Computational Optimization of Spectral Library Size Improves DIA-MS Proteome Coverage and Applications to 15 Tumors
通过计算优化谱图库的大小以提高DIA-MS蛋白质组学的鉴定率并应用在15种组织数据集中

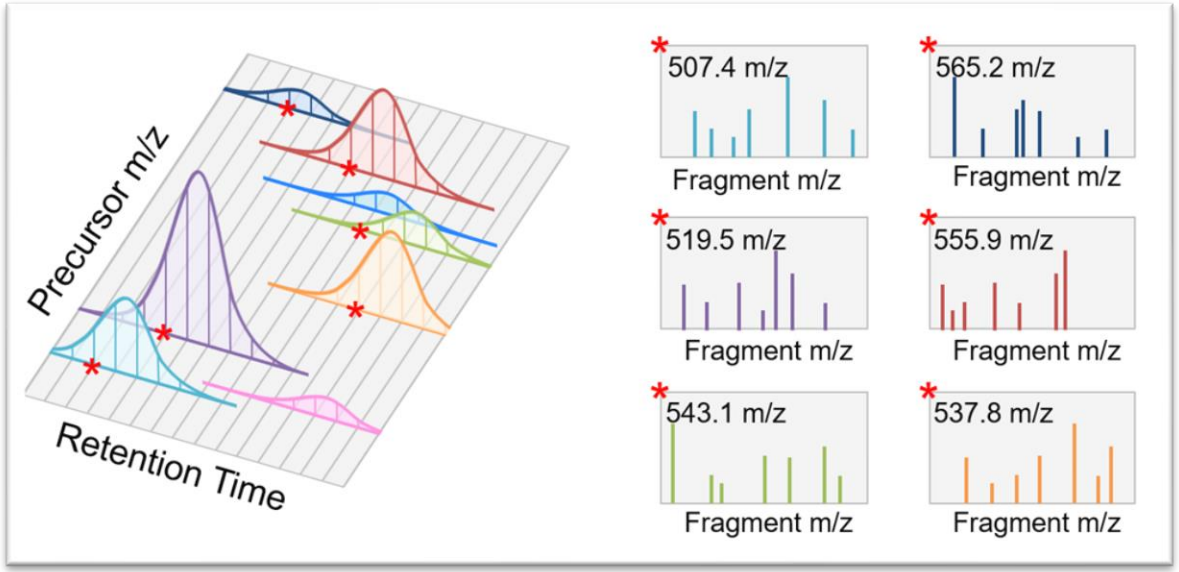
Weigang Ge
葛伟刚

蛋白质组大数据实验室
www.guomics.com

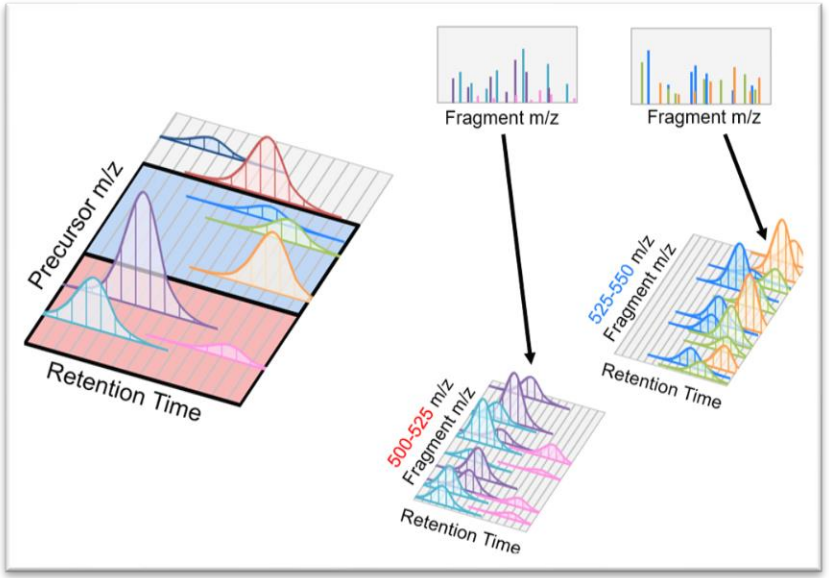
西湖大學
WESTLAKE UNIVERSITY

Data-independent acquisition (DIA)

- DDA

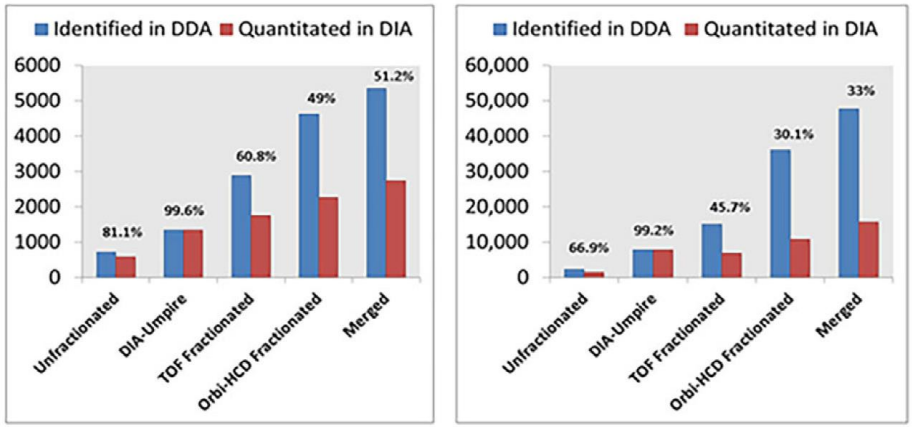


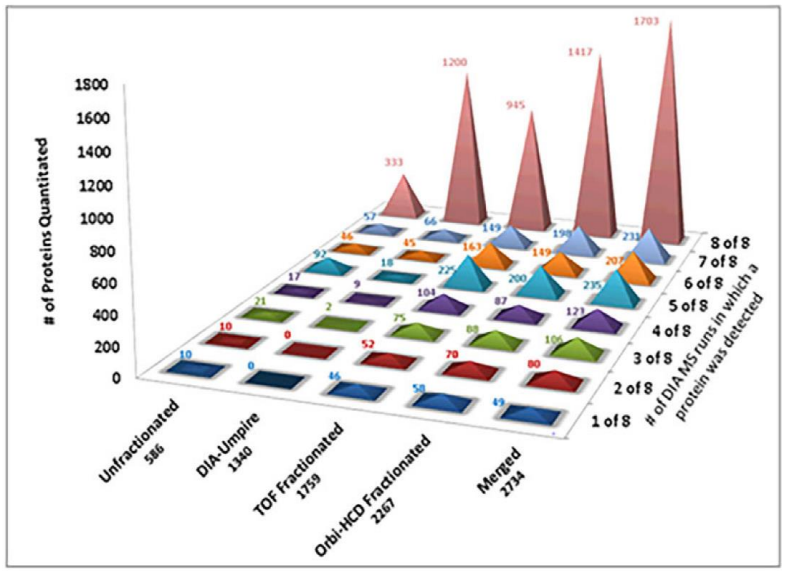
- DIA



[1] Searle et al. 2020; Emma Leith, et al, 2021 EncyclopeDIA (Galaxy Training Materials).

Background of Library





Both spectral library-free and library-based strategies are used to analyze DIA-MS data.

A large number of DDA-MS runs, particularly from fractionated samples, has been shown to lead to a more comprehensive spectral library, enabling the potential detection of a larger number of peptides and proteins from the DIA-MS datasets.

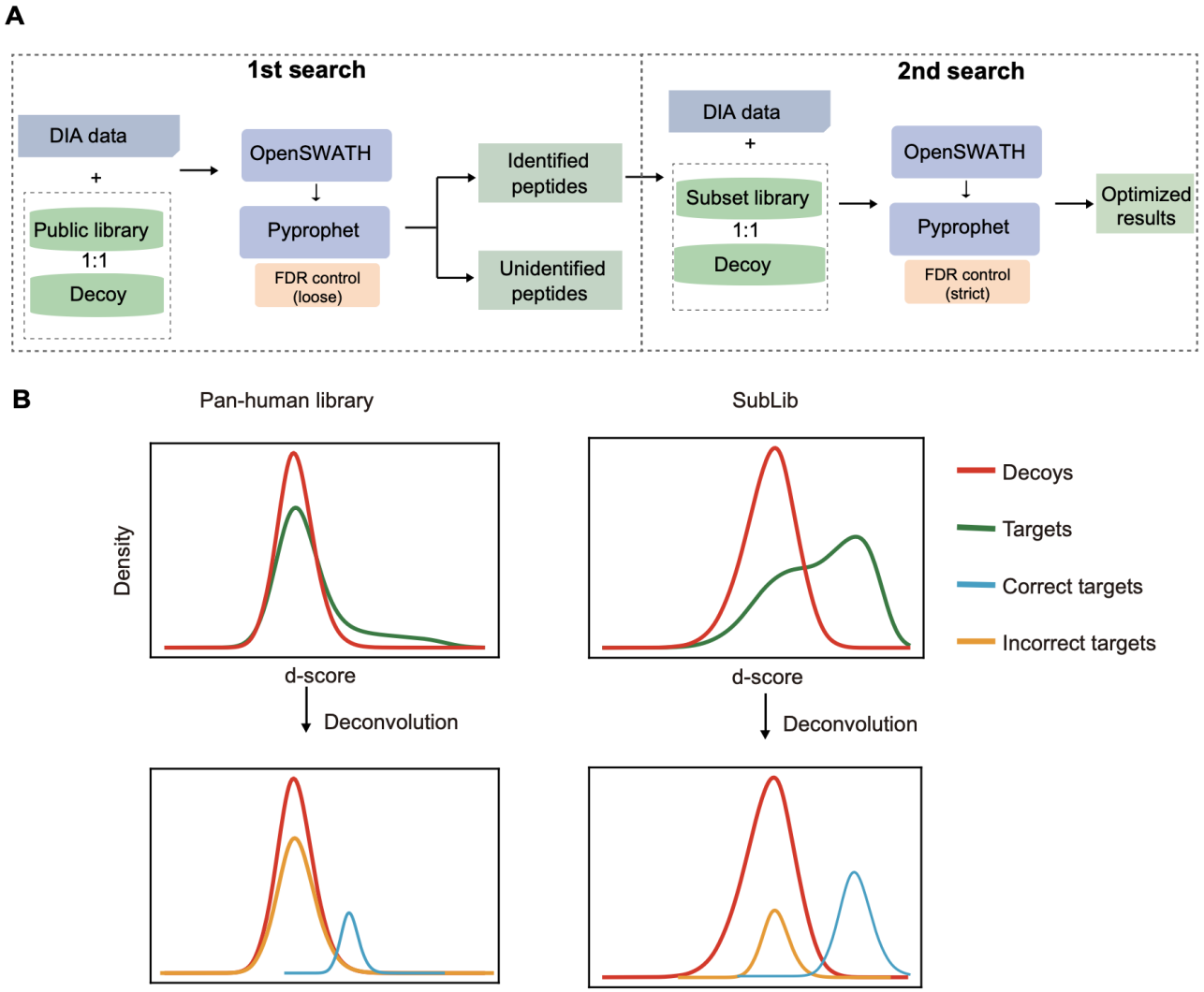
The size of the spectral library directly impacts the performance of the DIA-MS data analysis, specifically peptides and proteins identification and quantification, sensitivity, and specificity

[1] Parker, S. J.; Venkatraman, V.; Van Eyk, J. E., Effect of peptide assay library size and composition in targeted data-independent acquisition-MS analyses. Proteomics 2016, 16 (15-16), 2221-37.
[2] Zhang,F.; Ge, W.; Ruan, G.; Cai, X.; Guo, T., Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. PROTEOMICS 2020, 20 (17-18), 1900276.
[3] Rosenberger, G et al, Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. Nat Methods 2017, 14 (9), 921-927.

Objectives

- To perform a systematic evaluation of the size of the spectral library.
- To generate experiment-specific subset libraries using a priori analyses of the DIA data for the purpose of proteomics coverage improvement.

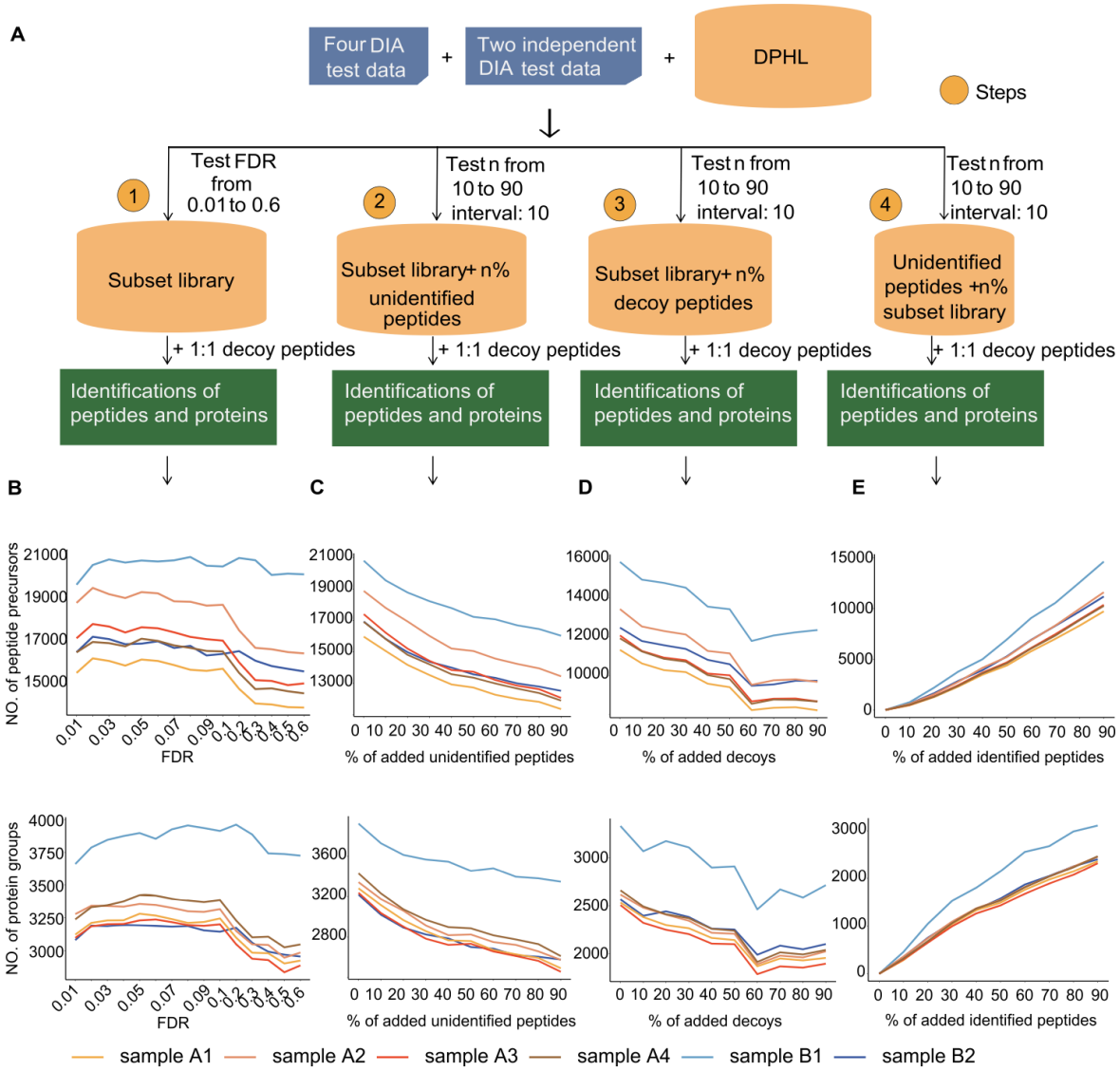
The workflow of subLib strategy



(A) The two-step workflow of subLib.

(B) The original distribution and deconvoluted distribution based on the Gaussian mix model of representative.

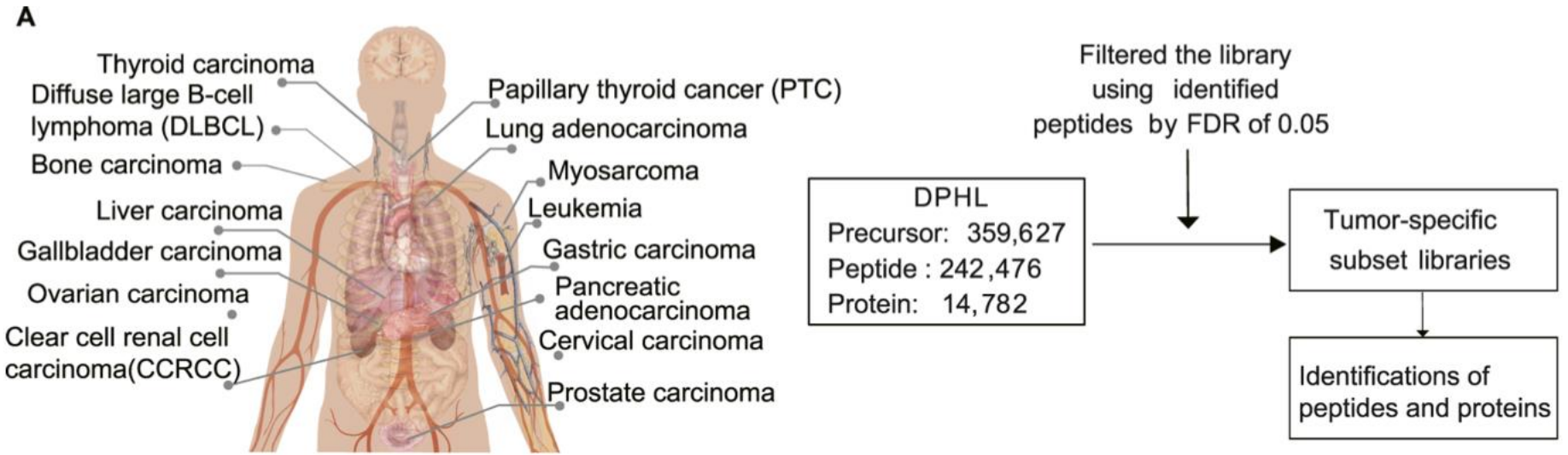
Optimization of colorectal cancer subset library by refining DPHL



(A) Workflow for spectral library optimization.

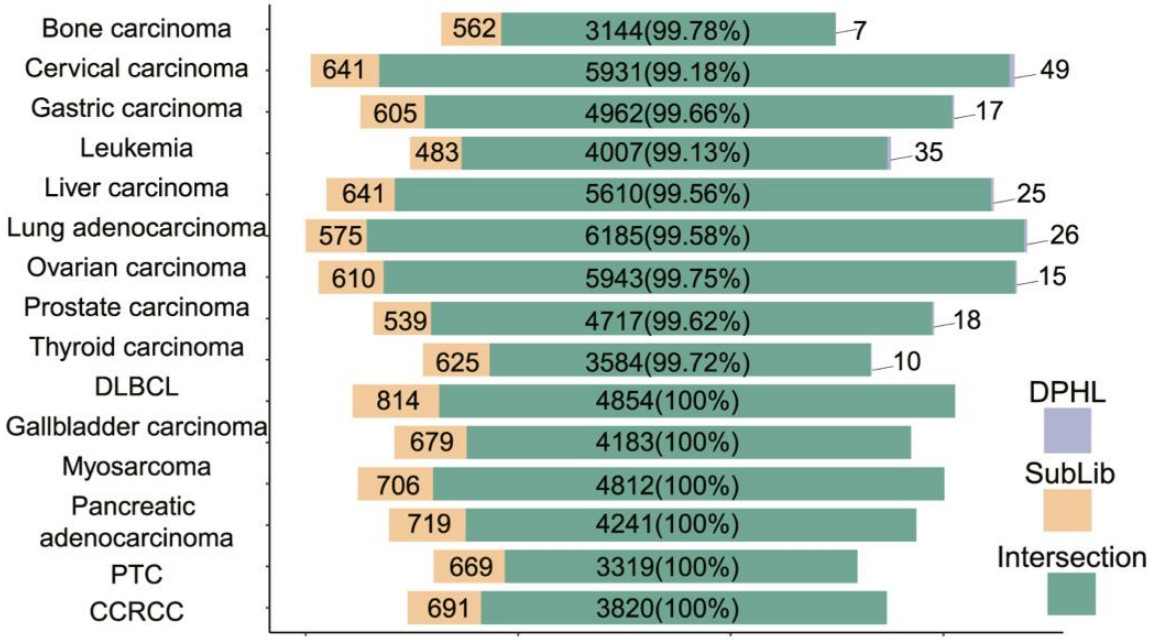
(B,C,D,E) The numbers of identified peptide precursors (upper panel) and protein groups (lower panel) based on the subset libraries.

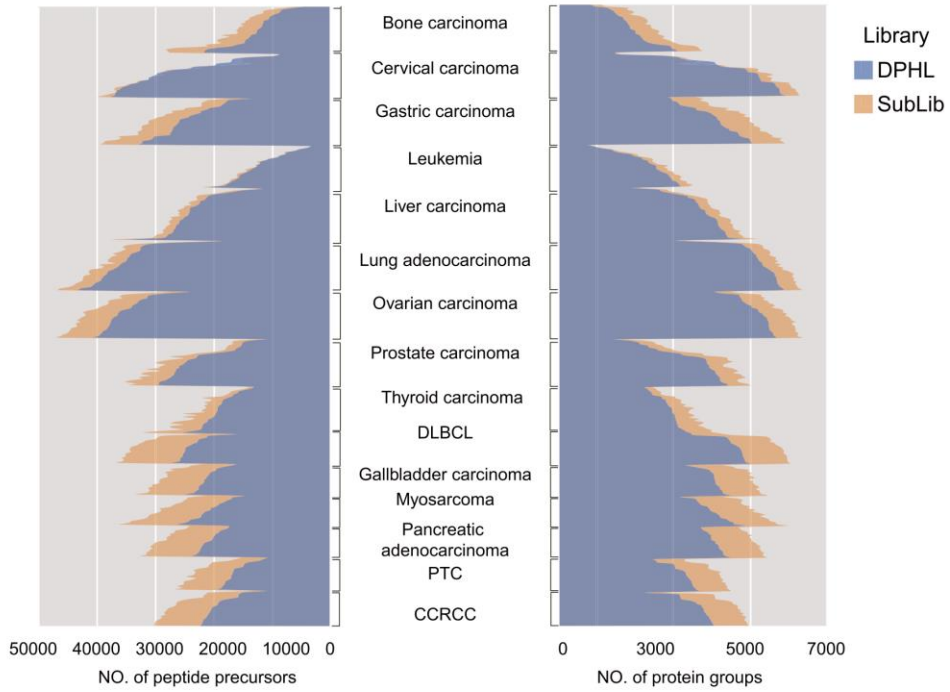
Applying subLib to the DIA-MS of 15 extended tumor sample types



SubLib for biomedical applications, we applied the subLib strategy to 15 different types of cancer samples, generated of 15 tumor type subset libraries with DPHL.

Applying subLib to the DIA-MS of 15 extended tumor sample types

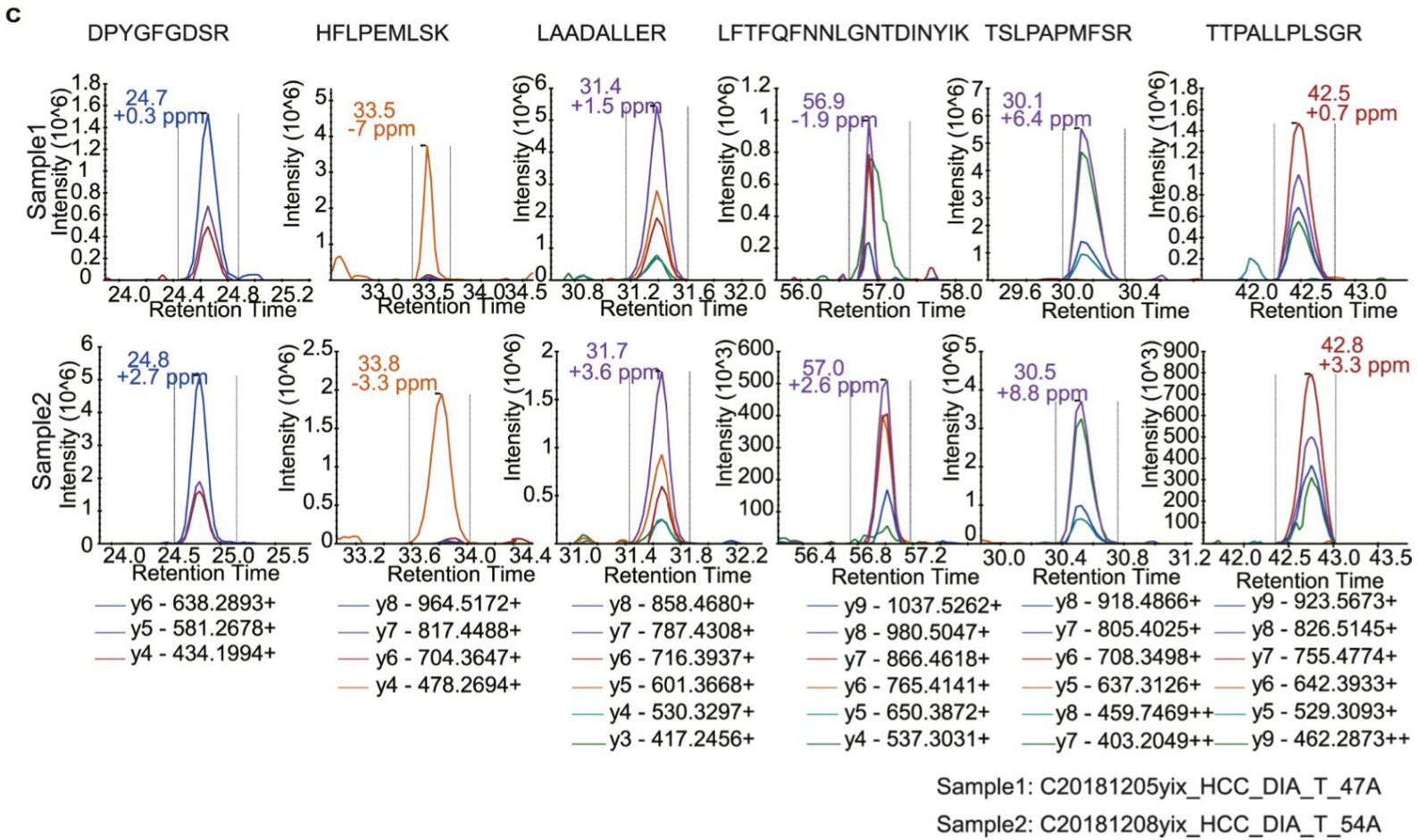




In all of the tumor sample types, peptide/protein identification numbers obtained using the subset library exceeded the numbers reached using DPHL, and over 99% of the protein identifications were overlapped

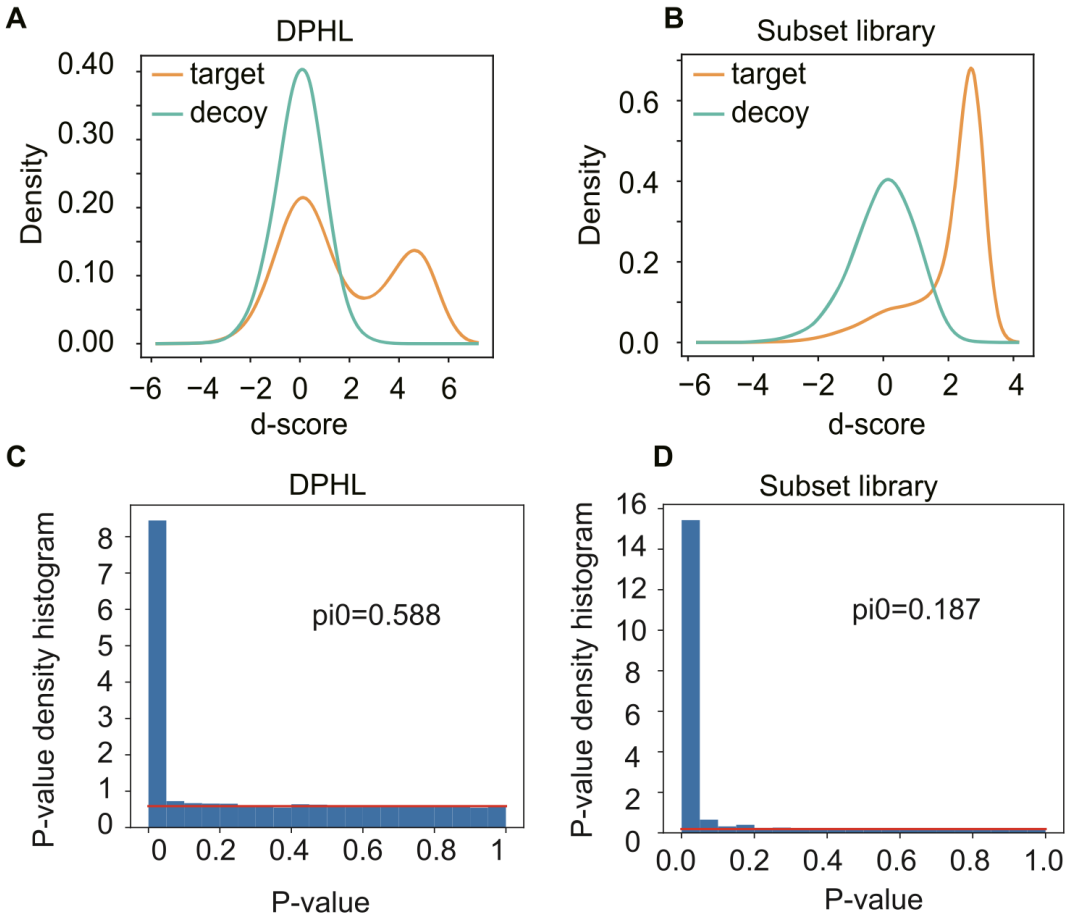
The identifications of individual DIA-MS data, found that subLib could increase peptide precursor and protein group identifications in almost all of the data files.

Applying subLib to the DIA-MS of 15 extended tumor sample types



We looked at identification results of two liver carcinoma samples and checked the spectra quality of six randomly chosen peptides that were identified from the subset library but not DPHL. They were with good mass accuracy, retention time consistency and peak group conformity, suggesting that subLib could increase the identification of correct targets.

subLib decreases the number of incorrect targets instead of removing all of them



SubLib did not completely remove all the incorrect targets after refinement. The decrease of incorrect targets also reduced the pi0 from 0.588 in the DPHL to 0.187 in the subset library, which improved peptide and protein dectection sensitivity.

Publication

Journal of proteome research

pubs.acs.org/jpr Article

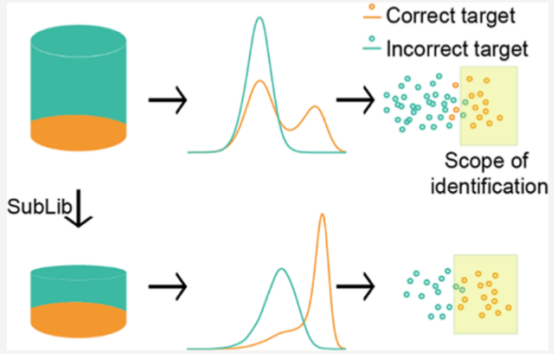
Computational Optimization of Spectral Library Size Improves DIA-MS Proteome Coverage and Applications to 15 Tumors

Weigang Ge, Xiao Liang, Fangfei Zhang, Yifan Hu, Luang Xu, Nan Xiang, Rui Sun, Wei Liu, Zhangzhi Xue, Xiao Yi, Yaoting Sun, Bo Wang, Jiang Zhu, Cong Lu, Xiaolu Zhan, Lirong Chen, Yan Wu, Zhiguo Zheng, Wangang Gong, Qijun Wu, Jiekai Yu, Zhaoming Ye, Xiaodong Teng, Shiang Huang, Shu Zheng, Tong Liu,* Chunhui Yuan,* and Tiannan Guo*

Cite This: https://doi.org/10.1021/acs.jproteome.1c00640 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Efficient peptide and protein identifications from data-independent acquisition mass spectrometric (DIA-MS) data typically rely on a project-specific spectral library with a suitable size. Here, we describe subLib, a computational strategy for optimizing the spectral library for a specific DIA data set based on a comprehensive spectral library, requiring the preliminary analysis of the DIA data set. Compared with the pan-human library strategy, subLib achieved a 41.2% increase in peptide precursor identifications and a 35.6% increase in protein group identifications in a test data set of six colorectal tumor samples. We also applied this strategy to 389 carcinoma samples from 15 tumor data sets: up to a 39.2% increase in peptide precursor identifications and a 19.0% increase in protein group identifications were observed. Our strategy for spectral library size optimization thus successfully proved to deepen the proteome coverages of DIA-MS data.



KEYWORDS: *data-independent acquisition, protein identification, pan-human library, spectral library optimization, target-decoy*

Acknowledgements

西湖大學 WESTLAKE UNIVERSITY
西湖欧米 WESTLAKE OMICS
UOMICS -since 2017

All labmates in Laboratory of Big Proteomic Data

ThermoFisher SCIENTIFIC
PBI Pressure BioSciences Inc.

Hospitals and clinical centers

the First Affiliated Hospital of College of Medicine, Zhejiang University;
the Second Affiliated Hospital of College of Medicine, Zhejiang University;
the Center for Stem Cell Research and Application, Union Hospital, Tongji Medical College;
Harbin Medical University Cancer Hospital;
the Cancer Hospital of the University of Chinese Academy of Sciences, Zhejiang Cancer Hospital;
Shengjing Hospital of China Medical University;

THANK YOU

西湖大學
WESTLAKE UNIVERSITY