fcs	FCS	Download	SCP0072_1
<input checked="" type="checkbox"/> Spectral_flow_thymus_live_ICS_metadata.csv	Metadata	Download	SCP0072_2
<input type="checkbox"/> Spectral_flow_thymus_live_surf_untrans_merged.fcs	FCS	Download	SCP0072_3
<input type="checkbox"/> Spectral_flow_thymus_live_surf_metadata.csv	Metadata	Download	SCP0072_4

Figure 2. A typical SingPro page describing the general information of a single-cell proteomics study. The information of each study & dataset is explicitly provided in the upper section, which includes: project ID, project title, description, research type, sample type (single-cell/small-cell-population), reference and external linkage of well-established *data processing & analysis* tools. Project files are established in the following section, including: file type, download linkage (for instant download of individual file), download ID and the corresponding staining panel. The user can select the desired file(s) in the checkbox and click 'Package Download' to activate the batch download. For user who want to batch download the files from different studies, a *data download tool* is also provided for enabling the download of multiple files from various studies.

specialized in offering the scientific storage of proteomic data, but *lack of* description on experimental details (such as study procedure, sample label, annotated cell type, and method for single-cell sorting and preparation) and *absent of* application of data processing and analysis, which makes it difficult for researcher, especially for those without a background in bioinformatics, to intuitively use the provided data and comprehend the protein expression profiles. Thus, a database that is specialized in providing SCP raw data and the explicit description on experimental details and expression profiles is urgently needed.

To address this gap, we developed 'SingPro', a database tailored for single-cell proteomics. *First*, a systematic litera-

ture review was conducted, which resulted in a total of 204 studies (129 case-control, 21 multi-class and 54 single-arm studies) containing the SCP raw data of >625 million cells and >16 000 proteins. *Second*, the experimental details (antibody panel, study procedure, sample label, annotated cell type, method for single-cell sorting and preparation, etc.) were manually retrieved and standardized based on the original publications. *Third*, all raw data were processed and analyzed using well-established tools to measure the expression profiles among sample groups for each protein. *Finally*, a user-friendly interface with quick search utility was constructed to facilitate the use of SCP data. All in all, SingPro database is unique in (a) systematically providing the SCP raw data for both *mass*

Downloaded from https://academic.oup.com/nar/article/52/D1/D552/7306671 by Zhejiang University Library user on 12 March 2024

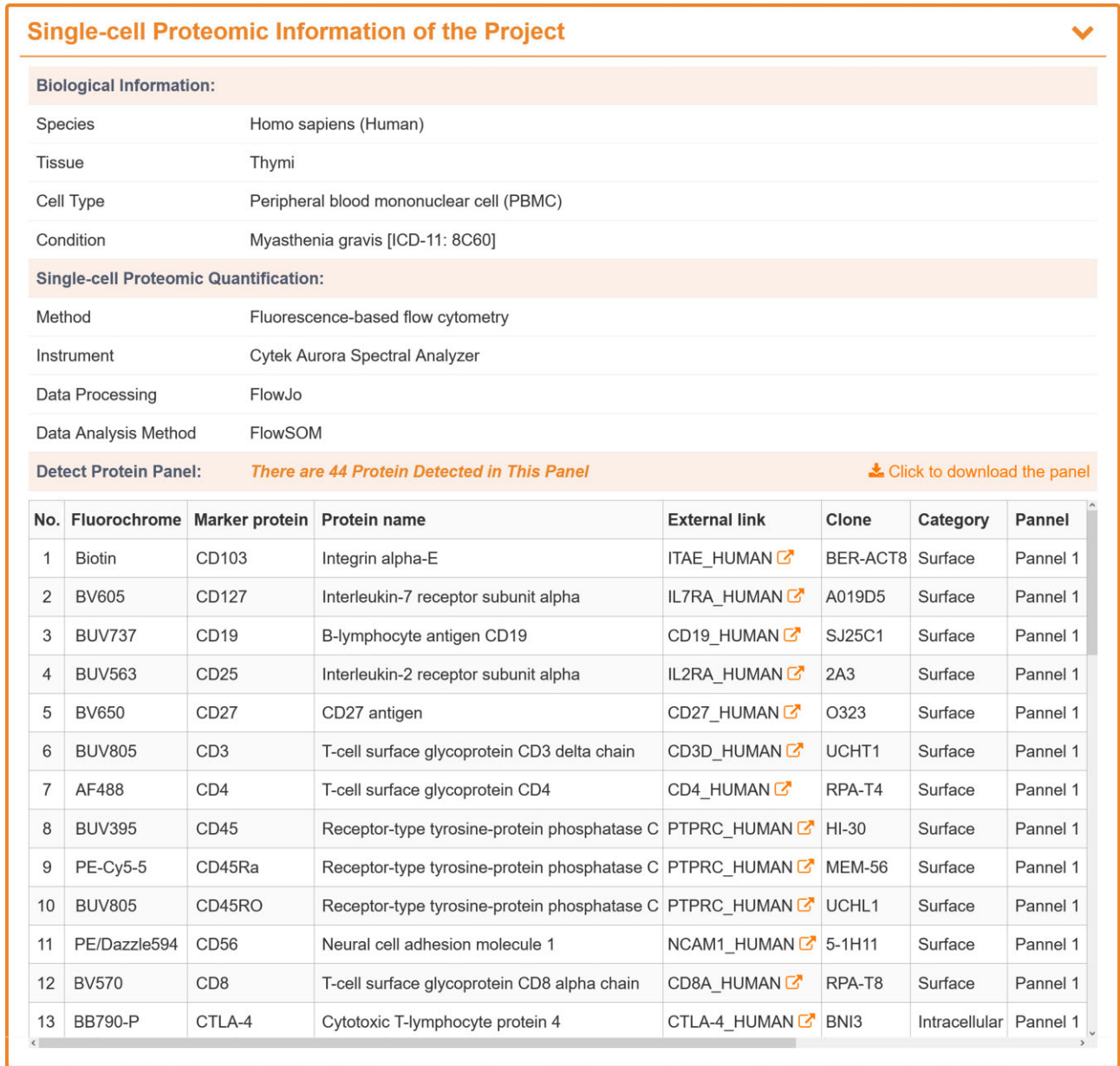


Figure 3. A typical SingPro page describing the quantification process for *flow cytometry*-based SCP. Each page is carefully organized to three sections: *Biological Information* (studied species, experiment tissue/organ, analyzed cell type, pathological/physiological conditions, etc.), *Single-cell Proteomic Quantification* (applied quantification approach, experimental platform, methods for data processing and analysis, etc.), and *Protein Panel* (fluorochrome, protein marker, external link, clone, category (surface/intracellular) and panel number).

spectrometry-based and *flow cytometry*-based studies and (b) explicitly describing the experimental detail of SCP studies and expression profiles of studied proteins. Due to the broad interest from research community, this database is highly expected to be a valuable repository facilitating OMIC-based biomedical studies.

Factual content and data retrieval

Systematic collection of the single-cell proteomic data

The SingPro’s single-cell proteomic data were systematically collated as outlined below. *First*, comprehensive literature review on single-cell proteomic data was conducted by search-

ing PubMed using such keyword combinations as: ‘mass cytometry + proteomics’, ‘flow cytometry + proteomics’, ‘single-cell + proteomics’, ‘single-cell + mass spectrometry’, ‘cytometry time-of-flight’, which resulted in a total of 5780 articles. *Second*, detailed information of each single-cell proteomic datum (such as studied species, disease indications, clinical status & experimental procedure) was systematically retrieved from original publications, and unified & crosslinked to well-established databases. *Finally*, a total of 204 studies (129 case-control, 21 multi-class & 54 single-arm studies) containing the SCP raw data of >625 million cells and >16 000 proteins were collected. As a result, SingPro provided SCP data from human and various model organisms (such as *Mus musculus*, *Xenopus laevis* and *Macaca mulatta*) and tissues/organs (such as peripheral blood, kidney and breast). Additionally, the

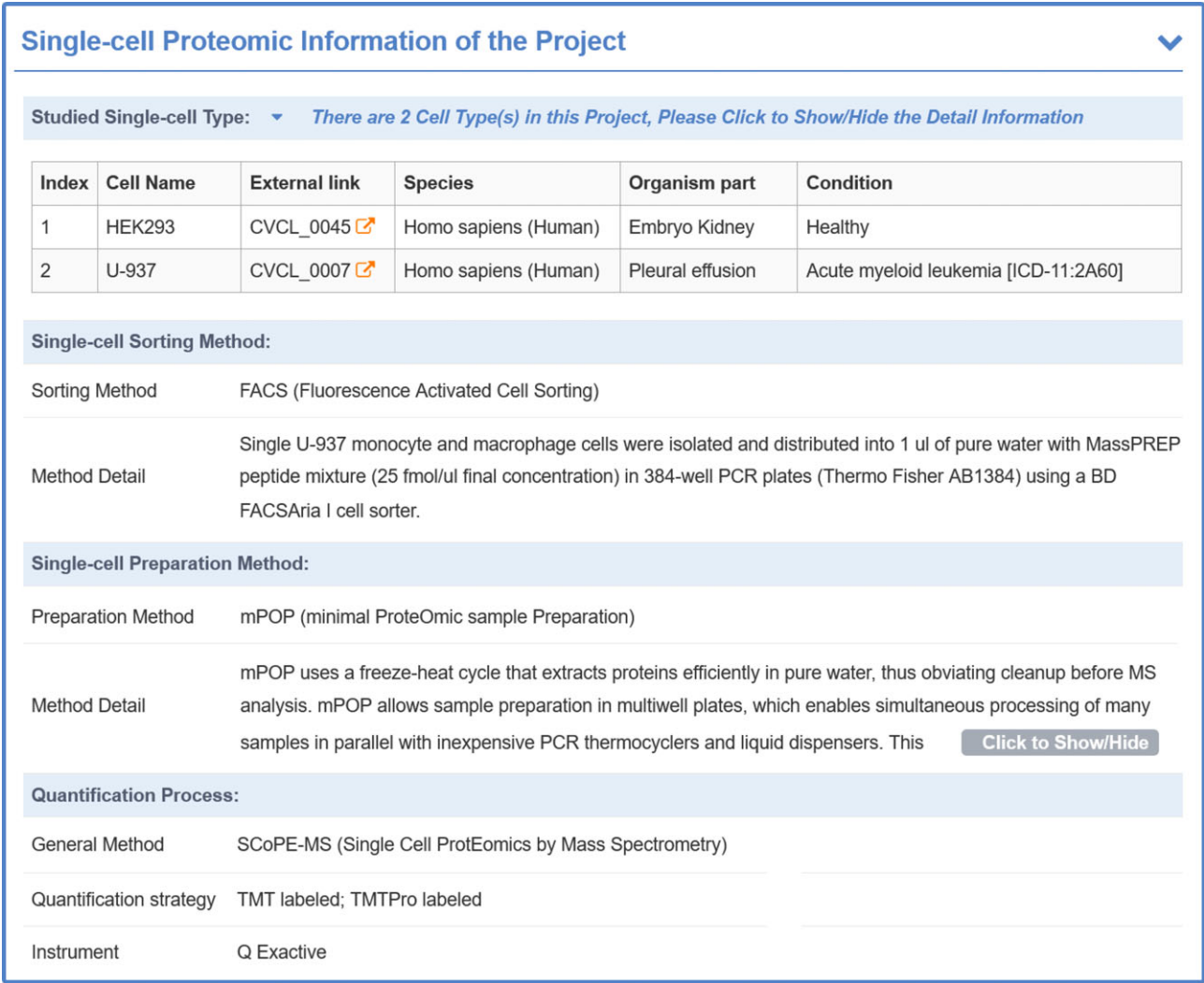


Figure 4. A typical SingPro page describing quantification process of *mass spectrometry*-based SCP. Each page is carefully organized to four sections: *Studied Single-cell Type* (studied species, cells, pathological/physiological condition, etc.), *Sorting Method* (method name & its application detail), *Preparation Method* (method name & its application detail) and *Quantification Process* (applied quantification approach, quantification strategy, experimental platform, etc.).

curated data cover an expansive range of diseases, encompassing not just cancer but also conditions like infections, digestive system ailments, and more.

General information of each SCP dataset in SingPro
For each SCP study, its general information was shown in the upper section of the corresponding SCP webpage, such as: project ID, project title, descriptions, research type, reference links to the original publications, data processing, and analytical tool (as illustrated in Figure 2). Two of the commonly adopted research types in SCP included cell subpopulation identification (which were applied to discover new marker protein (44–46)) and time trajectory interference (which had been adopted to reveal signal pathway and the mechanism underlying disease progression (47,48)). To make it convenient for users to identify the ideal data for their own research purposes and select suitable analytical algorithm, various research types of the collected SCP study were summarized, which were cell population identification, time trajectory interference (with clarified timepoints), comparative study (with de-

scription of different data groups) and novel method (with description of the experimental procedures and equipment innovations). Additionally, SingPro introduces and links to prominent data processing tools like ANPELA (49) and Cytobank (41), streamlining the process for users eager to repurpose the data.

Describing the quantification process of a SCP dataset
For each *flow cytometry*-based SCP study, its biological information, such as species, tissue, cell type and condition of the study were provided in SingPro database. According to the type of antibody, FC-SCP studies can be further divided into two quantification methods: *fluorescence-based flow cytometry* using fluorochrome labels, and *cytometry by time of flight* (CyTOF) using heavy metal isotopes label. For each method, various data processing and analysis tools were developed, such as CATALYST and CytoSpill were compensation tools specially for CyTOF. To enable users to choose subsequent analysis methods appropriate for that data, SingPro

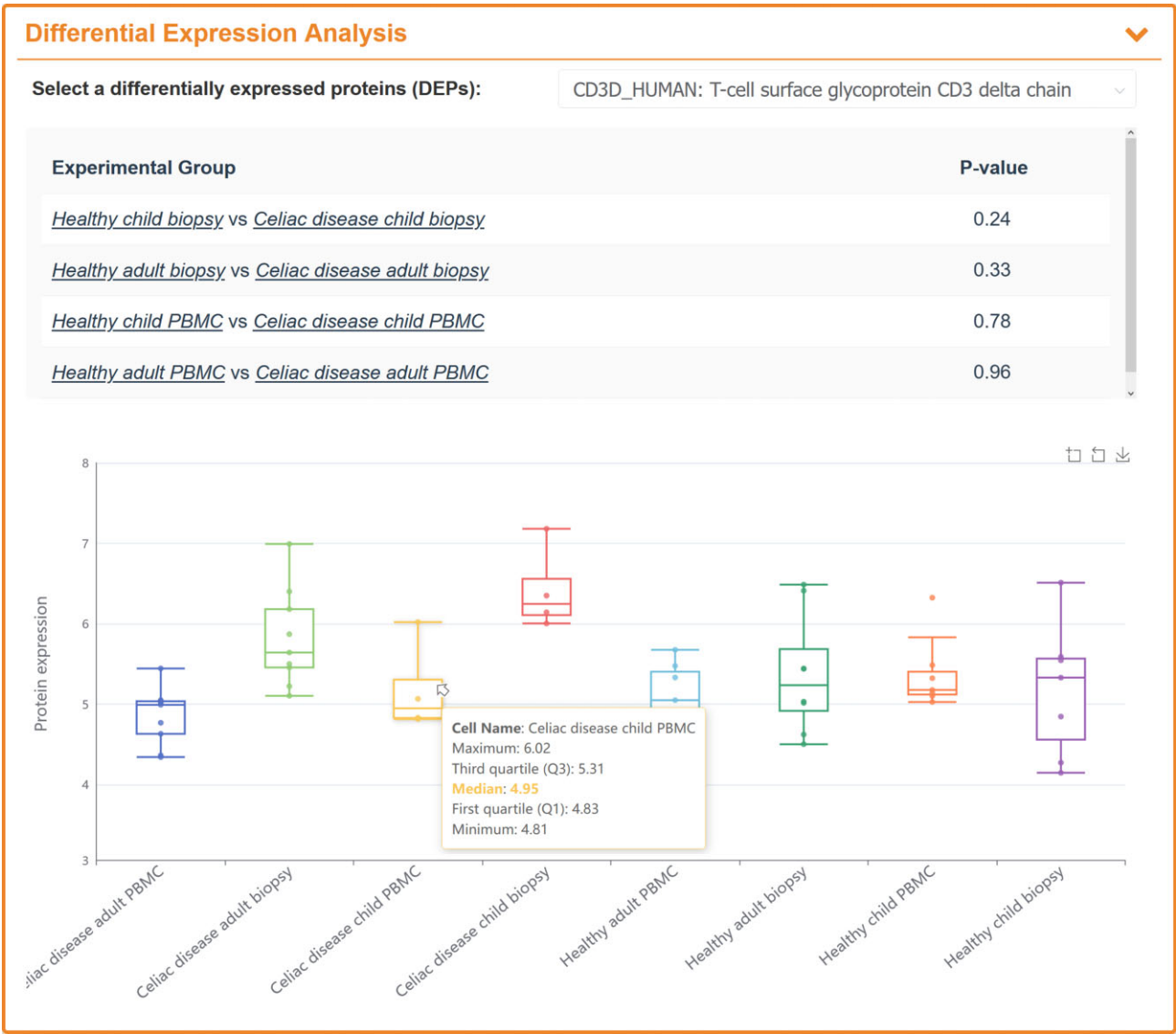


Figure 5. A typical SingPro page describing the expression variations of studied protein among multiple groups using *flow cytometry*-based SCP data. All proteins in staining panel are included into the drop-down-box where a user can select the protein of interest. The *P*-value of the selected protein between two groups is calculated and provided.

provided quantification process description, such as quantification methods, instrument, data processing method and data analysis method adopted in the original publications. The staining panel was also provided which allowed the researchers to directly determine whether the study contained their preferred proteins or whether the desired clustering could be achieved. The staining panel of each study contains information such as protein name, external link to uniprot, fluorochrome/metal isotopes, category (intracellular or surface protein) and clone number (as shown in Figure 3).

For each *mass spectrometry*-based SCP study, the cell type information was explicitly described, including cell line name, species, organism, condition (healthy or specific diseases), and external linkage to other well-established database, such as Cellosaurus (50). One of the major difficulties of the MS-SCP was its miniscule amount of proteins in each cell, proper sorting and subsequent preprocessing methods were essential for preserving the protein from digestion loss and surfaces adsorption (51). Therefore, the single-

cell sorting and preprocessing method of each dataset were manually collected and explicitly described in SingPro, such as CelleONE (52), nanoPOTS (53) and other popular preprocessing platforms. Furthermore, SingPro also described quantification methods used, such as LC-MS/MS (liquid chromatography-mass spectrometry), HPLC-FAIMS-MS/MS (high performance liquid chromatography-field asymmetric ion mobility spectrometry-MS), and CE-ESI-HRMS (capillary electrophoresis-electrospray ionization-high resolution MS), quantification strategy (dimethyl labeled, TMT labeled, label-free, data acquisition, etc.) and the instrument to facilitate the selection of appropriate analytical algorithms (shown in Figure 4).

SCP data processing and protein expression profiles

For *flow cytometry*-based SCP data, all data were imported into FlowJo (54) where the quality control was conducted

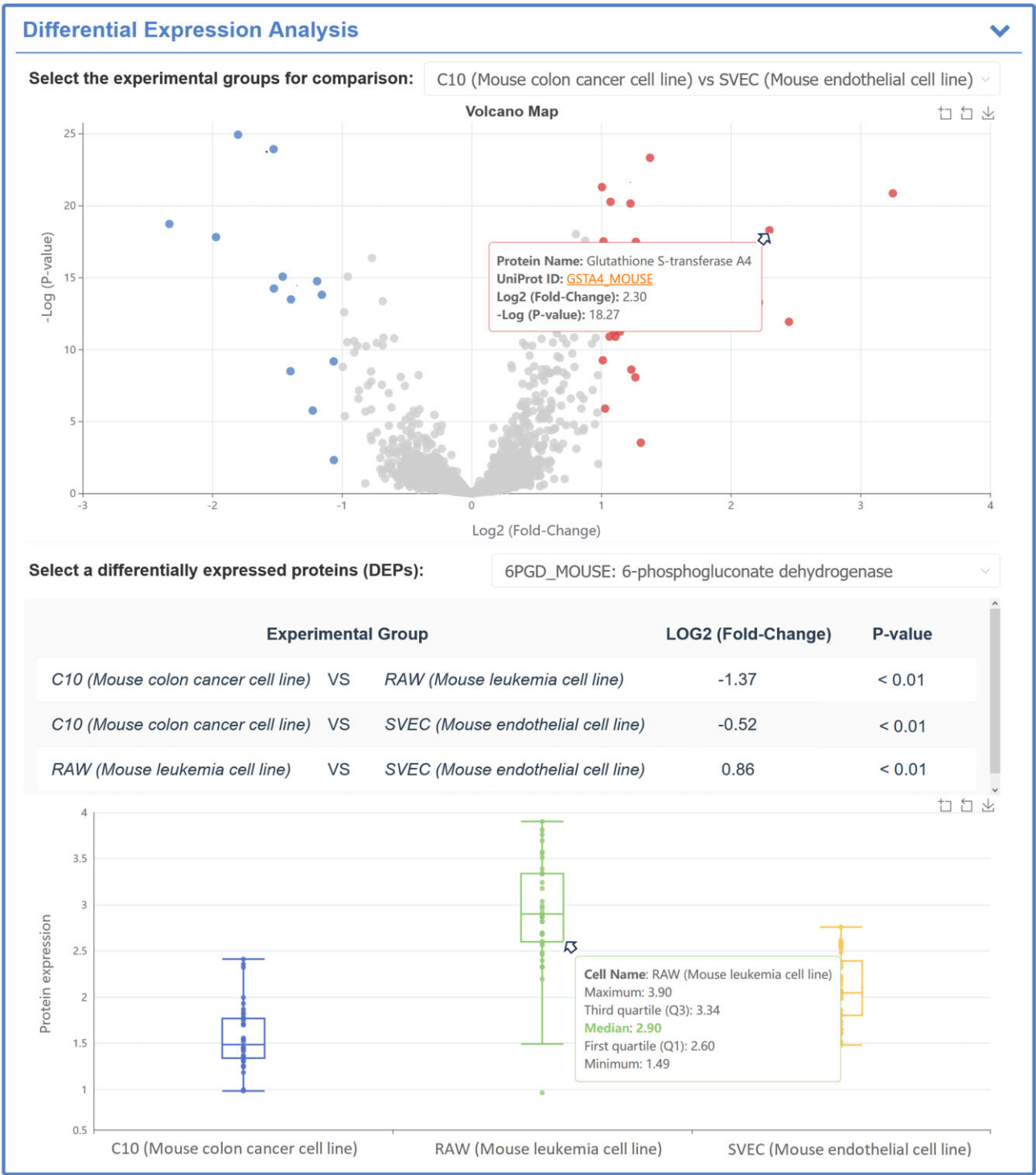


Figure 6. A typical SingPro page describing the expression variations of studied protein among multiple groups using *mass spectrometry*-based SCP data. Particularly, the volcanic map between two groups is calculated to provide the differential expression profiles for proteins (the horizontal coordinate indicates the *log2 fold change* (Log_2FC) and vertical one denotes $\log P$ -value ($\text{Log } P$); the proteins are colored in red and blue based on their Log_2FC & P -value ($\text{Log}_2\text{FC} > 1$ & P -value < 0.05 and $\text{Log}_2\text{FC} \leq -1$ & P -value < 0.05 , respectively). The differentially expressed proteins can be selected, and the P -value of selected protein between two groups is calculated and provided.

using FlowAI (55). After removing the anomalies, data were then manually gated for removing dead cells & other atypical events, and scaled with the arcsine transformation (56–59). The data were grouped according to the original publication, the statistical correlations of protein expression difference among groups were determined using two-way student *t*-test, and *P*-values <0.05 were considered statistically significant. The analytical result was displayed on the page in the form of box diagrams, user can select all the proteins in the dyeing panel through the drop-down box to view the expression level between groups (as shown in Figure 5).

For *mass spectrometry*-based SCP data, the raw data were processed using MaxQuant (version 2.4.0.0) (60). TMT channel, digestion enzymes, missed cleavage, variable modifications and many other parameters were set by referring to the original publication. Both peptide and protein were filtered with false discovery rate <1% to ensure the identification confidence. The corrected reporter ion intensities from MaxQuant were imported into Perseus (61). Reverse and contaminant proteins were filtered out and proteins containing over 70% valid values in each sample were considered. All data were then log-transformed and missing values were imputed based on standard distributions by setting width and downshift to 0.3 and 1.8, respectively (62). Fold changes and two-way student *t*-tests were applied to indicate the significant differences by setting fold change and *P*-value to >2 and <0.05, respectively). Since MS-SCP quantified much more proteins than FC-SCP, and only few of the thousand's proteins detected by MS-SCP were differentially expressed, SingPro provided the volcano maps to show which protein had differential expressed, and then the expression level of those proteins among multiple groups was shown using box maps (illustrated in Figure 6).

Standardization, access and download of the SCP data

To make the access and analysis of SingPro data convenient for the users, all the collected data were carefully cleaned up and then systematically standardized. These standardizations included: (a) all proteins, cell lines, species, and diseases in SingPro were cross-linked to well-established databases such as uniprot (63), Cellosaurus (50) and NCBI Taxonomy (64); (b) all diseases were standardized using the WHO ICD-11 (65). SingPro provided a user-friendly interface that can conveniently browse and search data, and the quick search utility was provided to allow users to find desired single cell proteomic data in main search frame or in a pull-down menu based on experiment accession numbers and the sample parameters, including quantification method, disease indication, species, tissue, marker proteins, etc. All data could be downloaded (including the MaxQuant analysis results, the raw data, and many other related files, such as the protein sequences in FASTA formats, and compensation matrix). Users can download all these data directly from the corresponding page or download and edit the desired file list then using the batch download tool constructed and provided by SingPro database.

Conclusion and prospectives

In this study, a new database, named SingPro, was introduced to provide comprehensive single-cell proteomic (SCP)

data. It was specialized in (a) systematically offering SCP raw data for both *mass spectrometry*- and *flow cytometry*-based studies and (b) explicitly describing experimental details of SCP studies and expression profiles of proteins. With the latest breakthrough of high-sensitivity mass spectrometry techniques, there will be an exponentially increasing amount of single-cell proteomic data. Therefore, SingPro will be updated in a timely fashion. Popular analysis and visualization tools, such as cell subpopulation analysis based on different clustering methods, time trajectory inference and pathway enrichment analysis will be added to keep pace with ongoing research. Due to the broad interest from research community, SingPro was highly expected to be a functional and popular complement to the existing molecular biological databases (63,66–75) in facilitating current OMIC-based studies.

Data availability

All single-cell proteomics data can be viewed, accessed, and downloaded from SingPro, which is freely accessible without any login requirement by all users at: <http://idrblab.org/singpro/>.

Funding

National Natural Science Foundation of China [82373790, 22220102001, U1909208, 81872798]; Natural Science Foundation of Zhejiang Province [LR21H300001]; National Key R&D Program of China [2022YFC3400501]; Leading Talent of the 'Ten Thousand Plan' National High-Level Talents Special Support Plan of China; The Double Top-Class Universities [181201*194232101]; Fundamental Research Funds for Central Universities [2018QNA7023]; Key R&D Program of Zhejiang Province [2020C03010]; Westlake Laboratory (Westlake Laboratory of Life Science & Biomedicine); Alibaba Cloud; Information Technology Center of Zhejiang University; Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare. Funding for open access charge: Natural Science Foundation of Zhejiang Province [LR21H300001].

Conflict of interest statement

None declared.

References

- Gohil,S.H., Iorgulescu,J.B., Braun,D.A., Keskin,D.B. and Livak,K.J. (2021) Applying high-dimensional single-cell technologies to the analysis of cancer immunotherapy. *Nat. Rev. Clin. Oncol.*, **18**, 244–256.
- Davis-Marcisak,E.F., Deshpande,A., Stein-O'Brien,G.L., Ho,W.J., Laheru,D., Jaffee,E.M., Fertig,E.J. and Kagohara,L.T. (2021) From bench to bedside: single-cell analysis for cancer immunotherapy. *Cancer Cell*, **39**, 1062–1080.
- Slavov,N. (2020) Unpicking the proteome in single cells. *Science*, **367**, 512–513.
- Leite Pereira,A., Tchitchek,N., Lambotte,O., Le Grand,R. and Cosma,A. (2019) Characterization of leukocytes from HIV-ART patients using combined cytometric profiles of 72 cell markers. *Front. Immunol.*, **10**, 1777.
- Gonzalez,H., Mei,W., Robles,I., Hagerling,C., Allen,B.M., Hauge Okholm,T.L., Nanjaraj,A., Verbeek,T., Kalavacherla,S., van

- Gogh, M., *et al.* (2022) Cellular architecture of human brain metastases. *Cell*, **185**, 729–745.
6. Kotliar, D., Lin, A.E., Logue, J., Hughes, T.K., Khoury, N.M., Raju, S.S., Wadsworth, M.H. 2nd, Chen, H., Kurtz, J.R., Dighe-Kemp, B., *et al.* (2020) Single-cell profiling of ebola virus disease in vivo reveals viral and host dynamics. *Cell*, **183**, 1383–1401.
 7. Schulte-Schrepping, J., Reusch, N., Paclik, D., Bassler, K., Schlickeiser, S., Zhang, B., Kramer, B., Krammer, T., Brumhard, S., Bonaguro, L., *et al.* (2020) Severe COVID-19 is marked by a dysregulated myeloid cell compartment. *Cell*, **182**, 1419–1440.
 8. Truong, T., Webber, K.G.I., Madisyn Johnston, S., Boekweg, H., Lindgren, C.M., Liang, Y., Nydegger, A., Xie, X., Tsang, T.M., Jayatunge, D., *et al.* (2023) Data-dependent acquisition with precursor coisolation improves proteome coverage and measurement throughput for label-free single-cell proteomics. *Angew. Chem.*, **62**, e202303415.
 9. Mund, A., Brunner, A.D. and Mann, M. (2022) Unbiased spatial proteomics with single-cell resolution in tissues. *Mol. Cell*, **82**, 2335–2349.
 10. Specht, H., Emmott, E., Petelski, A.A., Huffman, R.G., Perlman, D.H., Serra, M., Kharchenko, P., Koller, A. and Slavov, N. (2021) Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol.*, **22**, 50.
 11. Lombard-Banek, C., Moody, S.A., Manzini, M.C. and Nemes, P. (2019) Microsampling capillary electrophoresis mass spectrometry enables single-cell proteomics in complex tissues: developing cell clones in live xenopus laevis and zebrafish embryos. *Anal. Chem.*, **91**, 4797–4805.
 12. Stelzer, I.A., Ghaemi, M.S., Han, X., Ando, K., Hedou, J.J., Feytaerts, D., Peterson, L.S., Rumer, K.K., Tsai, E.S., Ganio, E.A., *et al.* (2021) Integrated trajectories of the maternal metabolome, proteome, and immunome predict labor onset. *Sci. Transl. Med.*, **13**, eabd9898.
 13. Kornej, J., Hanger, V.A., Trinquart, L., Ko, D., Preis, S.R., Benjamin, E.J. and Lin, H. (2021) New biomarkers from multiomics approaches: improving risk prediction of atrial fibrillation. *Cardiovasc. Res.*, **117**, 1632–1644.
 14. Wang, H., Luo, F., Shao, X., Gao, Y., Jiang, N., Jia, C., Li, H. and Chen, R. (2023) Integrated proteomics and single-cell mass cytometry analysis dissects the immune landscape of ankylosing spondylitis. *Anal. Chem.*, **95**, 7702–7714.
 15. Tajik, M., Baharfar, M. and Donald, W.A. (2022) Single-cell mass spectrometry. *Trends Biotechnol.*, **40**, 1374–1392.
 16. Budnik, B., Levy, E., Harmange, G. and Slavov, N. (2018) SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.*, **19**, 161.
 17. Perkel, J.M. (2021) Single-cell proteomics takes centre stage. *Nature*, **597**, 580–582.
 18. Labib, M. and Kelley, S.O. (2020) Single-cell analysis targeting the proteome. *Nat. Rev. Chem.*, **4**, 143–158.
 19. Cranney, C.W. and Meyer, J.G. (2021) CsoDIAq software for direct infusion shotgun proteome analysis. *Anal. Chem.*, **93**, 12312–12319.
 20. Lin, H., Himali, J.J., Satizabal, C.L., Beiser, A.S., Levy, D., Benjamin, E.J., Gonzales, M.M., Ghosh, S., Vasan, R.S., Seshadri, S., *et al.* (2022) Identifying blood biomarkers for dementia using machine learning methods in the framingham heart study. *Cells*, **11**, 1506.
 21. Choi, M., Carver, J., Chiva, C., Tzouros, M., Huang, T., Tsai, T.H., Pullman, B., Bernhardt, O.M., Huttenhain, R., Teo, G.C., *et al.* (2020) MassIVE.quant: a community resource of quantitative mass spectrometry-based proteomics datasets. *Nat. Methods*, **17**, 981–984.
 22. Wang, B., Lunetta, K.L., Dupuis, J., Lubitz, S.A., Trinquart, L., Yao, L., Ellinor, P.T., Benjamin, E.J. and Lin, H. (2020) Integrative omics approach to identifying genes associated with atrial fibrillation. *Circ. Res.*, **126**, 350–360.
 23. Lei, Y., Tang, R., Xu, J., Wang, W., Zhang, B., Liu, J., Yu, X. and Shi, S. (2021) Applications of single-cell sequencing in cancer research: progress and perspectives. *J. Hematol. Oncol.*, **14**, 91.
 24. Chen, K.L., Crane, M.M. and Kaerberlein, M. (2017) Microfluidic technologies for yeast replicative lifespan studies. *Mech. Ageing Dev.*, **161**, 262–269.
 25. Kocher, K., Delot-Vilain, A., Spencer, D., LoTempio, J. and Delot, E.C. (2021) Paucity and disparity of publicly available sex-disaggregated data for the COVID-19 epidemic hamper evidence-based decision-making. *Arch. Sex. Behav.*, **50**, 407–426.
 26. Dickinson, Q., Aufschnaiter, A., Ott, M. and Meyer, J.G. (2022) Multi-omic integration by machine learning (MIMaL). *Bioinformatics*, **38**, 4908–4918.
 27. Gray, G.K., Li, C.M., Rosenbluth, J.M., Selfors, L.M., Girnius, N., Lin, J.R., Schackmann, R.C.J., Goh, W.L., Moore, K., Shapiro, H.K., *et al.* (2022) A human breast atlas integrating single-cell proteomics and transcriptomics. *Dev. Cell*, **57**, 1400–1420.
 28. Lee, J., Hyeon, D.Y. and Hwang, D. (2020) Single-cell multiomics: technologies and data analysis methods. *Exp. Mol. Med.*, **52**, 1428–1442.
 29. Zhang, C., Miao, X., Wang, B., Thomas, R.J., Ribeiro, A.H., Brant, L.C.C., Ribeiro, A.L.P. and Lin, H. (2023) Association of lifestyle with deep learning predicted electrocardiographic age. *Front. Cardiovasc. Med.*, **10**, 1160091.
 30. Hedin, F., Konstantinou, M. and Cosma, A. (2021) Data integration and visualization techniques for post-cytometric analysis of complex datasets. *Cytometry A*, **99**, 930–938.
 31. Schoof, E.M., Furtwangler, B., Uresin, N., Rapin, N., Savickas, S., Gentil, C., Lechman, E., Keller, U.A.D., Dick, J.E. and Porse, B.T. (2021) Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nat. Commun.*, **12**, 3341.
 32. Vistain, L.F. and Tay, S. (2021) Single-cell proteomics. *Trends Biochem. Sci.*, **46**, 661–672.
 33. Deutsch, E.W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J.J., Kundu, D.J., Garcia-Seisdedos, D., Jarnuczak, A.F., Hewapathirana, S., Pullman, B.S., *et al.* (2020) The ProteomeXchange consortium in 2020: enabling big data approaches in proteomics. *Nucleic Acids Res.*, **48**, D1145–D1152.
 34. Perez-Riverol, Y., Bai, J., Bandla, C., Garcia-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., *et al.* (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.*, **50**, D543–D552.
 35. Moriya, Y., Kawano, S., Okuda, S., Watanabe, Y., Matsumoto, M., Takami, T., Kobayashi, D., Yamanouchi, Y., Araki, N., Yoshizawa, A.C., *et al.* (2019) The jPOST environment: an integrated proteomics data repository and database. *Nucleic Acids Res.*, **47**, D1218–D1224.
 36. Chen, T., Ma, J., Liu, Y., Chen, Z., Xiao, N., Lu, Y., Fu, Y., Yang, C., Li, M., Wu, S., *et al.* (2022) iProX in 2021: connecting proteomics data sharing with big data. *Nucleic Acids Res.*, **50**, D1522–D1527.
 37. Sharma, V., Eckels, J., Schilling, B., Ludwig, C., Jaffe, J.D., MacCoss, M.J. and MacLean, B. (2018) Panorama public: a public repository for quantitative data sets processed in skyline. *Mol. Cell. Proteomics*, **17**, 1239–1244.
 38. Farrah, T., Deutsch, E.W., Kreisberg, R., Sun, Z., Campbell, D.S., Mendoza, L., Kusebauch, U., Brusniak, M.Y., Huttenhain, R., Schiess, R., *et al.* (2012) PASSEL: the peptide atlas srm experiment library. *Proteomics*, **12**, 1170–1175.
 39. Spidlen, J., Breuer, K., Rosenberg, C., Kotecha, N. and Brinkman, R.R. (2012) FlowRepository: a resource of annotated flow cytometry datasets associated with peer-reviewed publications. *Cytometry A*, **81**, 727–731.
 40. Bhattacharya, S., Dunn, P., Thomas, C.G., Smith, B., Schaefer, H., Chen, J., Hu, Z., Zalocusky, K.A., Shankar, R.D., Shen-Orr, S.S., *et al.* (2018) ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci. Data*, **5**, 180015.

41. Chen,T.J. and Kotecha,N. (2014) CytoBank: providing an analytics platform for community cytometry data analysis and collaboration. *Curr. Top. Microbiol. Immunol.*, **377**, 127–157.
42. Yang,M., Derbyshire,M.K., Yamashita,R.A. and Marchler-Bauer,A. (2020) NCBI's conserved domain database and tools for protein domain analysis. *Curr. Protoc. Bioinformatics*, **69**, e90.
43. Vanderkaa,C. and Gatto,L. (2021) Replication of single-cell proteomics data reveals important computational challenges. *Expert Rev. Proteomic*, **18**, 835–843.
44. Geer,L.Y., Lapin,J., Slotta,D.J., Mak,T.D. and Stein,S.E. (2023) AIomics: exploring more of the proteome using mass spectral libraries extended by artificial intelligence. *J. Proteome Res.*, **22**, 2246–2255.
45. Fernandez,D.M., Rahman,A.H., Fernandez,N.F., Chudnovskiy,A., Amir,E.D., Amadori,L., Khan,N.S., Wong,C.K., Shamailova,R., Hill,C.A., *et al.* (2019) Single-cell immune landscape of human atherosclerotic plaques. *Nat. Med.*, **25**, 1576–1588.
46. Lavin,Y., Kobayashi,S., Leader,A., Amir,E.D., Elefant,N., Bigenwald,C., Remark,R., Sweeney,R., Becker,C.D., Levine,J.H., *et al.* (2017) Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. *Cell*, **169**, 750–765.
47. Pali,C.G., Cheng,Q., Gillespie,M.A., Shannon,P., Mazurczyk,M., Napolitani,G., Price,N.D., Ranish,J.A., Morrissey,E., Higgs,D.R., *et al.* (2019) Single-cell proteomics reveal that quantitative changes in co-expressed lineage-specific transcription factors determine cell fate. *Cell Stem Cell*, **24**, 812–820.
48. Mahdessian,D., Cesnik,A.J., Gnann,C., Danielsson,F., Stenstrom,L., Arif,M., Zhang,C., Le,T., Johansson,F., Schutten,R., *et al.* (2021) Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. *Nature*, **590**, 649–654.
49. Zhang,Y., Sun,H., Lian,X., Tang,J. and Zhu,F. (2023) ANPELA: significantly enhanced quantification tool for cytometry-based single-cell proteomics. *Adv. Sci. (Weinh)*, **10**, e2207061.
50. Bairoch,A. (2018) The cellSaurus, a cell-line knowledge resource. *J. Biomol. Tech.*, **29**, 25–38.
51. Zhang,L. and Vertes,A. (2018) Single-cell mass spectrometry approaches to explore cellular heterogeneity. *Angew. Chem.*, **57**, 4466–4477.
52. Croteck,C., Krssakova,G., Stejskal,K., Penninger,J.M., Mendjan,S., Mechtler,K. and Stadlmann,J. (2022) Comparative proteome signatures of trace samples by multiplexed data-independent acquisition. *Mol. Cell. Proteomics*, **21**, 100177.
53. Zhu,Y., Piehowski,P.D., Zhao,R., Chen,J., Shen,Y., Moore,R.J., Shukla,A.K., Petyuk,V.A., Campbell-Thompson,M., Mathews,C.E., *et al.* (2018) Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat. Commun.*, **9**, 882.
54. Team FlowJo (2021) *FlowJo™ Software for Windows, Version 10.8*. Becton, Dickinson and Company, Ashland, OR.
55. Monaco,G., Chen,H., Poidinger,M., Chen,J.M., de Magalhães,J.P. and Larbi,A. (2016) flowAI: automatic and interactive anomaly discerning tools for flow cytometry data. *Bioinformatics*, **32**, 2473–2480.
56. Cosma,A. (2020) The nightmare of a single cell: being a doublet. *Cytometry A*, **97**, 768–771.
57. Kramer,K.J., Wilfong,E.M., Voss,K., Barone,S.M., Shiakolas,A.R., Raju,N., Roe,C.E., Suryadevara,N., Walker,L.M., Wall,S.C., *et al.* (2022) Single-cell profiling of the antigen-specific response to BNT162b2 SARS-CoV-2 RNA vaccine. *Nat. Commun.*, **13**, 3466.
58. De Vargas Roditi,L., Jacobs,A., Rueschoff,J.H., Bankhead,P., Chevrier,S., Jackson,H.W., Hermanns,T., Fankhauser,C.D., Poyet,C., Chun,F., *et al.* (2022) Single-cell proteomics defines the cellular heterogeneity of localized prostate cancer. *Cell Rep. Med.*, **3**, 100604.
59. Zhang,H., Lu,M., Lin,G., Zheng,L., Zhang,W., Xu,Z. and Zhu,F. (2023) SoCube: an innovative end-to-end doublet detection algorithm for analyzing scRNA-seq data. *Briefings Bioinf.*, **24**, bbad104.
60. Tyanova,S., Temu,T. and Cox,J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.*, **11**, 2301–2319.
61. Tyanova,S., Temu,T., Sinitcyn,P., Carlson,A., Hein,M.Y., Geiger,T., Mann,M. and Cox,J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods*, **13**, 731–740.
62. Woo,J., Williams,S.M., Markillie,L.M., Feng,S., Tsai,C.F., Aguilera-Vazquez,V., Sontag,R.L., Moore,R.J., Hu,D., Mehta,H.S., *et al.* (2021) High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip. *Nat. Commun.*, **12**, 6246.
63. UniProt,C. (2023) UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.*, **51**, D523–D531.
64. Federhen,S. (2015) Type material in the NCBI Taxonomy Database. *Nucleic Acids Res.*, **43**, D1086–D1098.
65. Lancet,T. (2019) ICD-11. *Lancet*, **393**, 2275.
66. Kim,S., Chen,J., Cheng,T., Gindulyte,A., He,J., He,S., Li,Q., Shoemaker,B.A., Thiessen,P.A., Yu,B., *et al.* (2023) PubChem 2023 update. *Nucleic Acids Res.*, **51**, D1373–D1380.
67. Kanehisa,M., Furumichi,M., Sato,Y., Kawashima,M. and Ishiguro-Watanabe,M. (2023) KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.*, **51**, D587–D592.
68. Wang,Y., Zhang,S., Li,F., Zhou,Y., Zhang,Y., Wang,Z., Zhang,R., Zhu,J., Ren,Y., Tan,Y., *et al.* (2020) Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.*, **48**, D1031–D1041.
69. Li,F., Yin,J., Lu,M., Mou,M., Li,Z., Zeng,Z., Tan,Y., Wang,S., Chu,X., Dai,H., *et al.* (2023) DrugMAP: molecular atlas and pharma-information of all drugs. *Nucleic Acids Res.*, **51**, D1288–D1299.
70. Lu,S., Wang,J., Chitsaz,F., Derbyshire,M.K., Geer,R.C., Gonzales,N.R., Gwadz,M., Hurwitz,D.I., Marchler,G.H., Song,J.S., *et al.* (2020) CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.*, **48**, D265–D268.
71. Du,H., Gao,J., Weng,G., Ding,J., Chai,X., Pang,J., Kang,Y., Li,D., Cao,D. and Hou,T. (2021) CovalentInDB: a comprehensive database facilitating the discovery of covalent inhibitors. *Nucleic Acids Res.*, **49**, D1122–D1129.
72. Sun,X., Zhang,Y., Li,H., Zhou,Y., Shi,S., Chen,Z., He,X., Zhang,H., Li,F., Yin,J., *et al.* (2023) DRESIS: the first comprehensive landscape of drug resistance information. *Nucleic Acids Res.*, **51**, D1263–D1275.
73. Fu,T., Li,F., Zhang,Y., Yin,J., Qiu,W., Li,X., Liu,X., Xin,W., Wang,C., Yu,L., *et al.* (2022) VARIDT 2.0: structural variability of drug transporter. *Nucleic Acids Res.*, **50**, D1417–D1431.
74. Harding,S.D., Armstrong,J.F., Faccenda,E., Southan,C., Alexander,S.P.H., Davenport,A.P., Pawson,A.J., Spedding,M., Davies,J.A. and Nc,I. (2022) The IUPHAR/BPS guide to PHARMACOLOGY in 2022: curating pharmacology for COVID-19, malaria and antibacterials. *Nucleic Acids Res.*, **50**, D1282–D1294.
75. Li,W., O'Neill,K.R., Haft,D.H., DiCuccio,M., Chetvernin,V., Badretin,A., Coulouris,G., Chitsaz,F., Derbyshire,M.K., Durkin,A.S., *et al.* (2021) RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res.*, **49**, D1020–D1028.