

文件来源: <https://blog.csdn.net/Specally/article/details/145453939>

组学数据分析实操系列 | (一) 用 MaxQuant 软件对定量蛋白质组学数据进行数据库检索

前言

MaxQuant 是一款免费的蛋白质组学软件包，由马克斯-普朗克生物化学研究所开发。它内置了自己的搜索引擎 Andromeda，可以支持目前主流的蛋白质组学质谱仪厂商产生的原始数据格式。MaxQuant 支持标记定量和非标定量，2.0 版本后的 MaxQuant 还支持 DIA 的数据分析。MaxQuant 具备非线性质量校正和 Match Between Runs 功能，可以增加蛋白鉴定数量和提高定量准确性。下面我们主要以非标定量蛋白质组学 (LFQ) 数据为例，向大家演示 MaxQuant 数据库检索的过程。

一、非标定量数据搜库过程

01

下载 MaxQuant 软件

进入官方网站 www.maxquant.org 下载 MaxQuant 软件。软件无需安装，解压后双击 MaxQuant.exe 即可运行。

下载蛋白质序列数据库 fasta 文件

02

数据库检索 (简称搜库) 是将质谱采集到的原始谱图数据和蛋白质序列数据库进行比对解析的过程，因此大部分搜库软件都需要两类输入文件：蛋白质序列数据库和质谱原始数据，MaxQuant 也不例外。

MaxQuant 对蛋白质序列数据库 fasta 文件中的蛋白质序列进行模拟酶切和碎裂，将蛋白质序列信息转换为理论谱图数据。通过将理论谱图数据与实际采集到的原始谱图数据进行比对打分，实现搜库的过程。

蛋白质序列数据库 fasta 文件的下载方法如下：

(1) 打开 UniProt 网站 www.uniprot.org，以下载 human 蛋白质序列数据库为例，选择 Taxonomy，输入 Human 进行搜索，点击 Taxon ID 9606，进入以下页面。点击 View proteomes，

选择 Reference proteomes。

Taxonomy - Homo sapiens (species)

Download View proteins View proteomes

Mnemonic name HUMA Proteomes (2)

Taxon ID 9606 Reference proteomes (1)

Scientific name Homo sapiens

(2) 在参考蛋白质组数据库页面中，点击该数据库含有的蛋白数目 82685。

UniProt BLAST Align Peptide search ID mapping SPARQL Proteomes (organism_id:9606) Advanced | List

Proteome status

Reference proteomes (1)

Superkingdom

Eukaryota (1)

Taxonomy

9606

Filter by taxonomy

Proteomes 1 result

Download View: Cards Table Customize columns Share

Entry	Organism	Organism ID	Protein count
UP000005640	Homo sapiens (Human)	9606	82,685

BUSCO

n:13780 • primates_odb10

C:99.5% (S:37.4% D:62.1%) F:0% M:0.5%

(3) 跳转至新页面后，可以看到有些蛋白旁边带有黄色的图标，表示这些蛋白经过 Reviewed，即人工审阅，归属于 UniProt 下的 Swiss-Prot 数据库。对于研究比较多的模式物种如人和小鼠，可选择左侧 Reviewed (Swiss-Prot) 的蛋白条目，点击 Download，Format 可选择 FASTA(canonical)，Compressed 处选择 Yes，最后点击 Download 下载蛋白质序列数据库 fasta 文件。对于其它物种，Reviewed 蛋白可能偏少，一般优先选择 Reference Proteome 进行下载。

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB proteome:UP000005640 Advanced | List Search Help

Status

Reviewed (Swiss-Prot)

Popular organisms

Human (20,413)

Taxonomy

Filter by taxonomy

Group by

Taxonomy

UniProtKB 20,413 results

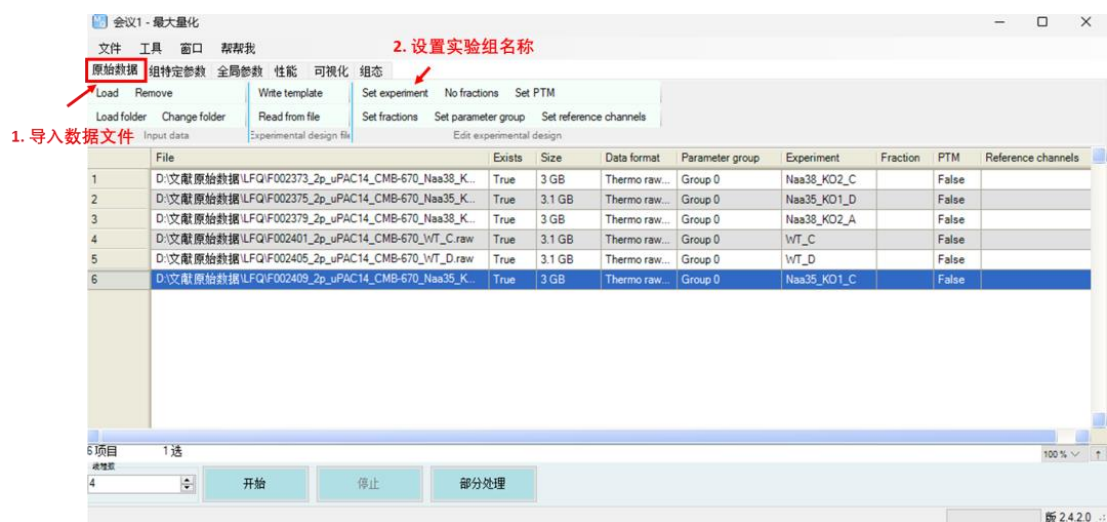
Download Add View: Cards Table Customize columns Share

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
A0A0C5B5G6	MOTSC_HUMAN	Mitochondrial-derived peptide MOTS-c[...]	MT-RNR1	Homo sapiens (Human)	16 AA
A0A1B0GTW7	CIROP_HUMAN	Ciliated left-right organizer metalloproteinase[...]	CIROP, LMLN2	Homo sapiens (Human)	788 AA
A0JNWS	BLTB3_HUMAN	Bridge-like lipid transfer protein family member 3B[...]	BLTP3B, KIAA0701, SHIP164, UHRF1BP1L	Homo sapiens (Human)	1,464 AA
A0JP26	POTB3_HUMAN	POTE ankyrin domain family member B3	POTB3	Homo sapiens (Human)	581 AA
A0PK11	CLRN2_HUMAN	Clarin-2	CLRN2	Homo sapiens (Human)	232 AA

打开 MaxQuant 软件，导入质谱原始数据文件

03

点击 Load 导入数据文件，或点击 Load folder 导入仅包含数据文件的文件夹。目前蛋白质组学质谱仪厂商的原始数据格式有 .raw (Thermo 公司)、.d (Bruker 公司)、.WIFF 和 .wiff.scan (SCIEX 公司) 等。可通过 Set experiment 设置实验组名称。这里我们用到的演示数据是从 ProteomeXchange 官网下载的 <https://proteomecentral.proteomexchange.org>，蛋白质数据集为 PXD034104，数据来源具体实验详见文献“N-terminal acetylation shields proteins from degradation and promotes age-dependent motility and longevity”。

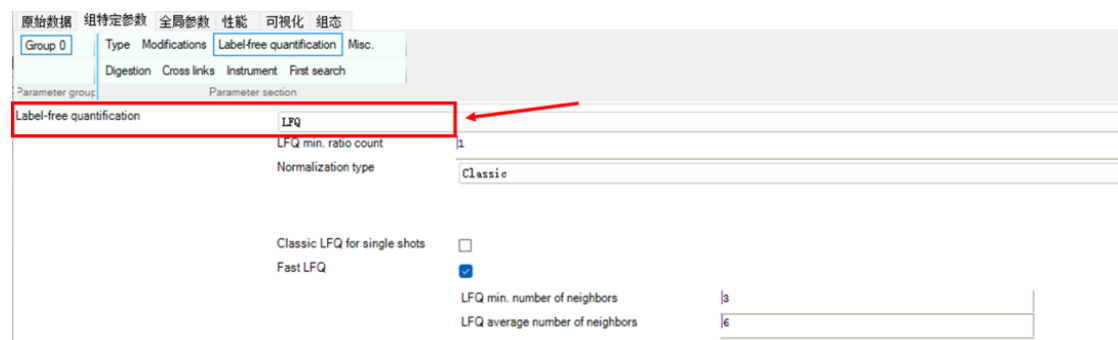


04

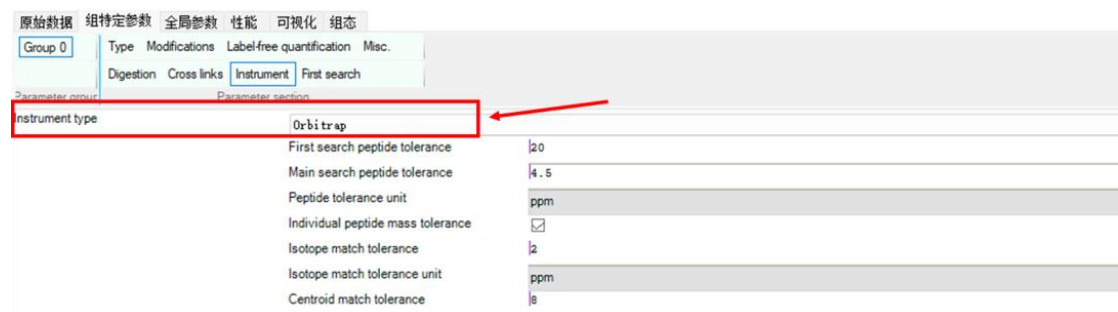
设定组特定参数

对于常规的非标定量检测，可直接使用软件默认的参数，包括：Type——选择 Standard，Modifications——选择固定修饰 Carbamidomethyl (C)、可变修饰 Oxidation (M)和 Acetyl (Protein N-term)，Digestion——选择 Specific 和 Trypsin/P。

需要修改的参数：Label-free quantification——选择 LFQ。



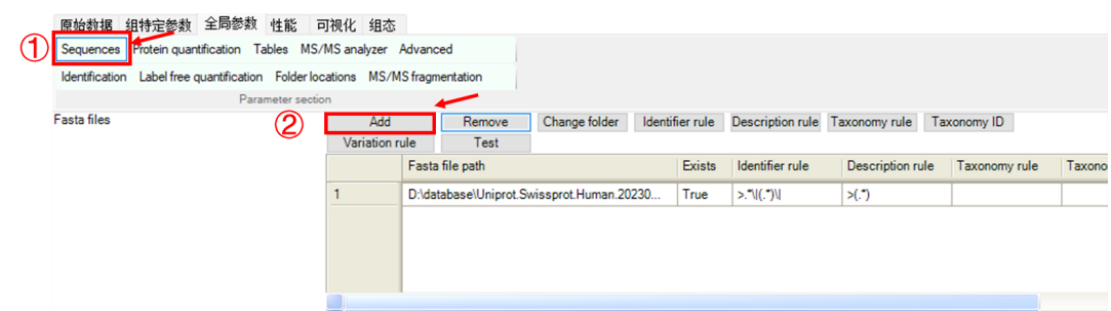
Instrument——根据实际质谱检测的仪器选择 Instrument type。



设定全局参数

05

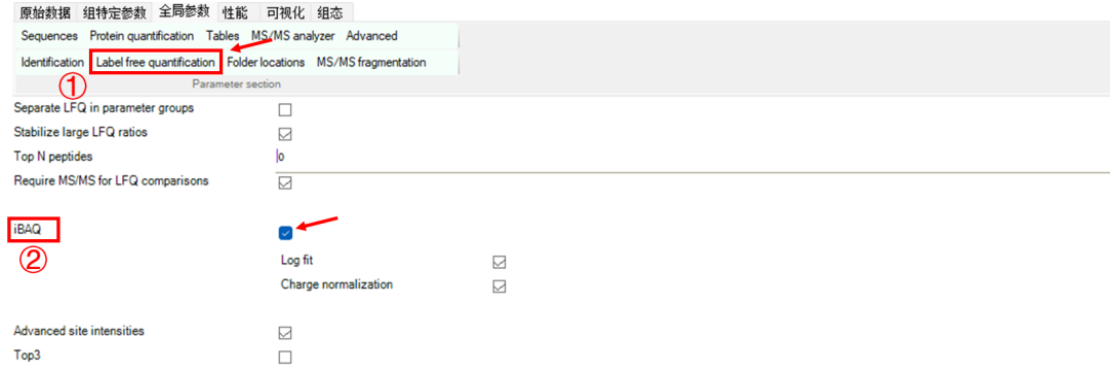
Sequence——导入参考的 Human 蛋白质序列数据库 fasta 文件。



Identification——勾选 Match between runs。由于缺失值的存在会削弱生物样本或实验条件之间真实定量差异的能力，Match between runs 功能可以改善缺失值问题。



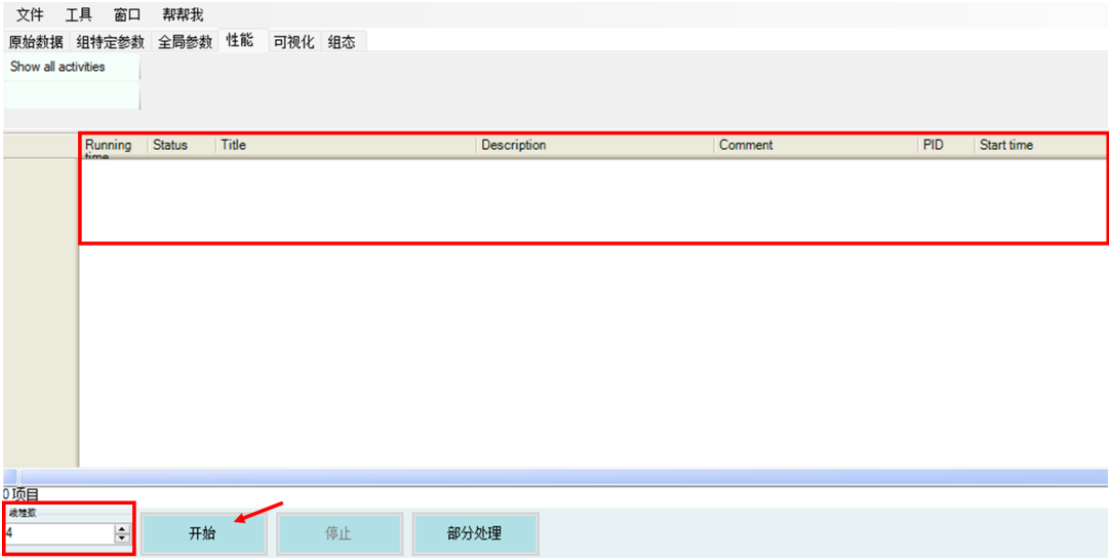
Label-free quantification——勾选 iBAQ。iBAQ 是基于 Intensity 的强度值，除以该蛋白的理论可被检测的肽段数目计算而来的定量值，主要用于不同蛋白的相互比较。



06

性能

左下角线程数可根据电脑实际线程进行修改，点击开始，MaxQuant 开始搜库，该页面会显示详细的搜库状态。



结果文件

07

搜库完成后，主要结果在 Combined 文件夹下的 txt 子文件夹，包含下图中的文件，常用的文件有蛋白结果信息 proteinGroups.txt 和肽段结果信息 peptides.txt。如果对于结果中的表头信息不太清楚，可查阅 tables.pdf 文件，其解释每个 txt 结果文件中各个表头的含义。

此电脑 > DATA1 (D:) > 文献原始数据 > LFQ > combined > txt

在 txt 中搜索

名称	修改日期	类型	大小
allPeptides	2023/11/10 7:24	文本文档	497,305 KB
evidence	2023/11/10 7:22	文本文档	108,699 KB
matchedFeatures	2023/11/10 7:22	文本文档	0 KB
modificationSpecificPeptides	2023/11/10 7:22	文本文档	18,333 KB
ms3Scans	2023/11/10 7:24	文本文档	0 KB
msms	2023/11/10 7:23	文本文档	565,585 KB
msmsScans	2023/11/10 7:24	文本文档	248,153 KB
mzRange	2023/11/10 7:24	文本文档	1,048 KB
Oxidation (M)Sites	2023/11/10 7:22	文本文档	1,162 KB
parameters	2023/11/10 7:22	文本文档	4 KB
peptides	2023/11/10 7:22	文本文档	21,356 KB
proteinGroups	2023/11/10 7:22	文本文档	7,308 KB
summary	2023/11/10 7:22	文本文档	4 KB
tables	2023/11/10 7:22	Foxit PDF Reade...	45 KB

二、TMT 标记定量数据搜库过程

如果是 TMT 标记定量的蛋白质组学数据，设置上有一些不同之处，我们这里把需要重点调整的参数拎出来，其它基本操作过程与上述步骤类似。

原始数据：导入质谱原始数据，如有多个组分，通过 Set fractions 进行设置；

组特定参数：此处根据标记实验和质谱采集过程选择对应参数，如 Type 选择 Reporter MS2，Isobaric labels 选择使用的标记试剂（选择后只保留使用的标记通道即可），其余参数可保持默认值；

原始数据	组特定参数	全局参数	样品	性能	可视化	组态
Group 0	Type Modifications Label-free quantification Misc.					
	Digestion Cross links Instrument First search					
Parameter group	Parameter section					
Type	Reporter MS2					
Isobaric labels						
	Add Remove Edit Import Export Change 4plex 8plex					
	2plex TMT 6plex TMT 8plex TMT 10plex TMT 11plex TMT 16plex 18plex iodo6plexT					
	Internal label Terminal label CF -2 [%] CF -1 [%] CF +1 [%] CF +2 [%] TMT-like					
1	TMT10plex-Lys126C TMT10plex-Nter126C 0 0 0 0 True					
2	TMT10plex-Lys127N TMT10plex-Nter127N 0 0 0 0 True					
3	TMT10plex-Lys127C TMT10plex-Nter127C 0 0 0 0 True					
4	TMT10plex-Lys128N TMT10plex-Nter128N 0 0 0 0 True					
5	TMT10plex-Lys128C TMT10plex-Nter128C 0 0 0 0 True					

全局参数：在 Sequences 中选择对应的 fasta 数据库文件，其余可保持默认值。

以上就是使用 **MaxQuant** 软件进行定量蛋白质组学数据搜库的过程啦，这一步我们只是拿到了基本的定量数据，下一篇我们继续讨论拿到 **MaxQuant** 搜库结果后该如何展开下一步的分析。小伙伴有疑问的话可以在文章下留言哦！

参考资料

1.Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nature Protocols 11, 2301 – 2319 (2016).

2.Sinitcyn, P., et al., MaxDIA enables library-based and library-free data-independent acquisition proteomics. Nature Biotechnology 39, 1563-1573 (2021).

3.UniProt 数据库网站 www.uniprot.org

4.Varland, Sylvia et al. N-terminal acetylation shields proteins from degradation and promotes age-dependent motility and longevity. Nature communications 14(1), 6774 (2023).

版权声明：本文为博主原创文章，遵循 CC 4.0 BY-SA 版权协议，转载请附上原文出处链接和本声明。

原文链接：<https://blog.csdn.net/Specally/article/details/145453939>