通过RNA-seq 量化的23926条RNA：S1-S12

通过LFQ 量化的3061个蛋白质: S13-S25

S1------ PvD\_trans.csv

S7,S8,S9------ Dataset S7-9.xlsx

S18------ PvD\_combine.csv

富集分析，筛选在远端或在近端富集的transcript和protein，作为候选记忆效应因子

根据23926条RNA (S1 distal vs. proximal (因为论文通过PCA排除了middle))

作出图1B 的transcriptome部分的cluster图

筛选发现在近端或者远端富集（即以近端或远端为基点，向另一个方向发生浓度递增或递减）的566个transcript: 存于S5中

根据3061个protein (S13)

作出图1B的protein部分的cluster图

筛选发现在近端或者远端富集（即以近端或远端为基点，向另一个方向发生浓度递增或递减）的238个protein: 存于S21中

根据23926条RNA (S1 distal vs. proximal (因为论文通过PCA排除了middle))

作出S2 A的transcriptome部分的cluster图

在量化筛选过程中出现了中段富集和中段缺失的情况的Transcript: 存于S6中

根据3061个protein？(S13?)

作出S2 B的protein部分的cluster图

在量化筛选过程中出现了中段富集和中段缺失的情况的protein: 存于S21中

差异化分析，筛选在近端和远端出现差异化表达的transcript和protein，作为候选记忆效应因子

根据23926条RNA?(S1 distal vs. proximal (因为论文通过PCA排除了middle))

作出图1E的transcriptome部分的火山图

筛选出在近端和远端出现差异化表达（FDR<1%, fold change>3）的1424个transcriptome: 存于S4中

根据3061条protein?(S13?)

作出图1F的proteome部分的火山图

筛选出在近端和远端出现差异化表达（FDR<5%）的113个protein: 存于S20中

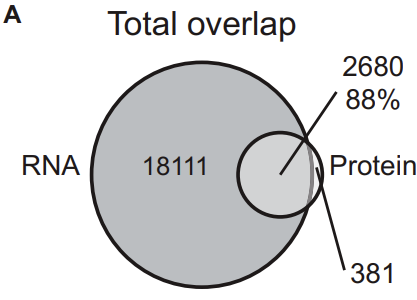
为了**创建一个在proximodistal axis上表达水平有梯度变化的分子列表**---🡪作者将近端RNA-seq与LFQ proteomics 的dataset与 远端 RNA-seq 与LFQ proteomics 的dataset进行比较

一. 首先缩小范围：（分析范围限定在两个数据集中以相似模式存在的transcript和protein）

目的是找到在远近两端的dataset中表现出similar pattern（均为梯度上升/梯度下降）的transcript与protein（目的：找出潜在同系统的位置记忆因子 因为为了构建完整作用机制）

结果：

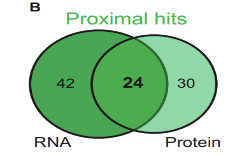
RNA-seq更具综合性，其量化了LFQ proteomics 所量化的88%的蛋白质（overlap）

LFQ proteomics仅仅量化了RNA-seq中所量化的13% 的transcript（overlap）

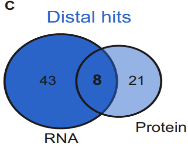
认为是用2680的重叠的gene，筛选出现富集表达的gene，在下方作出两个韦恩图

~~现在我认为上面这张图的total overlap没有作用，只是预览。因为如果是用重叠的2680的在下面找近端和远端富集的，那样下面的韦恩图则会完全重叠。~~

在RNA-seq 和LFQ proteomics 两种实验中所测定的protein与transcript（2680个重叠的protein与transcript---即对应着相同gene）中，根据富集分析结果找出在远端或近端富集的transcript和protein：



54 proteins，66 transcripts-----🡪在近端富集。 重叠的24为近端富集且有相同基因



29 proteins，51 transcripts-----🡪在远端富集。重叠的8为远端富集且有相同基因

在近端和远端区域，少于50%的在远近端差异化表达的transcripts被发现有类似的差异性表达的protein。

该发现与已知的其它系统中RNA和protein丰度比较的overlap形式一致， 表明在尾鳍这些区域存在高度的转录后调控。

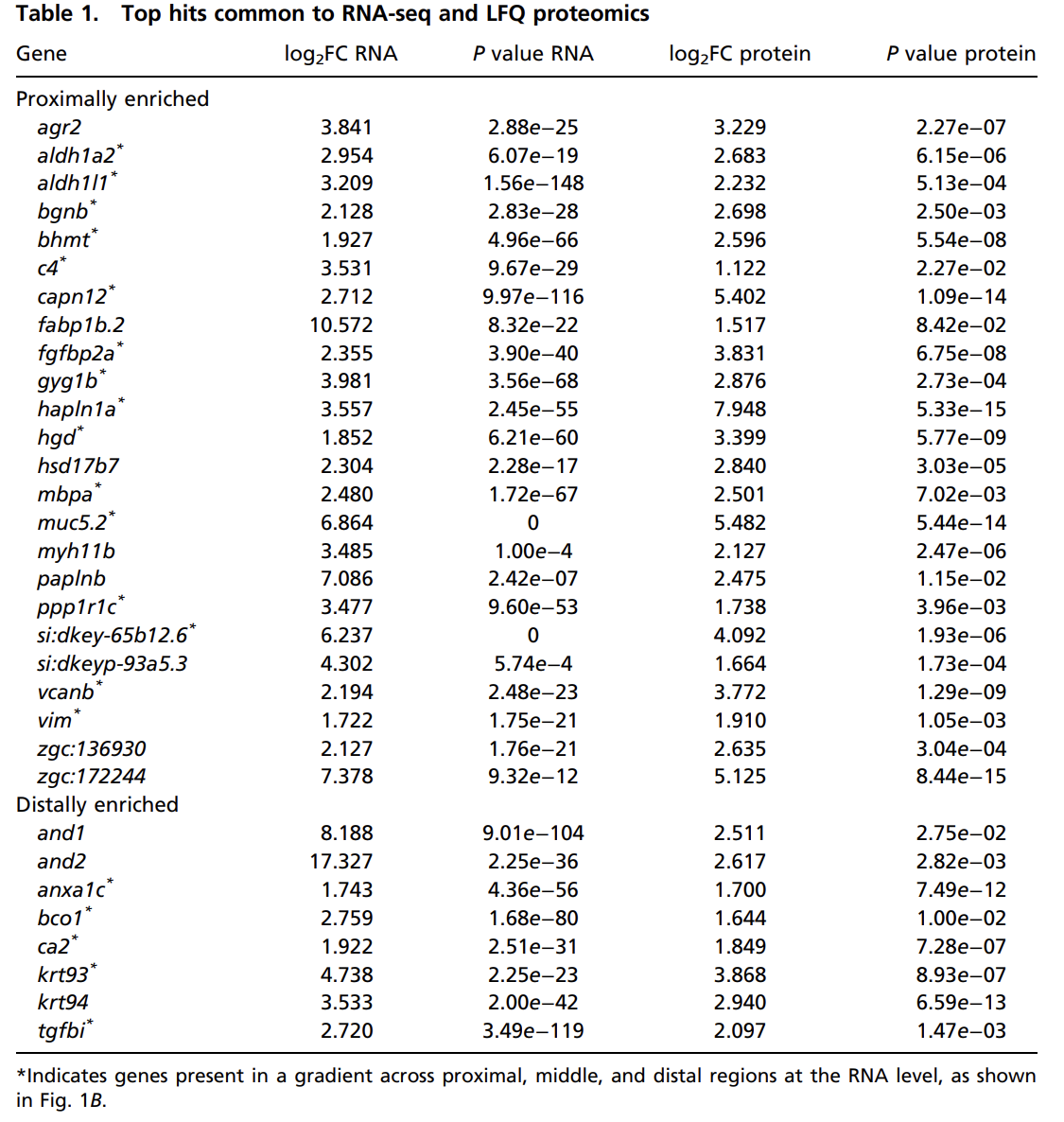
二. 找在近端富集处重叠的transcript & protein

和

和在远端富集处重叠的transcript & protein

这些列表之间的重叠（即所研究的在某端富集的protein是由某端富集的transcript翻译产生的，也就是都由一个gene表达而来，是一整个通路）产生了32(近端24+远端8)个gene的高置信度列表，表内gene表达的RNA和protein在尾鳍近端与远端区域的丰度上有significantly difference（潜在位置效应分子）。

下图为在远中近端表达的transcript和protein呈现显著性差异的gene

在以上32个gene中，有21个gene的transcript和protein在近端和远端区域的丰度存在梯度差

思路：在total overlap中的2680个是protein和transcript对应相同基因的，相当于缩小范围（因为这样的才有价值进行通路分析）