



```

1 from keras.layers.convolutional import Conv1D
2 from keras.layers.core import Dense, Dropout, Masking
3 from keras.layers.recurrent import LSTM
4 from keras.layers.wrappers import Bidirectional, TimeDistributed
5 from keras.models import load_model as keras_load_model
6 from keras.models import Sequential
7
8
9 from .options import PeptideMS2Options
10
11 def build_model(options, metrics=CosineSimilarity):
12     model = Sequential()
13     model.add(
14         Conv1D(
15             filters=64,
16             kernel_size=2,
17             activation="relu",
18             input_shape=(options.max_sequence_length, 20)),
19         model.add(Masking(mask_value=-1))
20     model.add(Bidirectional(LSTM(128, return_sequences=True)))
21     model.add(Dropout(0.5))
22     model.add(TimeDistributed(Dense(options.num_outputs)))
23     model.compile(
24         loss="mean_squared_error",
25         optimizer="adam",
26         metrics=metrics)
27     return model
28
29 def cosine_similarity(y_true, y_pred):
30     length = K.int_shape(y_pred)[1]
31     y_true = K.batch_flatten(y_true)
32     y_pred = K.batch_flatten(y_pred)
33     y_true = K.l2_normalize(y_true, axis=-1)
34     y_pred = K.l2_normalize(y_pred, axis=-1)
35     return K.sum(y_true * y_pred, axis=-1) / length
    
```



蛋白质组定量综合解决方案



沈诚频
 上海易算生物科技有限公司

www.omicsolution.com



关于我们

致 力 于 提 供 最 优 的 蛋 白 质 组 学 解 决 方 案



沈诚频

CEO

OMIC SOLUTION
上海易算生物科技有限公司

2004年起师从质谱领域权威专家杨芑原教授

2005-2010 复旦大学生物医学研究院 博士

从事蛋白质组学及生物信息学研究

2016年-至今

上海易算生物科技有限公司

PEAKS, Spectronaut等软件国内技术支持、销售

围绕蛋白质组质谱技术提供软件、试剂、耗材、培训、个性化服务



关于我们



科研院所

- 为国内数十个质谱实验室提供软件、检测方法、试剂、生信解决方案
- 共同申请、承担、研发科研项目

质谱公司

为Thermo, Bruker等主要质谱厂商提供软件和数据分析方案，合作拓展市场

检测公司

为国内多家CRO检测公司提供软件、方法开发、人员培训及分析合作

国外供应商

Biognosys、Bioinfor、PreOmics等多个国外质谱新技术供应商合作国内市场



科研协作



上海易算生物科技有限公司



杨芑原教授 乔亮教授

实验室派驻专业技术人员合作研发，目前有2名质谱技术人员派驻复旦，交大合作质谱检测技术；2名生信技术人员派驻复旦、交大合作生信平台开发；1名实验技术人员派驻交大附属医院合作临床样品处理；

In silico spectral libraries by deep learning facilitate data-independent acquisition roteomics.

DeepDIA DIA数据挖掘算法 IF13

2020 NATURE COMMUNICATIONS

Yi Yang¹, Xiaohui Liu¹, Chengpin Shen², Yu Lin³, Pengyuan Yang¹ see all 6 authors

¹ Fudan University, ² Shanghai Omicsolution Co., Ltd., Shanghai, 200000, China., ³ Australian National University

Proteomic and Metabolic Elucidation of Solar-Powered Biomanufacturing by Bio-Abiotic Hybrid System

蛋白、代谢多组学 IF18

2020 CHEM

Rutan Zhang¹, Ying He¹, Jia Yi¹, Lijuan Zhang², Chengpin Shen² see all 9 authors

¹ Fudan University, ² Shanghai Omicsolution Co., Ltd., Shanghai 201100, China

Metaproteomics characterizes human gut microbiome function in colorectal cancer.

DIA宏蛋白质组学 IF5

2020

Shuping Long^{1, 2}, Yi Yang¹, Chengpin Shen³, Yiwen Wang¹, Anmei Deng⁴ see all 7 authors

¹ Fudan University, ² Tongji University, ³ Shanghai Omicsolution Co., Ltd., Shanghai, China.,

⁴ Changhai Hospital, The Naval Military Medical University, Shanghai, China

易算生物与复旦大学IBS/化学系长期进行科研协作，在质谱检测技术和算法、临床样品分析方法开发等领域产出丰硕成果



质谱公司合作



- 提供全套DIA蛋白质组学定量方案
- 开发iOmics生信分析平台
- 合作进行客户蛋白质组及生信培训



- 提供全套DIA蛋白质组学定量方案
- 为布鲁克内部提供全套质谱数据分析软硬件平台和培训
- 合作进行客户蛋白质组及生信培训
- 合作进行客户数据测试



国外供应商合作



前沿DIA技术研发及CRO

- 代理全部软件、试剂产品
- 为国内客户提供全套售后、培训
- 目前已有40家国内质谱实验室客户



前沿De Novo及蛋白质组学软件、单抗、多肽组前沿技术、机器学习算法研发

- 代理全部软件、检测服务
- 为国内客户提供全套售后、培训
- 目前已有40家国内质谱实验室客户



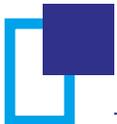
前沿蛋白质组样品处理试剂盒、自动处理工作站

- 代理试剂盒产品
- 为国内客户提供全套售后、培训
- 目前已有10家国内质谱实验室客户



最新研发色谱柱产品

- 代理国内销售和培训



DIA蛋白质组学全流程产品

The Omics Solution Provider
OMIC SOLUTION

样品制备

Biognosys、PreOmics蛋白提取试剂盒
DIA质控iRT试剂盒
PQ500 绝对定量试剂盒

质谱数据分析软件

Peaks Online、Studio、Antibody
Spectronaut、SpectroDive、SpectroMine
软件配套生信分析服务器

定制化分析

针对个性化需求定制分析流程和平台

分离检测

Omics 整体填料色谱柱
液相色谱柱温箱

生信平台

悟空综合生信分析平台
DeepDIA 谱图预测平台
OmicsCloud蛋白质组数据管理报告平台

培训

样品制备、质谱方法、数据分析



www.omicsolution.com

上海易算生物科技有限公司





样品制备和质控



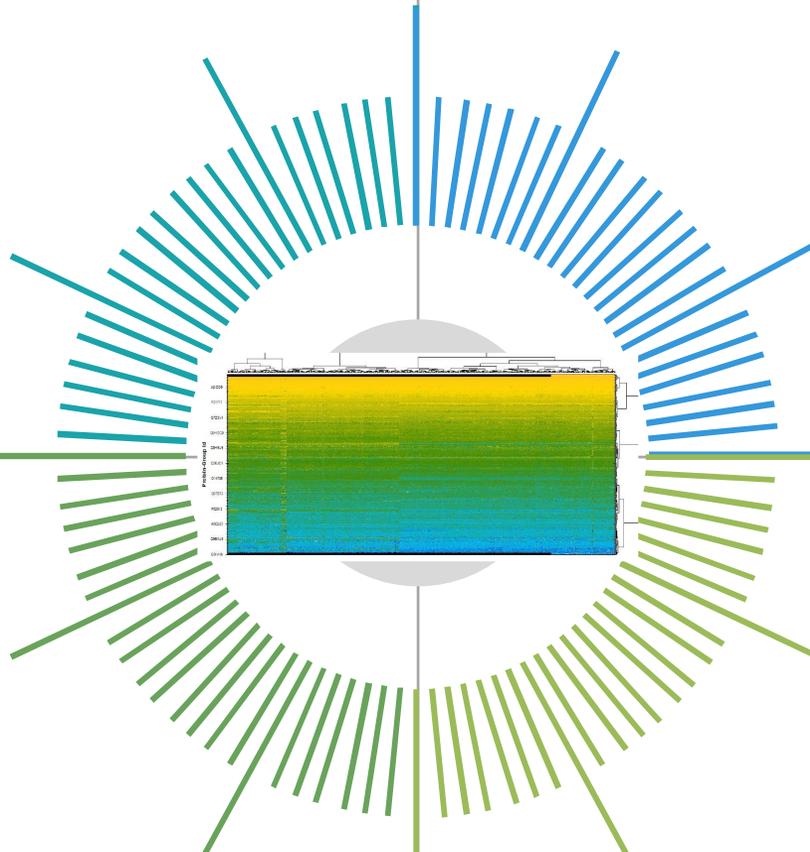
PreOmics iST系列

快速前处理试剂盒



Biognosys

样品前处理试剂盒



PQ500 绝对定量试剂盒

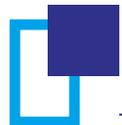
血液绝对定量



iRT质控标肽试剂盒

DIA辅助定量、液相-质谱
稳定性质控





分离检测

纳升色谱柱温箱



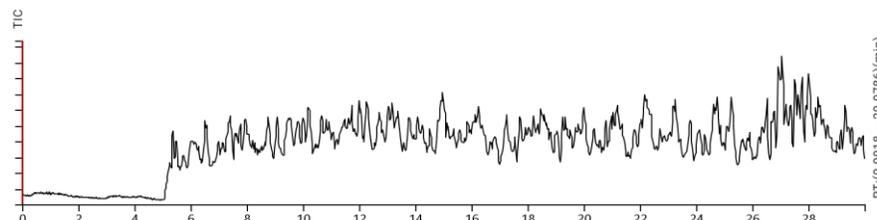
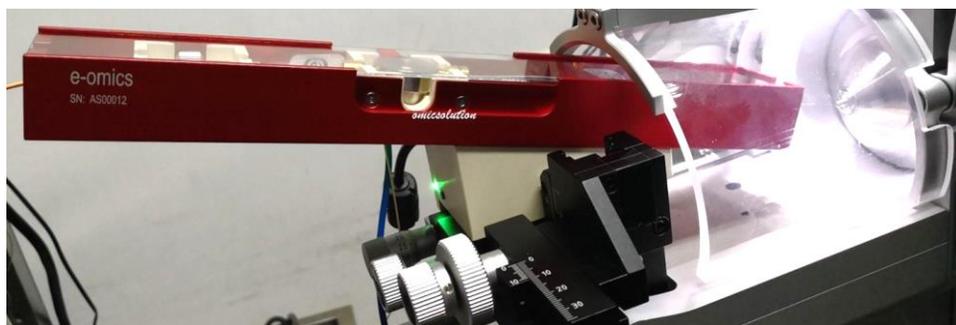
提高柱效、保证大样本量稳定性



介孔填料色谱柱



超长耐久、超低柱压、超高柱效



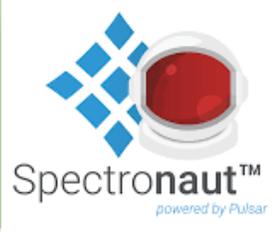
Hela

30min 1000nl/min 100bar
Fusion 2700 protein group
480F 3300 protein group

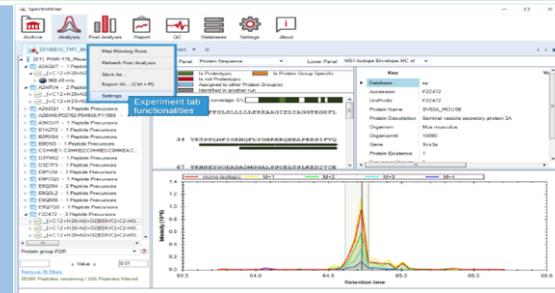
BGS

Spectronaut & SpectroDive & SpectroMine

DIA/SWATH



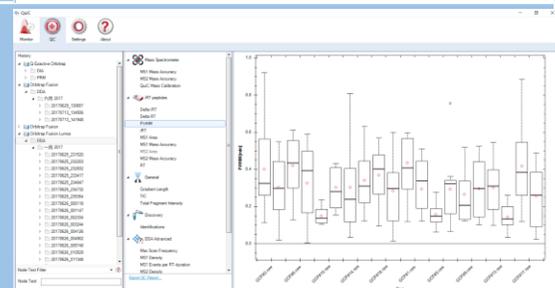
LFQ/iTRAQ/TMT



PRM/MRM



质谱数据质控

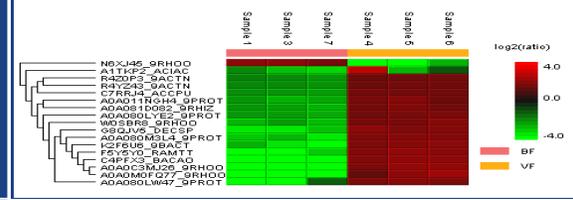


BSI

PEAKS & PEAKS AB

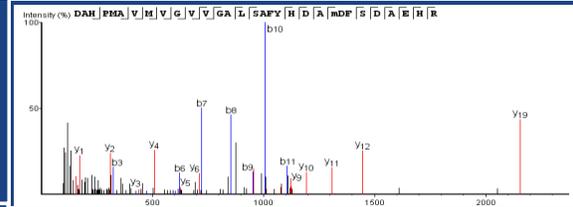
蛋白质组学

Label free、iTRAQ、TMT蛋白质鉴定、定量以及相应质控统计分析

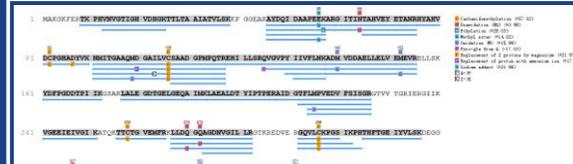


PEAKS Studio Online

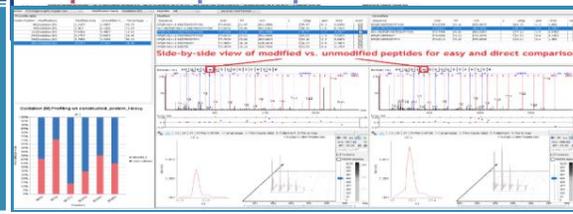
De novo 蛋白测序



蛋白质修饰、突变



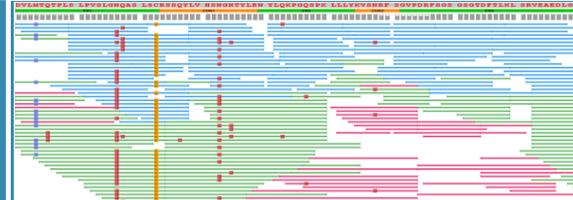
蛋白质修饰、突变



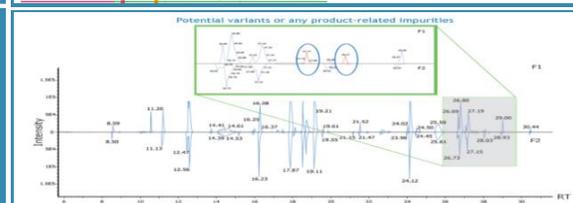
抗体测序

PEAKS AB

De novo 蛋白测序



抗体QA/QC





易算生信综合分析平台

OmicS Cloud

OmicS Cloud是易算自主开发的生信分析平台，旨在为有蛋白质组数据分析需求的客户提供从用户管理到数据分析、结果解读的一站式解决方案

Enter

PEAKS Online

PEAKS ONLINE由BSI公司开发，基于WEB界面为您提供最便利的质谱数据分析解决方案。为拥有海量质谱、多用户分析需求的实验室提供了最佳的解决方案。（**目前仅面向特定用户demo测试使用**）

Enter



Human Body Fluid Database

人体体液蛋白质组数据库是由协和基础所质谱平台的蛋白质组学实验数据分析产生，主要服务方向为生物学的基础和应用基础研究，主要用户为从事疾病标志物研发的科研和临床工作者。

Enter

DeepDIA

易算与复旦大学乔亮课题组、生物医学研究院杨芄原课题组、澳大利亚国立大学林宇课题组合作，利用深度学习技术开发了从肽段或蛋白质序列构建谱图库的工具deepDIA，实现了不依赖于DDA的DIA数据直接分析。

Enter

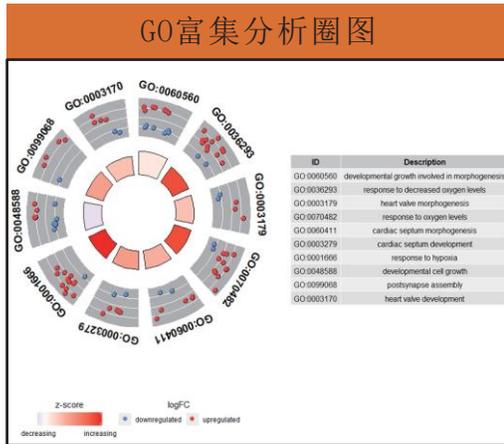


悟空：快速查看与分析数据云平台由易算与作者王诗盛合作，基于R语言开发，旨在以友好的人机交互辅助用户快速分析数据并时时展现结果。平台囊括各种回归模型，分类模型供用户学习使用。

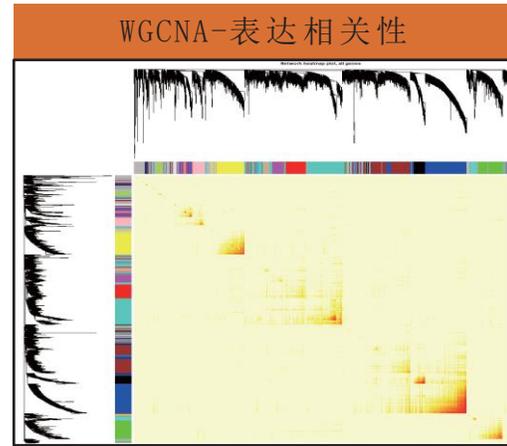
<https://project.omicsolution.com/home>



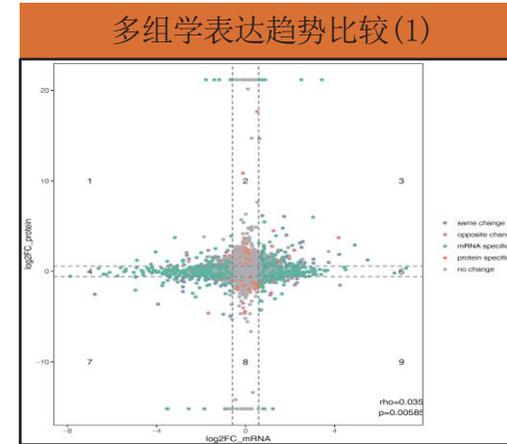
生信高级订制开发



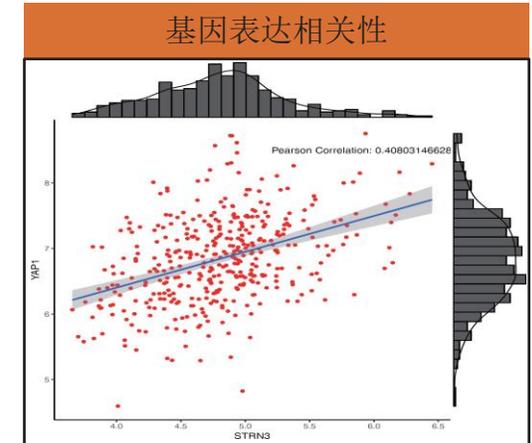
将显著富集的GO Term中的差异分子绘制在扇形中，形成圈图



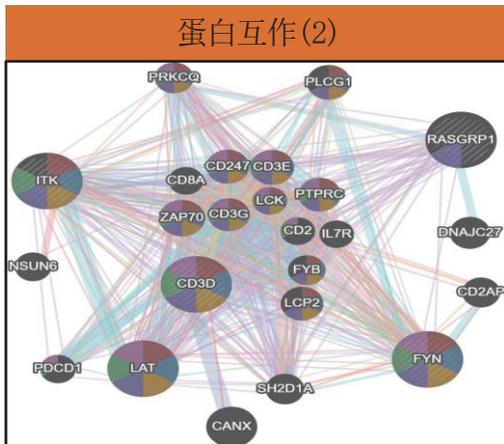
加权共表达网络分析(WGCNA)，所有基因聚类的模式



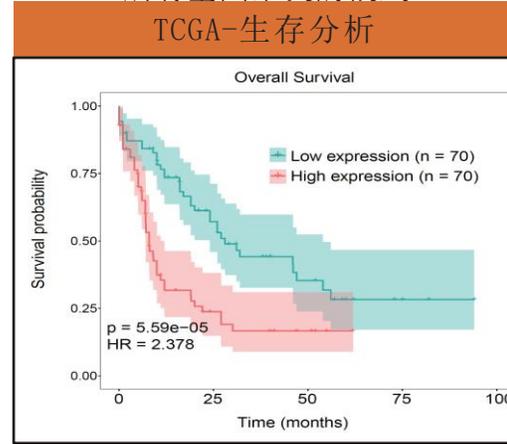
分析不同组学中所有基因的表达差异，并发现变化趋势异同



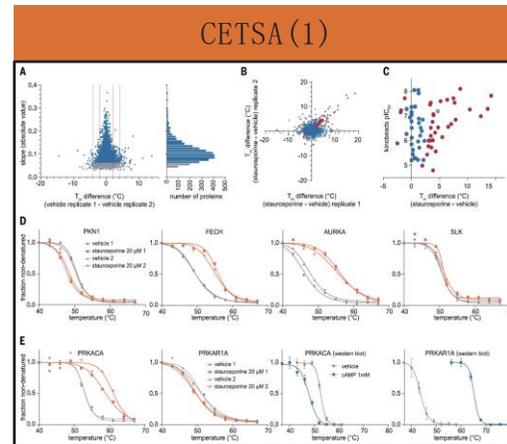
分析不同基因的表达相关性



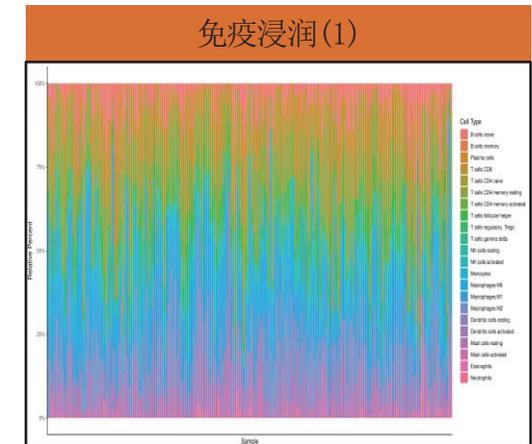
使用GeneMania分析差异分子的互作



结合TCGA数据库中的表达和临床信息，分析目的基因表达对预后的影响



细胞热转变分析(CETSA)通过拟合蛋白热稳定性曲线，揭示药物的靶标蛋白



基于表达矩阵和免疫细胞表达参考矩阵，分析样本中免疫细胞的浸润程度

部分分析展示

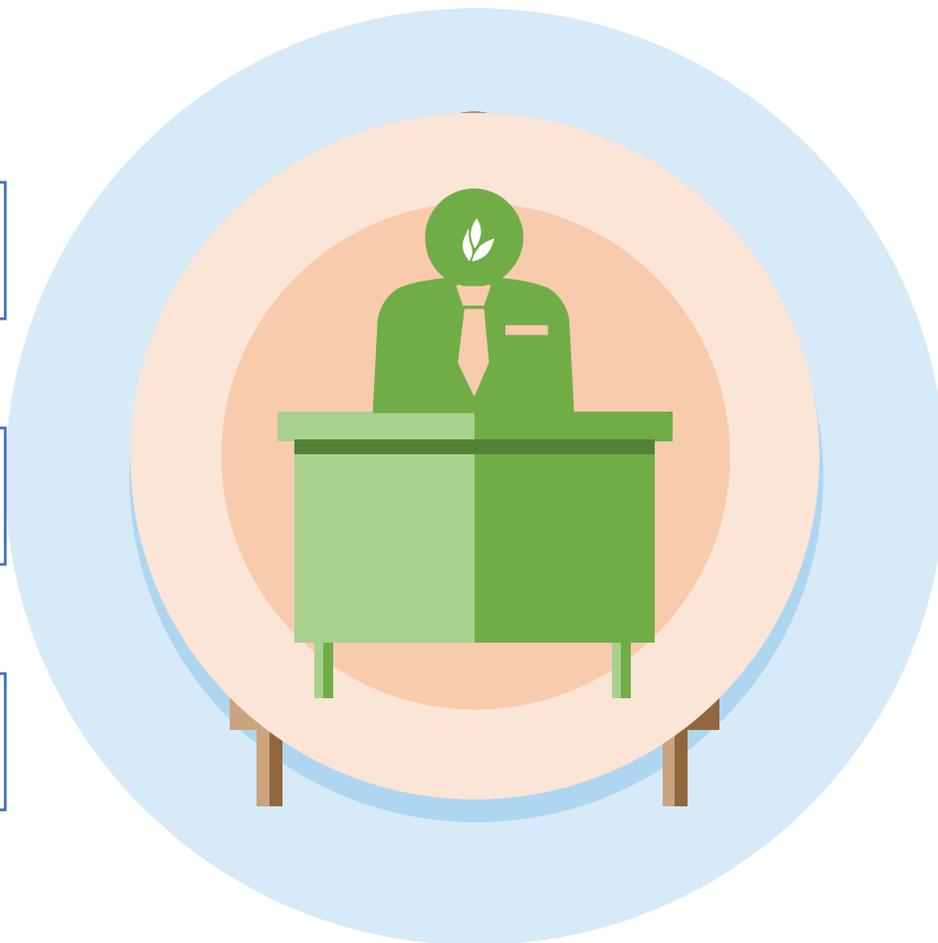


培训服务

新生入学

新技术入门

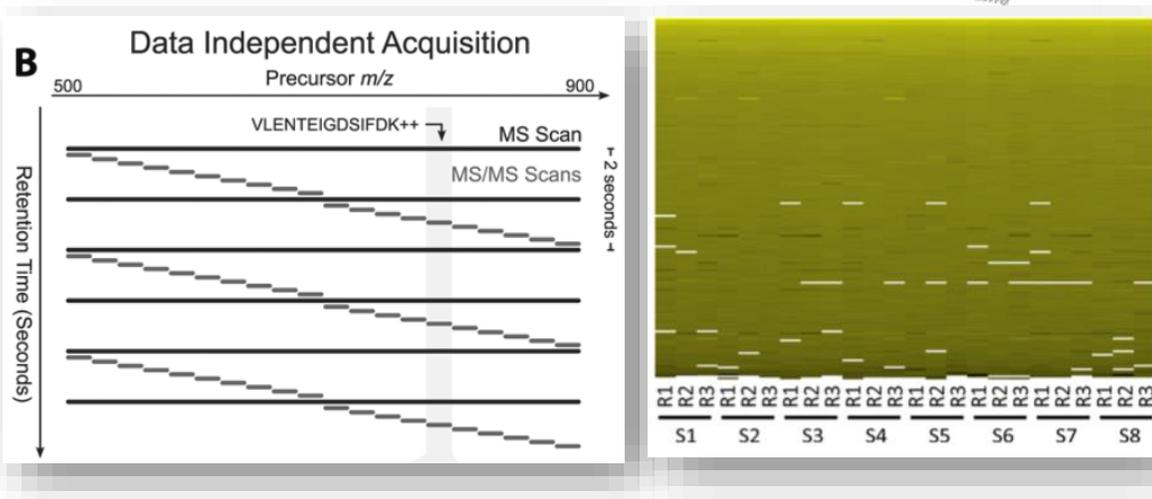
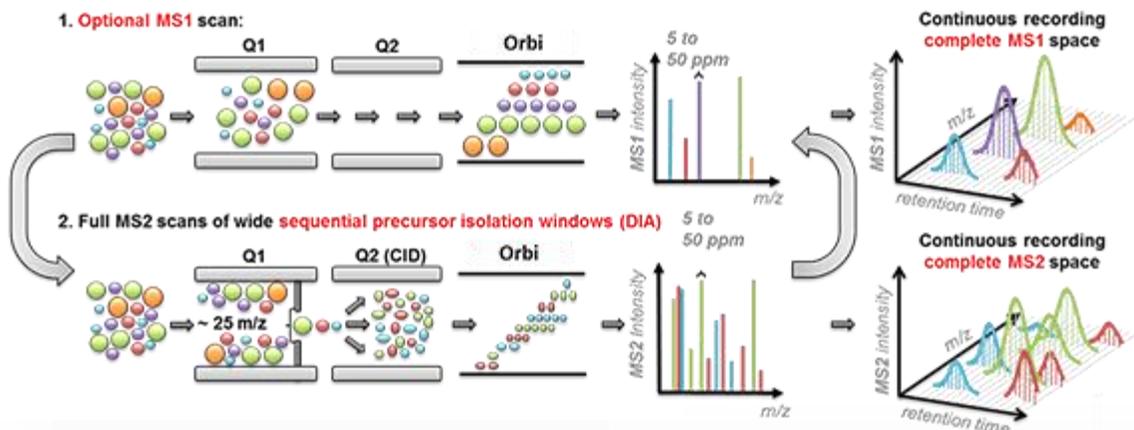
实验室推广



为多个质谱实验室提供常年定制化培训服务



DIA – data independent acquisition



Next-Generation Proteomics

DIA (数据非依赖采集)

蛋白质组学届巨擘**Ruedi Aebersold教授**开创了基于质谱的生物样本数字化的DIA/SWATH技术，帮助蛋白质组学迈向下一时代。通过他创立的Biognosys提供的整体解决方案，您可以轻松地实行大规模样本的组学定量流程。取决于您的研究方向，可以选择发现式研究方案，剖析不同条件及处理下样品的成千上万种蛋白；亦或是靶向研究方案，对选定的蛋白质进行高通量定量。

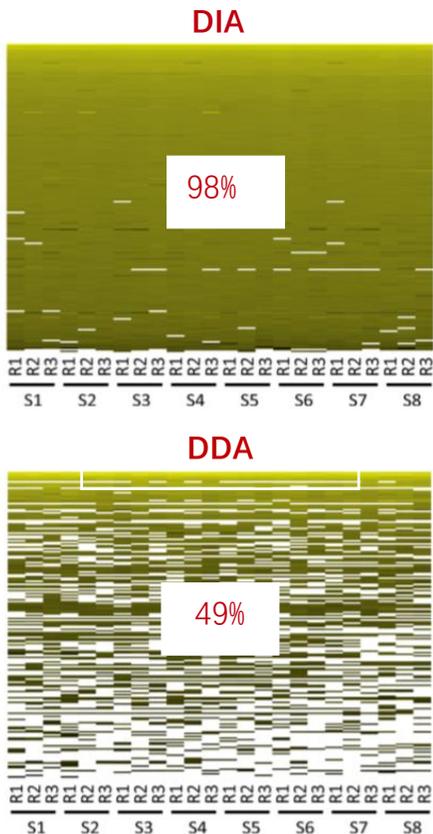
DIA技术融合了传统蛋白质组学“鸟枪法” (shotgun) 和质谱绝对定量“金标准”选择反应监测/多反应监测 (SRM/MRM) 技术的优势和特点。另外，DIA技术将质谱**整个全扫描范围分为若干个窗口**，高速、循环地对每个窗口中的所有离子进行选择、碎裂、检测，从而无遗漏、无差异地获得样本中所有离子的全部碎片信息。DIA技术具有**全景式扫描、数据可回溯**等优势。

DIA质谱技术被**Nature Methods**杂志评为**2015年最值得关注的技术**。



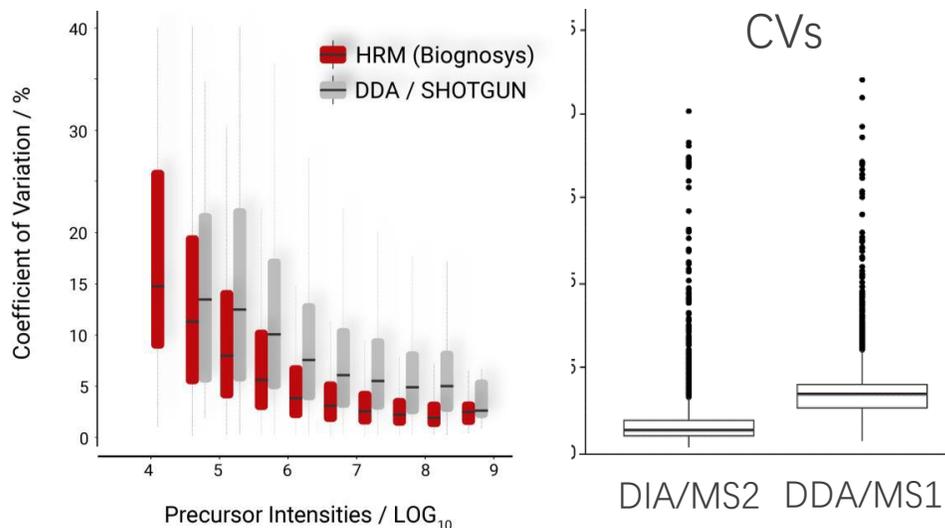
➤ 重复性大幅度提升

DIA的数据采集模式相比于常规非标定量DDA模式，有更少的缺失值，数据记录更为完整。



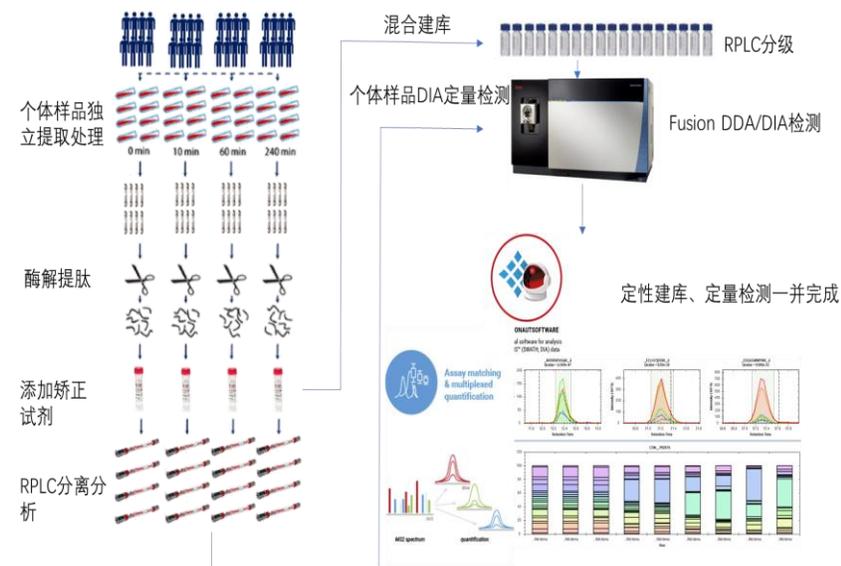
➤ DIA定量准确性更高

定量方式由通过一级质谱信号定量，升级为由每个肽段通过多个二级质谱峰面积信号进行定量。定量信息更为全面，定量准确性提升，CV值下降



➤ 实验操作简单，周期缩短

与传统的iTRAQ/TMT/LFQ相比，通过构建DDA谱图库的模式，样本信息采集使用DIA模式，样本无需分馏上机，极大缩短了每个样品的检测周期。





新策略——DpD策略

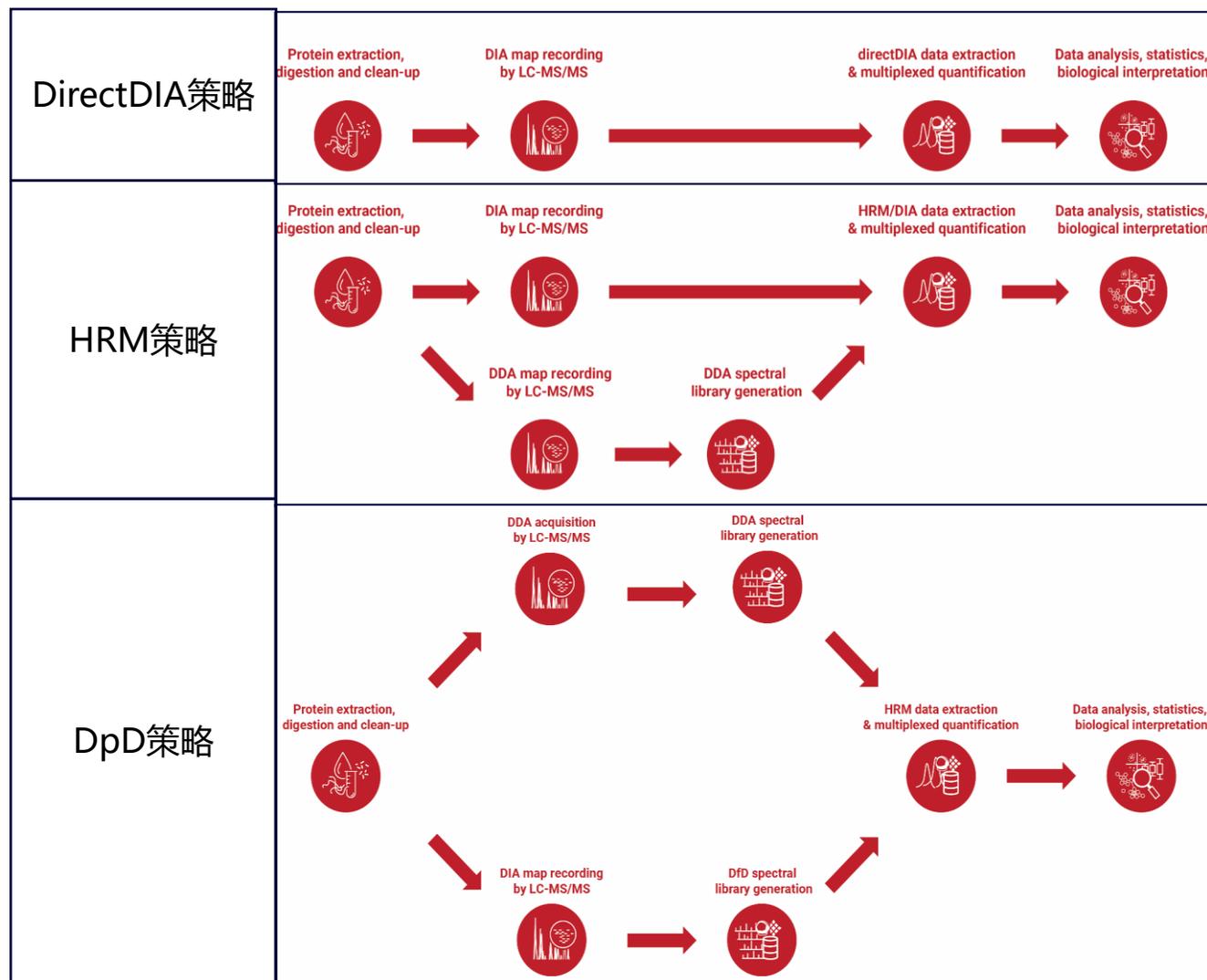
❖ 传统DIA策略——HRM策略

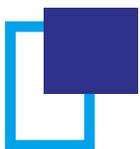
利用DDA数据建立谱图数据库，基于此谱图数据库进行DIA数据分析，利用iRT试剂进行校准

❖ DirectDIA策略 (DDA-free DIA分析策略, DfD)

❖ 基于兼具DDA和DIA 的谱图数据库 DIA数据分析 (DpD策略)

合并DDA以及DIA谱图数据库，提高鉴定率

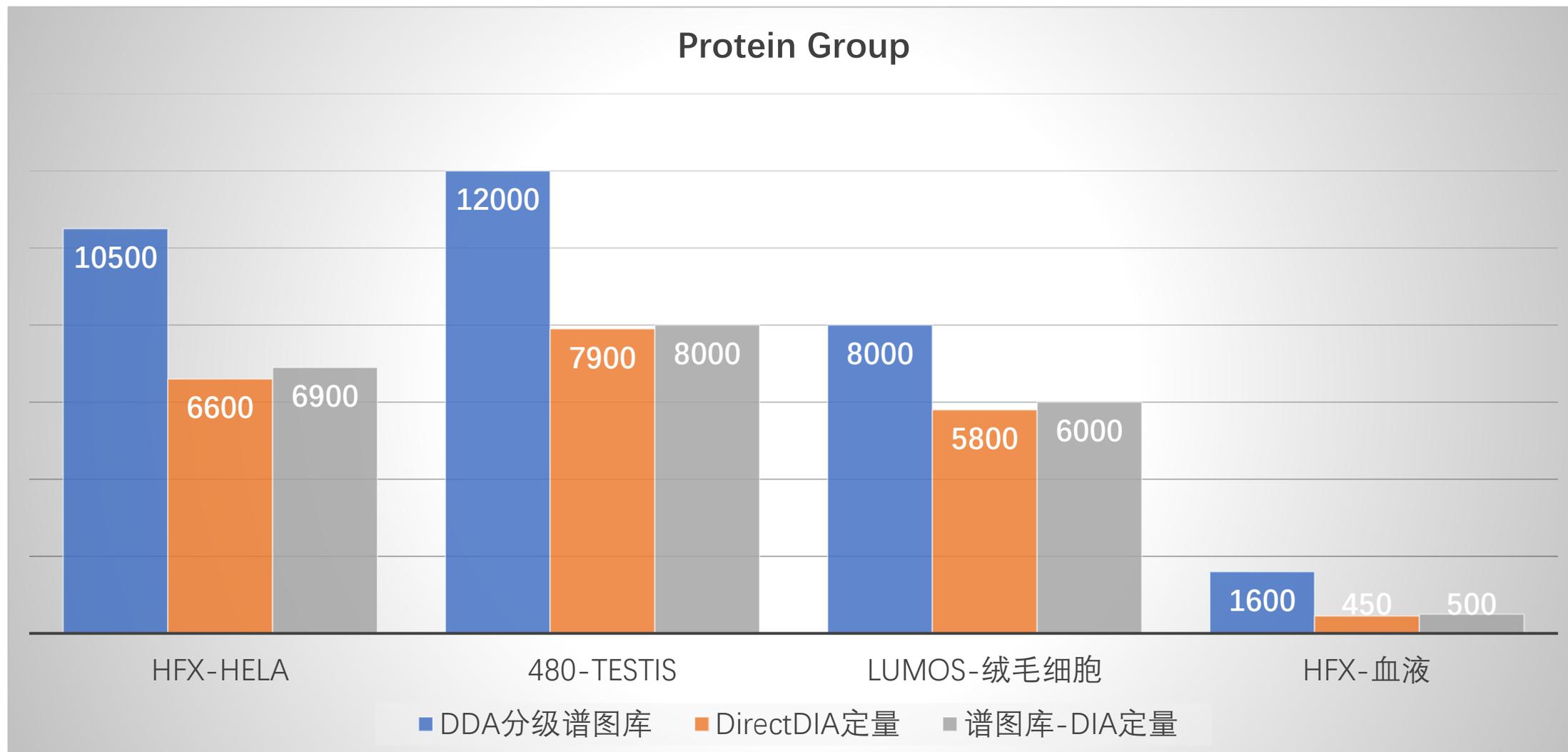


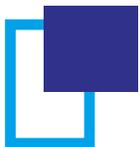


易算测试DIA数据结果

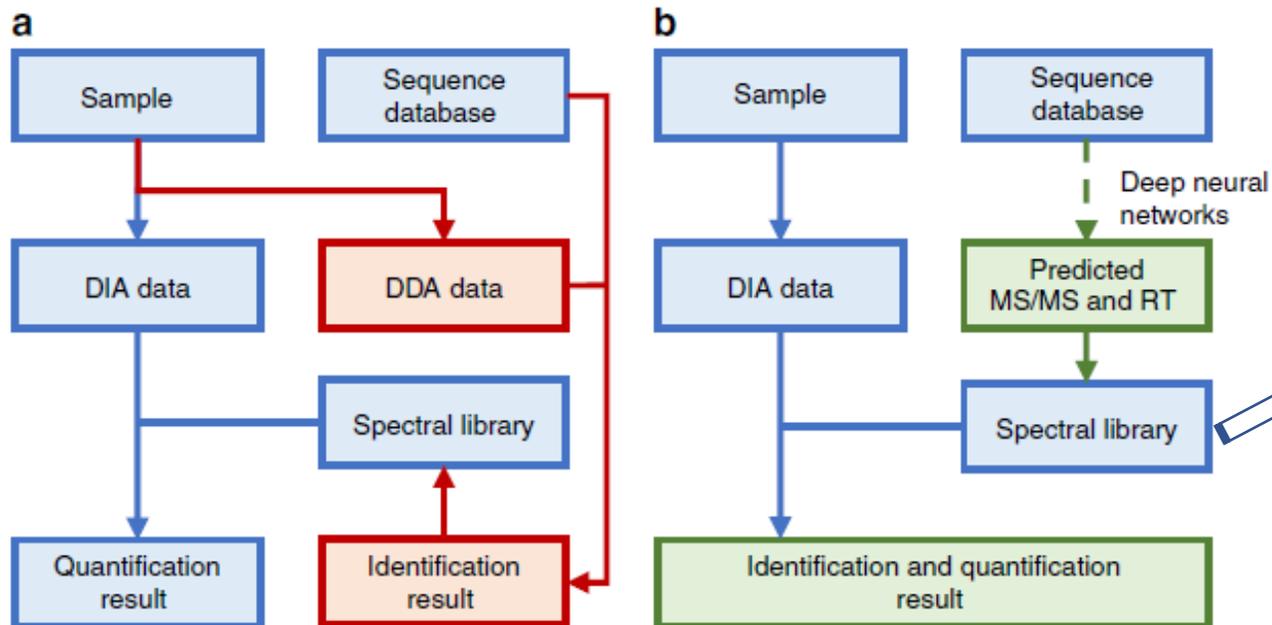


➤ 我们基于多种质谱测试发现DirectDIA拥有最佳的性价比表现





DeepDIA--In silico spectral library



- ❑ 降低时间、经济成本
- ❑ 节省稀有样本

DeepDIA Analysis

网站说明

建模预测分析

DeepDIA offers high quality MS2 train/predicted spectra for any organism fasta as well as iRT train/prediction. DeepDIA is part of the Omicsolution (www.omicsolution.com) shiny project and its models was trained on the project's/lab's High resolution mass spectrometry raw data or DDA Spectronaut search results (protein/peptide/fragment report). When using deepDIA is helpful for your research, please cite : Yang, Y., Liu, X., Shen, C. et al. In silico spectral libraries by deep learning facilitate data-independent acquisition proteomics. Nat Commun 11, 146 (2020). <https://doi.org/10.1038/s41467-019-13866-z>

复旦大学化学系乔亮课题组与生物医学研究院杨凡原课题组、上海易算生物科技有限公司沈诚频团队、澳大利亚国立大学林宇课题组合作，在《自然·通讯》上在线发表了题为“*In silico spectral libraries by deep learning facilitate data-independent acquisition proteomics*”的研究论文。该研究利用深度学习技术开发了从肽段或蛋白质序列构建谱图库的工具deepDIA，实现了不依赖于DDA的DIA数据直接分析。使用deepDIA构建专用于实验室或项目的特定仪器的模型，并且预测生成高质量的二级谱图结果和保留时间结果，效果接近DDA构建的谱图库。除此以外，我们还设计了预测肽段在质谱中的可检测性的模型，预测蛋白的理论酶切肽段的可检测性，筛选可检测性分数达到一定阈值的肽段来构建谱图库，最后可以直接将预测的数据库导入Spectronaut软件进行DIA分析。如果deepDIA对您的研究有帮助，请引用此内容：Yang, Y., Liu, X., Shen, C. et al. In silico spectral libraries by deep learning facilitate data-independent acquisition proteomics. Nat Commun 11, 146 (2020). <https://doi.org/10.1038/s41467-019-13866-z>



易算-复旦联合研发

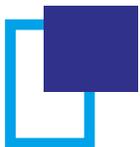
ARTICLE

<https://doi.org/10.1038/s41467-019-13866-z>

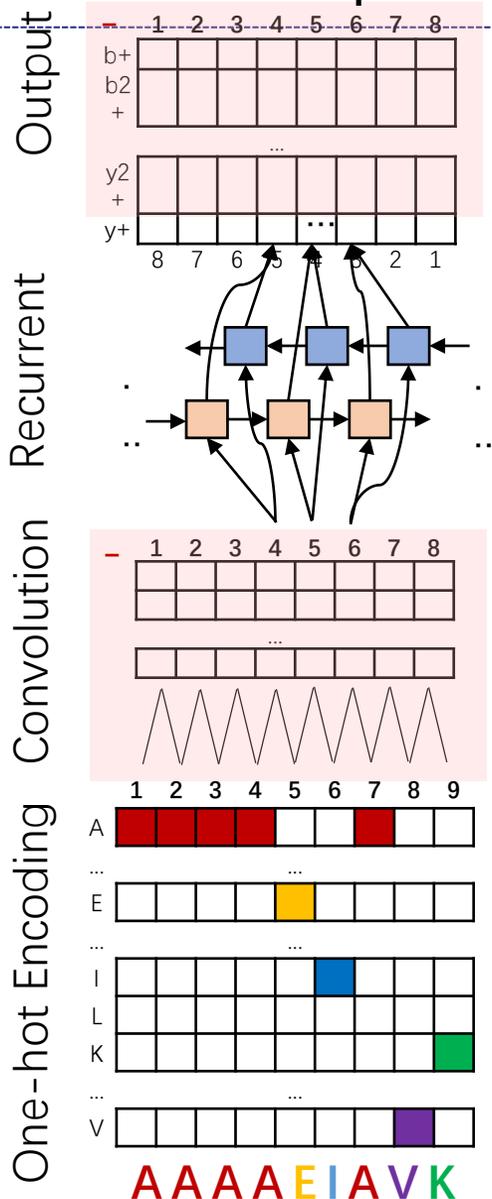
OPEN

In silico spectral libraries by deep learning facilitate data-independent acquisition proteomics

Yi Yang ^{1,4}, Xiaohui Liu ^{1,4}, Chengpin Shen ², Yu Lin ³, Pengyuan Yang¹ & Liang Qiao ^{1*}



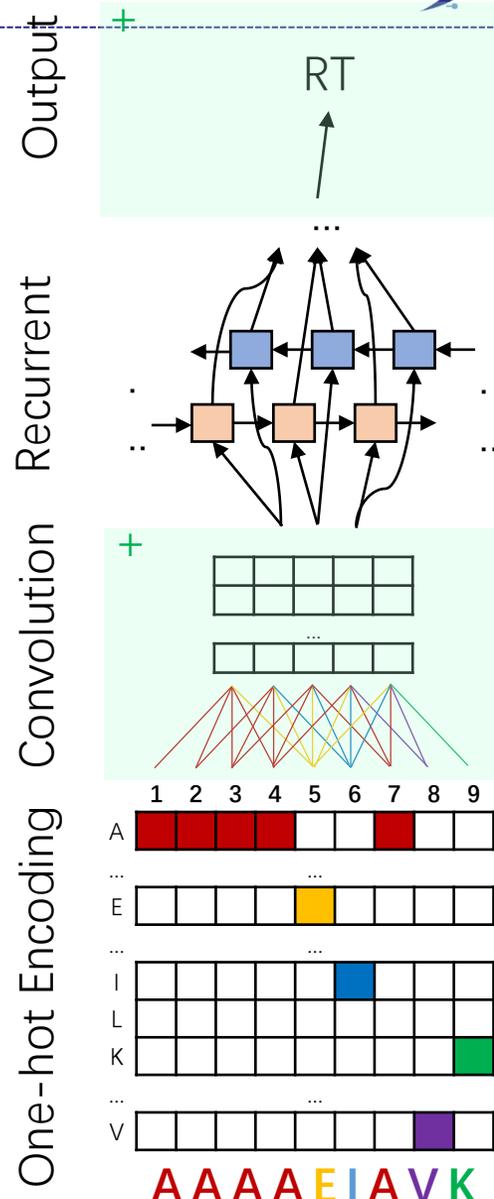
DeepDIA--In silico spectral library



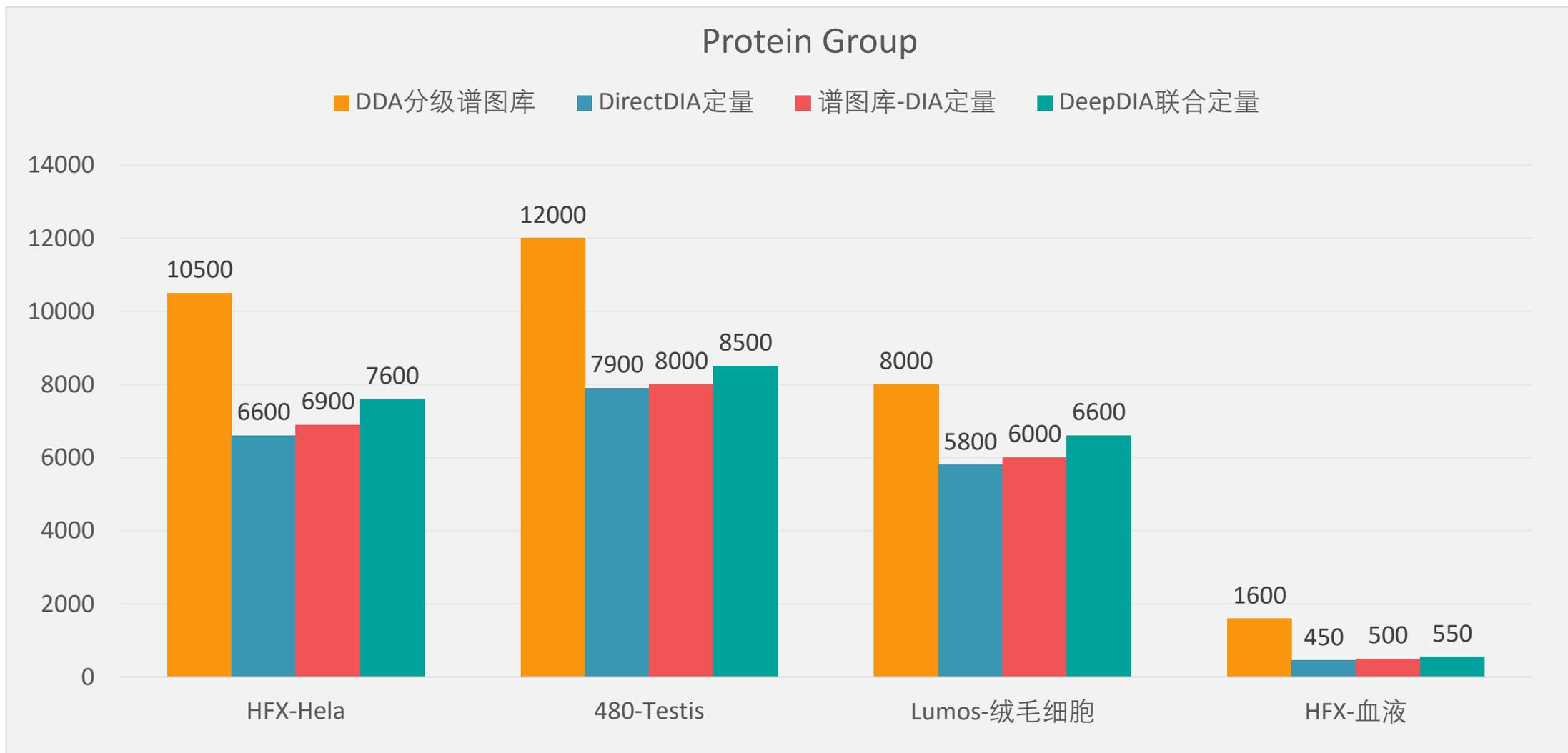
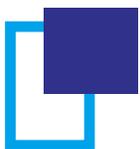
用于肽段 MS² 谱图预测的模型

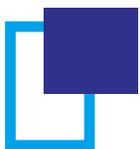


用于模型训练：一次HELA
细胞系分级建库的数据即可，
PSMs of 46,385 peptides
(38,132 doubly-charged
precursors and 18,312
triply-charged precursors)



用于肽段保留时间预测的模型

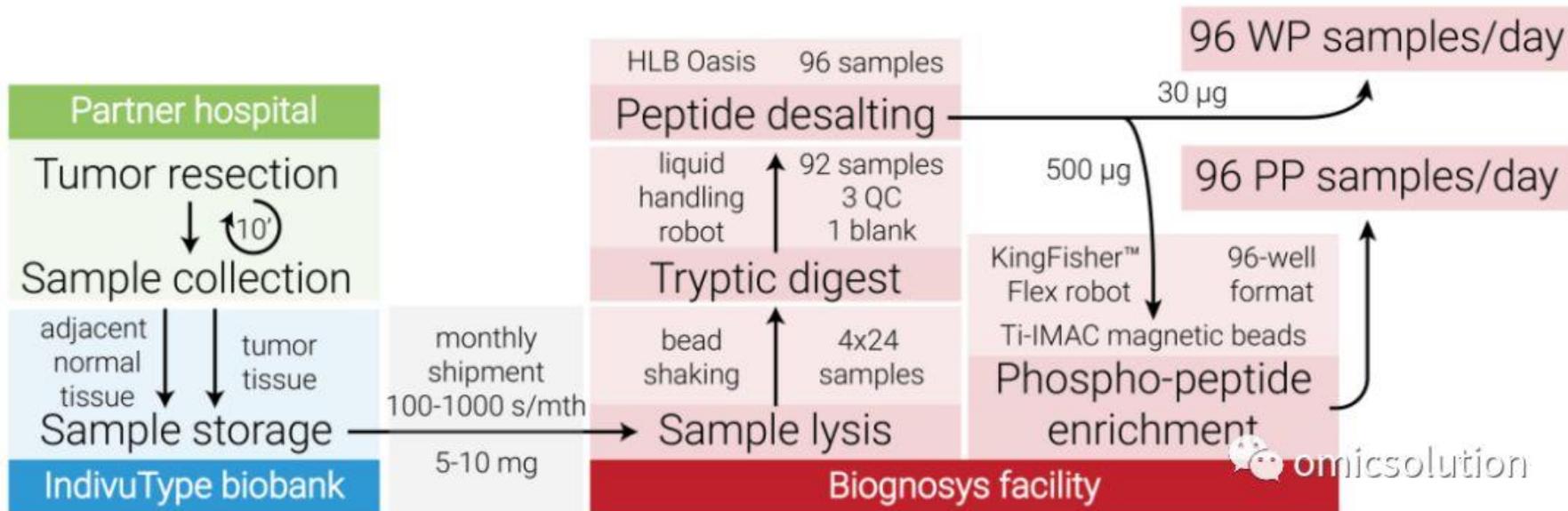




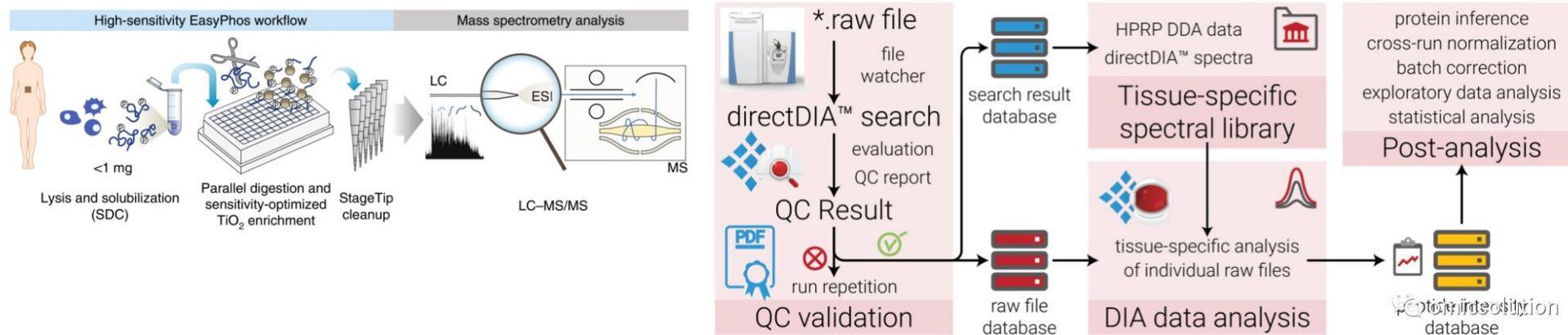
提升色谱质谱检测稳定性



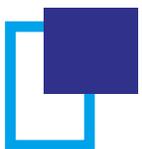
BIOGNOSYS
NEXT GENERATION PROTEOMICS



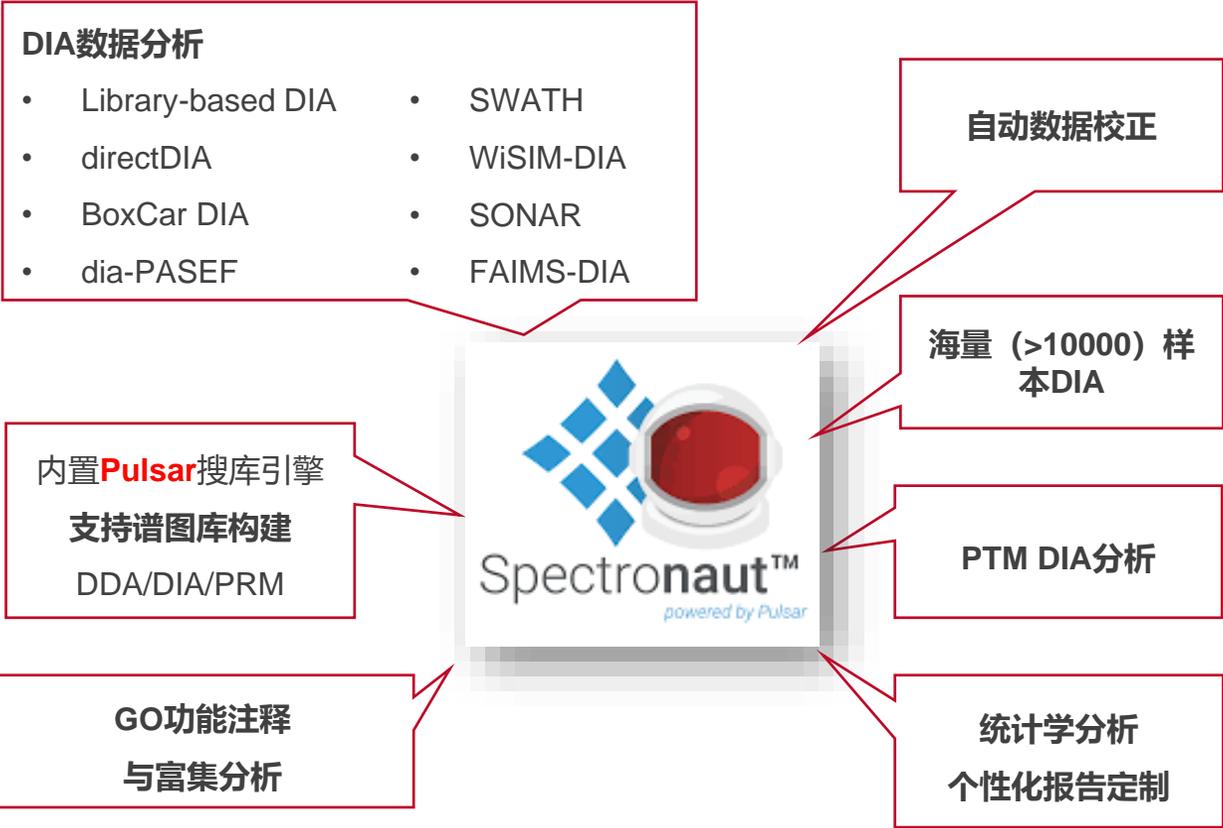
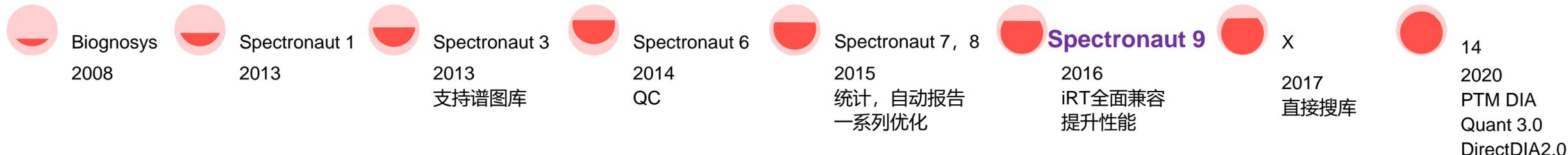
omicsolution

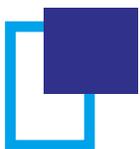


Biognosys 2019 1000例肿瘤组织



Spectronaut 介绍





搜库引擎

❖ DDA数据库检索

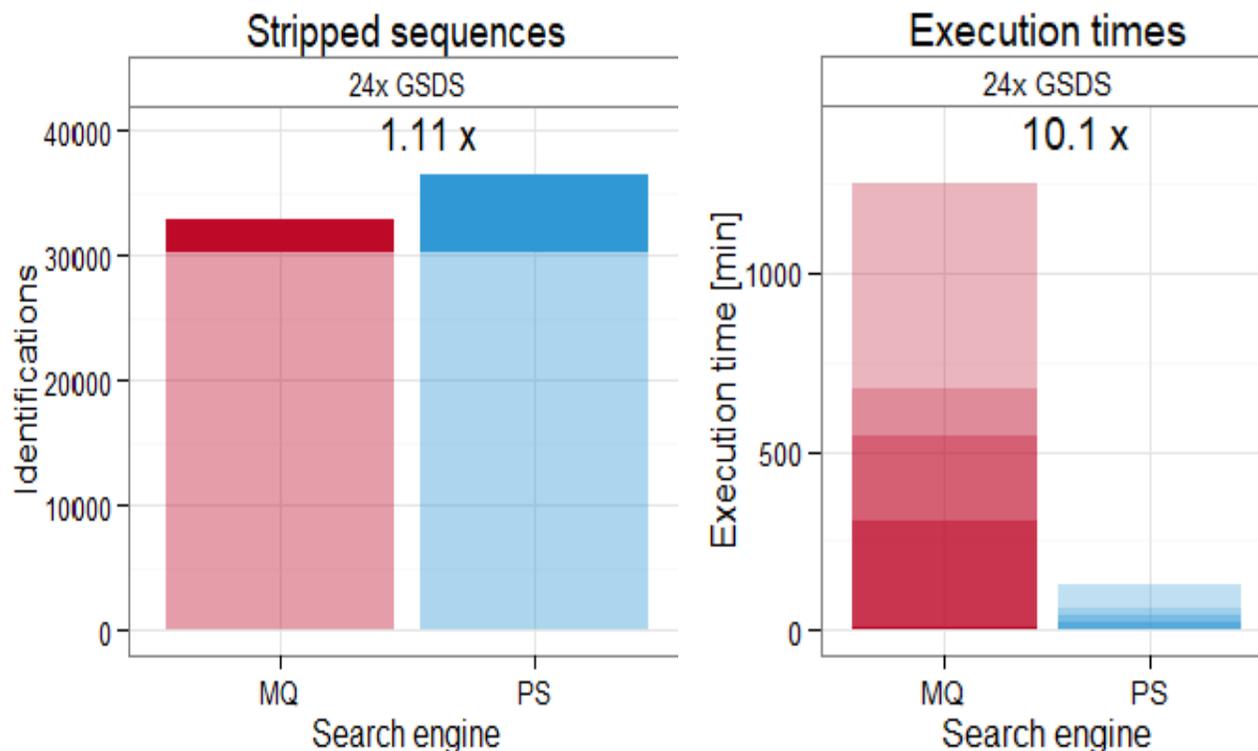
相较于MaxQuant软件，肽段数目提高6%-20%；
数据检索速度提高2-15倍

❖ DIA数据库检索

兼具DIA数据搜库功能
利用DIA数据建立谱图数据库

支持的数据类型：

- Thermo Scientific™ (DDA, DIA and PRM)
- SCIEX (DDA and DIA/SWATH™)
- Bruker (DDA/DIA)
- Waters (DDA)





基于Spectronaut的directDIA2.0

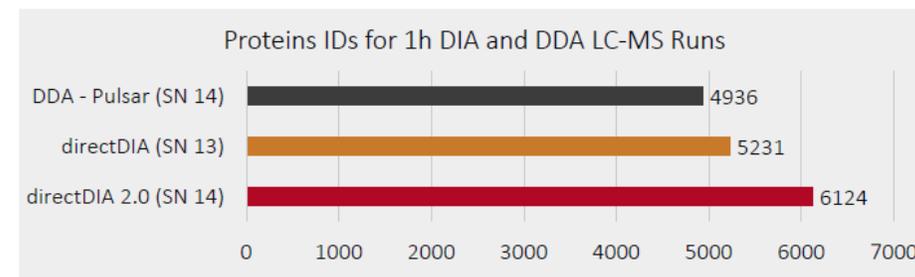
❖ DIA搜库引擎

DIA数据库检索功能使得DIA数据解析摆脱了依赖于DDA的谱图数据库，基于深度学习算法，检测灵敏度接近分级DDA建库数据

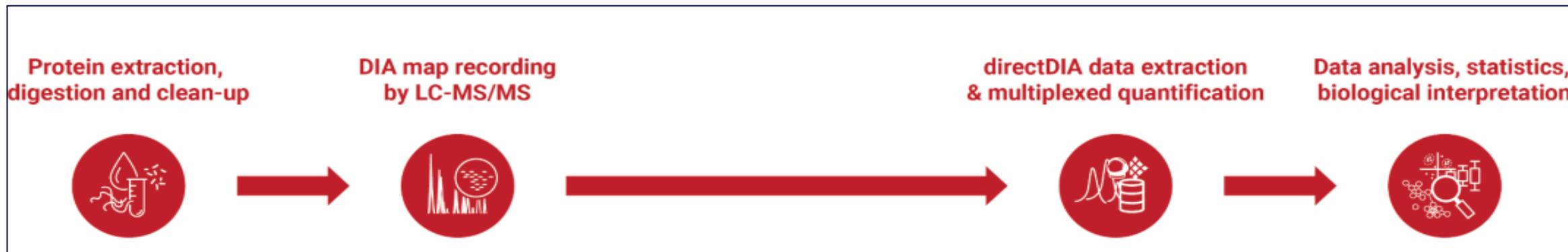
❖ 新一代Label Free蛋白质组学技术

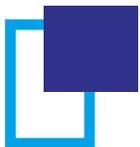
实验设计简单，类似于传统Label free技术

数据采集采用DIA模式，提高蛋白鉴定率以及实验重复性



- DDA / DIA acquisitions
- 1h gradient
- HeLa
- Thermo Exploris 480+ FAIMS





PTM DIA分析

挑战:

较宽的隔离窗口 → 复杂的混合谱

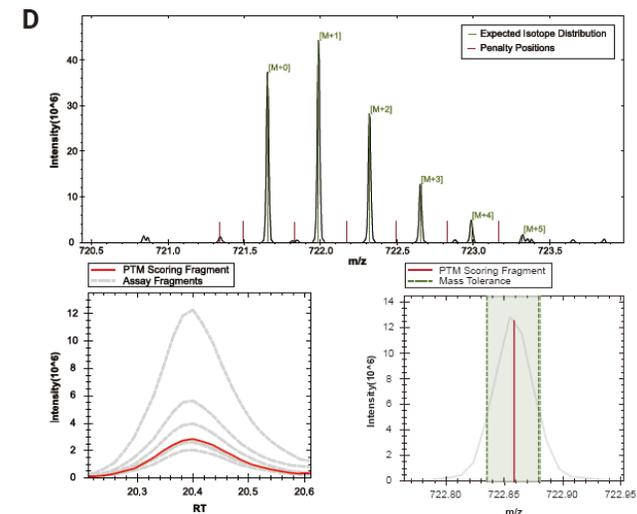
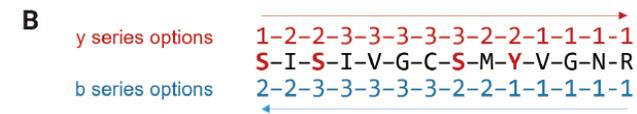
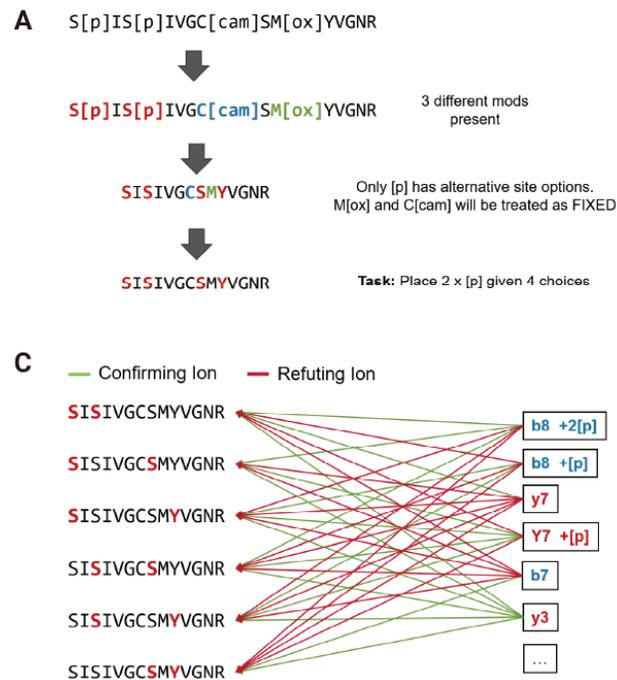


机遇:

1. 能获得所有碎片离子完整的同位素峰分布
2. 能获得所有碎片离子的洗脱曲线



Spectronaut



E

SISIVGCSMYVGNR $c_1 = \text{Score}(\text{cand1})$
 SISIVGCSMYVGNR $c_2 = \text{Score}(\text{cand2})$
 ...
 SISIVGCSMYVGNR $c_5 = \text{Score}(\text{cand5})$
 SISIVGCSMYVGNR $c_6 = \text{Score}(\text{cand6})$

$$p(\text{site}) = \frac{(c_1 + c_2 + c_3)}{c_{\text{all}}}$$

$$C_{\text{all}} = \sum_{i=1}^{N(c)} c_i$$

$$\frac{(c_3 + c_5 + c_6)}{c_{\text{all}}}$$

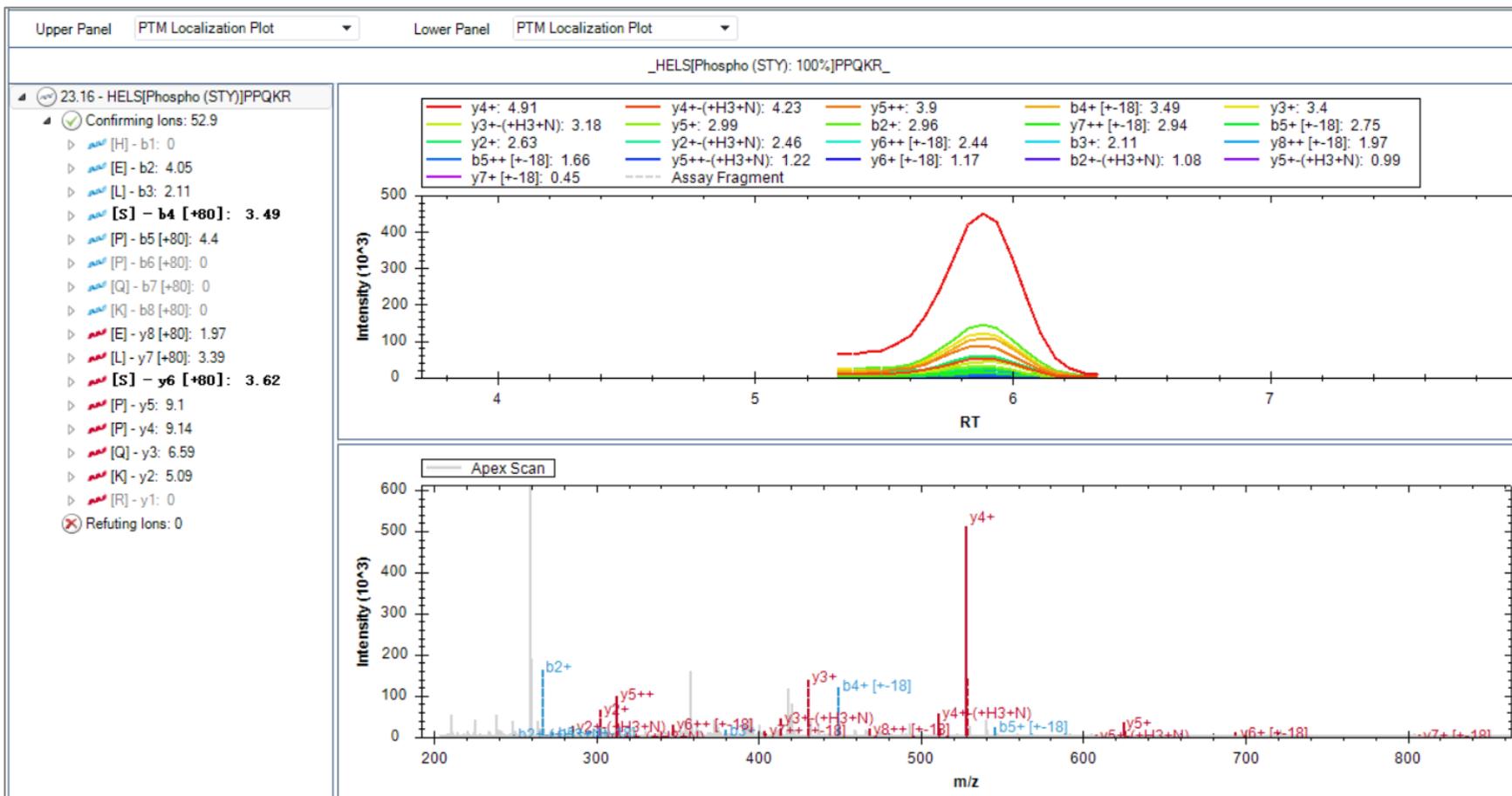
打分标准:

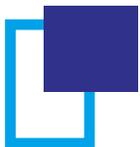
- 同位素分布
- 碎片离子强度
- 质量偏移
- XIC相关性



PTM DIA能够获得更高的组学覆盖度、定量准确度及灵敏度

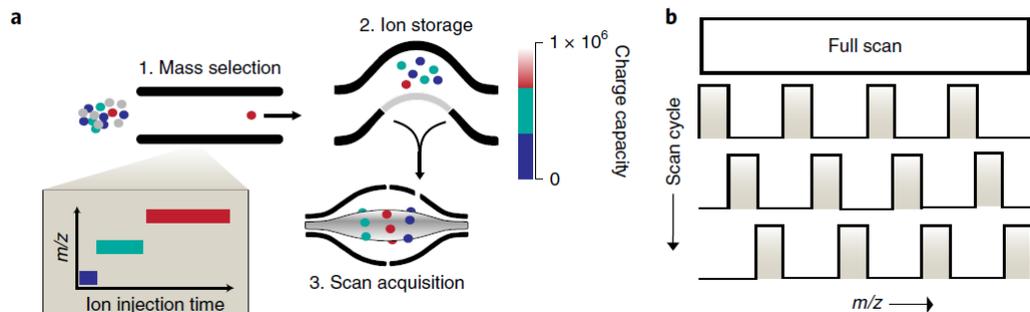
- 没有二级信号的丢失，所有precursor都进行碎裂
- MS2增加了保留时间维度的信息，提高了位点定位的可信度





BoxCar DIA

提高MS1的动态范围 → 提高DIA灵敏度

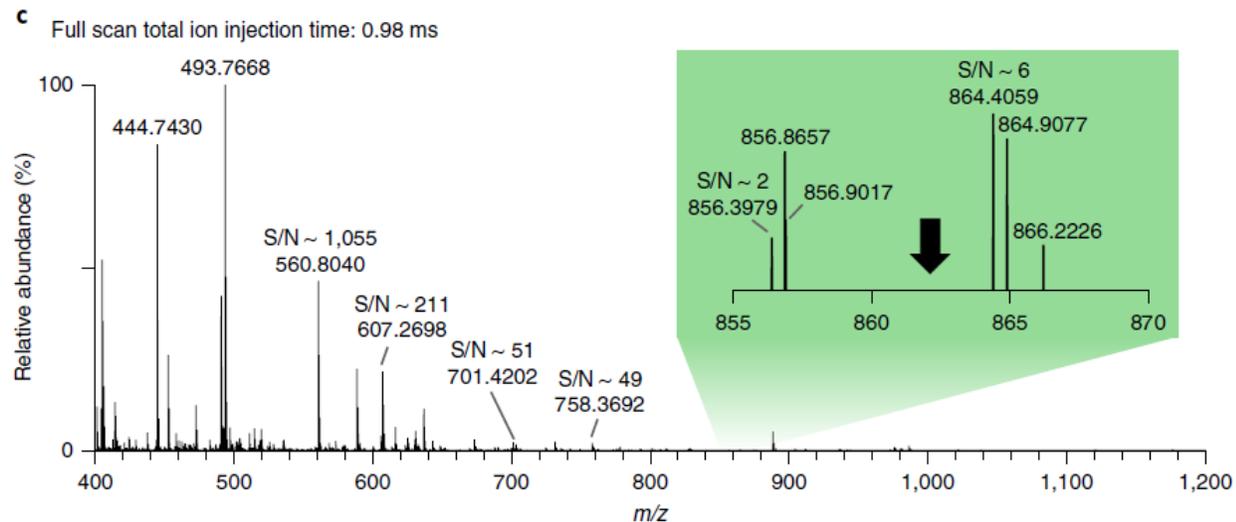


技术原理:

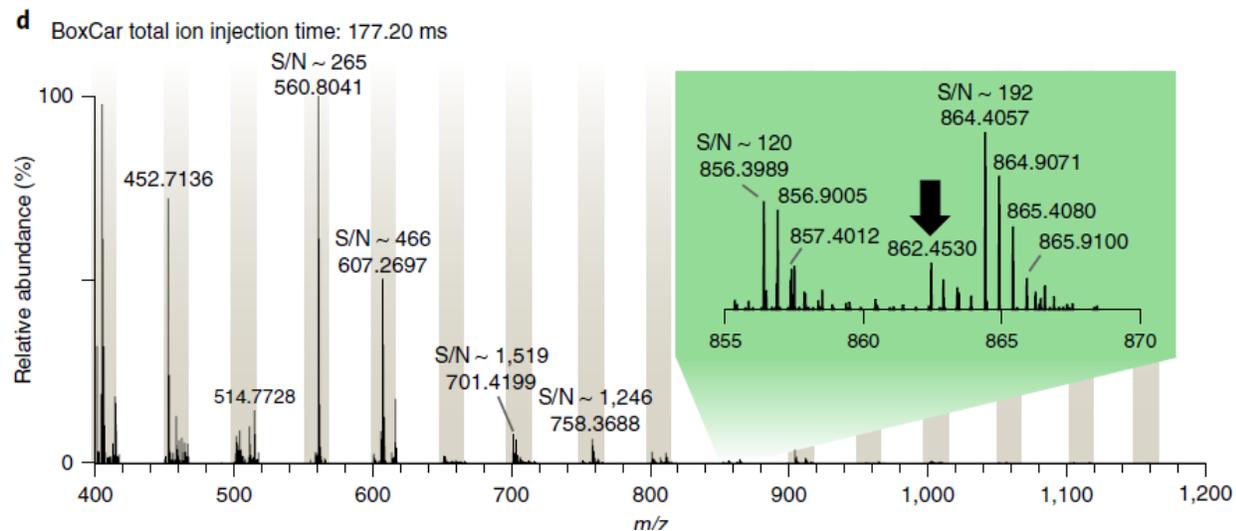
将一级扫描进行分段累积, 从而达到对低丰度信号的富集, 并且避免高丰度信号对于低丰度信号的抑制, 从而提升一级扫描的动态范围

适用范围:

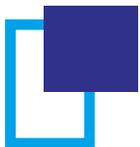
更适用于动态范围极高的样本类型的定量分析, 例如血液、脑脊液等体液样本



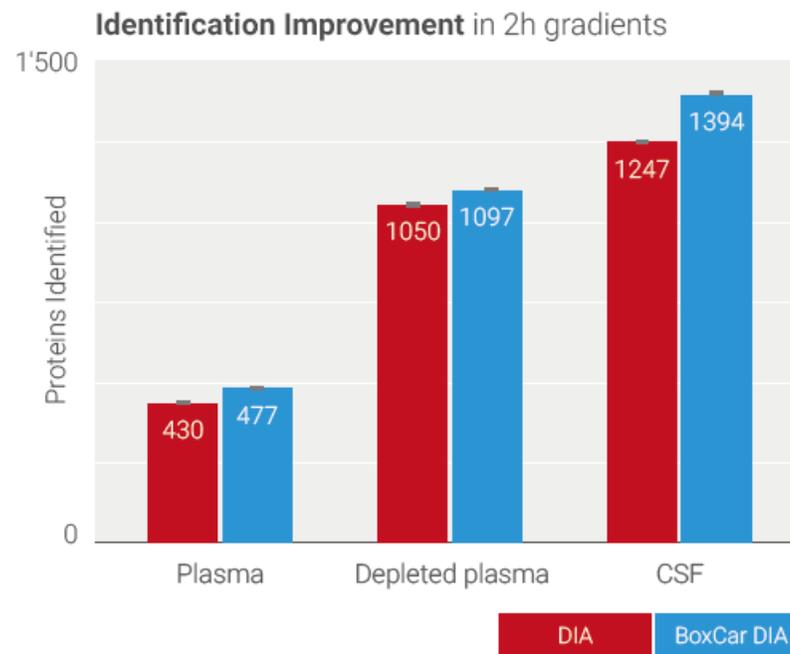
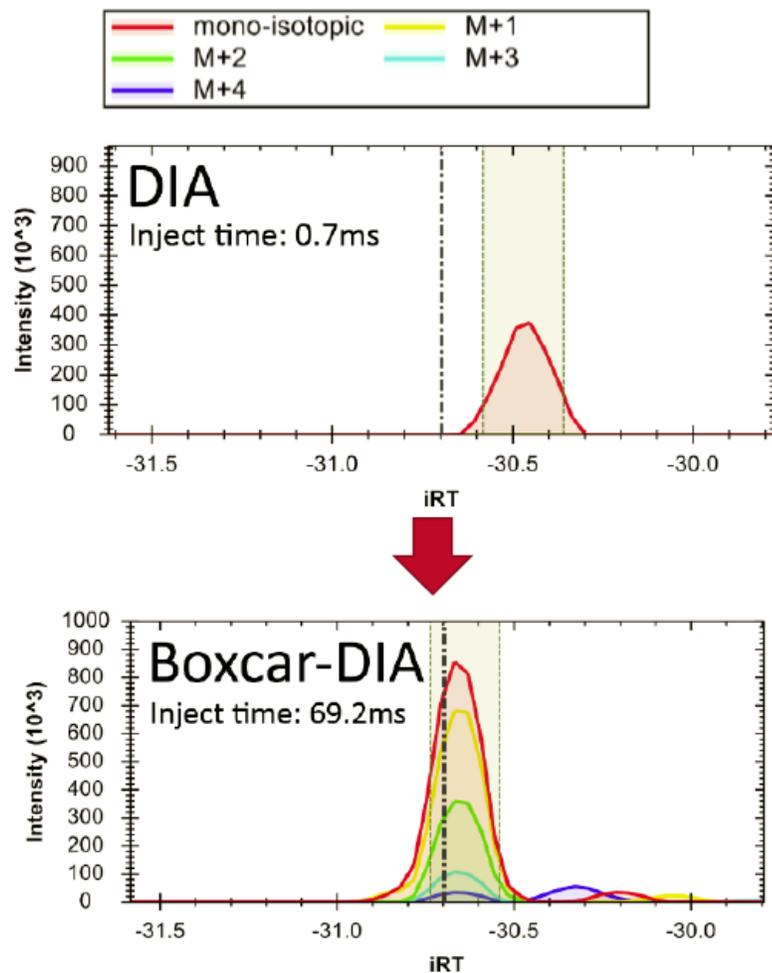
VS



*Meier, ..., Mann, Nature Methods, 2018

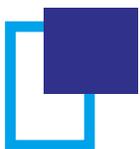


支持Boxcar DIA分析



BoxCar可提高样本鉴定的数量及定量准确性

BoxCar可采集到高于标准DIA3-5倍的母离子信号

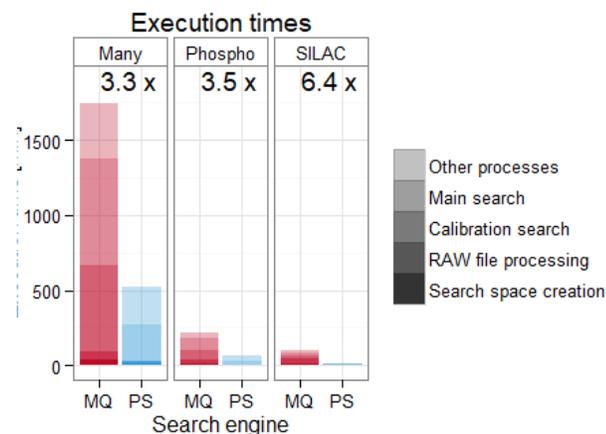
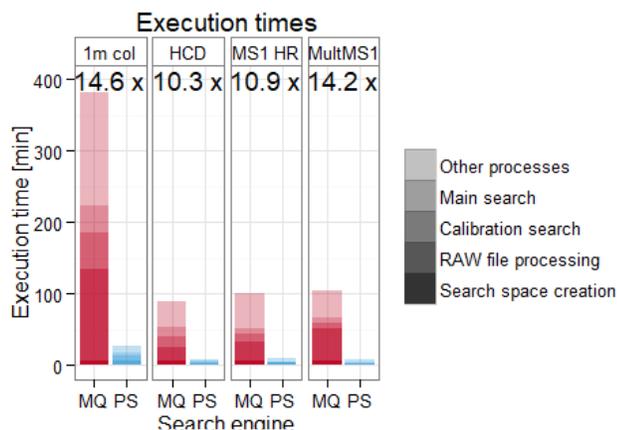
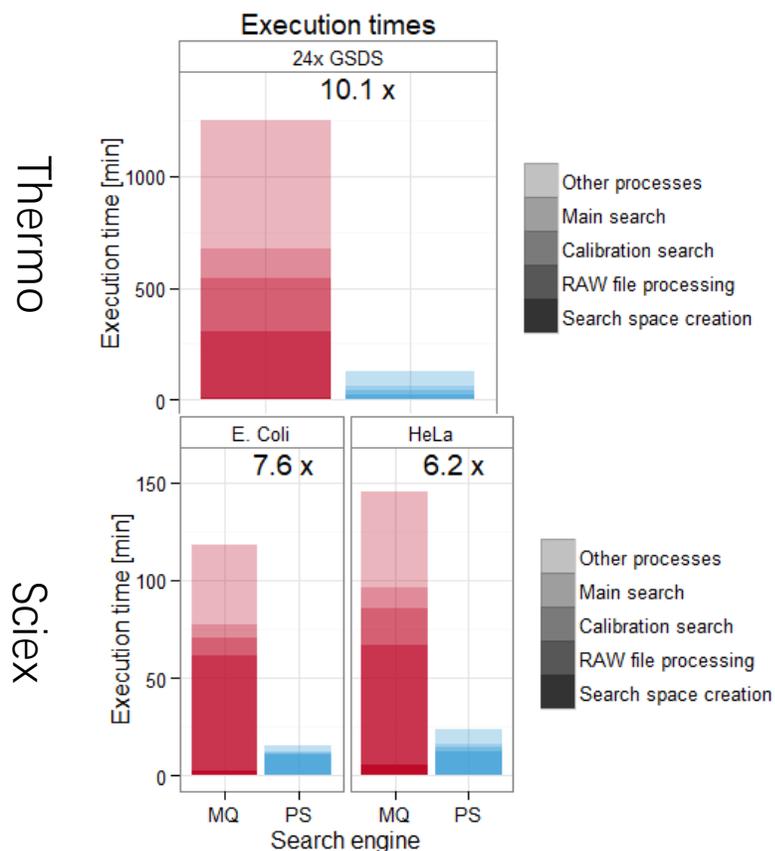


Spectronaut 技术优势——快



搜索快

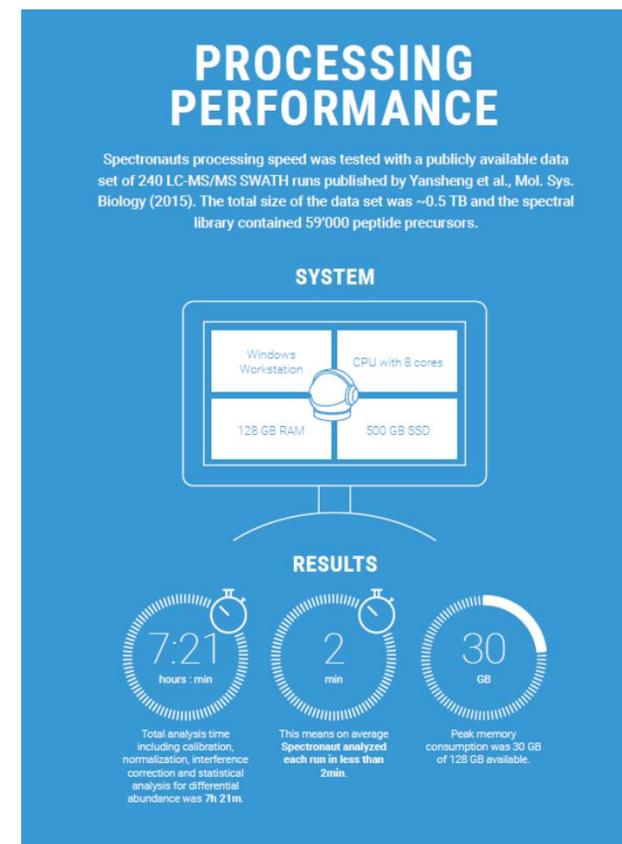
MaxQuant (>1.3) Proteome Discoverer™ (1.4-2.0)
and Protein Pilot (>4.0)

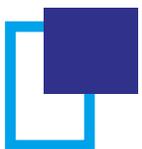


DIA数据处理快

原始数据导入时间低于10s

单个DIA数据分析时间低于2min





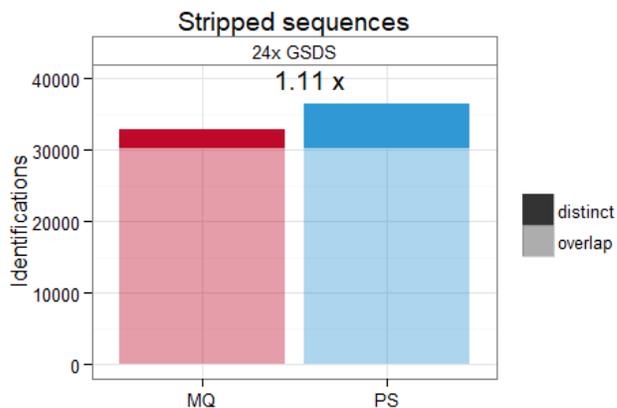
Spectronaut 技术优势——多



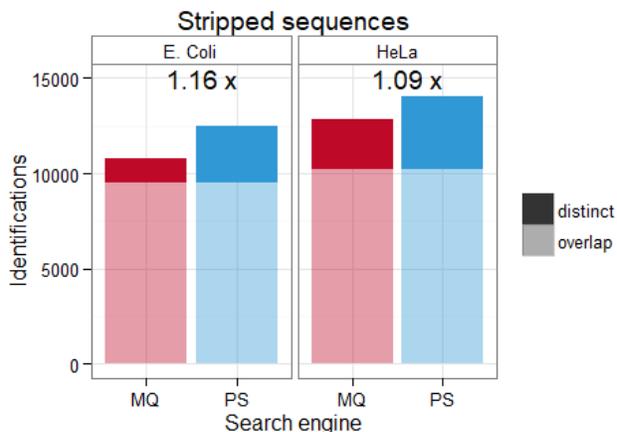
▣ 定性肽段多

相对于MaxQuant、ProteinPilot、PD、Mascot等软件肽段定性数量提高5%-10%

Thermo

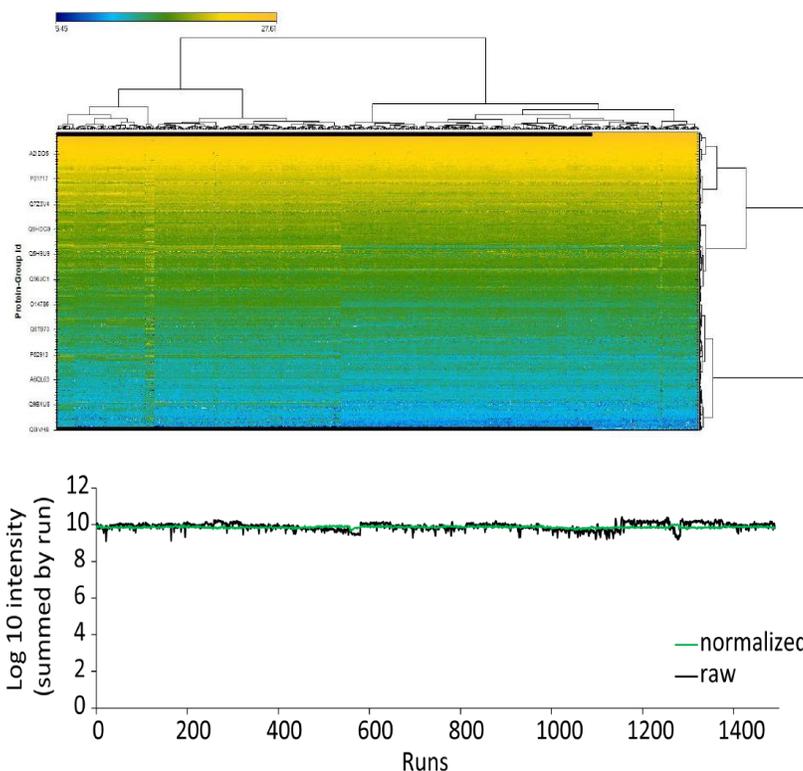


Sciex



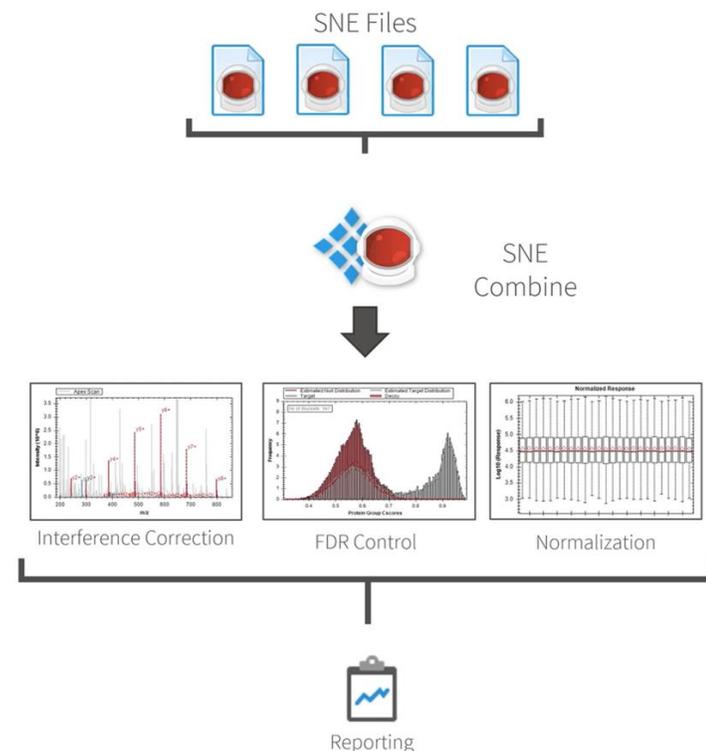
▣ 处理DIA数据多

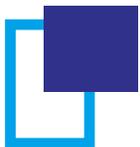
高效处理10000+样本DIA数据



▣ SNE Combine

大数据集分割分析，整合SNE生成报告





支持数据类型多

HRM-MS™
WiSIM-DIA
AIF
SWATH™

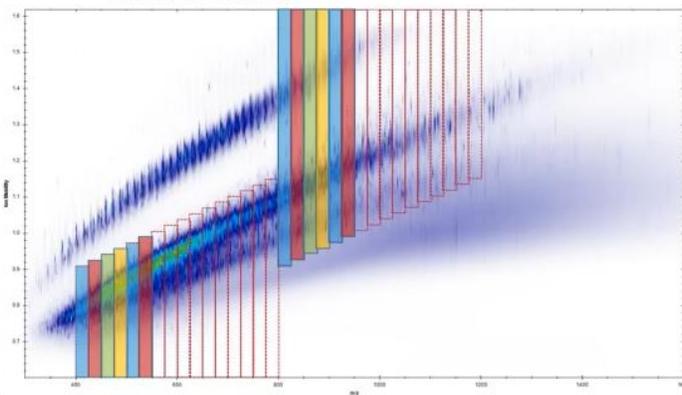
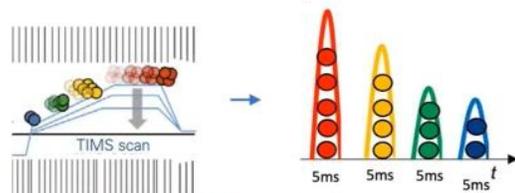
FAIMS
SONAR™
BoxCar
diaPASEF

支持仪器多

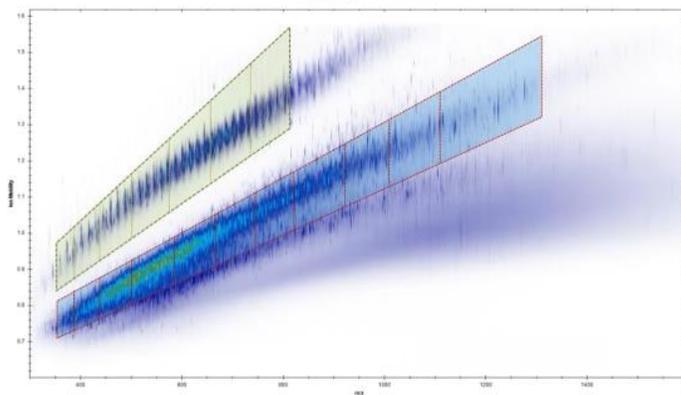
Thermo Q Exactive 系列
Thermo Orbitrap Exploris™ 480
Thermo Orbitrap Fusion系列

SCIEX TripleTOF系列 (5600, 5600+, 6600)
Bruker Q-TOF 系列、timsTOF、timsTOF Pro
WatersXevo® G2-XS QToF

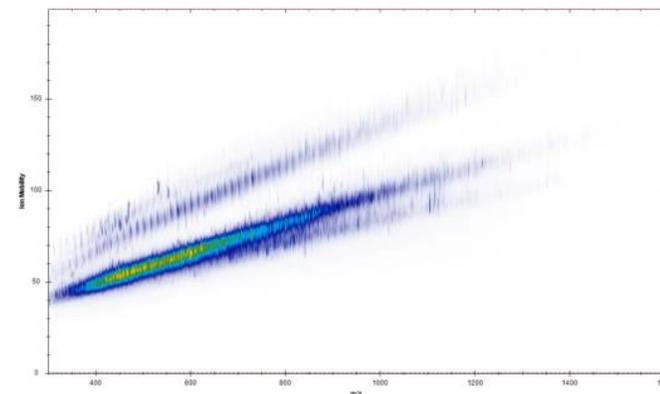
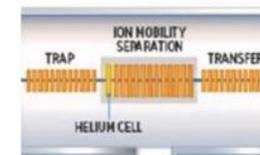
Bruker: (dia-)PASEF Ion mobility as a dimension

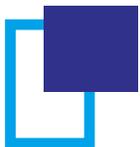


Thermo: FAIMS Pro Ion mobility filter



Waters: HDMS^E Ion mobility as a dimension





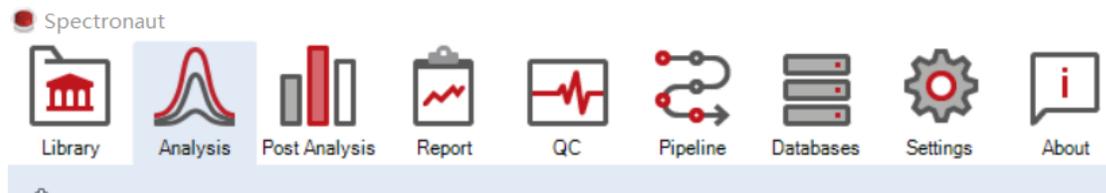
□ 操作简易

提供简易操作流程，实现一键式数据分析

□ 方法优化容易

优化DIA窗口数量及其分布

自动化动态峰提取window大小



数据库建立

数据分析以及查看

数据分组比较结果

数据导出

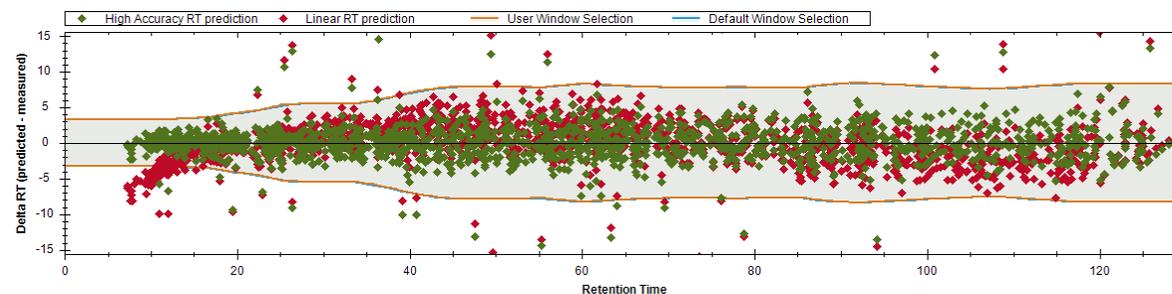
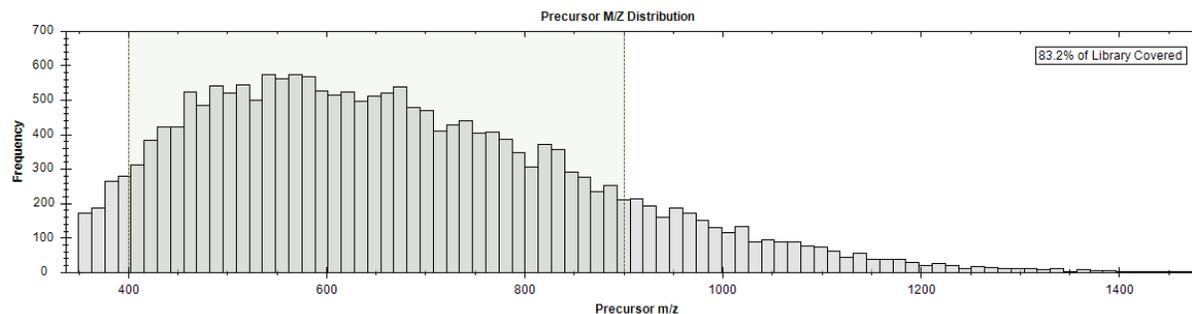
数据质量控制模块

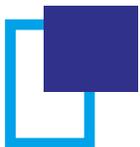
多组数据处理模块

数据库设置

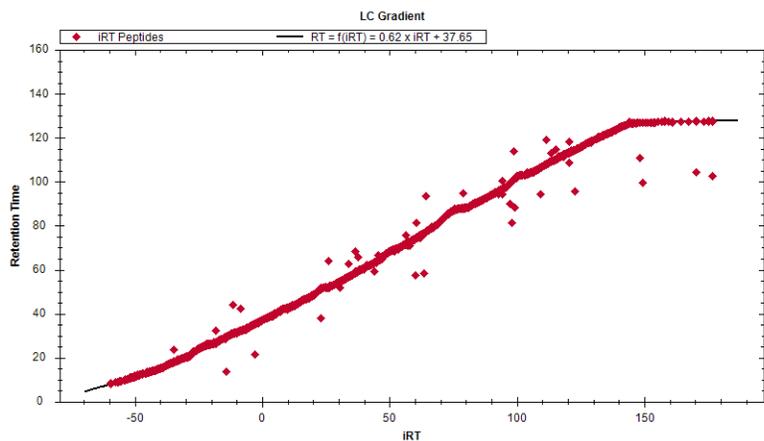
软件设置

$$DIA_{segments} \text{ optimized method} = \frac{(MS1 + MS2 \text{ segments}) * DPPP_{scouting \text{ method}}}{5} - 1$$

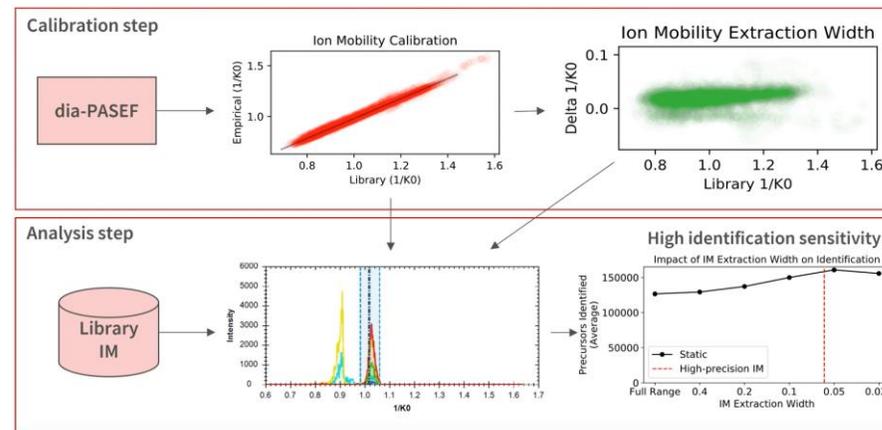




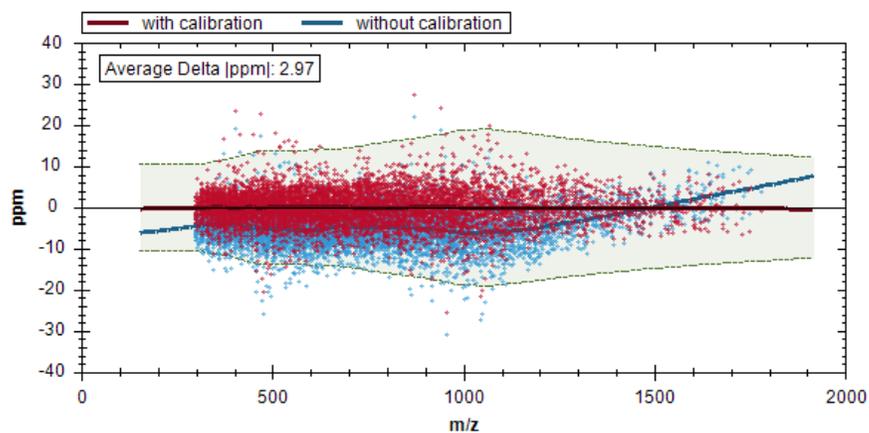
□ iRT保留时间校准



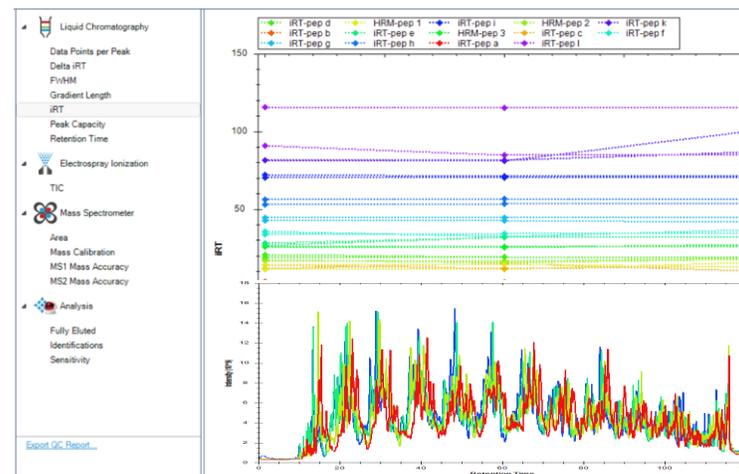
□ Ion mobility校准

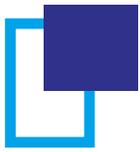


□ 质量误差校准

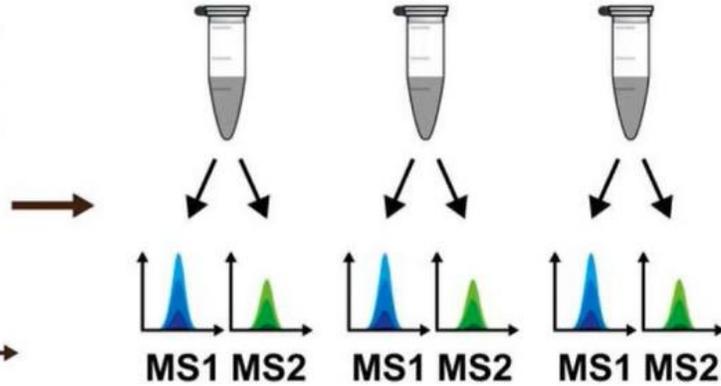
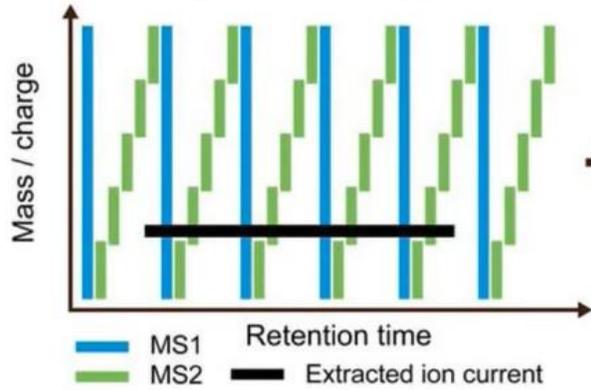


□ 实验QC

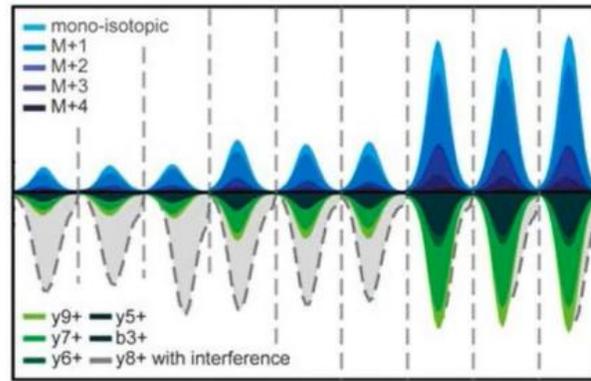




Data Independent Acquisition

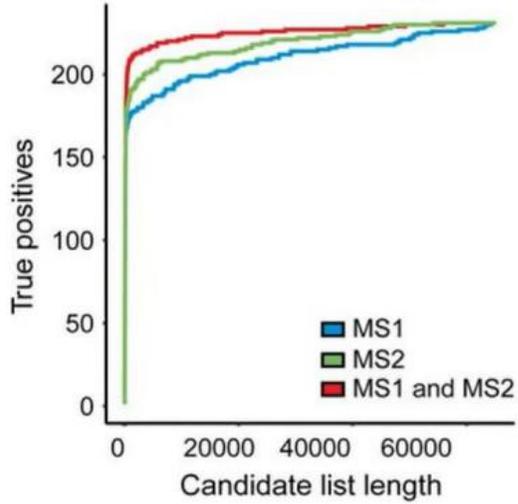


MS1



MS2

Statistical testing



QUANT 3.0

整合MS1和MS2定量结果进行差异检验提升差异分析的准确性

Huang et al., 2020

Molecular & Cellular Proteomics