每组内的\*\*\*\*\*\*\*-result.xlsx文件是GSEA分析的总体结果

前11列为分析数据，之后的为作图补充数据可以不看。

1，代表的KEGG中的信号通路，比如hsa04110就是KEGG中的Cell cycle的信号通路

2，对信号通路的描述，也就是信号通路的名字

3，该信号通路的基因个数

4，很熟悉的富集分数，就是ES

5，这是标准化以后的ES，全称是normalized enrichment score

6，P值

7，矫正以后的p值

8，Q值，也有的写FDR q-val(false discovery rate)，表示错误发现率

9，这里的rank就是ES在顶点时候，那个最幸运基因的位置，或者说是排名

10，这里的leading\_edge有点复杂，对富集贡献最大的基因成员,即领头亚集，用于定义Leading-edge subset的参数有：Tags，List，Signal，对于一个基因集而言，定义其中对Enrichment score贡献最大的基因为核心基因，也称之为leading edge subset，也就是ES顶点的之前的基因。tags表示核心基因占该基因集基因总数的比例，而list表示核心基因占所有基因总数的比例，signal利用这两个指标计算得到，公式如下：



N代表所有基因的数目，Nh代表该基因集下的基因总数。对于一个基因集而言，当核心基因的数目和该基因集下的基因总数相同，signal取值最大，当该基因集的基因数目和所有基因数目接近时，signal的取值接近于0。当然，我们希望的是signal越大越好。

11，这个core\_enrichment就是主要富集的基因，下面的数字其实是ENTREZID，就是一个一个基因，换句话说，就是做GSEA分析的基因列表中，Hit到该目的通路的基因列表。

————————————————

原文链接：https://blog.csdn.net/weixin\_32491317/article/details/112245840