谁说10X单细胞的可变剪切做不了？来找欧易吧

By: 张春媛

花了几万块做了10X的单细胞分析，数据还能进一步挖掘信息么？欧易公司的单细胞可变剪切分析来啦！！！！！

我们最新研发的单细胞可变剪切可应用于droplet-based的单细胞数据，以10X和M20两种产品为主要应用对象，通过检测junction占细胞表达总量的百分（PSI）比检测可变剪切，若junction占基因总表达量在两种不同类型细胞间的差异较大，则认为该基因可能存在可变剪切。

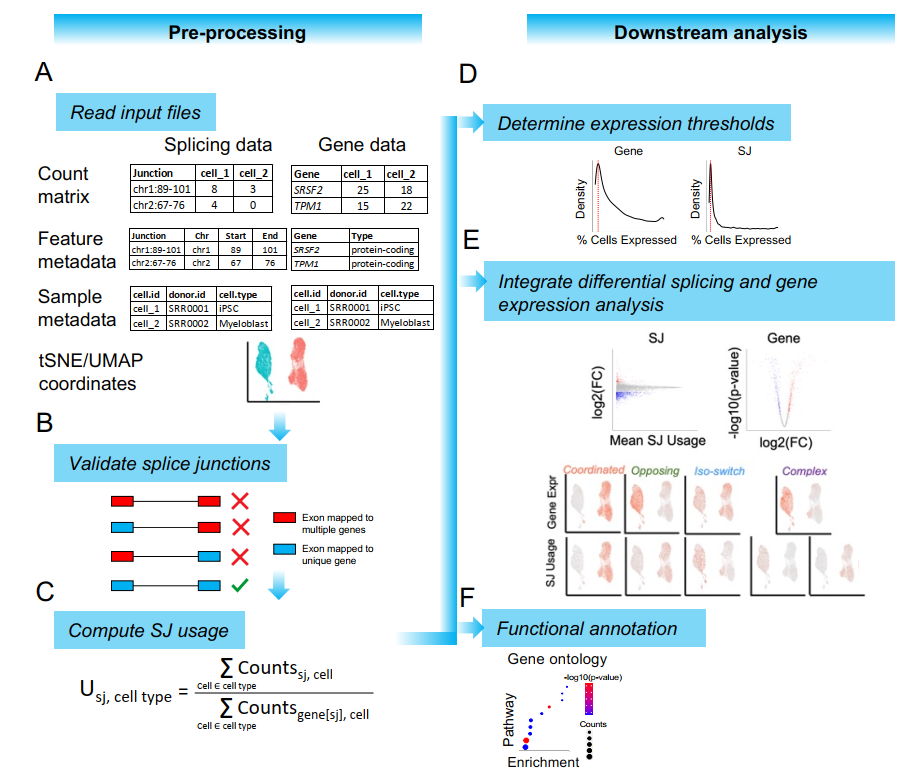


Figure 1．单细胞可变剪切分析原理[1]。

下面就跟随实际数据的分析结果来感受一下如何检测可变剪切的吧：

我们采用人类第2天（SRR9008755）和第4天（SRR9008752）的IPS细胞10X单细胞测序数据[2]做示例分析，重新比对后提取到所有junction的表达量，在两种细胞间对每个junction的表达量做差异分析，并以PSI绝对值大于5，p-value小于0.05为标准筛选显著差异的junction。Figure 2展示了所有junction差异表达的情况。

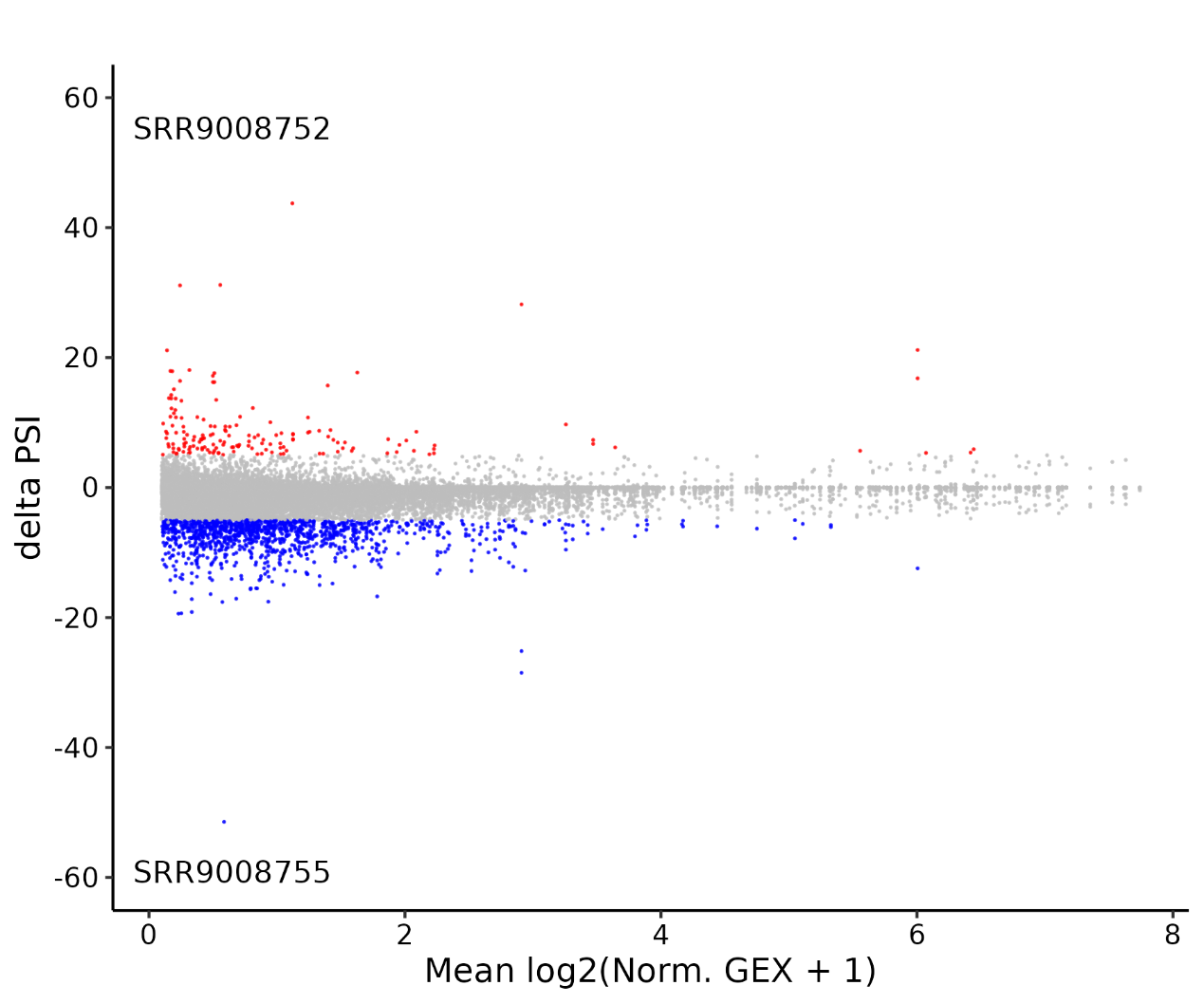


Figure 2．Junction差异表达火山图。每个点表示一个junction，一个基因可以含有多个junction。横坐标为基因在两种细胞内的平均表达量，纵坐标为junction两种细胞间PSI的差值。正值表示junction在第4天IPS细胞（SRR9008752）中高表达，负值表示在第2天的IPS细胞（SRR9008755）中高表达，其中标记为红色和蓝色的点为表达差异显著的junction，灰色点为表达差异不显著的junction。

结合基因、junction的表达差异的情况，我们将基因分为4种类型：

* Coordinated: 基因所有差异表达junction的差异方向和基因方向相同；
* Opposing: 基因所有差异表达junction的差异方向和基因方向相反；
* Iso-Switch: 基因差异表达junction差异方向相同，但是基因没有差异表达；
* Complex: 基因同时包含2个及以上差异表达junction, 且存在差异方向相反的情况；

其中Complex类型的基因最有可能检测到细胞类型间的可变剪切。在本例中，Iso-Switch类型的基因占了所有候选基因的大多数（Figure 3）。

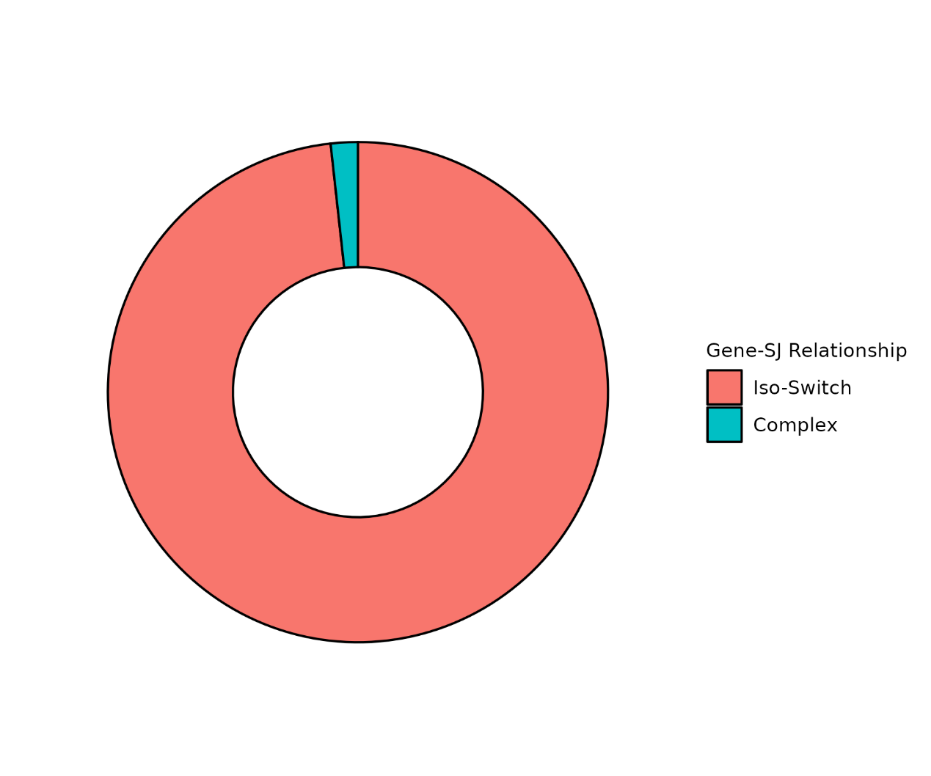


Figure 3．候选基因分类统计饼图。

分析结果我们发现，*ATP5F1C*基因可能存在可变剪切，在两种细胞类型间表达不同的转录本。*ATP5F1C*基因在两种细胞中的表达量虽然差异显著，但是其log2FoldChang仅有0.13（Table 1），通常情况下该基因并不会被认定为显著差异表达，但是却检测到三个junction显著差异表达（Figure 4，Figure 5，Table 1），其中chr10:7802855:7806973和chr10:7807011:7807658在两种细胞中均表达，但是在第2天的IPS细胞中表达量显著高于第4天IPS细胞，chr10:7802855:7807658在第4天的IPS细胞中表达量很低，在第2天的IPS细胞中依然有表达，这意味着第2天的IPS细胞和第4天的IPS细胞在表达*ATP5F1C*基因时可能表达了不同的转录本。

我们进一步依据注释信息绘制的转录本结构图（Figure 6），图中将junction按照位置进行投射，若转录本可能含有该junction，则进行标记，并不意味着该转录本一定含有对应junction。虽然如此，我们还是从结构图中读到了有效的信息，chr10:7802855:7806973和chr10:7807011:7807658两个在第2天IPS细胞中高表达的junction位于exon9的左右两侧，chr10:7802855:7807658跨过了exon9，这暗示着*ATP5F1C*基因在两种细胞中的可变剪切主要体现在是否完整保留exon9，同样的可变剪切早在1993年在肝脏和心肌组织中检测到（Figure 7）。

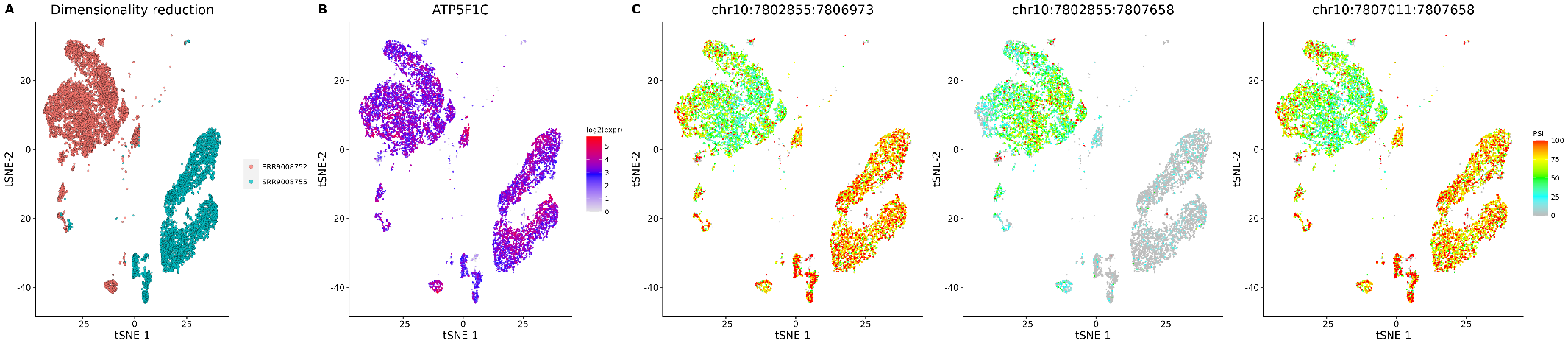


Figure 4．*ATP5F1C* 基因水平和junction水平的表达量散点图。（A）细胞分群；（B）细胞总体表达量；（C）基因在两种细胞之间的表达量；（C）三个差异junction的表达量。

Table 1．*ATP5F1C* 基因以及junction表达量统计表。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 注释 | junction | chr10:7802855:7806973 | chr10:7802855:7807658 | chr10:7807011:7807658 |
| 表达junction细胞比例 | pct.cells.expr.sj.g1 | 93.69 | 9.75 | 93.25 |
| junction占细胞表达量的比例（PSI） | psi.g1 | 74.24 | 1.38 | 71.12 |
| 表达基因细胞比例 | pct.cells.expr.gene.g2 | 97.5 | 97.5 | 97.5 |
| junction占细胞表达量的比例（PSI） | psi.g2 | 45.74 | 29.57 | 45.96 |
| 两个PSI的差值 | delta | -28.5 | 28.19 | -25.16 |
| junction差异显著性 | p-value.sj | 0 | 0 | 0 |
| 基因差异log2FoldChange | log2FoldChange.gene | 0.13 | 0.13 | 0.13 |
| 基因差异显著性 | p-value.gene | 6.56E-08 | 6.56E-08 | 6.56E-08 |

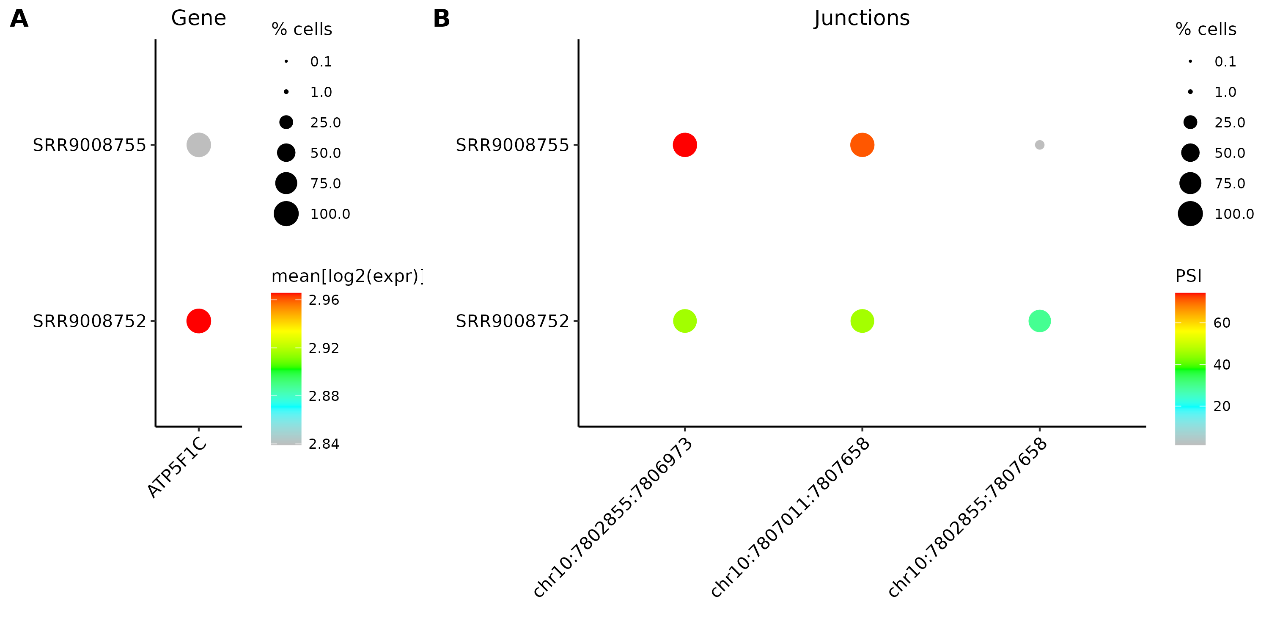


Figure 5．*ATP5F1C*基因以及三个差异表达的junction在所有细胞中表达量统计。

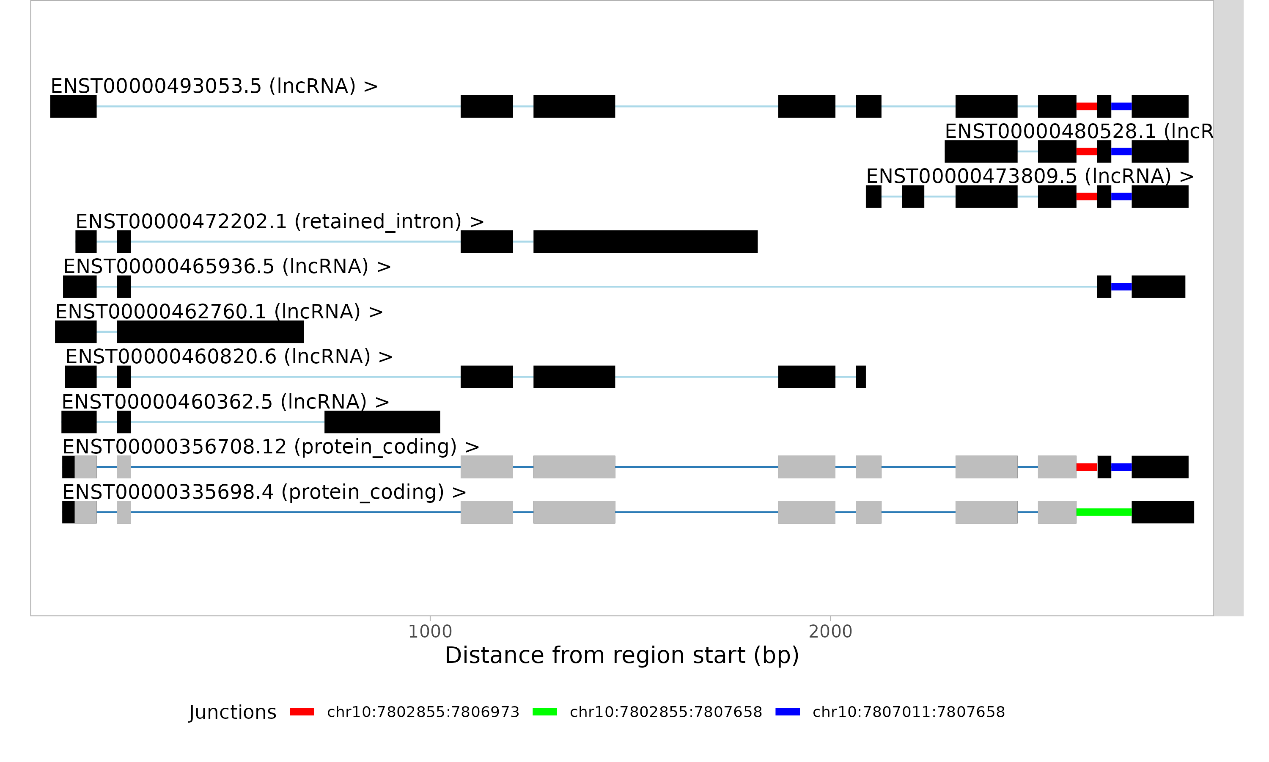


Figure 6．基于注释信息展示*ATP5F1C*基因所有转录本结构图。

每一行为一条转录本，同时标记了转录本的类型。方框表示外显子，其中灰色部分表示CDS区。给所有转录本标记可能存在的差异junction。从图中信息可以推断出*ATP5F1C*基因的可变剪切体现在是否包含倒数第二个外显子（exon9）上。

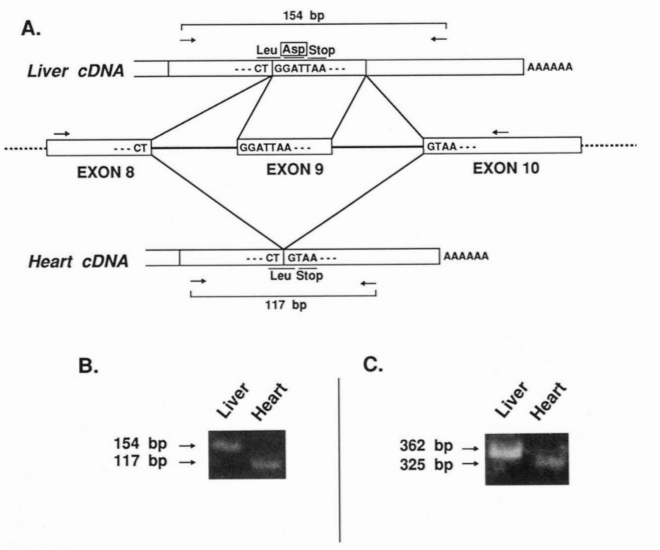


Figure 7．*ATP5F1C*基因在心脏和肝脏中表达不同的转录本[3]，主要表现为是否包含exon9，与我们检测到的可变剪切一致。

1. Wen, Wei X., A.J. Mead, and S. Thongjuea, *MARVEL: an integrated alternative splicing analysis platform for single-cell RNA sequencing data.* Nucleic Acids Research, 2023. **51**(5): p. e29-e29.

2. Ou, M., et al., *Single-cell sequencing reveals the potential oncogenic expression atlas of human iPSC-derived cardiomyocytes.* Biology Open, 2021. **10**(2): p. bio053348.

3. Matsuda, C., et al., *Gene structure of human mitochondrial ATP synthase gamma-subunit. Tissue specificity produced by alternative RNA splicing.* J Biol Chem, 1993. **268**(33): p. 24950-8.