位置记忆 -> 附肢再生过程的适当模式化

为什么会有位置记忆？-> 未损伤肢体的分子梯度

基于如上假设->如何调节位置记忆？ -> 只有极少数基因能调节(在未损伤肢体中梯度表达)

基因比较少 -> 确定新候选效应因子 -> 探究未损伤斑马鱼尾鳍上转录本、蛋白质和代谢物的丰度图 -> 高置信度的基因和代谢物清单

地点：未损伤斑马鱼尾鳍的近端、中间和远端区域

方法：

1. RNA 测序（RNA-seq）（23,926 个转录本）

2. 无标记定量（LFQ）蛋白质组学研究 （3,061 个蛋白质）

1. 任务1：确定沿近端轴水平不同的候选基因

为了生成沿近端轴水平不同的高置信度分子列表 -> 对近端和远端 RNA-seq 数据集和 LFQ 蛋白组学数据集进行了比较

-> 比较逻辑：两个数据集中存在相似模式的转录本和蛋白质

RNA-seq 筛查更为全面，量化了 LFQ 蛋白组学量化的 88% 的蛋白质

相反，LFQ 蛋白组学仅量化了 RNA-seq 数据中 13% 的转录本（图 S3A）

差异化表达的判断：近端和远端FDR和fold change

转录本: FDR<1% + fold change>3

蛋白质: FDR<5% + fold change>2

高度表达，潜在的记忆效应分子->即为“相似模式的转录本和蛋白质”

???鉴定和量化的转录本和蛋白质根据基因符号（gene symbols）进行分配。

HGNC机构 基因ID

结果： 66 个和 51 个转录本在近端和远端富集 -> RNA-seq 数据中识别重要基因的方法已在 BIO211 Lab4 中作了详细介绍

任务1.1:

如何鉴定近端和远端富集的转录本 -> 转录本的量化数据: 折变和 FDR -> PvD\_trans.csv

量化蛋白质的全套基因符号 -> PvD\_combine.csv (同一基因符号可能对应多个转录本)

任务1.2:

识别在两次实验中都测量到的近端和远端富集的蛋白质（在box里有代码）

任务1.3: 生成量化蛋白质的火山图 （报告主要结果并简要解释您的发现。另一个 R 脚本示例见下框。）

火山图->快速识别大型数据集中的变化

y 轴: 绘制显著性(FDR？？？)

x 轴: 折叠变化

火山图的更详细解释和 R 脚本示例 -> BIO211 Lab4

任务1.4: 比较近端和远端差异表达的转录本和蛋白质

两种筛选方法共同测定的 2,680 个基因 -> 有多少基因在近端和远端同时富集了转录本和蛋白质

简要说明您是如何进行分析的，并报告主要结果，如数字、维恩图等

2. 位置记忆可能由多种因素促成的证据

对位置记忆效应因子的实验操作 -诱发-> 再生过程中的异常模式化

确定沿尾鳍近侧轴差异表达的基因类型 -> 有助于发现位置记忆的新候选效应因子

通过 GO 和 KEGG 通路富集分析来实现

基因本体（GO）-> 旨在统一所有物种的基因和基因产物属性的表示

京都基因和基因组百科全书（KEGG）-> 从分子水平的信息中了解生物系统的高级功能和效用

富集分析 -> 可确定哪些途径、疾病和 GO 术语与相关基因/蛋白质有显著的统计学关联。

指纹

任务2: GO 或 KEGG 通路富集分析 -> 确定可能影响位置记忆的基因类型/特异基因

举例说明： 对转录本进行了 GO 生物过程术语注释。下图总结了 P < 0.05 的部分具有统计学意义的 GO 术语。从结果中我们可以得出结论：转录因子是细胞特性的主调节因子，因此是位置记忆效应因子的最佳候选者。

通过 RNA-seq 鉴定出的转录因子近端富集（绿色）和远端富集（蓝色）转录本（FDR < 1%，折叠变化 > 3）也汇总如下。

在 RNA-seq 数据中量化的 693 个编码转录因子的转录本（数据集 S7）中，有几个在鳍的近端和远端区域有差异表达。

这些转录因子包括 dlx5a、dlx6a、meis1a、msx1a、hoxb13a、raraa 和 lmx1bb。

它们直接参与了发育中的附肢的模式化，这与发育过程中的机制可能会调节成体结构的位置记忆这一预测是一致的。

3. 沿近模轴测量的相对代谢物丰度(开始研究代谢产物了)

维生素 D 代谢 -> 建立胸鳍的前后位置信息

RNA-seq和LFQ 数据 -> 包括 aldh1l1 在内的几种代谢酶沿尾鳍近端轴线有差异表达

关联性分析缩小研究的代谢物的范围

近端富集蛋白的 GO 项分析 -> 与氨基酸代谢和脂质运输相关的蛋白

探索可能参与位置记忆效应的代谢物 -> 使用质谱分析了未受伤尾鳍近端、中间和远端区域的代谢物 -> 对单个鱼鳍样本中的 151 种代谢物进行了量化 (P\_D\_meta.csv, 只关注近端和远端区域)

利用 MetaboAnalyst 提供的方法 -> 研究两组（未受伤尾鳍的近端区域和远端区域）之间存在显著差异的代谢物 -> 可作为潜在的效应物，将近端组织维持在代谢平衡状态，以实现更快的增殖

任务3.1

预处理和归一化处理 -> 原数据

只保留未受伤尾鳍近端和远端区域的代谢物定量(P\_D\_meta.csv)(注释数据已归一化，即无需额外的样本归一化或数据转换或缩放)

通过 MetaboAnalyst，尝试找到通过单变量分析和 PLS-DA 分析一致确定的候选代谢物。

在系统层面解读多组学数据仍然是一项重大挑战: 一种常见的策略: 单独分析每一组组学数据，然后利用从单个组学分析中发现的重要特征（基因、蛋白质、代谢物等）拼凑出 "全貌"

看trasncipts-protein-metabolome的一一对应组之间的作用机制

扩展解释范围：PROD1 MIER RA

gene transcripts protein metabolism之间存在互相作用的机理

任务3.2

对于共识结果，使用 MetaboAnalyst 途径分析模块确定其生物功能以及在途径中可能发挥的作用。报告分析的关键步骤和结果，并解释您的发现。

备注： 斑马鱼的 KEGG 通路可在 MetaboAnalyst 通路库中找到。

4. 位置记忆分子贡献的多组学综合解释

越来越多的组学数据显示了生物表型的调控复杂性。这种调控的复杂性需要系统层面的方法来理解。正如 "中心教条 "所描述的那样，在全局组学领域有一系列相互紧密联系的层次。因此，我们必须解释每一个全息层，并系统地整合多个组学层，以全面了解生物表型的调控情况。虽然每种特定的组学方法都有一个用于进行适当系统分析的基础生物网络，但只有将多组学数据结合起来并纳入适当的网络中，才能获得完整的调控图景。

MetaboAnalyst 中的 "联合途径分析 "模块旨在支持多组学分析和解释（即代谢组学、转录组学或蛋白质组学）。特别是，它允许用户将重要的基因/蛋白质（从基因表达或蛋白质组学研究中识别）与重要的代谢物（从代谢组学研究中识别）一起映射到代谢通路上，以进行功能富集分析和通路拓扑分析。该模块背后的工作假设是，通过整合基于基因/蛋白质表达和代谢物浓度变化的证据，人们更有可能精确定位基础生物过程所涉及的通路。

任务 4

对本项目中产生的 RNA-seq/LFQ 蛋白质组学/代谢组学数据进行联合通路分析，在系统水平上寻找位置记忆的综合分子线索。报告主要结果并解释您的发现。有关联合途径分析模块使用的详细介绍，请访问 MetaboAnalyst 网站。