

Skyline 中的 DIA/SWATH 数据分析

注意：本教程使用的数据采集于 Thermo Q-Exactive Plus 仪器。如果使用的数据采集于 Sciex TripleTOF 6600，请参阅本教程的另一版本 Skyline DIA TTOF。

本教程使用的数据是一个 SWATH-MS 数据集，它通过使用 18 个可变宽度窗口母离子隔离方案（Bruderer R. 等人，MCP 2015）和 1 小时梯度，采集于 QqOrbi 仪器（Q-Exactive Plus、Thermo）。通过本教程，您将了解如何使用 Skyline 对肽段以及推断的蛋白质检测和定量进行靶向分析。

用于采集数据的样品复制于 [LFQBench 研究](#)。在这项研究中，用于基准化定量分析的样品由三个物种的蛋白质组以确定的比例混合组成。

为了处理 DIA 数据集，您首先要在 Skyline 中设置所有相关参数；然后，您需要从原始数据文件提取定量的信息；最后，您要通过导入 DDA 数据的搜索结果创建一个质谱图库，从而产生肽段的查询参数。



[图片取自

Navarro, P. et al. A multicenter study benchmarks software tools for label-free proteome quantification. *Nature Biotech* 34, 1130–1136 (2016)

和

Bruderer, et al. Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues.” *Molecular & Cellular Proteomics* 14, (2015)]

入门指南

开始本教程之前，请下载下列 ZIP 文件：

<https://skyline.ms/tutorials/DIA-QE.zip>

将文件解压到您电脑上的某个文件夹，比如：

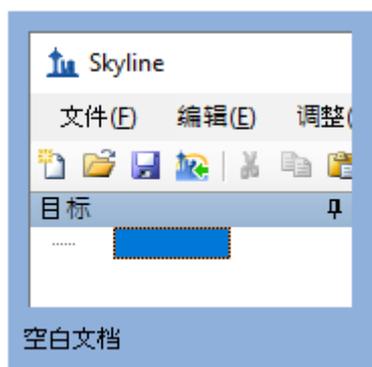
C:\Users\brendanx\Documents

该操作将创建一个新文件夹：

C:\Users\brendanx\Documents\DIA-QE

如果您在开始学习本教程之前就一直在用 Skyline，最好将 Skyline 恢复为默认设置。操作如下：

- 启动 Skyline。
- 在**起始页**上单击**空白文档**，显示如下：

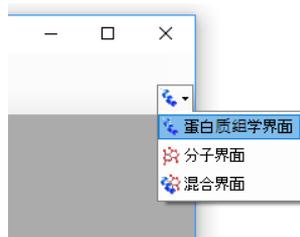


- 在**设置**菜单中单击**默认值**。
- 在询问您是否保存当前设置时，单击表单上的**否**。

现在，Skyline 当前实例中的设置已重置为默认值。

由于本教程涵盖蛋白质组学主题，因此务必将用户界面设置为“蛋白质组学界面”。

- 单击 Skyline 工具栏右上角的用户界面按钮，然后选择如下所示的**蛋白质组学界面**：



Skyline 将在蛋白质组模式下运行，Skyline 右上角 将随之显示蛋白质图标 。

对于 DIA 分析，您需要强制 Skyline 对提取的所有定量离子对的峰面积进行积分操作。操作如下：

- 在**设置**菜单上，单击**全部积分**。

使用“导入肽段搜索”向导

如果您知道所要查看的位置，可以使用 Skyline 中的各种菜单选项执行以下所有步骤。为简化 DIA 的正常工作流程，可以改为使用“向导”，它会逐步引导您完成 Skyline 运行分析所需要的

关键选择。先从 DDA 肽段谱图匹配结果构建谱图库，然后指定一组 DIA 运行以从中提取色谱图，接着指定各种设置，最后指定您想查询的目标本身。

您将对每个 A 和 B 样品进行一次 DDA 运行分析来构建谱图库。我们已使用 Comet 搜索引擎执行了 DDA 搜索，并使用 PeptideProphet 进行了后续处理 - 请参阅此链接，了解 Skyline 所支持的更多 DDA 搜索：

https://skyline.ms/wiki/home/software/Skyline/page.view?name=building_spectral_libraries。您将从 PeptideProphet 的输出文件 interact.pep.xml 开始，该文件包含这两个 DDA 文件的数据库搜索结果。

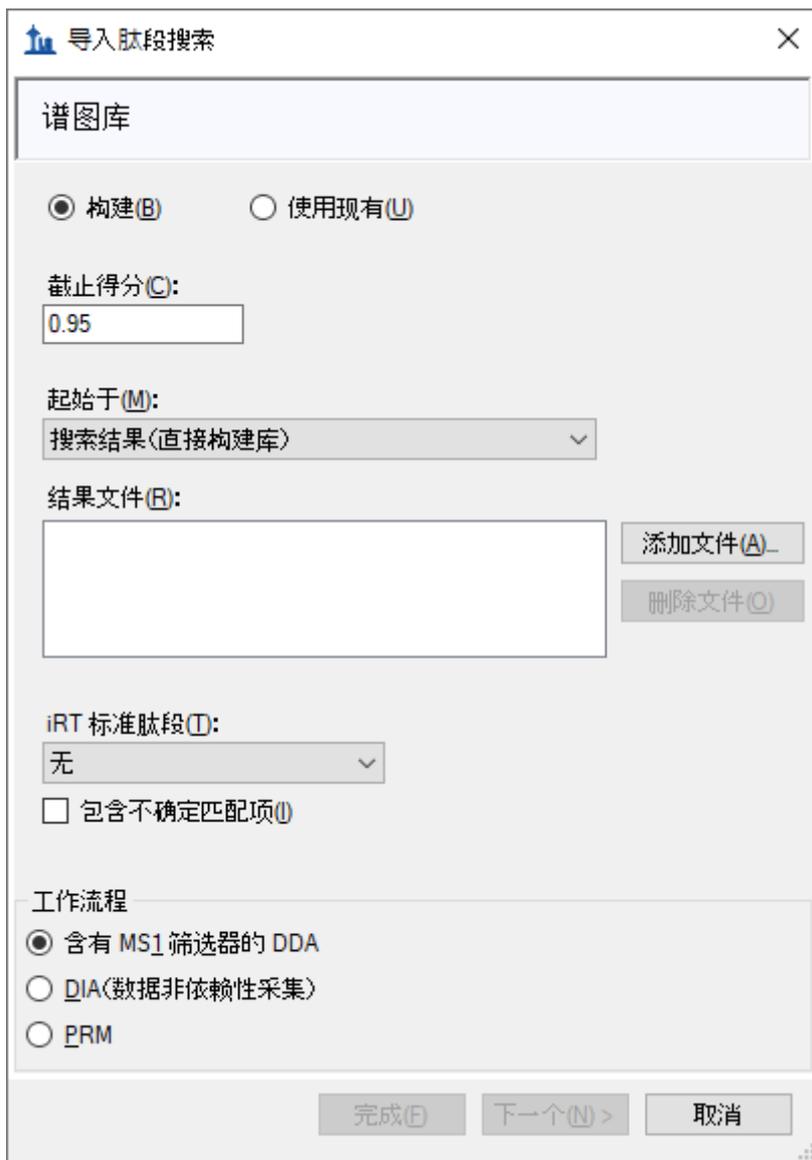
要开始，请执行如下操作：

- 在文件菜单中选择**导入**，然后单击**肽段搜索**。

Skyline 会指示您需要保存当前文档，这样它便能获知要在电脑磁盘中执行分析的位置。

- 单击**确定**按钮。
- 导航到先前创建的 DIA-QE 文件夹。
- 在**文件名**字段中键入 "DIA-QE-tutorial"。
- 单击**保存**按钮。

导入肽段搜索向导应显示如下：



注意：0.95 是 DDA 数据库搜索中每个肽段谱图匹配度计算的 PeptideProphet 概率阈值 - 在这个特定的数据集中，这相当于 0.2% 的 PSM 错误发现率，但此值在不同数据集中会有所不同，因此应在此处输入您要使用的错误发现率的概率阈值。

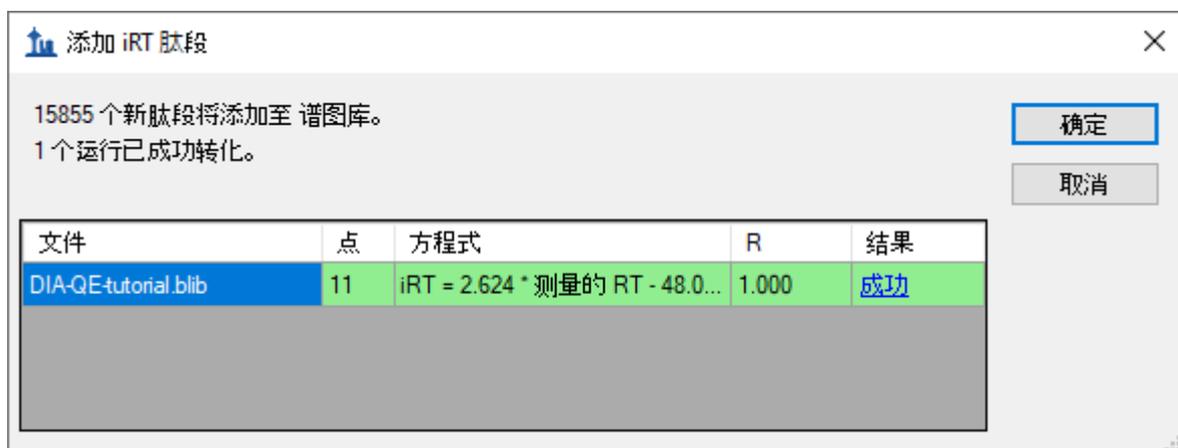
- 单击**添加文件**按钮。
- 导航到先前创建的 DIA-QE 文件夹，并进入 DDA_search 子文件夹。
- 双击 "interact.pep.xml" 文件。
- 在 **iRT 标准肽段**下拉列表中，单击 "Biognosys-11 (iRT-C18)"。
- 在**工作流程**框中，单击 **DIA** 选项。

向导中的**谱图库**页面现在应显示如下：

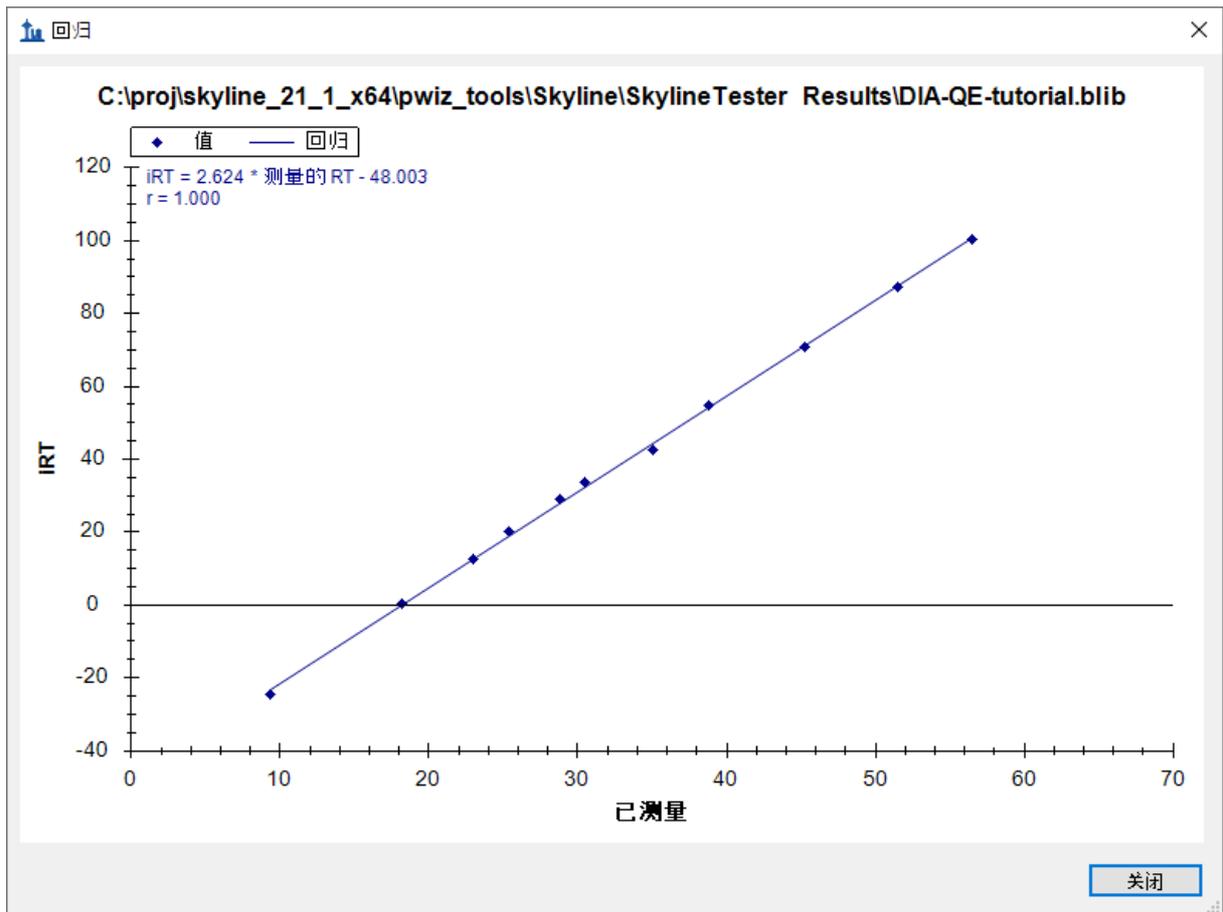


- 单击下一个按钮。

Skyline 开始创建谱图库，并在**构建肽段搜索库**表单中显示进度。创建的谱图库将通过加入样品中的 Biognosys iRT 标准肽段搜索结果校准 iRT 值。操作成功后，呈现如下所示的表单：



- 单击**成功**链接（突出显示蓝色）查看线性回归。



- 单击**关闭**按钮。
- 单击**确定**按钮。

Skyline 会询问您是否要重新校准 iRT 标准值到已计算并显示在上图中的回归线。

- 单击**否**按钮。

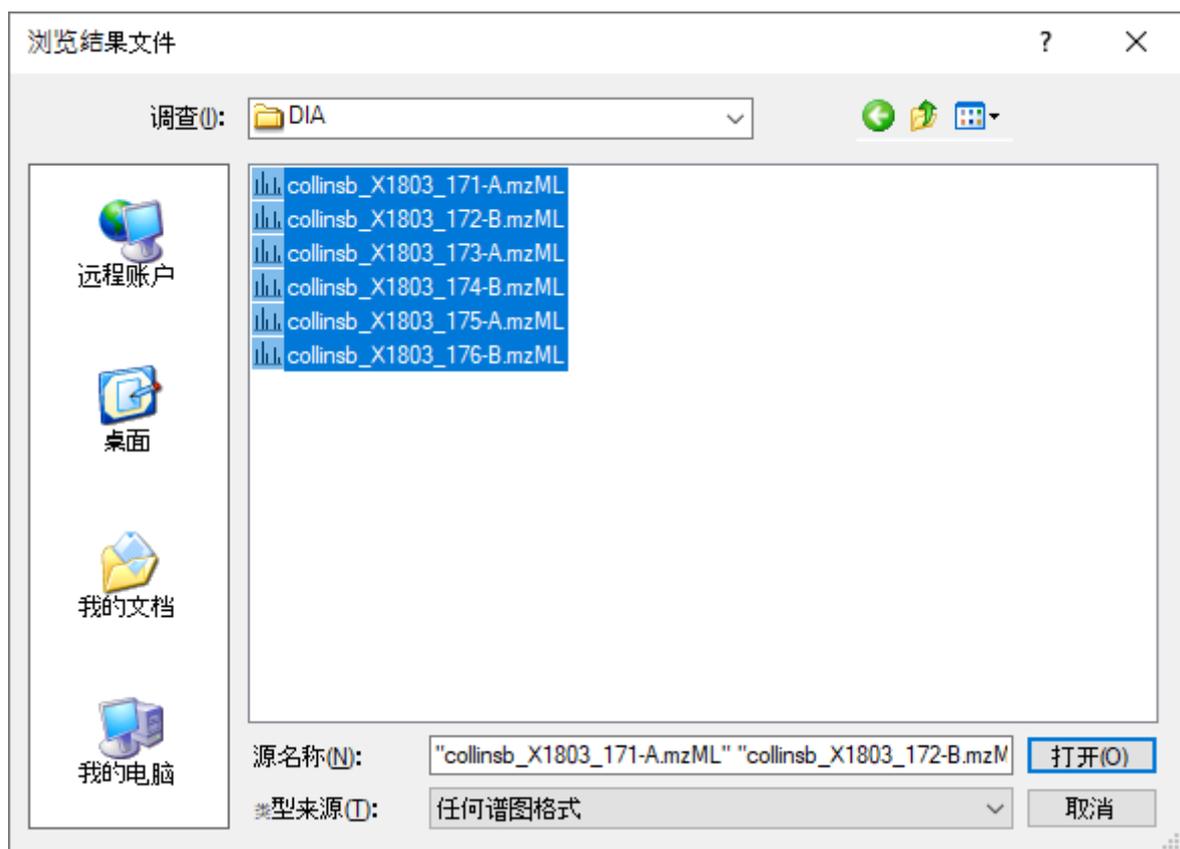
Skyline 呈现出一长串不确定匹配的肽段清单。换言之，这些肽段的唯一证据来自与多个肽段相匹配的谱图，从而导致这些谱图和匹配的肽段不适用于后续的靶向分析。

- 单击**确定**按钮。

在显示的**提取色谱图**页面中，您可以告知 Skyline 在哪里可以找到将用于色谱图提取、峰检测和峰面积计算的 DIA 数据文件。

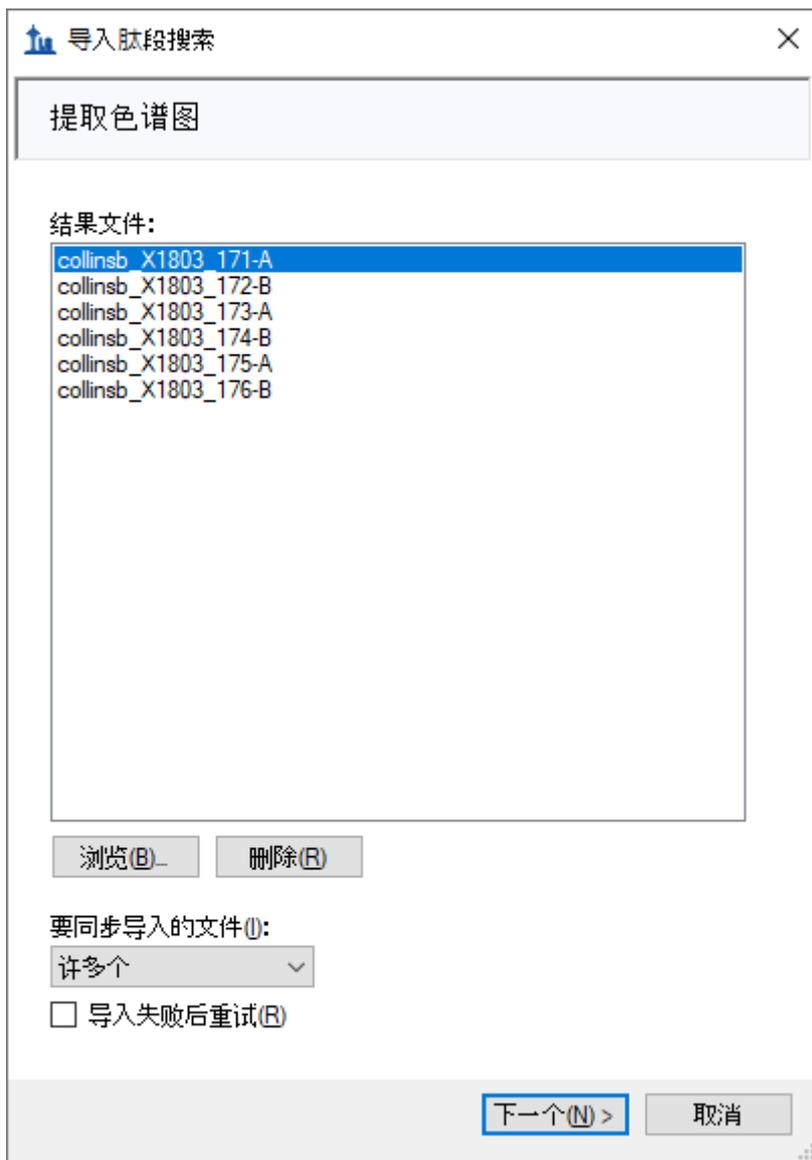
- 单击**浏览**按钮。
- 在 DIA-QE 文件夹中，双击 DIA 子文件夹。
- 将此文件夹中的 6 个文件全部选中。

浏览结果文件应显示如下：



- 单击打开按钮。

导入肽段搜索表单应显示如下：



- 单击下一个按钮。

Skyline 会显示一个表单，询问您是否移除公共前缀 "collinsb_X1803_17"，从而使 Skyline 界面中显示的重复测定名称更短，并且更易于使用。

- 单击**确定**按钮。

执行如下操作以跳过**添加修饰**页面：

- 单击下一个按钮。

在**配置离子对设置**页面中，对默认值进行以下更改：

- 将**母离子电荷**设置为 "2, 3, 4"。
- 将**离子类型**设置为 "y, b"。

注意：您可以保留表示母离子的离子类型 "p"，Skyline 将从 DIA 数据文件中的 MS1 谱图中提取覆盖母离子同位素分布的色谱图。不过，为了简化本教程并减少处理时间和输出文件大小，下面的描述中仅处理从 DIA MS/MS 谱图中提取的碎片离子色谱图。

- 选中使用 DIA 母离子窗口进行排除。

导入肽段搜索表单应显示如下：

- 单击下一个按钮。

Skyline 显示配置全扫描设置页面。

注意：这些提取设置取决于数据采集所用仪器的类型和设置。每个数据集的最佳提取设置可能会略有不同。在本项分析中，您将使用质心数据以节省空间。因此，您应当选择“质心”并指定质量精度为 "20" ppm 进行提取。仪器的分辨能力可以通过轮廓模式数据指定。

对默认值进行以下更改：

- 在两个**质量精度**字段中，均使用 "20" ppm。

现在，您需要根据仪器上定义的数据非依赖性采集的参数来定义新的隔离方案。

注意：在本例中，我们使用了 18 个可变宽度窗口，它们覆盖从 400 到 1220 质荷比的范围，采用 2 质荷比重叠，以支持三重四极杆范围边缘的 1 质荷比排除边距。

- 在**隔离方案**下拉列表中，单击"<添加...>"。

随即弹出**编辑隔离方案**表单，从中执行以下操作。

- 在**名称**字段中，输入 "ETH QE (18 variable)"。
- 单击**预指定隔离窗口**选项。
- 单击**导入**按钮。
- 导航到先前创建的 DIA-QE 文件夹，并进入其 DIA 子文件夹。
- 双击文件 "collinsb_X1803_171-A.mzML"。

此时**编辑隔离方案**表单应显示如下：

编辑隔离方案

名称(N): 确定 取消

使用结果数据隔离目标(U)

隔离宽度(I): 去卷积(D):

预指定隔离窗口(P) 计算(C) 导入(I) 图形(G)

测量

开始	结束	边距
400	430	1
428	459	1
457	483	1
481	506	1
504	531	1
529	554	1
552	576	1

去卷积(E): 指定边距(M) 指定 CE 范围(R)

每次扫描的窗口(W):

Skyline 已解出在质谱仪上测得的 DIA 隔离范围的**开始**和**结束**质荷比值。可以看出它们重叠了 2 质荷比，为补偿重叠，Skyline 在隔离窗口的每侧增加了 1 质荷比边距，在色谱图提取过程中它将会被排除在外。要在操作中查看此项：

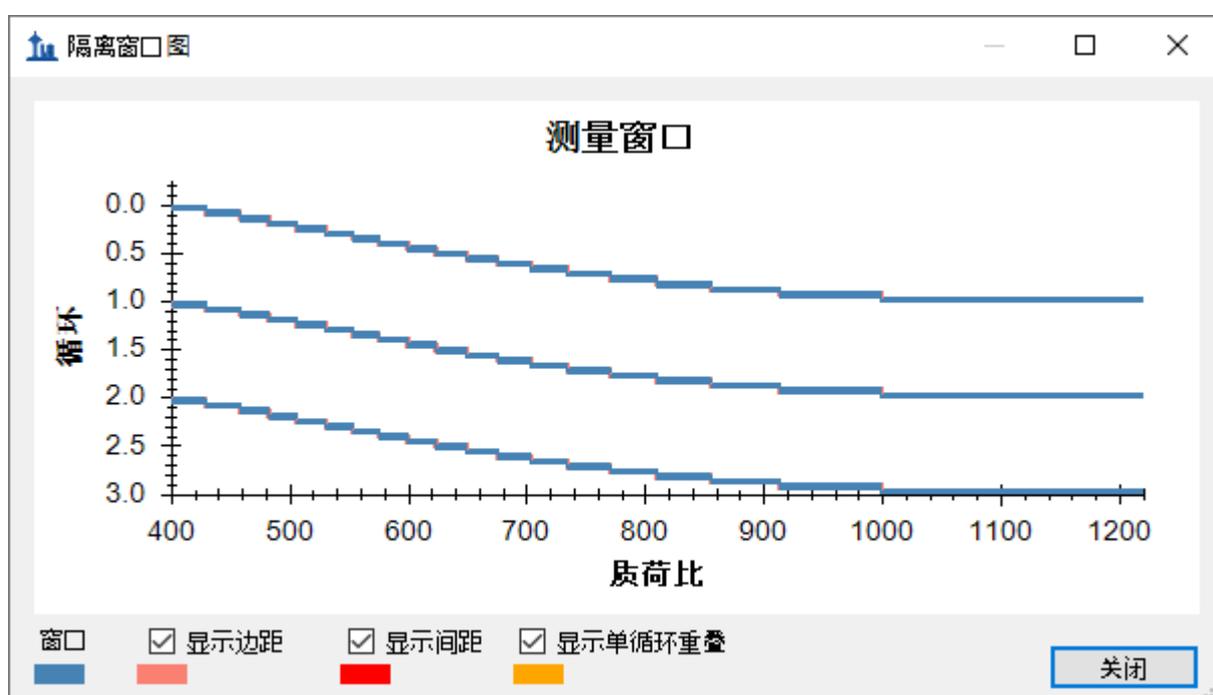
- 在**预指定隔离窗口**下方的下拉列表中，单击“提取”。

您会看到**开始**和**结束**值向内偏移 1 质荷比，并且范围不再重叠。这些是 Skyline 在将肽段母离子质荷比值与采集的谱图进行匹配，来进行碎片离子色谱图提取时将使用的范围。

要查看隔离方案的视觉效果：

- 单击**图形**按钮。

随即显示下面这样的表单：



您可以单击并拖动矩形进行放大，或是使用鼠标滚轮查看边距。如果取消选中**显示边距**后又重新选中，图形名称将在**测量窗口**和**提取窗口**之间切换，并且粉红色的边距条消失后又会重新出现。图中没有红色的**显示间距**或黄色的**显示单循环重叠**，二者均为 DIA 隔离方案设计中的错误。

- 单击**关闭**按钮。
- 单击**编辑隔离方案**表单中的**确定**按钮。

导入肽段搜索向导现在显示如下：



- 单击下一个按钮。

您会看到**导入 FASTA** 页面，从中执行下列操作：

- 单击**浏览**按钮。
- 导航到先前创建的 **DIA-QE** 文件夹，并进入其 **DIA** 子文件夹。
- 双击 "**target_protein_sequences.fasta**" 文件。

注意：此时您可以选择执行蛋白质组学范围分析，导航到 "**interact.pep.xml**" 文件所在的 **DDA_search** 子文件夹，然后双击肽段搜索中采用的完整 **FASTA** 序列文件 "**napedro_3mixed_human_yeast_ecoli_20140403_iRT_reverse.fasta**"。虽然此操作将产生更多目标，花费的处理时间也更多，但在大多数现代笔记本电脑上仍然可行。

- 在**诱饵生成方法**下拉列表中，单击“随机序列”。
- 选中**自动训练 mProphet** 模型。

导入肽段搜索向导现在显示如下：

导入肽段搜索

导入 **FASTA** (必需)

酶: 最大遗漏的酶解位点数:

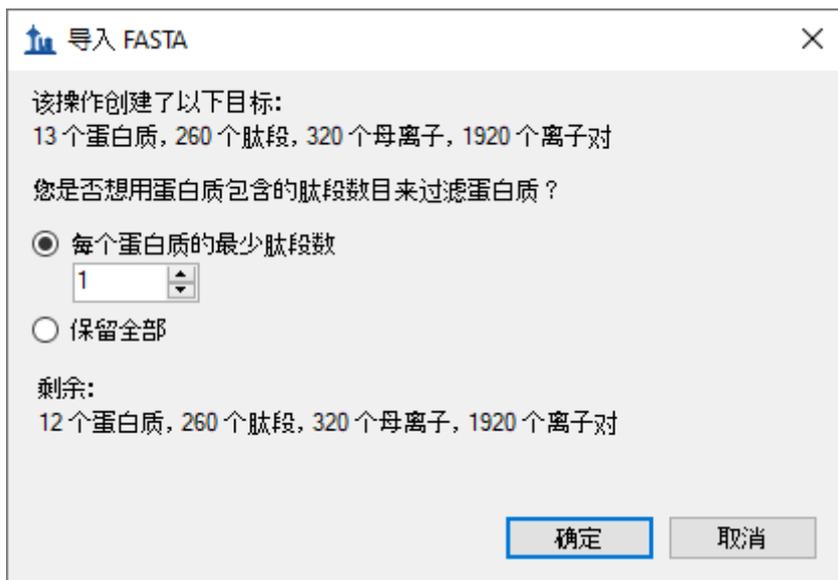
FASTA 记录以">"作为开头并含有蛋白质名称, 名称后可带有蛋白质描述。

诱饵生成方法: 每个目标的诱饵数:

自动训练 mProphet 模型

- 单击完成按钮。

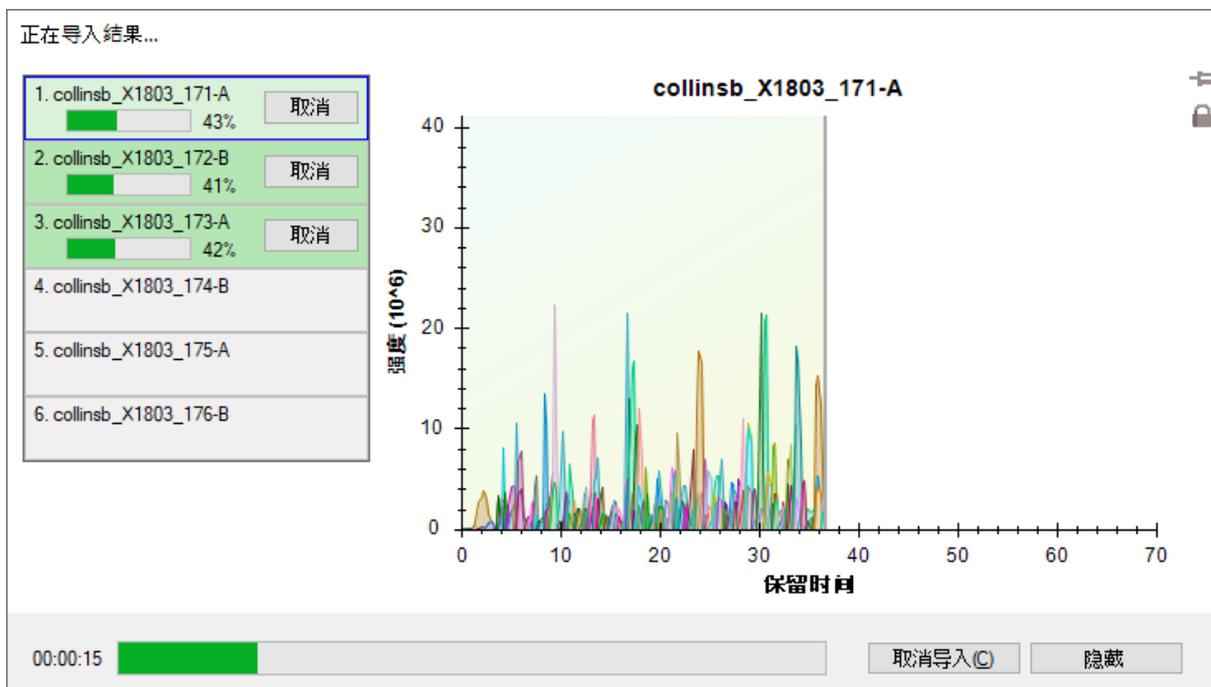
此时应当显示一张表单, 其中描述了根据您的设置和所提供的 FASTA 序列文本计算得出的目标, 如下所示:



可以发现，**每个蛋白质的最少肽段数**过滤器设置为 "1" 时，FASTA 文件中的 13 个蛋白质产生的**剩余蛋白质**为 12 个。这是因为有 1 个蛋白质没有在谱图库中找到肽段。如果单击**保留全部**，您会看到警告“将添加 1 个空蛋白质”，但请先切换回去，以避免将空蛋白质添加到目标列表中。

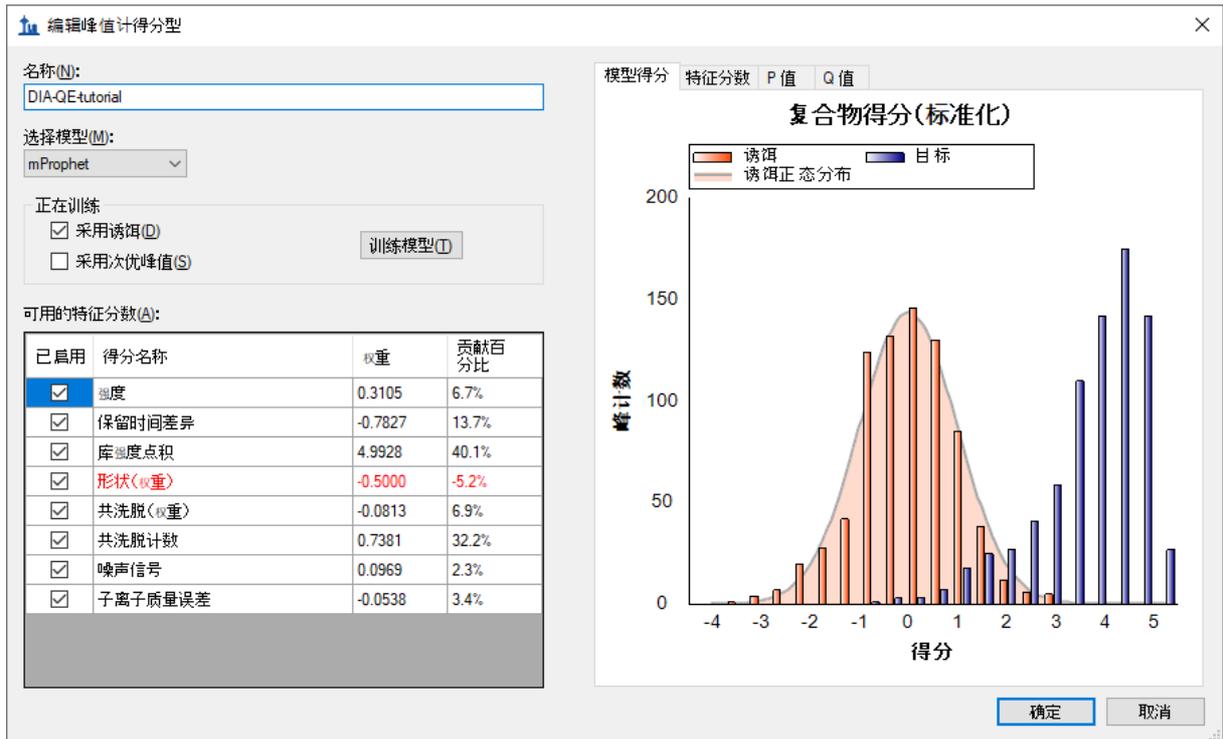
- 单击**确定**按钮。

Skyline 开始提取色谱图，在标准的 i7 4 核处理器上显示如下：



在 6 核及以下的处理器中，所有文件将并行处理；在大多数配备双核处理器的笔记本电脑上，导入过程中一次处理 2 个文件。

导入完成后，Skyline 将显示您在**导入肽段搜索**向导中的**导入 FASTA**页面上请求的 mProphet 模型。其显示如下：



系统采用默认的分数的 Skyline 在每组色谱图中找到的 10 个最优峰重新评分，并将这些峰重新选择为 mProphet 得分最高的峰。这些 mProphet 分数（在 Skyline 中叫做**检测 Z 得分**）进行扩展后，以使 1.0 表示与平均 mProphet 分数为 1 的标准偏差，从而实现所请求的随机排序诱饵肽段相似最优峰的分布。其中每个分数都被分配了一个 q 值（在 Skyline 中叫做**检测 Q 值**）。

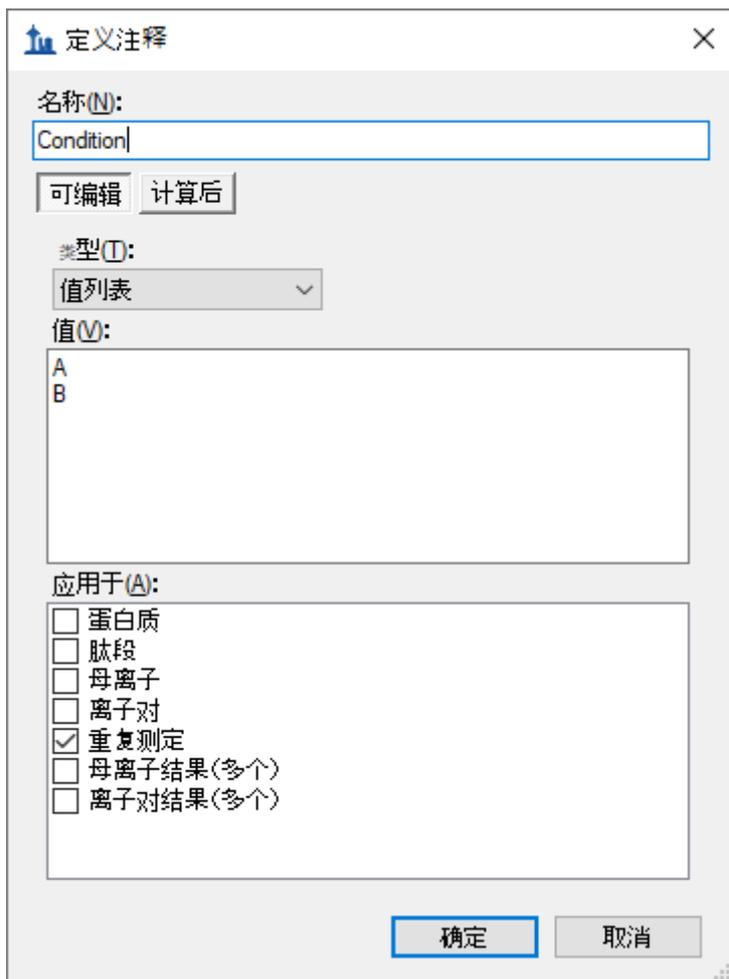
样品注释

接下来需要定义哪些样品属于哪个实验组：

- 在**设置**菜单中单击**文档设置**。
- 如果**注释**选项卡不是处于活动状态，请单击此选项卡。
- 单击**添加**按钮，以查看**定义注释**窗口。
- 在**名称**字段中输入 "Condition"。
- 在**类型**字段中选择“值列表”。
- 在**适用于**下面的列表中，选中**重复测定**。

在此实验中有两个条件：Condition A - 样本中蛋白质组组成为 20% 的大肠杆菌、15% 的酵母和 65% 的人类蛋白质，Condition B - 其组成为 5% 的大肠杆菌、30% 的酵母和 65% 的人类蛋白质。

定义注释窗口应如下所示：



- 单击**确定**按钮。
- 按照上述步骤执行操作，创建另一个名称为 "BioReplicate" 的注释。
- 在**类型**下拉列表中，为 BioReplicate 选择“文本”注释。
- 在**适用于**下面的列表中，选中**重复测定**。
- 单击**确定**按钮。

此时将返回**文档设置**窗口。

- 确保选中 **Condition** 和 **BioReplicate** 复选框。
- 单击**确定**按钮。

现在，您可以为导入的重复测定添加注释：

- 在**视图**菜单上，选择**文档网格** (Alt-3)。

随即打开**文档网格**窗口，如果您以前从未用过它，则会显示**蛋白质**报告：

- 在**文档网格：蛋白质**窗口中，单击**报告**菜单，然后单击**重复测定**。

如下面的截屏中所示，使用键盘上的 A, B, 1、2、3、Enter 和方向键对样品进行注释：

文档网格: 重复测定

报告 6 of 6 导出... 操作 查找:

重复测定	样品类型	分析物浓度	Condition	BioReplicate
1-A	未知		A	1
2-B	未知		B	1
3-A	未知		A	2
4-B	未知		B	2
5-A	未知		A	3
6-B	未知		B	3

- 单击表单右上角的红 X，即可关闭文档网格：重复测定窗口。

为便于查看，您现在可以按条件将数据拆分到 Skyline 窗口中的 2 个窗格。

- 在视图菜单中选择排列图形，然后单击分组。
- 在排列图形分组表单中，将窗格分组设置为 "2"。
- 选择选项在群组之间分配图形。
- 在显示字段中，选择“平铺”。
- 在排序字段中，选择“文档”。
- 单击确定按钮。

注意：如果此阶段仍在导入数据，需等待完成后以保存文档。

- 保存您的 Skyline 文档。

手动检查数据

接下来，您要手动检查一些色谱图和基础谱图。要执行这项操作，您需要显示保留时间和峰面积视图：

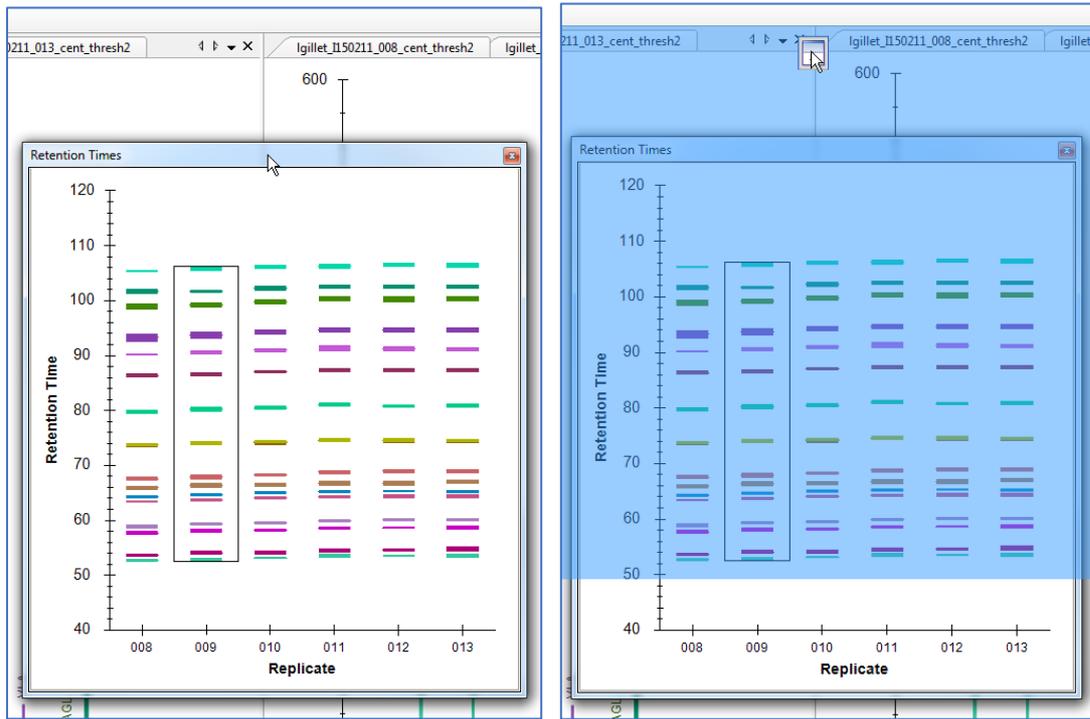
- 在视图菜单中选择保留时间，然后单击重复测定比较 (F8)。
- 在视图菜单中选择峰面积，然后单击重复测定比较 (F7)。

现在这两个视图应该都可见了。您可以执行以下操作在屏幕上进行最佳配置：

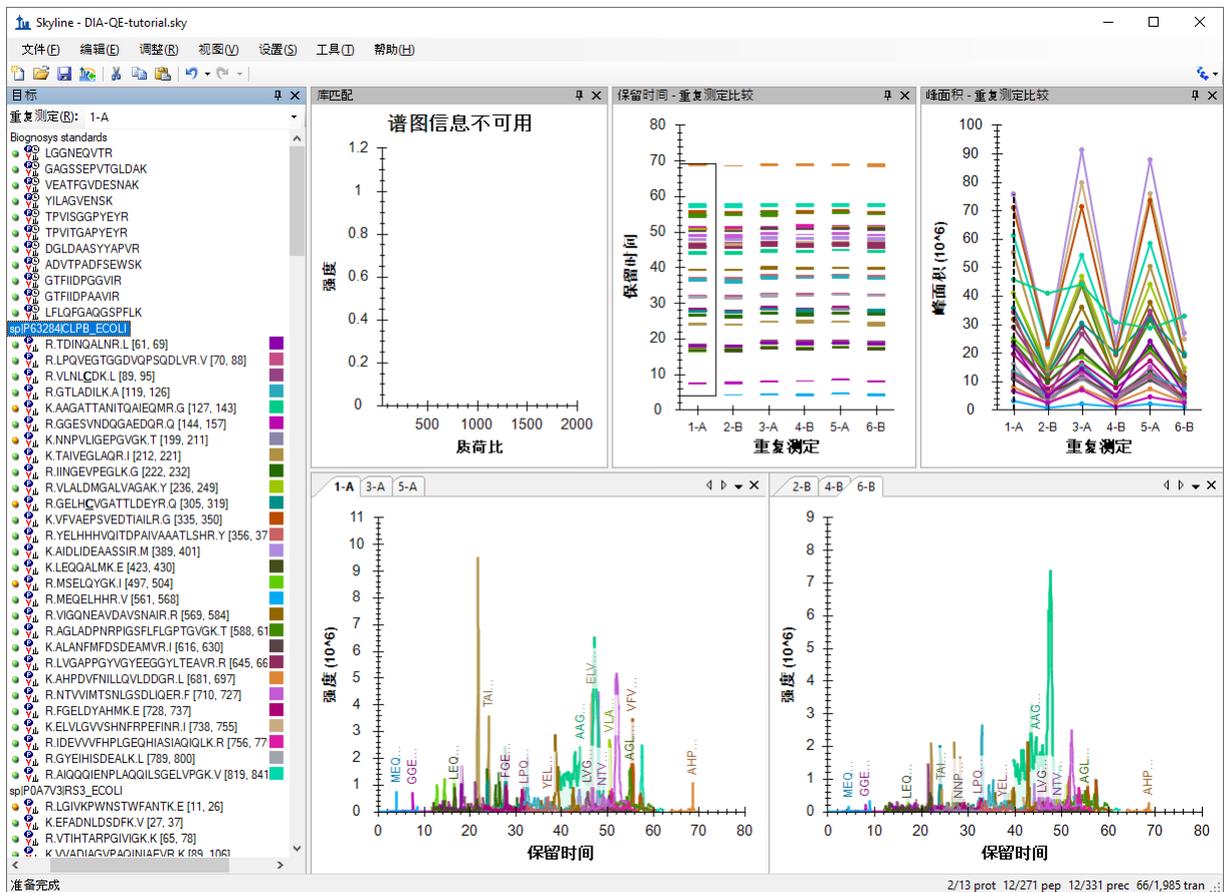
- 如果这些新窗口处于浮动状态，可以单击浮动窗口的顶部边框，然后按住鼠标左键并拖动该窗口，从而将它们停靠在主窗口中。

您可以在库匹配窗口中执行以下所示的相同操作。

注意：要停靠这样的窗口，必须将鼠标箭头光标放置在某个停靠图标的内部（本例中停靠在顶部）。一旦在按住鼠标左键的情况下开始拖动，如果此时释放鼠标左键，Skyline 将在放置窗口的位置显示停靠图标和一个透明的蓝色矩形。



- 以同样的方式停靠峰面积：重复测定比较和库匹配窗口，以便轻松查看所有信息，如下所示：



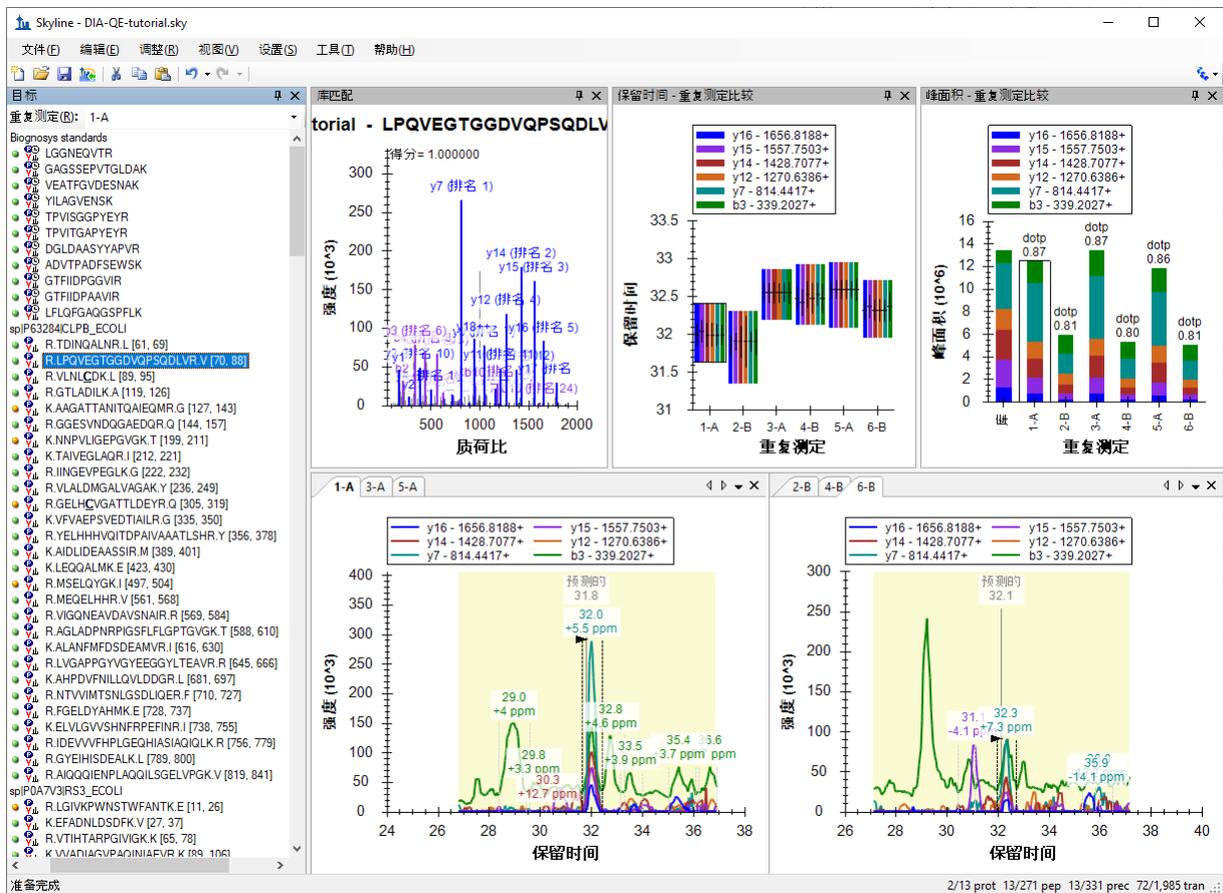
- 单击第一个大肠杆菌蛋白质: "sp|P63284|CLPB_ECOLI".

您应当会在各种图表（色谱图、峰面积和保留时间重复测定图）中看到该蛋白质的所有肽段。上面的截屏展示了一个蛋白质被选择后，组成它的所有肽段被汇总在每个视图中（除了库匹配窗口未显示任何内容）。

根据您在峰面积-重复测定比较图中看到的结果，该蛋白质是否受到差异化的调控？回想一下，对于大肠杆菌，A:B 之间的预期倍数变化率为 4:1。

- 选择此蛋白质中的 LPQVEGTGGDVQPSQDLVR 肽段。

您将在所有这些视图中获得该肽段的特定信息，如下所示：



检查属于该蛋白质的其余肽段的峰面积模式。蛋白质肽段的定量模式是否与预期的差异化调控模式 (4:1) 一致？

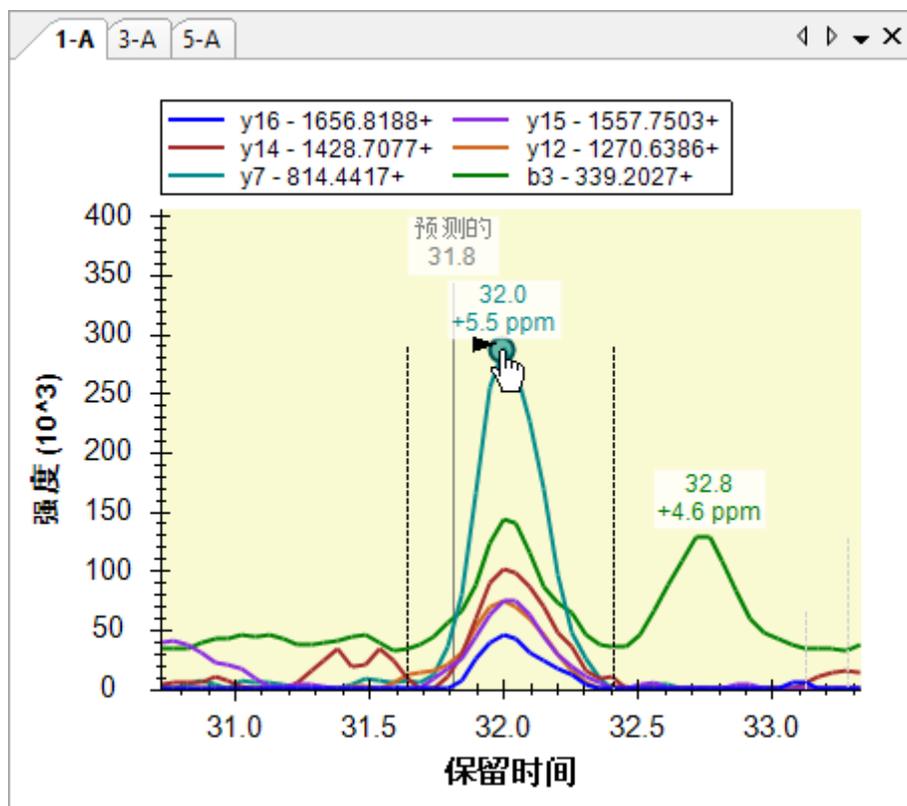
注意： 如果给定的肽段序列有多个母离子电荷状态，它们将被分别提取并评分。您可以单击肽段序列旁边的 +，然后单击各个电荷状态来进行查看。

- 单击文档中的某些人类蛋白质。

从蛋白质水平视图和肽段水平视图检查重复测定峰面积。峰面积是否与预期的比率 (1:1) 一致？酵母蛋白质呢（是否为预期的 1:2 比率）？

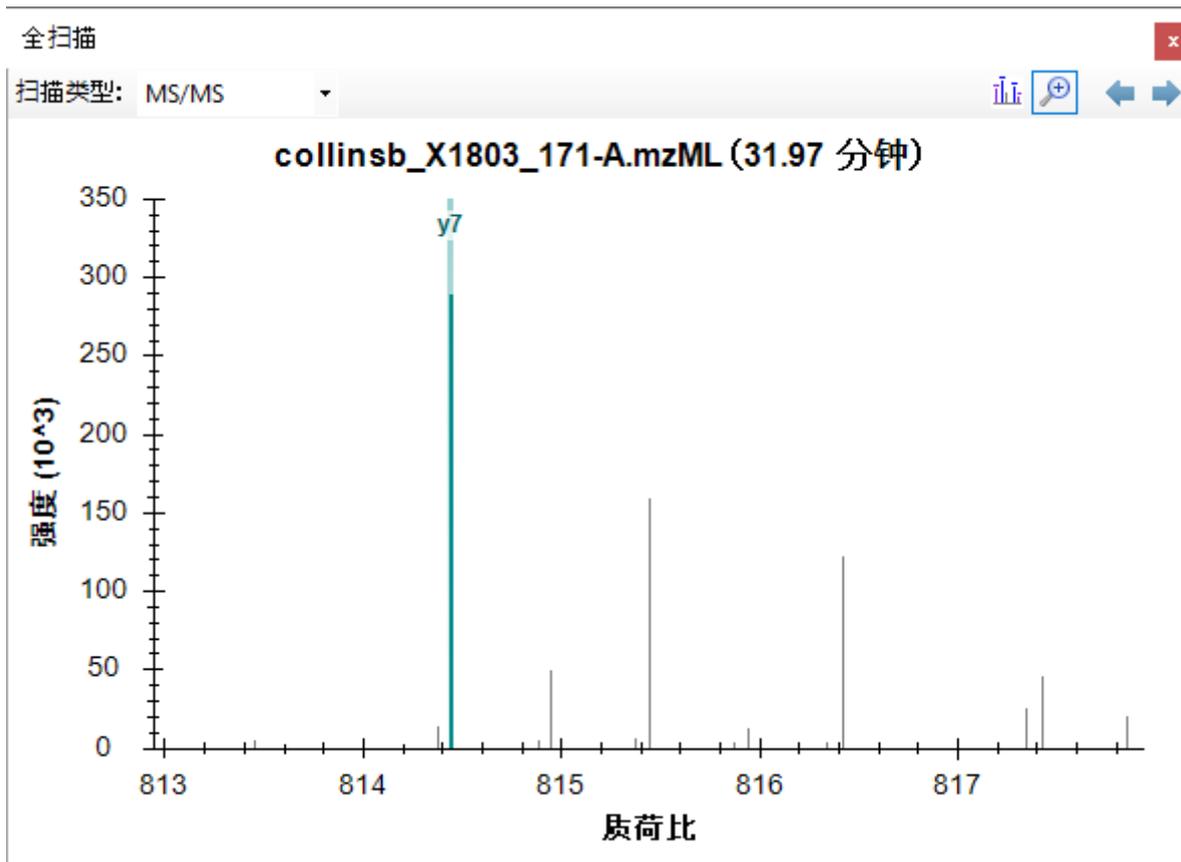
- 再次单击第一个蛋白质中的 LPQVEGTGGDVQPSQDLVR 肽段。
- 在视图菜单中选择**自动缩放**，然后单击**最优峰 (F11)**。

请注意，如果将光标悬停在色谱图曲线上，曲线上就会出现与该曲线同色的圆圈。

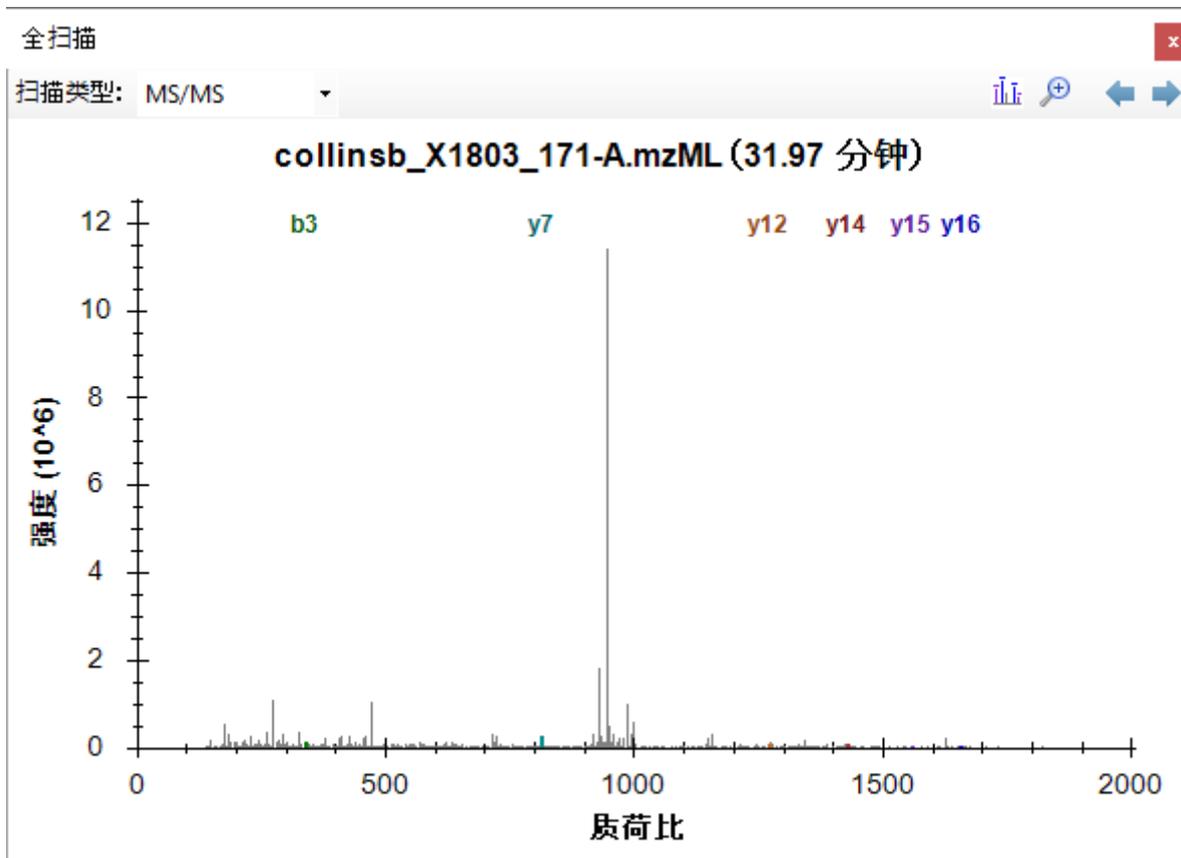


- 单击色谱图点上的圆圈。

Skyline 将在从中提取该色谱图点的谱图上打开**全扫描**视图，并放大您选择的离子。



- 单击**全扫描**工具栏中的放大镜按钮可以查看整个谱图。



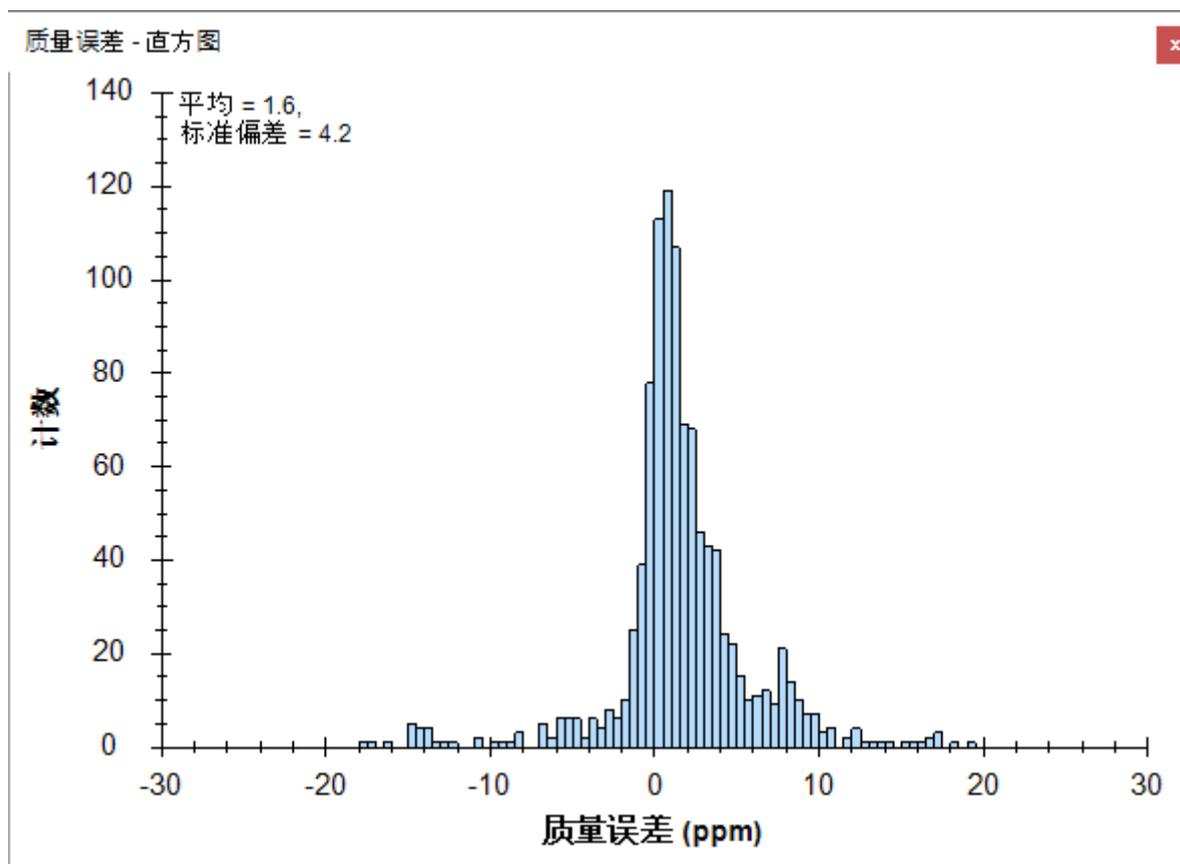
进一步手动探索数据（包括一些诱饵）。

质量误差和保留时间偏差

您可以检查质量精度和保留时间预测准确度，以确定是否使用了最佳提取参数，或是否需要进行某些调整以改善结果。

- 在视图菜单中选择**质量误差**，然后单击**直方图**。

这将显示质量误差在数据集中的分布。提取窗口 (± 20 ppm) 能否进一步优化？一般情况下，平均误差 ± 3 个标准偏差便足以应对。但是，如果显示的直方图被截止，并且在极端情况下计数很高，则可能需要增宽提取窗口。



请注意，此分布涵盖全部 6 次运行。为进一步理解每次运行的代表性，请执行以下操作：

- 右键单击该图，选择**重复测定**，然后单击**单一**。
- 单击**目标视图**顶部的**重复测定**下拉列表。
- 使用键盘上的向上和向下箭头键查看所有 6 个重复测定中的质量误差值。

您会看到平均值介于 1.3 到 1.8 之间，而标准偏差值介于 3.7 到 4.4 之间。在最极端的情况下使用简单的计算方式，即平均值 $+ 3 * SD = 1.8 + 3 * 4.4 = 15$ PPM，这似乎表明 15 PPM 的公差足以用于这些数据。

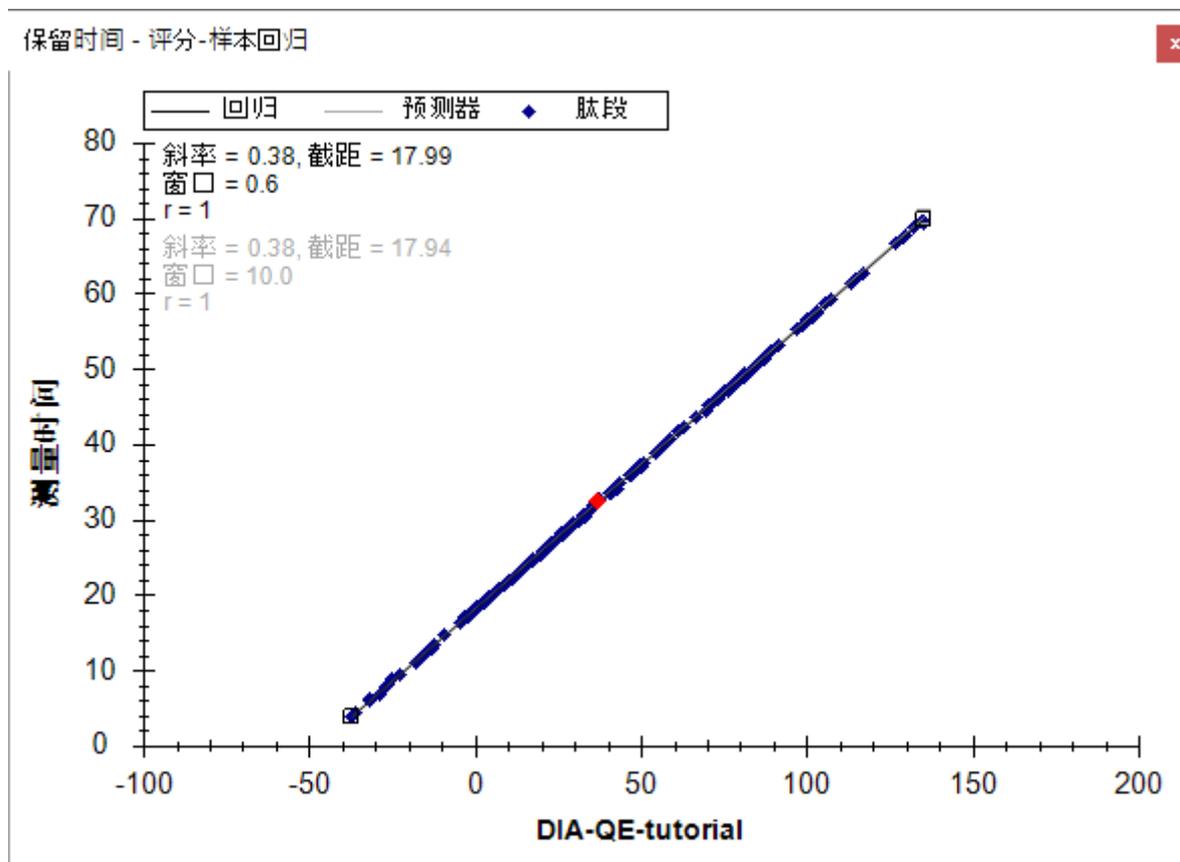
还要查看诱饵的质量误差：

- 右键单击该图，选择**点**，然后单击**诱饵**。

- 右键单击该图，选择**重复测定**，然后单击**全部**。
(然后切换回**目标**)

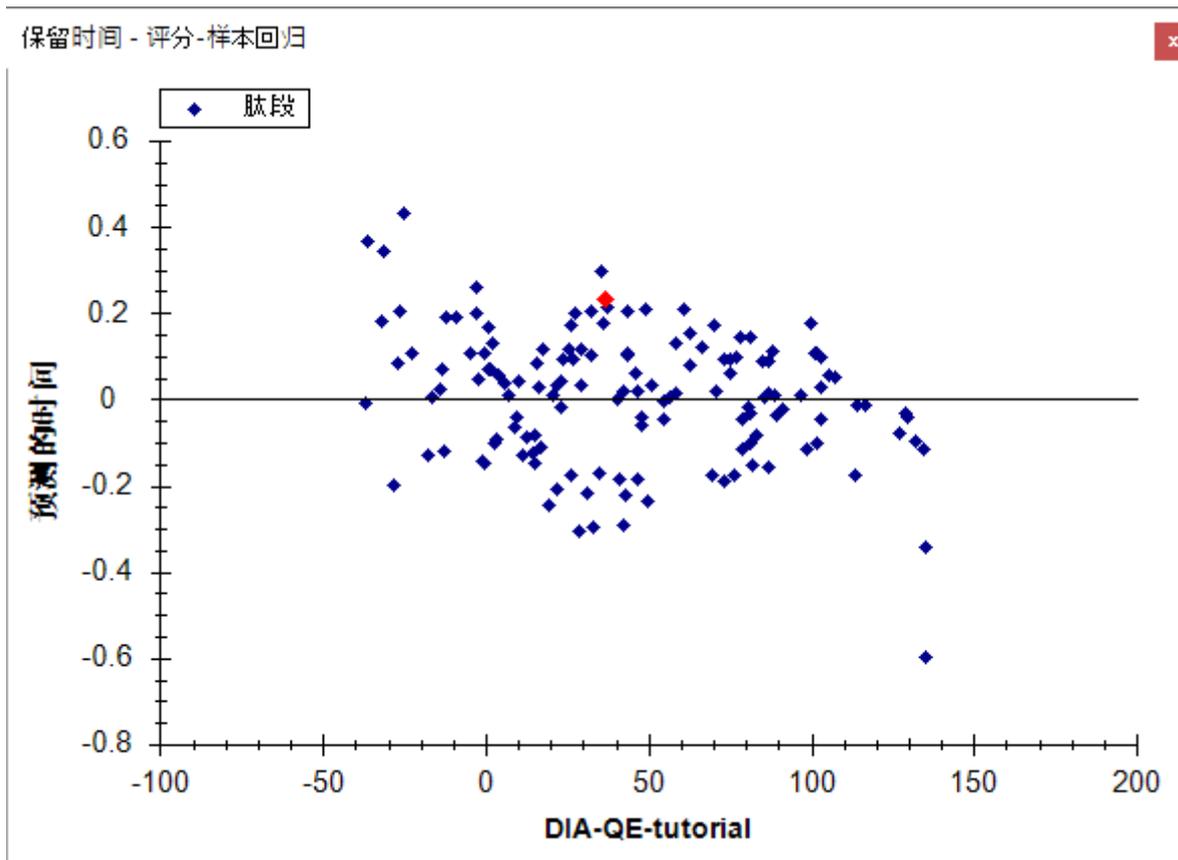
要查看基于 iRT 肽段和目标肽段库 iRT 值预测目标肽段保留时间而使用的线性回归：

- 在**视图**菜单中选择**保留时间**，然后单击**回归**，最后单击**运行分数**。



- 右键单击此图，选择**绘图**，然后单击**残差**。

随即显示该数据集中与预计保留时间的偏差。能否针对此分析进一步优化提取窗口 (± 5 分钟)? 这时请进一步检查诱饵。



注意：因为这项分析的谱图库是通过将相同样品进行并排分析构建而成，所以保留时间预测的准确性非常好。在对不同仪器不同时间不同样品分析创建的外部谱图库中，保留时间将导致这些预测中的误差较大。

定量比较

您已经进行了一些常规验证，证明使用 Skyline 进行的数据处理没有严重缺陷。mProphet 评分模型中包含所有可用的特征评分，并且该模型在目标和诱饵（用于模拟不可检测的随机目标）之间实现了合理的隔离。您或许觉得 RT 提取范围可以从 ± 5 收紧到 3。这些设置显然有用，并且最终的 mProphet 模型也看似可以接受。

这种对数百个肽段进行的分析通常需要针对每个目标进行手工分析，该操作既耗时，又容易出错。当前的分析则可以扩展到整个包含三个物种（人类、酵母、大肠杆菌）蛋白序列的 FASTA 文件，即全蛋白组水平分析，其中包含数以万计的肽段。研究人员通常不会单独考虑每个肽段，而是进行某些类型的群组比较，然后跟进那些以有趣的方式明显变化的肽段或蛋白质。

要在 Skyline 中执行简单的成对群组比较，请执行以下操作：

- 在视图菜单中选择**其他网格**，然后选择**群组比较**，最后单击**添加**。
- 在**名称**字段中输入 "By Condition"。
- 为**控制群组注释**字段选择 "Condition"。
- 为**控制群组值**字段选择 "A"。
- 为**比较值**字段选择 "B"。
- 单击**高级**按钮。

- 在 **Q 值界限** 字段中输入 "0.01"。

编辑群组比较表单应如下显示。

- 单击**确定**按钮。

要查看刚刚创建的群组比较：

- 在**视图**菜单中选择**其他网格**，然后选择**群组比较**，最后单击 **By Condition**。

这时会出现一个表格，显示肽段水平的倍数变化以及已调整的 p 值（错误发现率 - FDR 的估计值），这些值用于 A、B 样品混合物之间的比较。

- 展开**蛋白质**列标题的宽度，以便看到完整的蛋白质名称 - 包括相应的物种名称。
- 单击**倍数变化结果**列标题，然后单击**升序排列**。

检查表中某些肽段的倍数变化估计值，同时要记住它们来自哪个物种以及预期的比率（人类蛋白质 1:1，酵母 1:2，大肠杆菌 4:1）。考虑已调整的 p 值以及您可能期望的值。

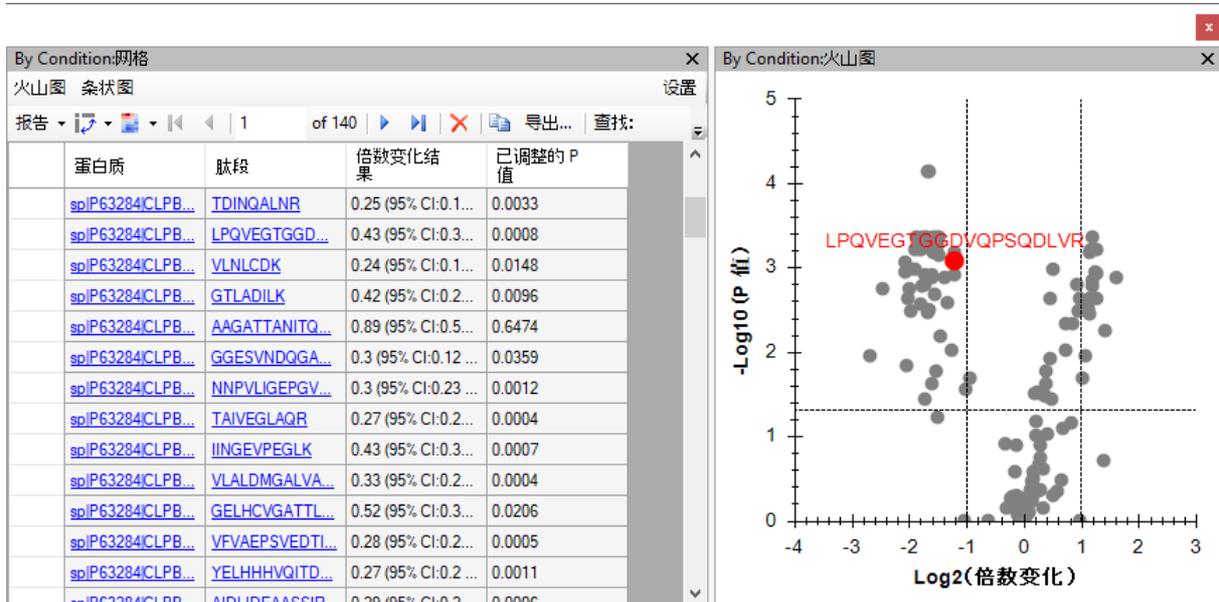
By Condition: 网络

火山图 桑状图

报告 1 of 140 导出... 查找:

蛋白质	肽段	倍数变化结果	已调整的 P 值
sp P63284 ICLPB_ECOLI	TDINQALNR	0.25 (95% CI:0.16 to 0.41)	0.0033
sp P63284 ICLPB_ECOLI	LPQVEGTGGD...	0.43 (95% CI:0.37 to 0.51)	0.0008
sp P63284 ICLPB_ECOLI	VLNLCDK	0.24 (95% CI:0.11 to 0.53)	0.0148
sp P63284 ICLPB_ECOLI	GTLADILK	0.42 (95% CI:0.27 to 0.64)	0.0096
sp P63284 ICLPB_ECOLI	AAGATTANITQ...	0.89 (95% CI:0.55 to 1.45)	0.6474
sp P63284 ICLPB_ECOLI	GGESVNDQGA...	0.3 (95% CI:0.12 to 0.74)	0.0359
sp P63284 ICLPB_ECOLI	NNPVLIGEPGV...	0.3 (95% CI:0.23 to 0.4)	0.0012
sp P63284 ICLPB_ECOLI	TAIVEGLAQR	0.27 (95% CI:0.23 to 0.32)	0.0004
sp P63284 ICLPB_ECOLI	IINGEVPEGLK	0.43 (95% CI:0.37 to 0.5)	0.0007
sp P63284 ICLPB_ECOLI	VLALDMGALVA...	0.33 (95% CI:0.29 to 0.38)	0.0004
sp P63284 ICLPB_ECOLI	GELHCVGATTL...	0.52 (95% CI:0.35 to 0.78)	0.0206
sp P63284 ICLPB_ECOLI	VFVAEPSVEDTI...	0.28 (95% CI:0.23 to 0.34)	0.0005
sp P63284 ICLPB_ECOLI	YELHHHVQITD...	0.27 (95% CI:0.2 to 0.36)	0.0011
sp P63284 ICLPB_ECOLI	AIDLIDEAASSIR	0.29 (95% CI:0.23 to 0.36)	0.0006

- 单击网络窗口左上角的火山图按钮。



对于大多数变化的肽段，已可以看到出现在预期比率附近和 5% FDR 水平线以上的点簇。为了更好地了解哪些点属于哪种生物体，请执行以下操作：

- 右键单击火山图，然后单击格式。
- 单击表达式列后面第一行中的 "..." 按钮。
- 在创建匹配表达式表单中，将匹配下拉列表设置为“蛋白质名称”。
- 在正则表达式字段中输入 "ECOLI"。

表单应显示具有匹配蛋白质名称的肽段列表，如下所示：

创建匹配表达式

正则表达式(R): ECOLI

匹配(M): 蛋白质名称

过滤器

倍数变化截止值(F): 无

P 值截止(P): 无

匹配的

- TDINQALNR (sp|P63284|CLPB_ECOLI)
- LPQVEGTGGDVQPSQDLVR (sp|P63284|CLPB_ECOLI)
- VLNLCDK (sp|P63284|CLPB_ECOLI)
- GTLADILK (sp|P63284|CLPB_ECOLI)
- AAGATTANITQAIEQMR (sp|P63284|CLPB_ECOLI)
- GGESVNDQGAEDQR (sp|P63284|CLPB_ECOLI)
- NNPVLIGEPGVGK (sp|P63284|CLPB_ECOLI)

确定 取消

- 单击**确定**按钮。
- 单击 **RGB** 列后面第一行中的 "..." 按钮。
- 选择紫色，然后单击**确定**按钮。
- 选择橙色，针对 "YEAS" 再次执行上述过程。
- 选择绿色，针对 "HUMAN" 再次执行上述过程。

火山图格式表单应显示如下：

火山图格式

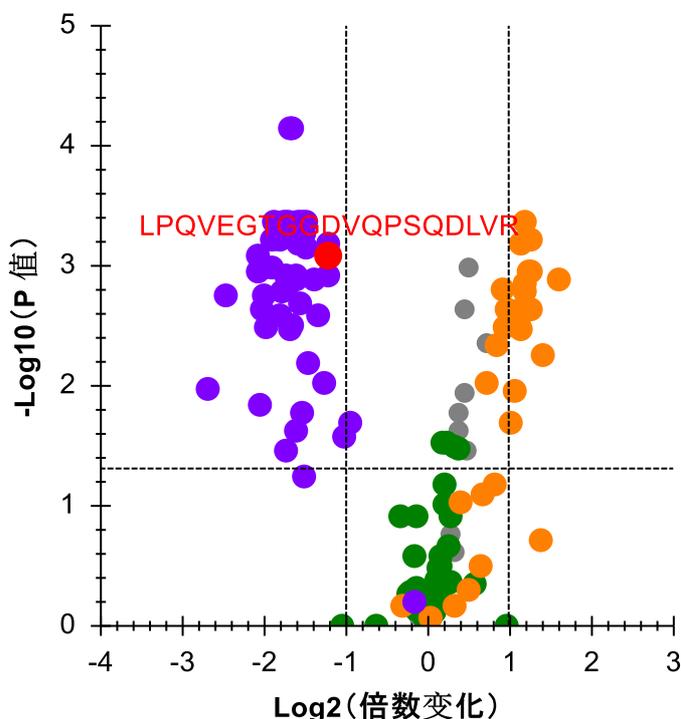
表达式	RGB	符号	大小
ProteinName: ECOLI	128, 0, 255	圆圈	正常
ProteinName: YEAS	255, 128, 0	圆圈	正常
ProteinName: HUMAN	0, 128, 0	圆圈	正常

RGB 高级

确定 取消

- 单击**确定**按钮。

火山图应显示如下：



请注意，图中仍有一些灰点。您可以单击其中任意的点，这样它们在**目标视图**中将处于选中状态，并且将在图中以红色显示测试。您也可以将鼠标光标悬停在任何点上以获取更多信息。通过这种方式，可以看到灰点属于 iRT 标准肽段。您可以执行以下操作，同时从网格和图表中过滤这些灰点：

- 单击**蛋白质**列标题，然后单击**过滤器**。
- 在**过滤器类型**下拉列表中，选择“不包含”。
- 在“标准”下方的字段中输入。
- 单击**确定**按钮。

此时将删除火山图中的灰点。您可以尝试对蛋白质名称使用不同类型的过滤器，从而只在此图中显示一种物种。

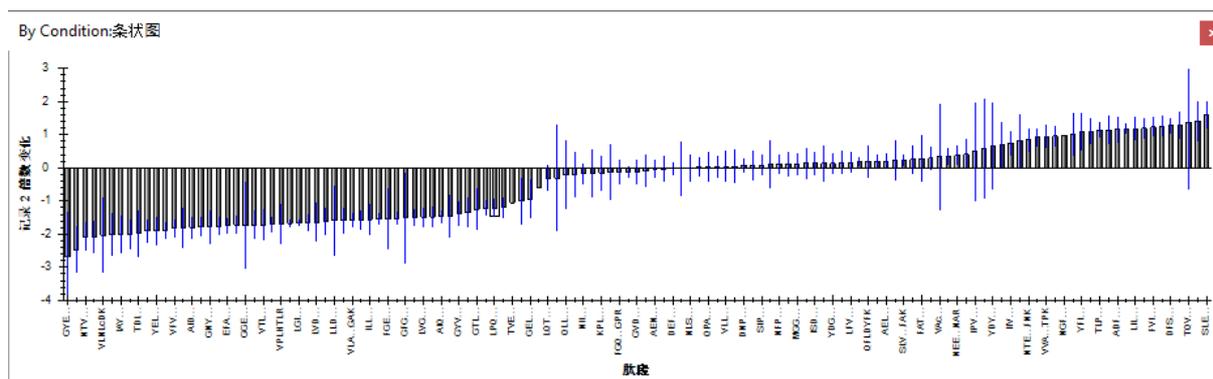
Skylines 还为倍数变化值提供了条状图视图，其中整合了（未调整的）置信区间误差线。对于所测试的多个假设而言，虽然它们并未经过调整，但它们仍有助于您了解度量值的方差。请执行以下操作，查看条状图中的倍数变化值：

- 单击网格窗口左上角的**条状图**按钮。

条状图将出现在火山图的顶部。若要为条状图提供更多的查看空间，请执行以下操作：

- 单击 **By Condition: 条状图**选项卡，按住并拖动到新位置，然后释放。
- 调整现在自由浮动的 **By Condition: 条状图**窗口的大小。
- 在 **By Condition: 网格**中，单击**倍数变化结果**列标题，然后单击**升序排列**。

该图现在应显示如下：



与火山图一样，您可以单击目标视图中的单个条形以将其选定（这样便可确认图中的不同肽段来自哪个物种）。

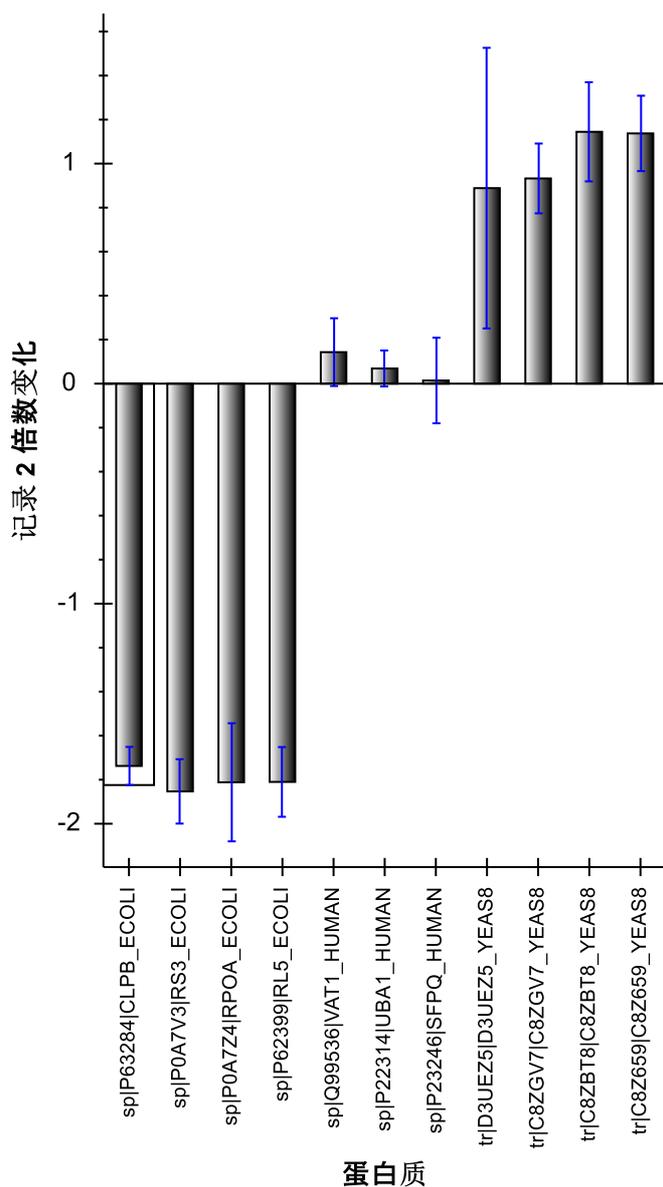
要查看 11 种靶向蛋白质（不包括 iRT 标准和诱饵）水平的倍数变化结果，请执行以下操作：

- 单击网格窗口右上角的**设置**按钮，以重新打开**编辑群组比较**表单。
- 将**范围**选项由**肽段**更改为**蛋白质**。

Skyline 应当会立即调整**条状图**和**火山图**。您在**条状图**中会看到只有 9 条，而不是您可能期望的 11 条。这是因为当前设置使得 Skyline 无从处理缺失的值（由于 q 值界限为 0.01），只能丢弃目标，因为没有足够的测量值来支持 T 测试。要解决此问题，请执行以下操作：

- 在**编辑群组比较**表单中，将**摘要**方法更改为 "Tukey's Median Polish"。

现在，**条状图**应显示所有 11 种蛋白质的倍数变化估算值，如下所示：



- 单击右上角的 X 方格，关闭编辑群组比较。
- 保存文档。

您可能还想使用更复杂的统计工具（例如 MSstats）来进行比较。为此，您可以安装 Skyline 工具商店中的 MSstats（使用 Skyline 菜单中的工具>商店，或单击 Skyline 安装网页上的“工具商店”按钮）。然后，可以直接从 Skyline 工具菜单运行 MSstats，或是执行以下操作准备稍后在 R 或 RStudio 中运行 MSstats：

- 在文件菜单中选择导出，然后单击报告。
- 从外部工具文件夹中选择 MSstats 输入。
- 另存为 "DIA_QE_MSstats_Input.csv"。

结语

祝贺您！您已经使用 Skyline 和 DIA 数据完成了第一次差异蛋白质组学比较。您了解了如何使用**导入肽段搜索**向导来简化初始数据处理，从根据 DDA 肽段搜索数据构建谱图库，到从定量 DIA 运行提取色谱图，创建并应用 mProphet 统计峰检测模型来改善峰的选择并分配 q 值（用于控制群组比较中使用的峰面积之间的错误发现率）。您先了解了使用质量误差和保留时间回归图评估 DIA 数据质量。最后，您了解了在 Skyline 中的两组样品之间直接进行简单的成对比较，从而获取交互式的群组比较网格、火山图和条状图，帮助您理解结果并与结果进行互动。导出 MSstats 输入报告后，即可使用 MSstats 外部工具执行进一步的统计测试。

现在，您不妨考虑使用全蛋白组水平的 FASTA 文件和/或操作中包括母离子来重新尝试执行这些步骤，以了解如何处理更广泛的目标集和/或其他维度（MS1 谱图中的母离子）。或者您也可以使用自己的数据来进行尝试。