The First Stage of π-HuB Project (2023-2032)

The scope of the π-HuB project is obviously unprecedented. To enable π-HuB as a broadly applicable project targeting its ultimate goals, it is necessary to set deliverables on staged programs in a relatively short timeframe to be relevant to the community. During the initiation and development stage (2023-2032, herein referred to as 'the first stage'), π-HuB will build an international cooperation network and achieve the following milestones.

- ***Cell-Type Resolved Proteome Atlases*** π-HuB will ultimately generate single-cell resolution atlases of all major human organs or tissues (e.g., liver, stomach and lung) of different races (e.g., Negroid, Mongoloid, Australoid). Currently, however, proteomic analysis of each single cell in any organ is not feasible due to the technical limitations of existing methods (i.e., relatively high cost and low throughput as compared to single cell transcriptomics). During the first stage, we therefore plan to initially build up the reference cell-type proteome atlases for all major organs using the combination of state-of-the-art flow sorting[57], lase-capture microdissection[58], paralleling MS acquisition technologies[59]. Moreover, rapidly evolving spatial proteomics technologies will provide additional insights to the secreted proteins in the microenvironment and subcellular localization of the proteome[60]. Compare with many genome/transcriptome-centric projects, these proteome-centric atlases with cell-type resolution will offer unique insights in the structures and functions of different cell types and tissue. In this regard, Synergies of the π-HuB project with the HuBMAP[61], the Human Cell Atlas[62], the Human Tumor Atlas Network[63], and the LifeTime Initiative[64] are clearly visible. In addition, the π-HuB project will intrinsically synergize with other initiatives in the field of proteomics. For instance, the Grand project, the HUPO's flagship initiative for future HPP, is going to get better understanding of a function for every protein[65]. In fact, few proteins act alone. Instead, they typically exert biological functions by constituting pathways and networks in certain space and time. In this regard, the high-resolution spatial information for every protein in human body should be of great value for understanding protein function.
- ***Life-Oriented Proteome Atlases*** The proteome of each individual is not only highly dynamic across different stages during the lifetime, but also associated with a number of

non-genetic factors, such as lifestyles, environments and the symbiotic microbiomes. Currently, however, the cost of proteomic analysis is not affordable to most populations, even in developed countries. During the first stage, we therefore plan to focus on the most dominate factors that shape or remodel the human proteome. Specifically, we plan to accumulate the first one million perturbed proteome of bulk tissue or biofluids from large-scale cohort studies, aiming 1) to trace the proteomic trajectory in lifetime by quantifying dynamic changes in the proteomes of the five major prebiotic cycles (e.g., gametogenesis, fertilization, embryonic development, fetal development, and delivery) and the five major postbiotic cycles (e.g., adolescence, puberty, gestation, menopause, and old age); 2) to analyze the effects of four major dietary nutrition patterns (i.e., Western, Japanese, Mediterranean and their subsistence) on the human proteome; 3) to map the proteomes of populations in six major ecological environments (e.g., hot/warm/cold climate, arid, polar/highland climate), and analyze the trajectory of human proteome to habitat patterns-perception-response-acclimatization-adaptation; 4) to map the human microbiome from internal and external environment and its interaction with human proteome, and to construct the adaptation trajectory of human proteome under the symbiotic mode.

- ***Implementation of Proteomics-Driven Precision Medicine.*** In the past decade, mounting evidence have already demonstrated that MS-based proteomic approaches can facilitate mechanistic understanding of diseases as well as its related biomarker discovery and therapy development. For instance, MS-based proteomics has joined forces with genomics and transcriptomics for comprehensive molecular characterization of numerous human tumors[28, 66, 67]. Moreover, MS-based proteomics has solely identified potential biomarkers or potential therapeutic targets for many tumors as well (e.g., early-stage hepatocellular carcinoma[30], melanoma[68], lung adenocarcinoma[31], pancreatic ductal adenocarcinoma[34], ovarian cancer[69]) and a variety of other diseases, such as Coronavirus disease 2019[32], non-alcoholic fatty liver disease[70], Parkinson's disease[71], type 2 diabetes[72], Alzheimer's disease[73] and so on. During the first stage of this project, establishing paths to further implement proteome-driven approaches for the better understanding of complex diseases is therefore to be a natural extension. As such, we

plan to map the proteomes of all major organs at different pathophysiological stages, with a focus on 3-5 representative diseases for each related organ. Such analysis, together with aforementioned tissue proteome atlases with cell-type resolution, is able to construct the proteomic evolutionary trajectories of the occurrence and development of these diseases and their association pathways with specific life stages, life states, and survival conditions. Furthermore, π-HuB will actively collaborate with clinicians, policy makers and industrial partners to catalyze the discovery of novel protein-based biomarkers that can be applied in clinic to diagnose diseases, and discover novel drug targets. More importantly, all the π-HuB-generated proteome data will be projected into a virtual space and processed with cutting-edge AI-based algorithms that have emerged as one of the most promising methods in helping explain the complex relationships between molecular layers and phenotypes[74, 75]. These efforts will result in an initial version of the 'navigator' which, to certain extent, could depict and predict the physiological and pathological processes inside the human body. Newly obtained proteomic data could therefore be fed into the 'navigator', leading to in-depth interpretation of the data, facilitating the discovery and development of intervention and therapeutic strategies for proteomics-driven precision medicine.