Perspective

π-HuB: the proteomic navigator of the human body

https://doi.org/10.1038/s41586-024-08280-5

Received: 19 October 2023

Accepted: 23 October 2024

Published online: 11 December 2024

Check for updates

Fuchu He[1,2]✉, Ruedi Aebersold[3]✉, Mark S. Baker[4], Xiuwu Bian[5], Xiaochen Bo[6], Daniel W. Chan[7], Cheng Chang[1], Luonan Chen[8], Xiangmei Chen[9], Yu-Ju Chen[10], Heping Cheng[11], Ben C. Collins[12], Fernando Corrales[13], Jürgen Cox[14], Weinan E[15,16], Jennifer E. Van Eyk[17], Jia Fan[18], Pouya Faridi[19,20], Daniel Figeys[21], George Fu Gao[22], Wen Gao[23,24], Zu-Hua Gao[25], Keisuke Goda[26,27,28], Wilson Wen Bin Goh[29], Dongfeng Gu[30], Changjiang Guo[31], Tiannan Guo[32,33,34], Yuezhong He[2], Albert J. R. Heck[35,36], Henning Hermjakob[37], Tony Hunter[38], Narayanan Gopalakrishna Iyer[39], Ying Jiang[1], Connie R. Jimenez[40], Lokesh Joshi[41], Neil L. Kelleher[42], Ming Li[43,44], Yang Li[1], Qingsong Lin[45], Cui Hua Liu[46], Fan Liu[47], Guang-Hui Liu[48], Yansheng Liu[49], Zhihua Liu[50], Teck Yew Low[51], Ben Lu[52], Matthias Mann[53], Anming Meng[54], Robert L. Moritz[55], Edouard Nice[56], Guang Ning[57,58,59], Gilbert S. Omenn[60], Christopher M. Overall[61,62], Giuseppe Palmisano[63], Yaojin Peng[64,65,66], Charles Pineau[67], Terence Chuen Wai Poon[68], Anthony W. Purcell[69], Jie Qiao[70], Roger R. Reddel[71], Phillip J. Robinson[71], Paola Roncada[72], Chris Sander[73,74], Jiahao Sha[75], Erwei Song[76,77], Sanjeeva Srivastava[78], Aihua Sun[1], Siu Kwan Sze[79], Chao Tang[80], Liujun Tang[1], Ruijun Tian[81], Juan Antonio Vizcaíno[37], Chanjuan Wang[1], Chen Wang[82,83], Xiaowen Wang[1], Xinxing Wang[31], Yan Wang[1], Tobias Weiss[84], Mathias Wilhelm[85], Robert Winkler[86], Bernd Wollscheid[87], Limsoon Wong[88,89], Linhai Xie[1], Wei Xie[54], Tao Xu[90,91], Tianhao Xu[2], Liying Yan[70], Jing Yang[90], Xiao Yang[1], John Yates[92], Tao Yun[93], Qiwei Zhai[94], Bing Zhang[95,96], Hui Zhang[97], Lihua Zhang[98], Lingqiang Zhang[1], Pingwen Zhang[99,100], Yukui Zhang[98], Yu Zi Zheng[2,25], Qing Zhong[71], Yunping Zhu[1] & The π-HuB Consortium*

The human body contains trillions of cells, classified into specific cell types, with diverse morphologies and functions. In addition, cells of the same type can assume different states within an individual's body during their lifetime. Understanding the complexities of the proteome in the context of a human organism and its many potential states is a necessary requirement to understanding human biology, but these complexities can neither be predicted from the genome, nor have they been systematically measurable with available technologies. Recent advances in proteomic technology and computational sciences now provide opportunities to investigate the intricate biology of the human body at unprecedented resolution and scale. Here we introduce a big-science endeavour called π-HuB (proteomic navigator of the human body). The aim of the π-HuB project is to (1) generate and harness multimodality proteomic datasets to enhance our understanding of human biology; (2) facilitate disease risk assessment and diagnosis; (3) uncover new drug targets; (4) optimize appropriate therapeutic strategies; and (5) enable intelligent healthcare, thereby ushering in a new era of proteomics-driven phronesis medicine. This ambitious mission will be implemented by an international collaborative force of multidisciplinary research teams worldwide across academic, industrial and government sectors.

The Human Genome Project (HGP) has provided a comprehensive map of the human genome and identified the species' approximately 20,300 protein-coding genes[1,2]. This demonstrates the power of data-driven, large-scale coordinated 'omics' projects in transforming biomedical research, giving rise to genomics-driven precision medicine. The human body contains approximately 37 trillion cells of distinctive types, with diverse morphologies and functions, organized in tissues and organs, all of which share essentially the same genome. Moreover, during an individual's lifetime, tissues/organs and cells within their bodies have often undergone extensive or reversible/irreversible changes in response to changing conditions. Experience over the (approximately) 24 years since the publication of the human genome sequence has shown that the observed cellular and organismic complexities cannot be predicted from genomic information alone.

Human biology, with all its intricate complexities, is profoundly interconnected with the vast expanse that is often termed the protein 'universe' or, more scientifically, the proteome (the complete set of proteins expressed by a genome in any cell or tissue at a specific point

A list of affiliations appears at the end of the paper.

in time)[3]. Serving as the fundamental functional elements of cellular mechanisms, proteins are involved in essentially any biological process within an organism. Their significance extends beyond normal physiology: proteins have been crucially implicated as major contributors to the onset and progression of various diseases. They emerge as central figures in the field of therapeutics, being the primary molecular targets for a large majority of drugs. Hence, far beyond the static view provided by genomics, proteomics provides information about the dynamic aspects of the human body and its adaption to changing conditions. Following the HGP era, research based on the human proteome is one of the most exciting, yet challenging, topics in life sciences and medicine[4,5].

In 2001, coinciding with the publication of the human genome sequence, a group of proteomics researchers founded the International Human Proteome Organization (HUPO)[6]. In September 2010, HUPO took the first step towards an international collaborative effort, termed the Human Proteome Project (HPP), with its aim of finding high-quality evidence for the expression of all human protein-coding genes using mass spectrometry, and making them routinely and reliably measurable. Since then, HUPO has stimulated and coordinated many workshops to work within the HPP. Ten years later, in 2020, the HUPO HPP project teams described the first high-stringency HPP proteome map, covering 90.4% of the canonical human proteome and paralleling similar decadal progress made by the HGP[7]. By 2023, detection of 18,397 (93%) of the predicted 19,750 canonical proteins encoded in the human genome had been achieved[8], and a compendium of validated reference spectra for the highly specific targeted mass spectrometric measurement of over 99% of the annotated human proteins had been generated[9]. Since the formation of HUPO, there has been an expansion of biology/disease-centric initiatives under the HPP umbrella. These aim to spatially measure and interpret human proteome data under a range of physiological and pathological conditions, including protein abundance, post-translational modifications (PTMs), interaction partners and localization.

The first proteomics project dedicated to a human organ (the Human Liver Proteome Project) was actually launched several years previously in China in 2003, as a forerunner of the HPP[10–12]. This led to the characterization of liver protein expression profiles and protein–protein interactions (PPIs) in this metabolic organ[13], as well as the discovery of a major functional role of acetylation in metabolic regulation[14,15]. Subsequently, the proteomes of other tissues or organs (for example, brain[16], heart[17], stomach[18], skin[19] and immune cells[20]) have been characterized, creating an initial version of the organ/tissue-based human proteome map[21,22]. Meanwhile, an increasing number of disease-related organ/tissue proteomes have been analysed, as exemplified by the Chinese Human Proteome Project[23], the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium in the USA[24], the Tumor Profiler Project in Switzerland[25], the Human Protein Atlas in Sweden[26,27] and ProCan in Australia[28]. Moreover, recent advances in non-mass spectrometry-based approaches provide versatile opportunities for biomarker discovery in body fluids, which is thought to reflect a person's health or disease status[29–31]. All these efforts are providing a significant boost in advancing the field towards an era previously termed proteomics-driven precision medicine[32]. Despite notable technological and computational advances, we are only just beginning exploration of the complexities of the human proteome, and exploitation of its full potential for biomedical breakthroughs has yet to be fully harnessed (Box 1).

In 2020, the Chinese Ministry of Science and Technology funded a collaboration of around 40 proteomics research teams worldwide to envision future HPP-related projects. Since then, several multidisciplinary working groups have been established and numerous on-site meetings and webinars organized, which have been communicated with government and private funding bodies. These activities have allowed us to propose a 'big-science' project called the proteomic navigator of the human body (π-HuB). The project is now forming a consortium of Chinese and international scientists to generate megaproteomic datasets from all major human tissues/organs and cell types, and to subject the data to integrative analysis at an unprecedented scale. The aim is to build an intelligent computational engine, called the π-HuB navigator, which will integrate multimodality proteomic datasets to enhance our understanding of human biology, to facilitate disease risk assessment and diagnosis, to uncover new drug targets, to optimize appropriate therapeutic strategies and to enable intelligent healthcare.

Box 1

Complexities of the human proteome

First, for a human being at any time, there is immense molecular diversity of proteins in the human proteome at multiple scales, including their level of expression and degradation, their functional state as indicated by PTMs, PPIs and shapes and their cellular and subcellular location. Collectively, proteins and their attributes shape approximately 37 trillion cells in the human body with a wide range of morphology and function. Second, for any human society, the genomic diversity of the human population leads to a larger diversity of proteomes in that population, because each person will have a special private proteome and therefore a special private functional state. Third, during the human lifetime, an individual's proteome is highly dynamic and can be affected by disparate external and internal factors, including somatic mutations, the human microbiome (which can be defined as the microbial ecosystems that reside in various habitats of the body—for example, the human gut), the type of lifestyle (for example, diet, food, nutritional supplements, physical activity and drugs), the occurrence of somatic mutations and the state of the external environment, all of which are intimately related to human health and diseases.

Three central goals of the π-HuB project

The π-HuB project has the overriding mission to support, with an investment of billions of Ren Min Bi (currency of the People's Republic of China), the international and interdisciplinary collaboration of scientists for a period of three decades towards three specific goals (Fig. 1).

Discover principles of the human body

The π-HuB project will first dissect the human body into a hierarchy of digital proteomic anatomy spaces. By harnessing rapidly evolving techniques such as single-cell and spatial proteomics[33,34], the project will digitize, and continually refine, the complete quantitative proteomic and cellular composition of the human body, including cell composition of all major tissues/organs, protein composition of individual cell types and single cells, and proteome-centric molecular networks within cells (for example, PTMs and PPIs). Taking advantage of recent advances in multimodal data fusion/integration technology—in particular, the rapid development of deep-learning or foundational models—the high-resolution, anatomy-based proteomic data will provide an unprecedented opportunity to decode the essential molecular/cellular building principles of cells/tissues/organs, and to uncover the critical molecular/cellular mechanisms of biological processes—that is, to demonstrate causal relations from a protein network to a phenotype.

Develop the Meta Homo Sapiens model

The π-HuB project will conduct in-depth investigations into the dynamics of the human proteome throughout an individual's lifespan, exploring at a population level how the human proteome adapts in response

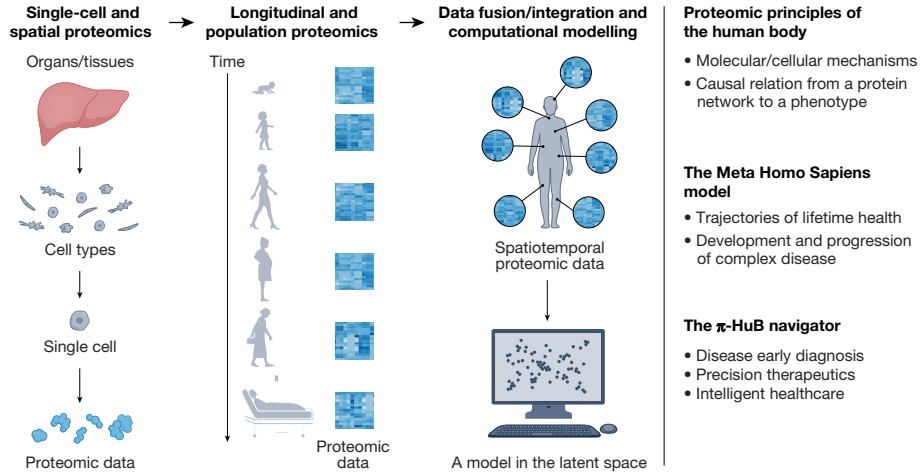

Fig. 1 | Overall goals of the π-HuB project. The schematic depicts the development and integration of major technological strategies for achieving the scientific milestones for each goal. The π-HuB project will start with the extensive measurement of human samples and the generation of data across digital proteomic anatomy spaces and individuals' state spaces in their lifetime. Then, it will incorporate the advantages of the latest advances in data and computational sciences to uncover the composing principles of the human body, to generate a digitized model called Meta Homo Sapiens and to build a global positioning system for the human body and its states.

to various factors impacting health outcomes. The entire human body state space will thus be transformed into multiple subspaces that are further dissected through various dimensions. Specifically, the aims are to (1) trace proteome-centric trajectories during major prenatal and postnatal stages, (2) profile longitudinal dynamic changes in complex proteomes during the development and progression of representatives of complex diseases and (3) determine the effects of non-genetic factors (for example, symbiotic microbiomes, lifestyles and different environments) on the human proteome. These state-contextual proteomic data will be integrated with other human omics data from complementary efforts (for example, HuBMAP[35], Human Cell Atlas[36], Human Tumor Atlas Network[37] and LifeTime Initiative[38]), and projected into a digitized model called Meta Homo Sapiens. Building such a model will be facilitated by composing principles of the human body; it will be formulated using a three-dimensional anatomical hierarchy that records digital features of organs, tissues, body fluids and cells at each level, and it consists of time-sequential frames, with each containing proteomic data measured and augmented within a unit period to represent the human body state at a given timestamp.

Build the π-HuB navigator

The goal of the π-HuB project is to instantiate proteomics-driven phronesis medicine, which is a concept similar to the practical wisdom (*phrónesis*, from ancient Greek) that is relevant to practical action in a particular situation. Unlike traditional and current paradigms of medicine, phronesis medicine aims to develop the ability to provide temporally precise control of the human body state to prevent disease. This ability should include accurate, efficient monitoring, diagnosis and treatment capabilities and highly robust decision-making capabilities for disease prediction, early warning, prevention, control and healthcare. It will provide temporally precise control of the human body state to prevent disease, by establishing a medical model of popularization and normalization of monitoring, diagnosis and treatment decisions and health management. Undertaken with a very keen eye towards realizing this goal, we aim to develop the π-HuB navigator that will be a virtual state space instrument, created by the convergence of physiological phenotypes and proteomic-oriented, spatiotemporal biochemical/biophysical information in cells, body fluids, tissues and organs. It can transfer the prototype Meta Homo Sapiens model from primary body conditions to different secondary states to obtain realistic models. This will be followed by the creation of a state space covering all key states of the human body, by simulating body dynamics with each model for specific periods and thereby tackling the most beneficial approaches to prediction of outcomes based on non-invasive proteomic snapshots and longitudinal proteomic measurements. Ultimately, causal inference will be used to identify underlying triggers that induce transitions between adjacent key states. Each state space can be regarded as a topological navigation map in which each node is the key state defined by the corresponding biomarkers, and each edge between two nodes records the triggers to transform fr om one state to the other. Thus, building such a navigator will grant an opportunity to track trajectories of wellness and health, to define factors important in disease risk assessment and early diagnosis and to drive the development of new therapeutic interventions and intelligent healthcare approaches for redirection of unhealthy transitions towards a long and prosperous life.

Pillars for building the π-HuB navigator

To achieve the above goals, the project is being supported by six key pillars (Fig. 2).

Human biospecimens

Human biospecimens are the fundamental component of the π-HuB project. To achieve the objectives outlined above, samples for π-HuB can be grouped into the following categories. (1) Anatomy-based samples will consist of freshly prepared organs, tissues and live human samples, obtained following the highest ethical standards, from post-mortem examinations. (2) Twin cohorts will allow calculation of the genetic component of observed variability in a population and benefit on controlling confounding factors in aetiological studies on complex diseases. (3) Population-based cohorts will be cross-sectional collections of high-quality biospecimens from a large number of individuals from diverse geographical regions of the world with different lifestyles and subjected to different environments. (4) Longitudinal cohort studies will apply non- or less invasive approaches, with relatively high sampling frequency, to sample healthy individuals or patients with defined exposures who have health or therapeutic implications or outcomes.

In practice, the π-HuB project will first use samples from existing state-of-the-art biobanks around the world, and is open to working closely with other resources being built. All samples in this project will be required to be well and consistently annotated with clinical and

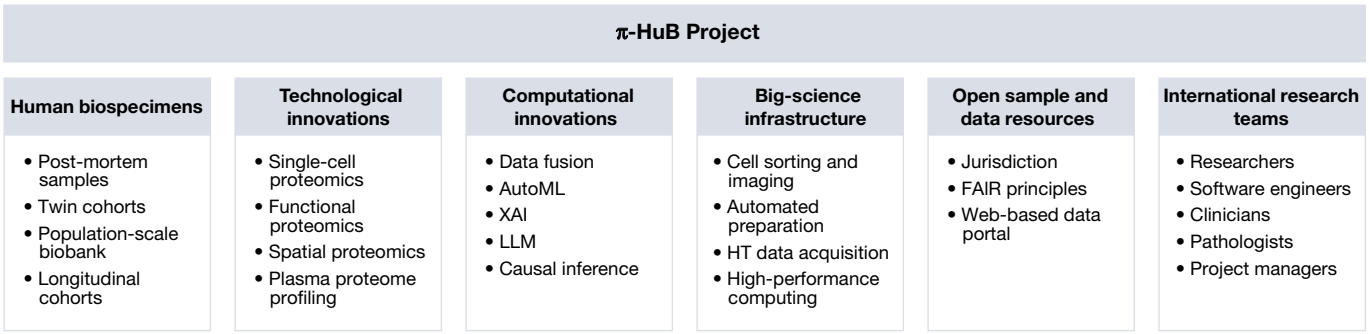

Fig. 2 | Key pillars for implementation of the π-HuB project. To achieve the π-HuB project goals, the project is being supported by six pillars, for which the key components are listed. HT, high throughput.

demographic information obtained from multiple sources, such as questionnaires, physical measurements, biochemical tests, medical imaging data, records of genetic variants implicated in disease susceptibility, and wearable device-based records, among others. Furthermore, annotations should be performed by using agreed metadata standards that are key for data accessibility and interoperability, and for artificial intelligence-based data integration across bioinformatic resources[39].

Innovations in measurement technology

Considering that the project aims to profile the human proteome at unprecedented resolution and scale and that it has an intended time horizon of 30 years, the advancement of measurement technologies is crucial to its success (Box 2). For example, there is a pressing need to identify and measure minute amounts of proteins from single cells. In recent years, we have witnessed tremendous progress in the development of mass spectrometry-based, single-cell proteomic (SCP) technologies[40]. However, a real-world, large-scale application of existing technologies for profiling millions to billions of human cells is still very far from being state of the art. Whereas current mass spectrometry-based SCP technologies are able to measure up to around 4,500 proteins per single cell[41], limited sample throughput due to the absence of multiplexing strategies remains a major shortcoming of SCP analyses. In this regard, the project will start by profiling a cell type-resolved human proteome atlas (see below) and, in parallel π-HuB, will coordinate and support an international community effort to accelerate and benchmark mass spectrometry-based SCP technologies across different platforms and laboratories. Once we can achieve a relatively high analytical performance for SCP across π-HuB data collection centres (for example, over 3,000 proteins per cell at a throughput of roughly 1,000 cells per day), the project will then launch an initiative to collect SCP data from human samples. Meanwhile, we will closely track new concepts and technologies for single-molecule protein sequencing that have substantial potential to enable broad sequence coverage in single-cell profiling[42].

In addition, the π-HuB project will also develop, integrate and apply robust technologies for the generation of multidimensional proteomic data indicating the functional states of the proteome. These technologies are exemplified by PTMs, structural states, localization and interactions, and the context-specific interdependencies of these attributes. To facilitate this, the π-HuB project will develop a specific technology scouting division separate from its own technology hub, seeking to identify and fund emerging technologies.

Because technologies will be evolving rapidly, data acquired in the early stage of this project may become superseded by those collected at a later stage. However, these early data will be invaluable for the development and benchmarking of data analysis and integration tools, for providing training opportunities for researchers in the field, for demonstrating data and sample interoperability procedures within the consortium and for supporting pilot studies. Moreover, we reasoned that, like HGP, the project will also proceed in stages at which 'drafts' of the high-fidelity human proteome atlases will become available at regular intervals, which will be further iterated to more accurate and complete versions using newer technologies.

Computational technology innovations

Beyond data collection, our ambition also extends to developing methods and tools for data integration, analysis and interpretation (Box 2). Data-driven modelling approaches, such as automated machine learning (autoML), have proved powerful in approximating many virtual and real-world systems[43]. However, transforming a biological 'black box' into a digital system does not typically provide us with any intellectual knowledge or insight that would enable it to be trusted in clinical practice. Therefore, the π-HuB project will push the boundaries in biomedicine by unveiling the molecular reconstruction of the human body. Inspired by success in mathematical intuition guidance and hypothesis proposal[44], explainable artificial intelligence (XAI) methods, large language models (LLMs)[45] and other yet-to-be-conceived approaches will be exploited to interpret a fit-for-purpose deep-learning model of the human body with escalating resolutions, from molecular to cellular to organ and systems levels, enabling the discovery of knowledge about biological events and the establishment of construction theories[46]. In this context, it will be key to attract artificial intelligence practitioners to the proteomics field. With this new knowledge, several 'white box' prototype Meta Homo Sapiens models will be constructed to serve as the critical preliminary conditions on which π-HuB will build the foundation model of the system. We believe that the π-HuB project is ideally positioned to meet data science challenges, because of the availability of proteomic datasets acquired by the project consortium that are unique in terms of the size, consistency of collection, annotation and processing, and their coverage of multiple layers of the proteome.

Big-science infrastructure

Because very large numbers of human samples (for example, sorted single cells from human organs and biospecimens from clinical cohorts) will be analysed in the π-HuB project, ultrahigh-throughput facilities for data manufacture, collection and processing will be required. Ideally, such facilities require expertise and streamlined, reproducible pipelines to process human samples, profile proteome-centric molecular data in samples and store, transfer, process and interpret those data. Therefore, the π-HuB project will be establishing national facilities/centres as the big-science infrastructure for the collection and processing of multilayer proteomic data. Minimally, during the first stage of the π-HuB project, such an infrastructure should be able to process 1,000–2,000 samples per day and generate 1 TB of mass spectrometry raw data per day (as of today, although this amount is expected to increase in parallel with developments in technology and instrumentation). In China, only a few existing programs possess such analytical capacities, including automation workstations for 'one-stop'

Box 2

Key technologies for the π-HuB project

- SCP: π-HuB will fully benchmark state-of-the-art SCP methods (for example, nanoPOTS[74], SCoPE-MS[75] and scPiMS[76]) and determine the appropriate time to launch a large-scale initiative to collect SCP data from human samples. To further increase the throughput of SCP analysis, it is also important to encourage engineering-level innovations for fully integrated/automated nanogram-level sample preparation technologies[77] and single-molecule protein-sequencing technologies[42].
- Spatial proteomics: π-HuB will initially apply deep visual proteomics technology[78] or its derivatives to spatially profile proteomes across different cell types. Nonetheless, new concepts and technologies for spatial proteomics are warranted by integration of artificial intelligence-based tissue imaging navigation, high-throughput and pixel-format sampling, multimodal data acquisition and integration[34].
- Plasma proteome profiling: π-HuB will apply mass spectrometry- or affinity-based technologies for plasma proteome profiling, which have been demonstrated to simultaneously analyse thousands of proteins in many thousands of plasma samples with high throughput[79,80].
- Functional proteomics: π-HuB will focus on new chemical biological and biophysical approaches for targeting and enriching native functional states of the human proteome. For example, recent technological advances have enabled the direct detection of subcellular localization, dynamic changes and interactions of proteins in vivo[60,61,81,82].
- AutoML: π-HuB aims to automate the end-to-end process of applying machine learning to the analysis and interpretation of large-scale proteomics data[83,84], which involves tailoring the selection and optimization of machine learning models, facilitating non-expert access to complex multimodal data analysis and addressing challenges on the study of proteomics such as protein identification, quantification and biomarker discovery, early disease diagnosis, optimal therapeutic interventions and the dynamics of biological processes.
- XAI: π-HuB will develop XAI methods that provide clear and understandable explanations of their findings in proteomics analysis[85] and that, in particular, can validate artificial intelligence-driven hypotheses in proteomics, ensuring that artificial intelligence conclusions are scientifically sound and interpretable, enhancing trust and collaboration between computational scientists and experimental biologists by providing transparent decision-making processes. Furthermore, XAI can bridge the gap between artificial intelligence models and practical applications (for example, predictions of disease risk at the individual level and the effect of drugs or drug combinations on the state of cells).
- LLM: π-HuB will build advanced artificial intelligence models trained on extensive biomedical literature, to understand and generate language specific to the proteomics field[45], which are expected to analyse and synthesize vast amounts of biomedical text, extract insights from unstructured data sources such as research papers relevant to proteomics and assist in identifying patterns, potential therapeutic targets and new connections within the complex human proteomics data.

sample preparation, more than 40 cutting-edge, high-resolution mass spectrometers and a high-performance computing system called 'Tianhe-II'[23]. In addition, many other big-science infrastructures and National Laboratories across China have pledged support for the π-HuB project, bringing to the project state-of-the-art single-cell technologies, multimode trans-scale biomedical imaging technologies and a cloud-based, high-performance artificial intelligence computing system. Furthermore, the π-HuB project is partnering with existing infrastructures from research entities attached to universities or other institutions worldwide, such as the Netherlands Proteomics Centre and ProCan in Australia.

Open resources

The π-HuB project will emphasize highly efficient, international, open resources, including standards, samples and their annotations, data and key analysis tools. Like other large community resources with a broad utility, the project will require an open sharing framework to ensure transparent global collaboration between researchers, funding agencies and stakeholders. Within this framework, π-HuB will maximize the importance of reusing collected human samples and reanalysing previously generated data to maximize the benefits from scientific advances, while minimizing risks to participant privacy and acknowledging the contributions of researchers. For example, all π-HuB-generated (non-sensitive) raw data will be directly available to the international scientific community through several well-established data portals implementing the findable, accessible, interoperable and reusable (FAIR) data principles, such as those established by the ProteomeXchange Consortium, including PRIDE and iProX[47,48]. In addition, bioinformatics infrastructure will be developed to integrate proteomics atlases into UniProt, the most popular protein knowledgebase in the world, so that information will be available to the whole life science community. Moreover, the project will enable clinicians and patients to freely enquire about medical intervention strategies by developing a web-based Meta Homo Sapiens computational framework based on π-HuB molecular and spatial data.

International research teams

The sixth pillar is 'people', including researchers, software engineers, clinicians, pathologists, project managers, administrators, financial staff, lawyers, commercial entities and others. Implementation of the π-HuB project requires synergy between a plethora of people globally to work collaboratively under the guidance of a decision-making body, and also clear governance and accountability guidelines. Specifically, the π-HuB project will be steered and governed by an executive committee and overseen by an advisory board. In addition, capacity building and cultural interchange will benefit international researchers in regard to both the exchange of ideas and results and research and intellectual culture. To foster this, π-HuB will establish several scholarship/fellowship programmes to promote these exchanges, through which it will attract additional early-career scientists to participate in this visionary international project.

Challenges

Ethics

Within the framework of the π-HuB project, several measures are proposed to mitigate the identified ethical and regulatory challenges in proteomic research[49]. Especially with regard to human samples, it is essential for the project to establish a common, flexible and generally accepted framework with respect to ethics approval criteria, patient/donor consent, sample annotation ontology material transfer agreements and non-disclosure agreements that can be accepted by

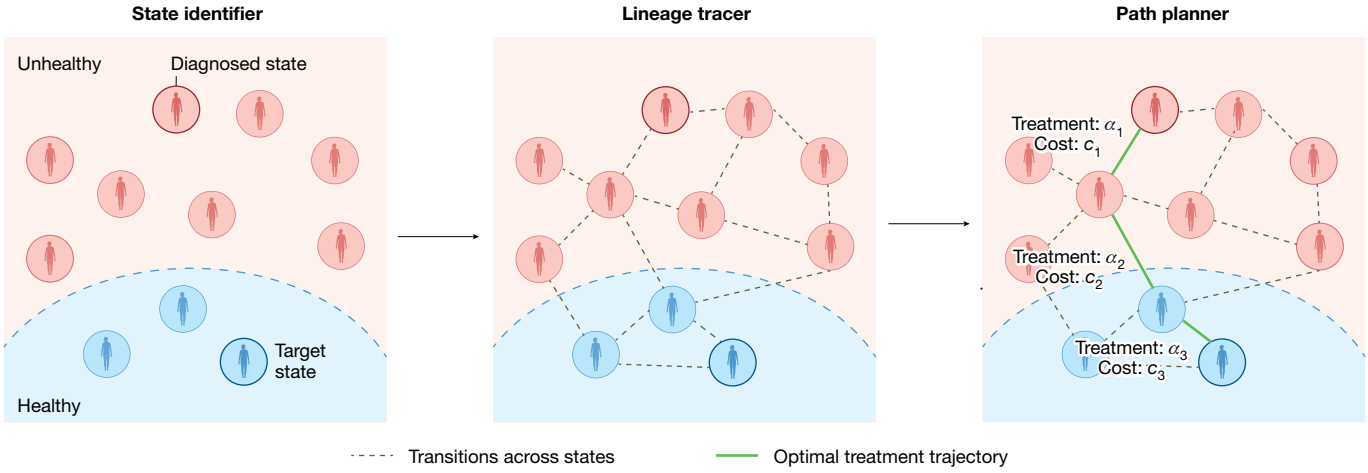

Fig. 3 | Basic modules of the π-HuB navigator. A state identifier generates key states of the human body based on massive hierarchical measurements from large cohorts. This is followed by a dynamic, model-based lineage tracer to detect potential transitions between states under all available treatments. Given the established state space, a path planner can search for a treatment trajectory while considering various objectives and constraints.

governments from different geographical regions. In addition, the risk of reidentifying individuals through their proteomic features demands meticulous assessment and management. To safeguard the ethical integrity and foster societal acceptance of the π-HuB project, we will establish a specialized ethics committee whose primary role will be to oversee every facet of the project, encompassing biospecimen collection, analysis, data management and information dissemination. Adopting this proactive stance towards ethical and regulatory compliance will not only enhance the scientific credibility of the π-HuB project but also strengthen public trust and participation.

Big data

π-HuB will work closely with existing international data centres, such as the ProteomeXchange consortium, for the sake of consistency of data standards and management rules widely adopted in the field of proteomics. Nonetheless, a fit-for-purpose data centre specifically designed for the π-HuB project is needed. First, new upgrades will be required to store and manage more metadata (for example, clinical information) that can meet the data management requirements of the project. Second, it should support multiomics and multimodel data management and applications. Last but not least, it needs to be easily accessible for researchers worldwide. This will be achieved by the establishment of subcentres in different countries that can meet the legal constraints of each country[50], while meeting the requirements of the π-HuB project by using data management system software and technical support from the headquarters centre.

Data generation and integration

The massive amount of data from different modalities will be collected across the international teams. Thus, it is necessary to ensure that data generated across different teams are comparable and integrable. The π-HuB Consortium will first ensure that all human samples are processed using standard operating procedures (SOPs) with respect to collection, annotation, handling, storage and tracking. In addition, to ensure that each data collection team can produce high-quality and unified proteome-centric datasets, the π-HuB project will adopt the HUPO Proteome Standards Initiative[51] principles to standardize both state-of-the-art mass spectrometry-based and non-mass spectrometry-based approaches in terms of sample preparation, methodological settings, data acquisition, processing and error control, and develop the SOPs for each step. Given the anticipated rapidity of methodological development, the consortium will also develop and share standards, test samples and benchmarking data to help each research centre update SOPs for new techniques.

Furthermore, new computational methods and machine learning models with strong generalization ability[52] are needed to further develop proteomic data analysis (for example, quality control[53], data cleansing, normalization[54] and missing-value imputation), and to specifically address questions that can benefit from multimodal and inter-centre/laboratory measurement. As such, we will develop a centralized, cloud-based, interactive platform for data sharing and analysis that will host standardized tools and pipelines for data processing, integration and interpretation. Lastly, the consortium will provide comprehensive training and support to all consortium members to ensure familiarity and compliance with the SOPs and computational tools. These efforts will develop approaches to support the mobilization of π-HuB data, support the discovery of new insights by providing new algorithms and develop new models for unification of multiomics layers.

Modelling

The π-HuB navigator will build on the computationally driven Meta Homo Sapiens model of the human proteome, which is an extremely complex task. In our initial plan, this model will consist of three basic modules (Fig. 3): (1) a state identifier to encode different states of the human body in state space through proteome-centric measurements, followed by integration of phenotypic information of the human body through a multimodality LLM; (2) a lineage tracer to quantify transition probability between each pair of states under different physiological/pathological/therapeutic conditions—for example, between each pair of states that can be estimated by Monte Carlo methods[55]; and (3) a path planner to search for the optimal treatment trajectory by balancing various objectives, such as efficacy, against financial costs and individual compliance.

Democratizing proteomics

Compared with genomics and its related sequencing technologies, the power of proteomics is far from being fully appreciated by the public and, in fact, by many clinicians. Thus, building the interface of the π-HuB project for clinicians and the public is a major aim, in addition to the research goals of the project, to raise public awareness and participation. Such advocacy is needed as a driving force in both sample procurement and addressing the most impactful and pressing needs in disease-focused research. The π-HuB Consortium will also provide training and education of clinicians, pathologists and patients to interpret and use proteomics data, and push proteomics-driven discoveries towards both clinics and healthcare. In particular, the tools hosted in this interface will assist researchers and clinicians in understanding the biological pathways of specific state changes, guiding daily research and clinical practice.

Table 1 | Major outcomes of π-HuB phase 1

Expected outcomes	Principles of cell type-based tissue organization	Proteomics-driven lifestyle guidelines	Proteomics-driven precision medicine
Biospecimen inputs	All major organs from postmortem of healthy donors	Biobanks of natural populaiton	Large-scale international multicentre patient cohorts
Key measurements	Protein expression, subcellular localization, PTMs and PPIs in each cell type	Protein expression and PTMs in body fluids or other non-invasive human samples	Protein expression, PTMs and PPIs in diseased tissues and/or body fluids
Major deliverables	(1) A cell type-resolved, multidimensional human proteome atlas (2) New molecular/cellular mechanisms of biological processes	(1) A resource of human proteome traits associated with lifetime states (2) A proteomic health score	(1) A resource of human proteome traits associated with major diseases (2) New biomarkers and therapeutic targets

Major outcomes π-HuB phase 1

To enable π-HuB to become a broadly applicable project, it is necessary to maximize relevance to the community by setting deliverables and expected outcomes as a series of staged programmes undertaken within a relatively short time frame. During the initiation and development stage (2024–2033, herein referred to as 'phase 1'), we will build an international cooperative network to lay the technical foundation of this project by promoting methodological advances, benchmarking state-of-the-art technologies for standardization, building the computational infrastructures for data integration and modelling and so on. Meanwhile, it is also important for the project, in a relatively short-term frame, to achieve major outcomes as follows (Table 1).

Principles of cell type organization

The π-HuB project will eventually support studies that generate single-cell-resolution atlases of all major human organs and tissues from people who identify as emanating from different ancestral populations/origins. However, during phase 1, we will initially build reference cell type proteome atlases for all major organs using a combination of state-of-the-art flow sorting[56] and parallel mass spectrometry acquisition platforms[57]. Moreover, rapidly evolving spatial proteomics technologies will provide additional insights into secreted proteins in the surrounding microenvironment and subcellular localization of the proteome at the tissue, cellular and molecular levels. In this regard, multiple cutting-edge and synergistic approaches will be implemented, including mass spectrometry-based, multiplex immunoaffinity-based and super-resolution-imaging-based methods[58,59]. In addition, the emerging proximity labelling and in vivo crosslinking approaches will enable the profiling of protein subcellular localization, protein complexes and PPIs in diverse human cell types[60,61]. These analyses will provide versatile opportunities to uncover new molecular/cellular mechanisms of biological processes in shaping diverse cell types and cell states within each organ. Together, we envision that these cell type-resolved, multidimensional proteome atlases, in combination with cutting-edge computational and bioinformatic approaches, will be able to uncover the building principles of cell type organization of major tissues/organs.

Proteomics-driven lifestyle guidelines

During phase 1, we will focus on the most dominant factors that shape/remodel the proteome of healthy individuals. Specifically, we will accumulate a large number of biofluid proteomes from large-scale natural populations, aiming to (1) map quantitative trait loci for circulating/tissue proteins and protein allelic variants associated with genetic variants implicated in disease susceptibility, allowing us to construct disease-causing pathways; (2) trace the proteomic trajectory over the human lifespan by quantifying dynamic changes in the biofluid proteomes of five major prenatal cycles (gametogenesis, fertilization, embryonic development, fetal development and delivery) and five major postnatal cycles (adolescence, puberty, gestation, menopause and old age); (3) analyse the effects of four major dietary nutrition patterns (Western, Japanese, Mediterranean and subsistence)[62] on the human biofluid proteome; (4) map the proteomes of populations in six major ecological environments classified by the Köppen–Geiger map (hot, warm, cold, arid, polar and highland)[63] and analyse the trajectory of the human biofluid proteome during acclimatization and adaptation; (5) map interactions of the human gut and skin proteome with representative microbiomes from internal and external environments, and construct the adaptation trajectory of the human proteome in response to microbiomes; and (6) map the responses of the human proteome to various clinical intervention strategies such as medication, diet and exercise. Together, these analyses will generate a resource of human proteome traits associated with the lifetime states as indicated above. Such a resource will provide opportunities to develop a proteomic health score through a neural network or LLMs, thereby formulating proteomics-driven lifestyle guidelines.

Generalization of proteomics-driven precision medicine

In the past decade, there has been growing evidence that proteomic approaches can facilitate the mechanistic understanding of diseases, as well as facilitate biomarker discovery and optimize therapy development. In particular, proteomics alone has been able to identify potential biomarkers, or potential therapeutic targets, for many tumour types[32,64–67] and a variety of other diseases[68–73]. Despite these advances, most proteomic findings in the context of human diseases have yet to be validated, and treatment suggestions arising from the data have yet to be approved. For example, most potential biomarkers identified by proteomic studies are generated from small-scale, retrospective studies lacking the basis for the subsequent generalization of using such biomarkers in a wider population. The π-HuB Consortium, consisting of multi-interdisciplinary researchers and clinicians, provides an unprecedented opportunity to reimagine biomarker discovery through a proteomic approach. We therefore reason that, under the umbrella of π-HuB, it will be more feasible to organize large-scale, international, multicentre cohort studies for the validation of new biomarkers for early and companion diagnosis of major diseases. As such, we plan to map the proteomes of ten major organs and corresponding biofluids at different pathophysiological stages, focusing on between three and five representative diseases for each related organ. Such analyses, together with the aforementioned tissue proteome atlases with cell type resolution and biofluid proteome atlases with life-oriented adaptive proteome atlases, will allow the construction of proteomic evolutionary trajectories mapping the occurrence and development of these diseases, and the pathways associated with specific life stages and survival conditions. Furthermore, the π-HuB project will actively collaborate with clinicians, policymakers and industrial partners to catalyse the discovery of new protein-based biomarkers and drug targets that can be applied in clinics to diagnose disease and drug development, driving a paradigm shift towards proteomics-driven precision medicine.

Outlook

Since its inception in 2020, the π-HuB Consortium has grown to become an international collaborative force of more than 100 members,

mobilizing scientists worldwide across academic, industrial and government sectors in the protein and health sciences. The π-HuB project will probably foster further global collaboration and discussion by integrating the results from a worldwide community of multidisciplinary scientists to better understand human biology, and to advance medicine from disease trajectory predictions to new treatment options. We anticipate that the π-HuB project will make a major contribution to biomedical research in the coming decades, facilitating disease prevention and diagnosis, accelerating drug discovery and, ultimately, ushering in an era of proteomics-driven phronesis medicine.

1. Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
2. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
3. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).
4. Abbott, A. And now for the proteome. *Nature* **409**, 747 (2001).
5. Fields, S. Proteomics. Proteomics in genomeland. *Science* **291**, 1221–1224 (2001).
6. The proteome isn't genome II. *Nature* **410**, 725 (2001).
7. Adhikari, S. et al. A high-stringency blueprint of the human proteome. *Nat. Commun.* **11**, 5301 (2020).
8. Omenn, G. S. et al. The 2022 Report on the Human Proteome from the HUPO Human Proteome Project. *J. Proteome Res.* **22**, 1024–1042 (2023).
9. Kusebauch, U. et al. Human SRMAtlas: a resource of targeted assays to quantify the complete human proteome. *Cell* **166**, 766–778 (2016).
10. Cyranoski, D. China takes centre stage for liver proteome. *Nature* **425**, 441 (2003).
11. Jia, H. & Louet, S. China pushes liver proteomics. *Nat. Biotechnol.* **22**, 136 (2004).
12. He, F. Human liver proteome project: plan, progress, and perspectives. *Mol. Cell. Proteomics* **4**, 1841–1848 (2005).
13. Wang, J. et al. Toward an understanding of the protein interaction network of the human liver. *Mol. Syst. Biol.* **7**, 536 (2011).
14. Wang, Q. et al. Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. *Science* **327**, 1004–1007 (2010).
15. Zhao, S. et al. Regulation of cellular metabolism by protein lysine acetylation. *Science* **327**, 1000–1004 (2010).
16. Sharma, K. et al. Cell type- and brain region-resolved mouse brain proteome. *Nat. Neurosci.* **18**, 1819–1831 (2015).
17. Doll, S. et al. Region and cell-type resolved quantitative proteomic map of the human heart. *Nat. Commun.* **8**, 1469 (2017).
18. Ni, X. et al. A region-resolved mucosa proteome of the human stomach. *Nat. Commun.* **10**, 39 (2019).
19. Dyring-Andersen, B. et al. Spatially and cell-type resolved quantitative proteomic atlas of healthy human skin. *Nat. Commun.* **11**, 5587 (2020).
20. Rieckmann, J. C. et al. Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat. Immunol.* **18**, 583–593 (2017).
21. Wilhelm, M. et al. Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).
This paper, together with Kim et al. (2014), shows the initial version of the tissue/organ-centric human proteome by applying mass spectrometry-based approaches.
22. Kim, M. S. et al. A draft map of the human proteome. *Nature* **509**, 575–581 (2014).
23. Cyranoski, D. China pushes for the proteome. *Nature* **467**, 380 (2010).
24. Rodriguez, H., Zenklusen, J. C., Staudt, L. M., Doroshow, J. H. & Lowy, D. R. The next horizon in precision oncology: proteogenomics to inform cancer diagnosis and treatment. *Cell* **184**, 1661–1670 (2021).
25. Irmisch, A. et al. The Tumor Profiler Study: integrated, multi-omic, functional tumor profiling for clinical decision support. *Cancer Cell* **39**, 288–293 (2021).
26. Uhlén, M. et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics.* **4**, 1920–1932 (2005).
27. Uhlén, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
28. Tully, B. et al. Addressing the challenges of high-throughput cancer tissue proteomics for clinical application: ProCan. *Proteomics* **19**, e1900109 (2019).
29. Eldjarn, G. H. et al. Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **622**, 348–358 (2023).
30. Xu, Y. et al. An atlas of genetic scores to predict multi-omic traits. *Nature* **616**, 123–131 (2023).
31. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
32. Jiang, Y. et al. Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* **567**, 257–261 (2019).
This study showcases the concept of proteomics-driven precision medicine.
33. Kelly, R. T. Single-cell proteomics: progress and prospects. *Mol. Cell. Proteomics* **19**, 1739–1748 (2020).
34. Mund, A., Brunner, A. D. & Mann, M. Unbiased spatial proteomics with single-cell resolution in tissues. *Mol. Cell* **82**, 2335–2349 (2022).
A review summarizing state-of-the-art spatial proteomics for human samples.
35. Hu, B. C. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187–192 (2019).
36. Elmentaite, R., Dominguez Conde, C., Yang, L. & Teichmann, S. A. Single-cell atlases: shared and tissue-specific cell types across human organs. *Nat. Rev. Genet.* **23**, 395–410 (2022).
37. Rozenblatt-Rosen, O. et al. The Human Tumor Atlas Network: charting tumor transitions across space and time at single-cell resolution. *Cell* **181**, 236–249 (2020).
38. Rajewsky, N. et al. LifeTime and improving European healthcare through cell-based interceptive medicine. *Nature* **587**, 377–386 (2020).
39. Mann, M., Kumar, C., Zeng, W. F. & Strauss, M. T. Artificial intelligence for proteomics and biomarker discovery. *Cell Syst.* **12**, 759–770 (2021).
40. Perkel, J. M. Single-cell proteomics takes centre stage. *Nature* **597**, 580–582 (2021).
41. Guzman, U. H. et al. Ultra-fast label-free quantification and comprehensive proteome coverage with narrow-window data-independent acquisition. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-023-02099-7 (2024).
42. MacCoss, M. J. et al. Sampling the proteome by emerging single-molecule and mass spectrometry methods. *Nat. Methods* **20**, 339–346 (2023).
A review summarizing next-generation proteomics technologies.
43. Bi, K. et al. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **619**, 533–538 (2023).
44. Davies, A. et al. Advancing mathematics by guiding human intuition with AI. *Nature* **600**, 70–74 (2021).
45. Thirunavukarasu, A. J. et al. Large language models in medicine. *Nat. Med.* **29**, 1930–1940 (2023).
46. Kang, M., Ko, E. & Mersha, T. B. A roadmap for multi-omics data integration using deep learning. *Brief. Bioinform.* **23**, bbab454 (2022).
47. Chen, T. et al. iProX in 2021: connecting proteomics data sharing with big data. *Nucleic Acids Res.* **50**, D1522–D1527 (2022).
48. Deutsch, E. W. et al. The ProteomeXchange consortium at 10 years: 2023 update. *Nucleic Acids Res.* **51**, D1539–D1548 (2023).
49. Fierro-Monti, I., Wright, J. C., Choudhary, J. S. & Vizcaino, J. A. Identifying individuals using proteomics: are we there yet? *Front. Mol. Biosci.* **9**, 1062031 (2022).
50. Bandeira, N., Deutsch, E. W., Kohlbacher, O., Martens, L. & Vizcaino, J. A. Data management of sensitive human proteomics data: current practices, recommendations, and perspectives for the future. *Mol. Cell. Proteomics* **20**, 100071 (2021).
51. Deutsch, E. W. et al. Proteomics standards initiative at twenty years: current activities and future work. *J. Proteome Res.* **22**, 287–301 (2023).
A perspective paper summarizing the 20-year-long community effort for the proteomics community with respect to data formats, quality control and annotation.
52. Sharifi-Noghabi, H., Harjandi, P. A., Zolotareva, O., Collins, C. C. & Ester, M. Out-of-distribution generalization from labelled and unlabelled gene expression data for drug response prediction. *Nat. Mach. Intell.* **3**, 962–972 (2021).
53. Olivella, R. et al. QCloud2: an improved cloud-based quality-control system for mass-spectrometry-based proteomics laboratories. *J. Proteome Res.* **20**, 2010–2013 (2021).
54. Chawade, A., Alexandersson, E. & Levander, F. Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. *J. Proteome Res.* **13**, 3114–3120 (2014).
55. James, F. Monte Carlo theory and practice. *Rep. Prog. Phys.* **43**, 1145 (1980).
56. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630 (2013).
57. Goncalves, E. et al. Pan-cancer proteomic map of 949 human cell lines. *Cancer Cell* **40**, 835–849 (2022).
58. Qiao, C. et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nat. Methods* **18**, 194–202 (2021).
59. Qiao, C. et al. Rationalized deep learning super-resolution microscopy for sustained live imaging of rapid subcellular processes. *Nat. Biotechnol.* **41**, 367–377 (2023).
60. Liu, Z. et al. Bioorthogonal photocatalytic proximity labeling in primary living samples. *Nat. Commun.* **15**, 2712 (2024).
61. Zhang, Z. et al. Progress, challenges and opportunities of NMR and XL-MS for cellular structural biology. *JACS Au* **4**, 369–383 (2024).
62. Scarmeas, N., Anastasiou, C. A. & Yannakoulia, M. Nutrition and prevention of cognitive impairment. *Lancet Neurol.* **17**, 1006–1015 (2018).
63. Kottek, M., Grieser, J., Beck, C., Rudolf, B. & Rubel, F. World Map of the Köppen-Geiger climate classification updated. *Meteorol. Z.* **15**, 259–263 (2006).
64. Harel, M. et al. Proteomics of melanoma response to immunotherapy reveals mitochondrial dependence. *Cell* **179**, 236–250 (2019).
65. Xu, J. Y. et al. Integrative proteomic characterization of human lung adenocarcinoma. *Cell* **182**, 245–261 (2020).
66. Shi, Y. et al. Targeting LIF-mediated paracrine interaction for pancreatic cancer therapy and monitoring. *Nature* **569**, 131–135 (2019).
67. Eckert, M. A. et al. Proteomics reveals NNMT as a master metabolic regulator of cancer-associated fibroblasts. *Nature* **569**, 723–728 (2019).
68. Shen, B. et al. Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* **182**, 59–72 (2020).
69. Nie, X. et al. Multi-organ proteomic landscape of COVID-19 autopsies. *Cell* **184**, 775–791 (2021).
70. Niu, L. et al. Noninvasive proteomic biomarkers for alcohol-related liver disease. *Nat. Med.* **28**, 1277–1287 (2022).
71. Virreira Winter, S. et al. Urinary proteome profiling for stratifying patients with familial Parkinson's disease. *EMBO Mol. Med.* **13**, e13257 (2021).
72. Wigger, L. et al. Multi-omics profiling of living human pancreatic islet donors reveals heterogeneous beta cell trajectories towards type 2 diabetes. *Nat. Metab.* **3**, 1017–1031 (2021).
73. Johnson, E. C. B. et al. Large-scale deep multi-layer analysis of Alzheimer's disease brain reveals strong proteomic disease-related changes not observed at the RNA level. *Nat. Neurosci.* **25**, 213–225 (2022).
74. Zhu, Y. et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells. *Nat. Commun.* **9**, 882 (2018).
75. Petelski, A. A. et al. Multiplexed single-cell proteomics using SCoPE2. *Nat. Protoc.* **16**, 5398–5425 (2021).
76. Su, P. et al. Single cell analysis of proteoforms. *J. Proteome Res.* **24**, 1883–1893 (2024).
77. Chen, W. et al. Simple and integrated spintip-based technology applied for deep proteome profiling. *Anal. Chem.* **88**, 4864–4871 (2016).

78. Mund, A. et al. Deep visual proteomics defines single-cell identity and heterogeneity. *Nat. Biotechnol.* **40**, 1231–1240 (2022).
79. Geyer, P. E. et al. Plasma proteome profiling to assess human health and disease. *Cell Syst.* **2**, 185–195 (2016).
80. Deutsch, E. W. et al. Advances and utility of the human plasma proteome. *J. Proteome Res.* **20**, 5241–5263 (2021).
81. Buljan, M. et al. A computational framework for the inference of protein complex remodeling from whole-proteome measurements. *Nat. Methods* **20**, 1523–1529 (2023).
82. Mackmull, M. T. et al. Global, in situ analysis of the structural proteome in individuals with Parkinson's disease to identify a new class of biomarker. *Nat. Struct. Mol. Biol.* **29**, 978–989 (2022).
83. Tsamardinos, I. et al. Just Add Data: automated predictive modeling for knowledge discovery and feature selection. *NPJ Precis. Oncol.* **6**, 38 (2022).
84. Bai, Y. et al. AutoDC: an automatic machine learning framework for disease classification. *Bioinformatics* **38**, 3415–3421 (2022).
85. Elmarakeby, H. A. et al. Biologically informed deep neural network for prostate cancer discovery. *Nature* **598**, 348–352 (2021).

Acknowledgements This work was supported by the Ministry of Science and Technology of the People's Republic of China (grant no. 2020YFE0202200), the National Natural Science Foundation of China (no. 32088101), National Institutes of Health grants P30ES017885-11-S1 and U24CA271037 (G.S.O.) and the Big-Science Infrastructure of Phronesis Medicine, of which the pilot phase is funded by Guangzhou Development District.

Author contributions F.H. conceived the concept of π-HuB and designed its scientific goals, and contributed ideas for phronesis medicine with L.X., F.H., R.A., M.S.B., X.W.B., X.C.B., D.W.C., C.C., L.C., X.C., H.C., F.C., W.E., J.F., P.F., D.F., G.F.G., W.G., Z.-H.G., K.G., W.W.B.G., D.G., C.G., T.G., A.J.R.H., H.H., T.H., N.G.I., Y.J., C.R.J., L.J., N.L.K., M.L., Y.L., Q.L., C.H.L., F.L., G.-H.L., Y.S.L., Z.L., T.Y.L., B.L., M.M., A.M., R.L.M., E.N., G.N., G.S.O., G.P., Y.P., C.P., T.C.W.P., A.P., J.Q., R.R., P.J.R., P.R., C.S., J.S., E.S., S.S., A.S., S.K.S., C.T., L.T., R.T., J.V.E., J.A.V., C.W., X.W.W., X.X.W., Y.W., T.W., M.W., R.W., B.W., L.W., L.X., W.X., Tao Xu, L.Y., J.Y., X.Y., J.R.Y., Q.W.Z., L.H.Z., L.Q.Z., Y.K.Z., Q.Z. and Y.P.Z. contributed ideas and suggestions for the conception and design of this project. T.G., L.T. and Y.W. contributed coordination of the π-HuB Consortium. J.Y. wrote the first draft of the manuscript, and created the figures with F.H., T.G., Y.L. and L.X. F.H., R.A., M.S.B., F.C., P.F., D.F., Z.-H.G., K.G., W.W.B.G., T.G., H.H., T.H., N.G.I., C.R.J., L.J., M.L., Q.L., F.L., Y.S.L., T.Y.L., R.L.M., G.S.O., T.C.W.P., A.P., R.R., P.J.R., C.S., S.K.S., J.A.V., T.W., R.W., B.W., L.W., J.Y., J.R.Y. and Q.Z. provided important edits to the manuscript. All authors contributed to review and editing of the manuscript. The π-HuB Consortium contributed to the discussion of strategic π-HuB research plans.

Competing interests R.A. holds shares of Biognosys AG, which operates in the field covered by the article. D.F. is co-founder of MedBiome Inc., a precision nutrition company. K.G. is a shareholder of CYBO, LucasLand, and FlyWorks. T.G. is the founder of Westlake Omics Inc. M.M. is an indirect investor in EvoSep. R.T. is a founder of BayOmics. The other authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Fuchu He or Ruedi Aebersold.
Peer review information *Nature* thanks Tamar Geiger and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]State Key Laboratory of Medical Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing, China. [2]International Academy of Phronesis Medicine (Guangdong), Guangdong, China. [3]Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland. [4]Macquarie Medical School, Macquarie University, Sydney, New South Wales, Australia. [5]Institute of Pathology and Southwest Cancer Center, Southwest Hospital, Third Military Medical University (Army Medical University) and Key Laboratory of Tumor Immunopathology, Ministry of Education of China, Chongqing, China. [6]Institute of Health Service and Transfusion Medicine, Beijing, China. [7]Department of Pathology and The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD, USA. [8]Key Laboratory of Systems Biology, Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China. [9]Department of Nephrology, First Medical Center of Chinese PLA General Hospital, Nephrology Institute of the Chinese People's Liberation Army, State Key Laboratory of Kidney Diseases, National Clinical Research Center for Kidney Diseases, Beijing Key Laboratory of Kidney Disease Research, Beijing, China. [10]Institute of Chemistry, Academia Sinica, Taipei, China. [11]National Biomedical Imaging Center, State Key Laboratory of Membrane Biology, Institute of Molecular Medicine, Peking-Tsinghua Center for Life Sciences, College of Future Technology, Peking University, Beijing, China. [12]School of Biological Sciences, Queen's University of Belfast, Belfast, UK. [13]Functional Proteomics Laboratory, Centro Nacional de Biotecnología-CSIC, Madrid, Spain. [14]Computational Systems Biochemistry Research Group, Max-Planck Institute of Biochemistry, Martinsried, Germany. [15]AI for Science Institute, Beijing, China. [16]Center for Machine Learning Research, Peking University, Beijing, China. [17]Advanced Clinical Biosystems Research Institute, Cedars Sinai Medical Center, Los Angeles, CA, USA. [18]Department of Liver Surgery and Transplantation, Key Laboratory of Carcinogenesis and Cancer Invasion (Ministry of Education), Liver Cancer Institute, Zhongshan Hospital, Fudan University, Shanghai, China. [19]Centre for Cancer Research, Hudson Institute of Medical Research, Clayton, Victoria, Australia. [20]Monash Proteomics and Metabolomics Platform, Department of Medicine, School of Clinical Sciences, Monash University, Clayton, Victoria, Australia. [21]School of Pharmaceutical Sciences and Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada. [22]The D. H. Chen School of Universal Health, Zhejiang University, Hangzhou, China. [23]Pengcheng Laboratory, Shenzhen, China. [24]School of Electronic Engineering and Computer Science, Peking University, Beijing, China. [25]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada. [26]Department of Chemistry, University of Tokyo, Tokyo, Japan. [27]Department of Bioengineering, University of California, Los Angeles, California, USA. [28]Institute of Technological Sciences, Wuhan University, Wuhan, Hubei, China. [29]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. [30]School of Medicine, Southern University of Science and Technology, Shenzhen, China. [31]Department of Nutrition, Tianjin Institute of Environmental and Operational Medicine, Tianjin, China. [32]School of Medicine, Westlake University, Hangzhou, China. [33]Westlake Center for Intelligent Proteomics, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, China. [34]Research Center for Industries of the Future, School of Life Sciences, Westlake University, Hangzhou, China. [35]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, the Netherlands. [36]Netherlands Proteomics Center, Utrecht, the Netherlands. [37]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridge, UK. [38]Molecular and Cell Biology Laboratory, Salk Institute for Biological Studies, La Jolla, CA, USA. [39]Department of Head & Neck Surgery, Division of Surgery & Surgical Oncology, Division of Medical Sciences, National Cancer Centre Singapore, Singapore, Singapore. [40]OncoProteomics Laboratory, Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam UMC, Amsterdam, the Netherlands. [41]Advanced Glycoscience Research Cluster, School of Biological and Chemical Sciences, University of Galway, Galway, Ireland. [42]Departments of Molecular Biosciences, Departments of Chemistry, Northwestern University, Evanston, IL, USA. [43]David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada. [44]Central China Institute of Artificial Intelligence, Henan, China. [45]Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore, Singapore. [46]CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China. [47]Department of Structural Biology, Leibniz-Forschungsinstitut für MolekularePharmakologie (FMP), Berlin, Germany. [48]State Key Laboratory of Membrane Biology, Key Laboratory of Organ Regeneration and Reconstruction, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. [49]Cancer Biology Institute, Yale University School of Medicine, West Haven, CT, USA. [50]State Key Laboratory of Molecular Oncology, National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [51]UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia. [52]Department of Critical Care Medicine and Hematology, The Third Xiangya Hospital, Central South University; Department of Hematology and Critical Care Medicine, The Third Xiangya Hospital, Central South University, Changsha, China. [53]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. [54]School of Life Sciences, Tsinghua University, Tsinghua-Peking Center for Life Sciences, Beijing, China. [55]Institute for Systems Biology, Seattle, WA, USA. [56]Clinical Biomarker Discovery and Validation, Monash University, Clayton, Victoria, Australia. [57]Department of Endocrine and Metabolic Diseases, Shanghai Institute of Endocrine and Metabolic Diseases, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [58]Shanghai National Clinical Research Center for Metabolic Diseases, Key Laboratory for Endocrine and Metabolic Diseases of the National Health Commission of the PR China, Shanghai, China. [59]Shanghai Key Laboratory for Endocrine Tumor, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [60]Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [61]Department of Oral Biological and Medical Sciences, University of British Columbia, Vancouver, British Columbia, Canada. [62]Yonsei Frontier Lab, Yonsei University, Seoul, Republic of Korea. [63]Glycoproteomics Laboratory, Department of Parasitology, University of São Paulo, Sao Paulo, Brazil. [64]Institute of Zoology, Chinese Academy of Sciences, Beijing, China. [65]Beijing Institute for Stem Cell and Regenerative Medicine, Beijing, China. [66]University of the Chinese Academy of Sciences, Beijing, China. [67]Institut de Recherche en Santé Environnement et Travail, Univ. Rennes, Inserm, EHESP, Irset, Rennes, France. [68]Pilot Laboratory, MOE Frontier Science Centre for Precision Oncology, Centre for Precision Medicine Research and Training, Institute of Translational Medicine, Faculty of Health Sciences, University of Macau, Macau, China. [69]Infection and Immunity Program and Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia. [70]State Key Laboratory of Female Fertility Promotion, Center for Reproductive Medicine, Department of Obstetrics and Gynecology, Peking University Third Hospital, Beijing, China. [71]ProCan®, Children's Medical Research Institute, Faculty of Medicine and Health, The University of Sydney, Westmead, New South Wales, Australia. [72]Department of Health Sciences, University Magna Græcia of Catanzaro, Catanzaro, Italy. [73]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. [74]Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[75]State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, China. [76]Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Medical Research Center, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China. [77]Breast Tumor Center, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China. [78]Indian Institute of Technology Bombay, Mumbai, India. [79]Department of Health Sciences, Faculty of Applied Health Sciences, Brock University, St. Catharines, Ontario, Canada. [80]Center for Quantitative Biology, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. [81]Department of Chemistry, Southern University of Science and Technology, Shenzhen, China. [82]State Key Laboratory of Respiratory Health and Multimorbidity, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing, China. [83]Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, National Clinical Research Center for Respiratory Diseases, China-Japan Friendship Hospital, Beijing, China. [84]Department of Neurology, Clinical Neuroscience Center, University Hospital Zurich and University of Zurich, Zurich, Switzerland. [85]Technical University of Munich, Freising, Germany. [86]Advanced Genomics Unit, Center for Research and Advanced Studies, Irapuato, Mexico. [87]Institute of Translational Medicine, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland. [88]Department of Computer Science, National University of Singapore, Singapore, Singapore. [89]Department of Pathology, National University of Singapore, Singapore, Singapore. [90]Guangzhou National Laboratory, Guangzhou, China. [91]School of Biomedical Engineering, Guangzhou Medical University, Guangzhou, China. [92]The Scripps Research Institute, La Jolla, CA, USA. [93]China Science and Technology Exchange Center, Beijing, China. [94]CAS Key Laboratory of Nutrition, Metabolism and Food Safety, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. [95]Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX, USA. [96]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. [97]Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD, USA. [98]State Key Laboratory of Medical Proteomics, National Chromatography R. & A. Center, CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China. [99]School of Mathematical Sciences, Peking University, Beijing, China. [100]Wuhan University, Wuhan, China. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: hefc@bmi.ac.cn; aebersold@imsb.biol.ethz.ch

The π-HuB Consortium

Ruedi Aebersold[3], Mark S. Baker[4], Xiuwu Bian[5], Xiaochen Bo[6], Daniel W. Chan[7], Cheng Chang[1], Luonan Chen[8], Chris Soon Heng Tan[81], Xiangmei Chen[9], Yu-Ju Chen[10], Heping Cheng[11], Ben C. Collins[12], Fernando Corrales[13], Jürgen Cox[14], Weinan E[15,16], Jennifer E. Van Eyk[17], Jia Fan[18], Pouya Faridi[19,20], Daniel Figeys[21], George Fu Gao[22], Mingxia Gao[101], Qiang Gao[18], Wen Gao[23,24], Yushun Gao[82], Zu-Hua Gao[25], Keisuke Goda[26,27,28], Wilson Wen Bin Goh[29], Dongfeng Gu[30], Changjiang Guo[31], Tiannan Guo[32,33,34], Xuejiang Guo[75], Fuchu He[1,2], Jie He[82], Jun He[102], Qing-Yu He[103], Yuezhong He[2], Albert J. R. Heck[35,36], Henning Hermjakob[37], Jinlin Hou[104], Canhua Huang[105], Tony Hunter[38], Chenxi Jia[1], Narayanan Gopalakrishna Iyer[39], Ying Jiang[1], Connie R. Jimenez[40], Lokesh Joshi[41], Neil L. Kelleher[42], Bernhard Kuster[85], Chaoying Li[1], Dong Li[1], Ming Li[43,44], Yan Li[106], Yanchang Li[1], Yang Li[1], Qingsong Lin[45], Cui Hua Liu[46], Fan Liu[47], Guang-Hui Liu[48], Siqi Liu[107], Xiaonan Liu[108], Ya Liu[108], Yansheng Liu[49], Zhihua Liu[50], Zhongyang Liu[1], Teck Yew Low[51], Ben Lu[52], Haojie Lu[101], Matthias Mann[53], Anming Meng[54], Robert L. Moritz[55], Edouard Nice[56], Yongzhan Nie[108], Guang Ning[57,58,59], Mariko Okada[109], Gilbert S. Omenn[60], Christopher M. Overall[61,62], Giuseppe Palmisano[63], Yaojin Peng[64,65,66], Charles Pineau[67], Terence Chuen Wai Poon[68], Anthony W. Purcell[69], Guojun Qian[110], Jie Qiao[70], Hongqiang Qin[98], Yu Rao[111], Zihe Rao[112], Roger R. Reddel[71], Xianwen Ren[113], Yan Ren[107,114], Phillip J. Robinson[71], Paola Roncada[72], Chris Sander[73,74], Jiahao Sha[75], Feng Shen[110], Lin Shen[106], Erwei Song[76,77], Sanjeeva Srivastava[78], Shicheng Su[76,77], Aihua Sun[1], Siu Kwan Sze[79], Minjia Tan[115], Ben Zhong Tang[116], Chao Tang[80], Liujun Tang[1], Sheng-Ce Tao[117], Ruijun Tian[81], Juan Antonio Vizcaíno[37], Chanjuan Wang[1], Chen Wang[82,83], Jian Wang[1], Tong Wang[103], Xiaowen Wang[1], Xinxing Wang[31], Yan Wang[1], Liming Wei[101], Tobias Weiss[84], Mathias Wilhelm[85], Robert Winkler[86], Bernd Wollscheid[87], Catherine C. L. Wong[82], Limsoon Wong[88,89], Linhai Xie[1], Wei Xie[54], Xiaoliang Sunney Xie[113,118], Li Xu[119], Ping Xu[1], Tao Xu[91], Tianhao Xu[2], Liying Yan[70], Huanming Yang[107], Jianjun Yang[108], Jing Yang[90], Xiao Yang[1], John Yates[92], Mingliang Ye[98], Wantao Ying[1], Xiaobo Yu[1], Yaxiang Yuan[120], Qingcun Zeng[121], Qiwei Zhai[94], Qimin Zhan[106,122], Bing Zhang[95,96], Hui Zhang[97], Lihua Zhang[98], Lingqiang Zhang[1], Pingwen Zhang[99,100], Xiaofei Zhang[102], Xu Zhang[123], Ying Zhang[101], Yukui Zhang[98], Yu Zi Zheng[2,25], Nan-Shan Zhong[90,124], Qing Zhong[71], Feng Zhou[101], Yi Zhu[32,33,34], Yunping Zhu[1] & Zemin Zhang[113,118]

[101]Institutes of Biomedical Sciences, Fudan University, Shanghai, China. [102]Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China. [103]College of Life Science and Technology, Jinan University, Guangzhou, China. [104]Department of Infectious Diseases, Nanfang Hospital, Southern Medical University, Guangzhou, China. [105]State Key Laboratory of Biotherapy, West China Hospital, and West China School of Basic Sciences & Forensic Medicine, Sichuan University, Chengdu, China. [106]Peking University Cancer Hospital & Institute, Beijing, China. [107]BGI Group, Shenzhen, China. [108]Xijing Hospital, Fourth Military Medical University, Xi'an, China. [109]Institute for Protein Research, Osaka University, Osaka, Japan. [110]Eastern Hepatobiliary Surgery Hospital, Second Military Medical University (Naval Medical University), Shanghai, China. [111]School of Pharmaceutical Sciences, Tsinghua University, Beijing, China. [112]School of Medicine, Tsinghua University, Beijing, China. [113]Changping Laboratory, Beijing, China. [114]Experiment Center for Science and Technology, Shanghai University of Traditional Chinese Medicine, Shanghai, China. [115]State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China. [116]School of Science and Engineering, Shenzhen Institute of Aggregate Science and Technology, The Chinese University of Hong Kong, Shenzhen, China. [117]Shanghai Center for Systems Biomedicine, Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Jiao Tong University, Shanghai, China. [118]Biomedical Pioneering Innovation Center (BIOPIC), Peking University, Beijing, China. [119]Department of Liver Surgery, Sun Yat-sen University Cancer Center, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China. [120]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. [121]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China. [122]Peking University International Cancer Institute, Beijing, China. [123]Department of Urology, The Third Medical Center, Chinese PLA General Hospital, Beijing, China. [124]State Key Laboratory of Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China.