1. **研究方法**

* **数据来源**：

从TCGA数据库下载基因表达和临床注释数据。没有完整生存数据的患者被排除在外。TCGA-肝细胞癌(TCGA-LIHC)数据集用于进一步分析。最后，本研究选择了TCGA-LIHC队列中的总共371名患者。对于TCGA数据集，用于分析基因组数据共享(GDC)数据的R包TCGA biolinks (Colaprico et al., 2016)用于下载基因每千碱基每百万映射读数(FPKM)值的片段来自GDC的表达。FPKM值进一步转换为每千碱基百万(TPM)值的转录本。使用SVA包中的参数和非参数经验贝叶斯框架算法校正了由与任何生物变化无关的因素产生的批次效应。与体细胞突变相关的数据是从TCGA数据库下载的。

* **m5C聚类分析**

从数据集中提取m5C甲基化修饰调节因子：11个writers（NSUN2，NSUN3，NSUN4，NSUN5，NSUN6，NSUN7，DNMT1，TRDMT1，DNMT3A，DNMT3B和NOP2），3个readers（TET1，TET2，TET3），1个erasers（ALYREF）。用无监督聚类分析来区分不同的m5C修饰，然后对患者进行分类以进行后续分析。

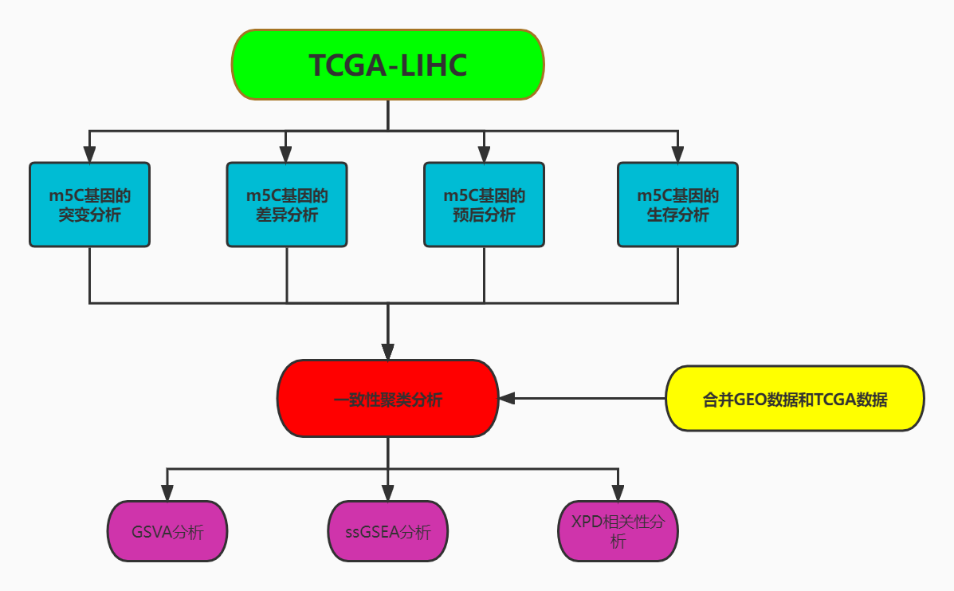
* **基因集变异分析和功能注释**

为了探索m5C修饰模式中生物过程的差异，使用基因集变异分析(“GSVA”) R包来执行GSVA。该软件包基于非参数和无监督算法，广泛用于估计表达数据集中基因集富集的变化（Hänzelmann等人，2013年）。GSVA是使用从分子特征数据库(MsigDB)获得的“c2.cp.kegg.v6.2.symbols”基因组实现的。小于0.05的调整后P值被认为具有统计学意义。我们应用“ClusterProfiler”R包在<0.05的错误发现率(FDR)阈值下对m5C相关基因进行功能注释。

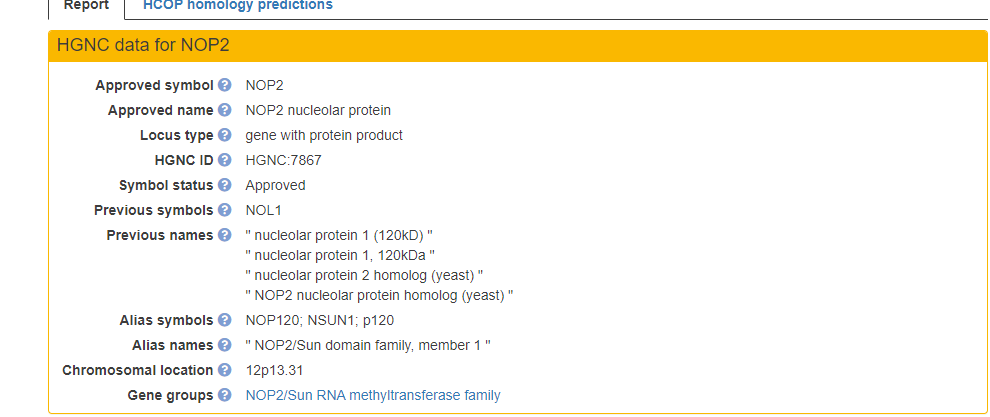
* **单样本基因集富集分析**

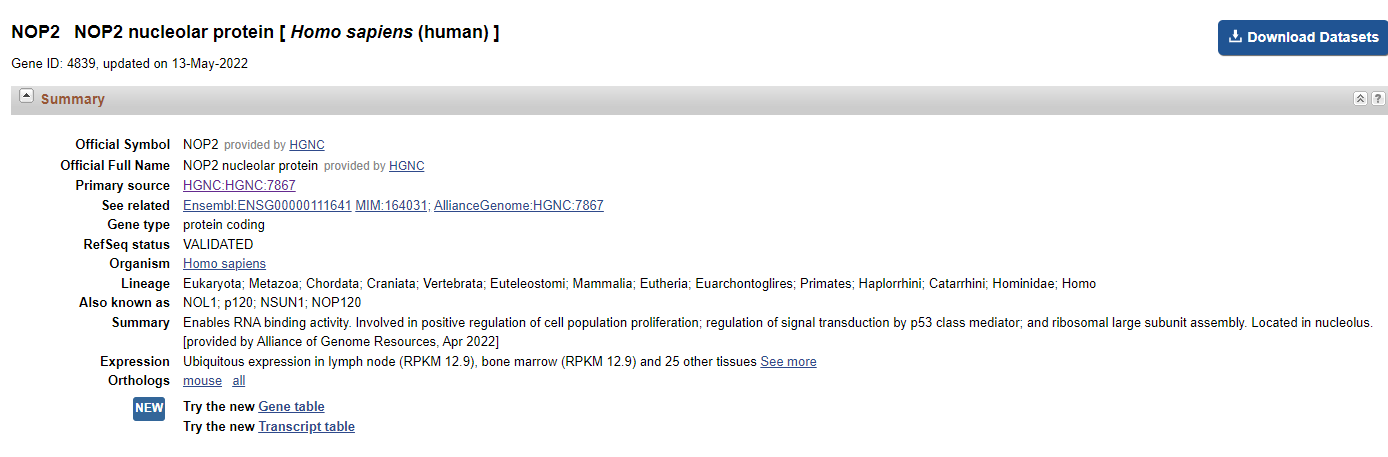
单样本基因集富集分析(ssGSEA)算法用于确定TME中细胞浸润的相对丰富度。我们从Charoentong获得了与TME中每种浸润性免疫细胞类型相关的基因集，Charoentong存储了各种人类免疫细胞的信息，包括CD8T细胞、树突细胞(DC)、自然杀伤(NK) T细胞、巨噬细胞、调节性T细胞细胞等。ssGSEA用于确定富集分数并定义相应样品中每种TME浸润细胞类型的相对丰度。

1. **分析流程**



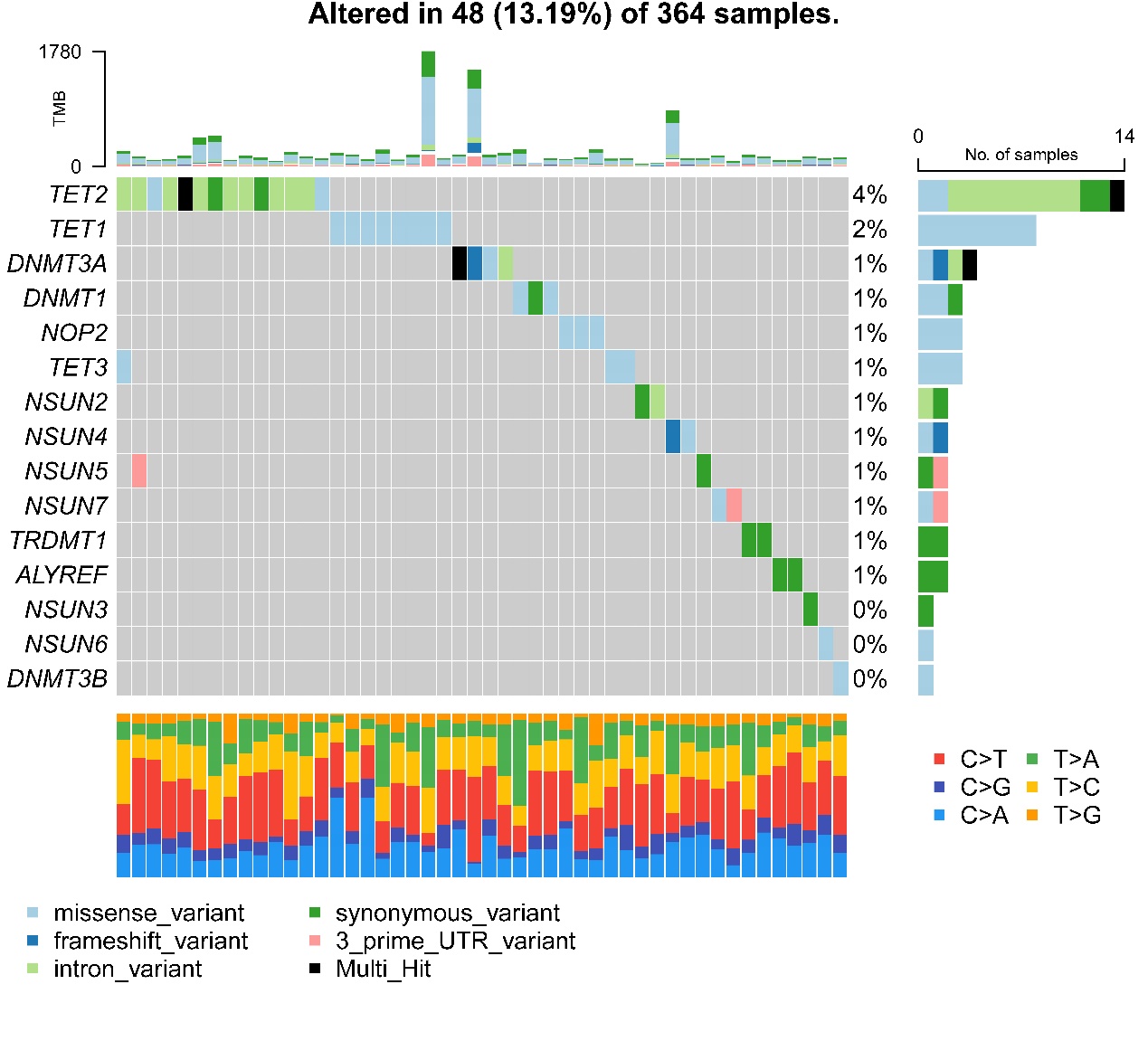
本文中NSUN1在NCBI以及HGNC中显示的基因名字为NOP2





结果

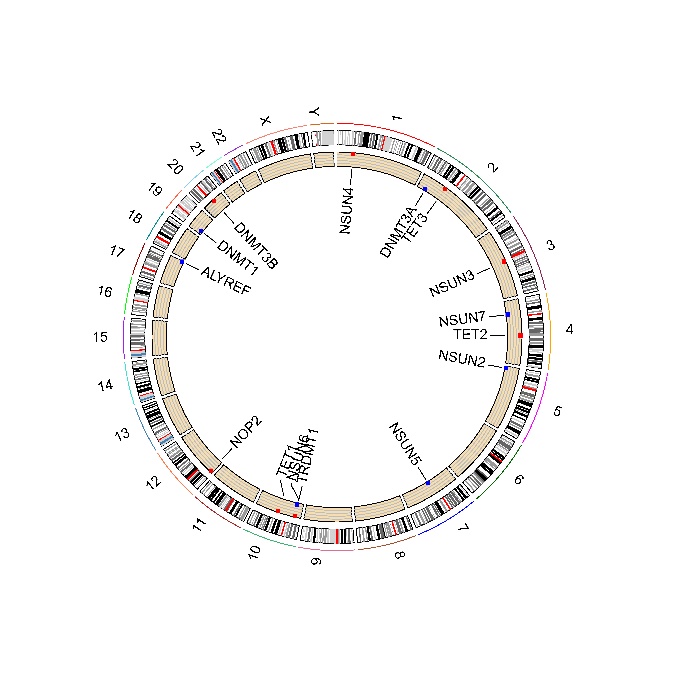
共鉴定出15个m5C调节因子，包括11个11个writers、1个erasers和3个readers。首先，总结了 HCC 中 CNV 的发生率和调节因子中的体细胞突变。364份样本中，48份m 5 C调节子发生突变，发生率为13.19%。发现TET2突变的频率最高4%，其次是 TET2为2%，而 DNMT3B、NSUN3 和 NSUN6则没有突变（图 1A）。其余调节子突变频率为1%。



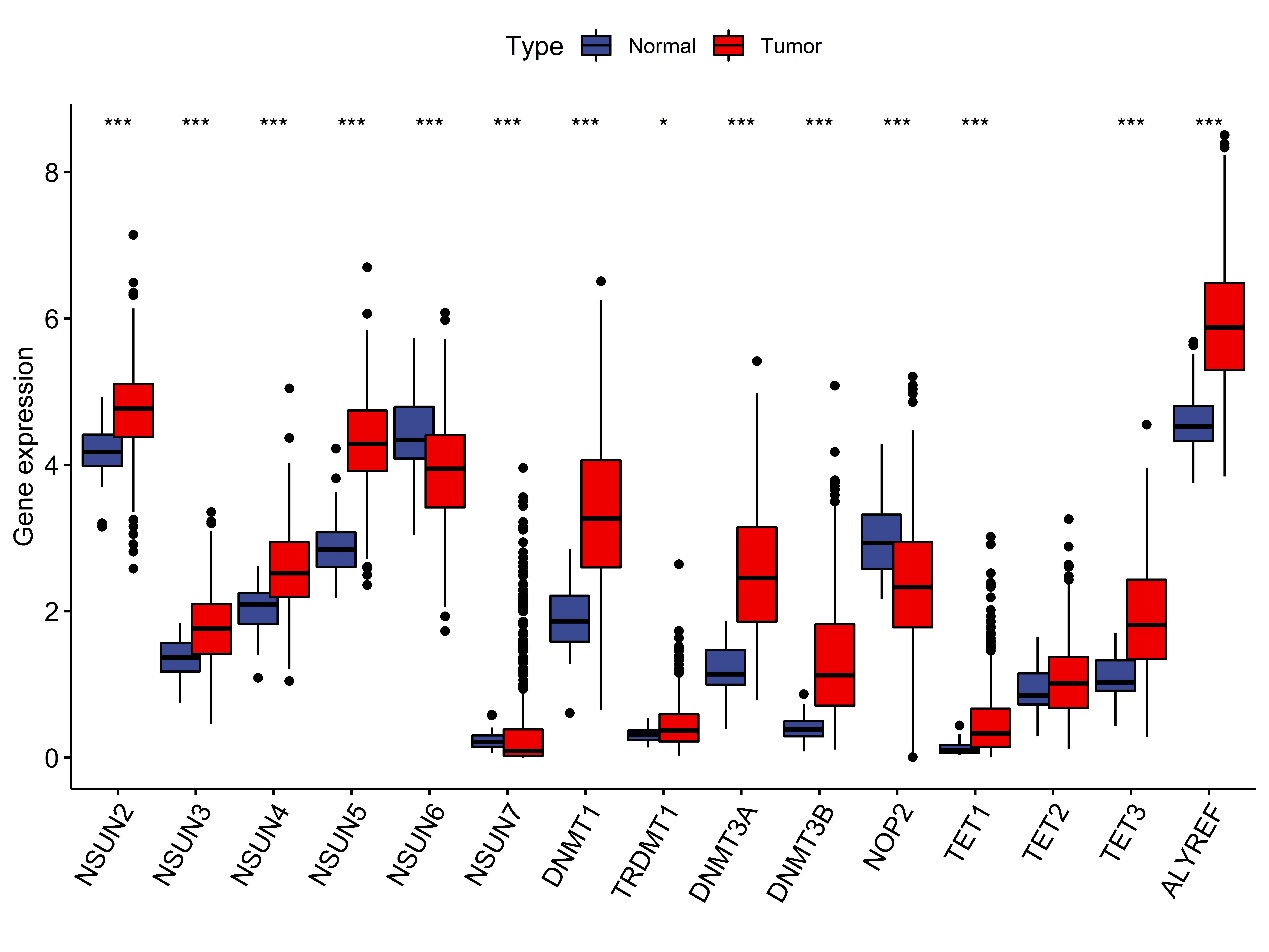
在探索其修饰频率后，还在 15 个m5C调节因子中检测到 CNV。大多数修改涉及拷贝数增加，但 TET2、NOP2 、NSUN4和DNMT1 出现广泛缺失（图 1B）。



m 5 C 调节子的染色体位点显示在图 1C(此处需要修图)



为了确定m 5 C 调节因子的表达是否受到上述基因突变的影响，我们研究了调节剂的mRNA表达。我们发现 m 5 C 的变化是导致 m 5 C 调节剂表达扰动的重要因素。与正常肝组织相比，具有CNV扩增的m 5 C调节因子的表达显着低于HCC组织（例如，ALYREF和NSUN2）（图1B，E ））。具有广泛缺失如NOP2和NSUN26显著高于HCC组织。

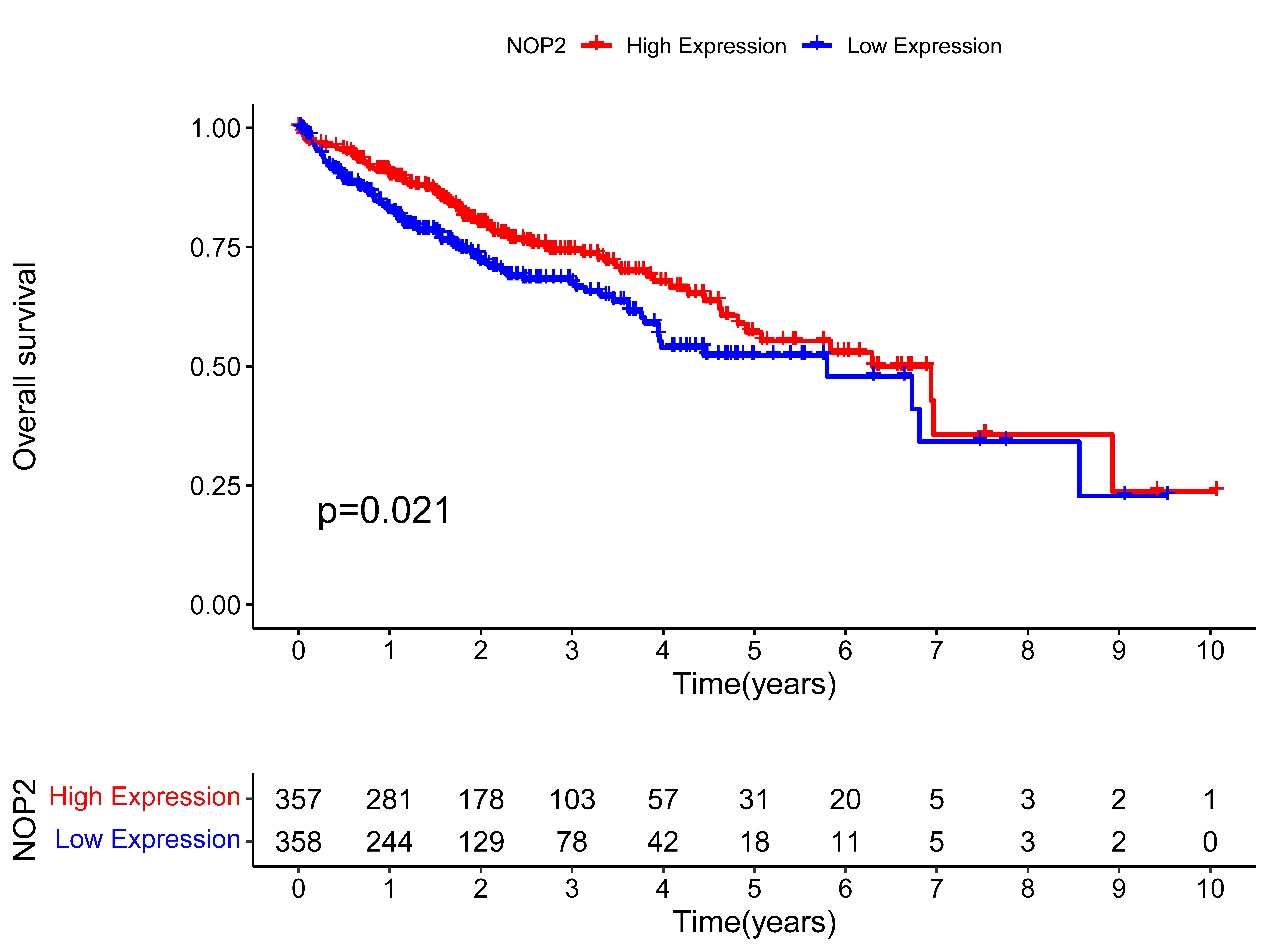


由 15 种调节剂介导的m 5 C 甲基化改变模式

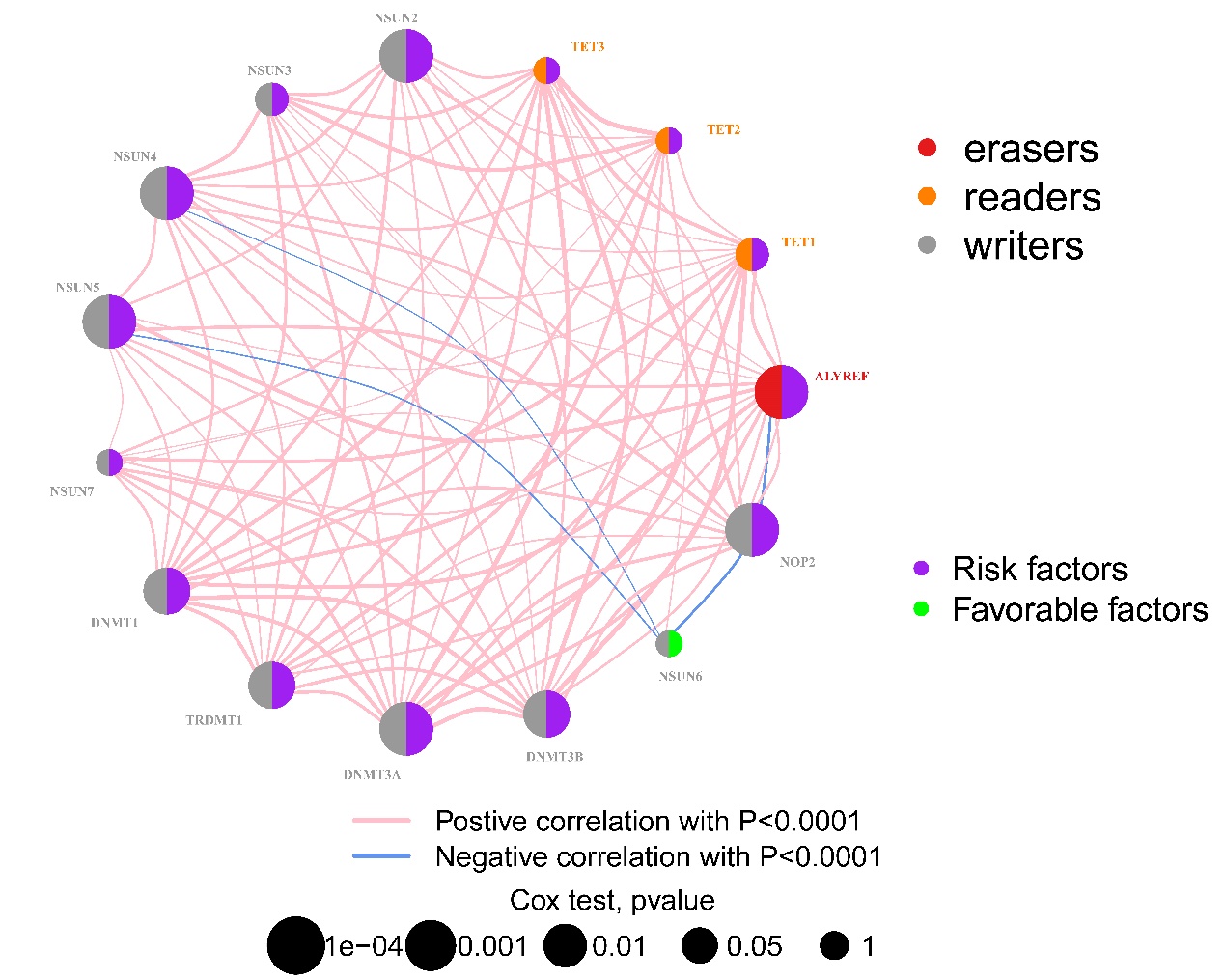
单变量 Cox 回归分析显示 13 m 5 C 调节剂在 HCC 患者中具有预后意义（表一）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id | HR | HR.95L | HR.95H | pvalue | km |
| NSUN2 | 1.712684 | 1.309638 | 2.239769 | 8.48E-05 | 5.16E-05 |
| NSUN3 | 1.531225 | 1.101138 | 2.129298 | 0.01132 | 2.16E-05 |
| NSUN4 | 1.995773 | 1.488238 | 2.676394 | 3.92E-06 | 4.01E-06 |
| NSUN5 | 1.615134 | 1.272541 | 2.049961 | 8.10E-05 | 5.97E-08 |
| NSUN6 | 0.9332 | 0.769735 | 1.13138 | 0.481652 | 0.034182 |
| NSUN7 | 1.03159 | 0.834932 | 1.274568 | 0.773191 | 0.048964 |
| DNMT1 | 1.337036 | 1.148314 | 1.556775 | 0.000183 | 8.95E-06 |
| TRDMT1 | 2.328204 | 1.48965 | 3.638797 | 0.000208 | 2.76E-07 |
| DNMT3A | 1.5029 | 1.239007 | 1.822998 | 3.54E-05 | 6.15E-07 |
| DNMT3B | 1.451218 | 1.176145 | 1.790624 | 0.000515 | 4.88E-05 |
| NOP2 | 1.588984 | 1.259614 | 2.00448 | 9.33E-05 | 1.05E-10 |
| TET1 | 1.531209 | 1.096639 | 2.137987 | 0.012363 | 0.002173 |
| TET2 | 1.022355 | 0.746085 | 1.400925 | 0.890594 | 0.156398 |
| TET3 | 1.260372 | 0.997683 | 1.592227 | 0.052317 | 0.001437 |
| ALYREF | 1.567842 | 1.329861 | 1.848411 | 8.61E-08 | 6.14E-08 |

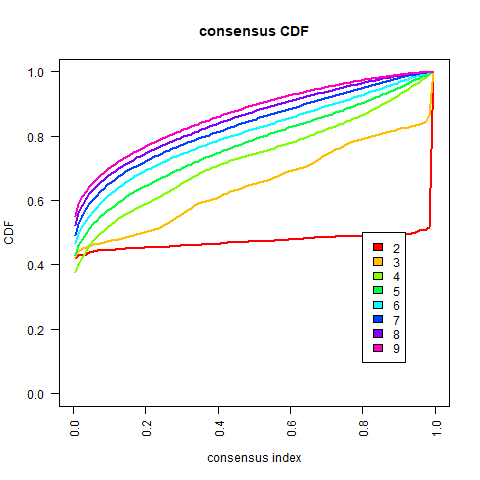
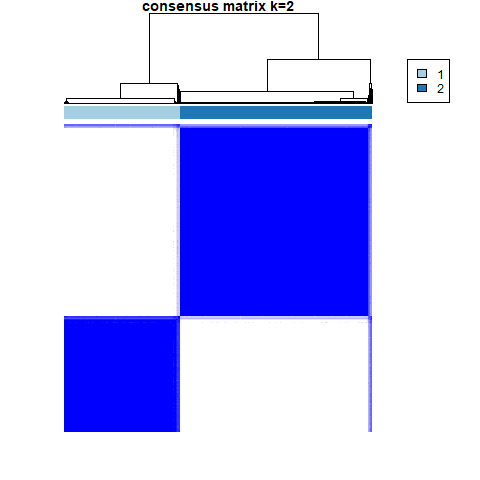
m 5 C 调节网络揭示了 m 5 C 调节剂相互作用、调节剂连接及其对患者的预后意义，同时对NOP2进行生存分析，使用中位值来区分患者NOP2基因的高低表达，可以发现高表达的NOP2患者预后要更好。

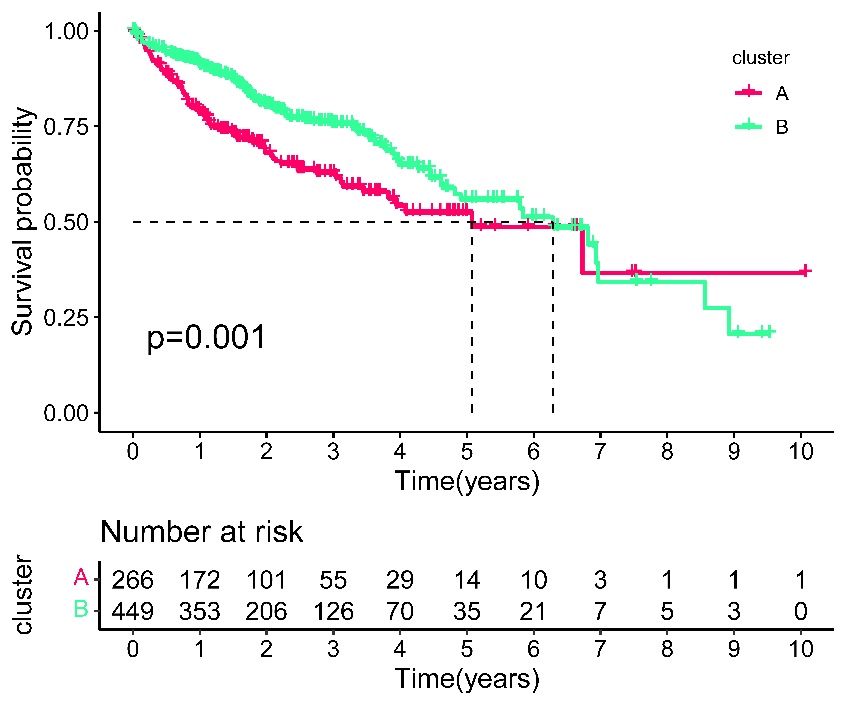


经过单因素cox分析后我们根据HR值进行m5C预后网络的构建，除NSUN6不是风险因子，其余都是风险因子。肝癌中m5C调节器之间的相互作用。圆圈大小代表各调节因子对预后的影响，通过Log-rank检验计算的数值范围分别为P＜0.0001、P＜0.001、P＜0.01、P＜0.05和P＜1。圈内的颜色代表m5C的基因类型和风险基因类型。连接调节器的线表示它们的相互作用，厚度表示调节器之间的相关强度。负相关用蓝色标记，正相关用粉色标记。这里我们发现NOP2与erasers ALYREF和writers NSUN6呈负相关。

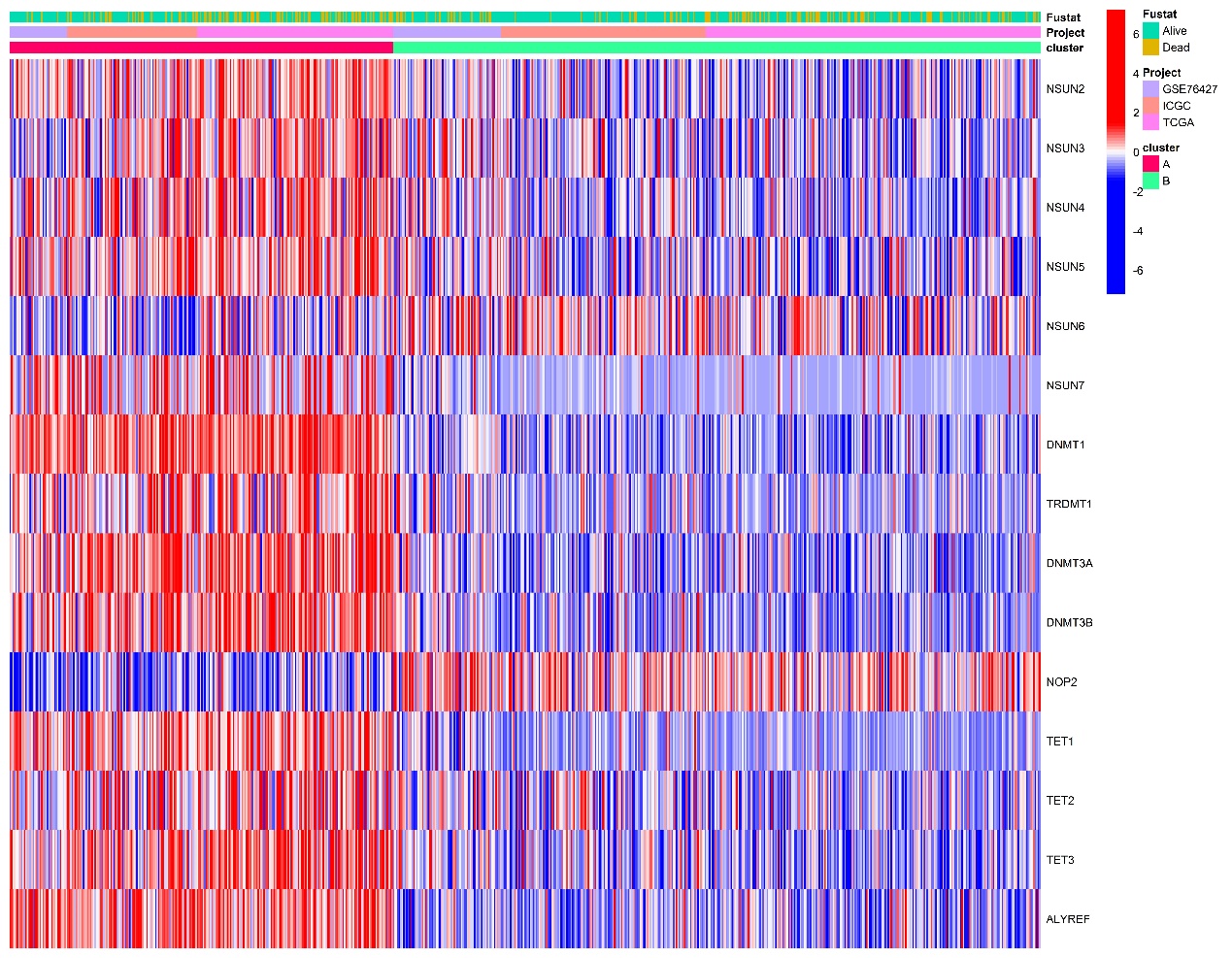


基于GEO队列GSE76427和ICGC队列 LIRI-JP以及TCGA-LIHC的732名HCC患者，我们使用ConsensusClusterPlus的R软件包，根据15个m5C调节器的表达情况，对具有定性不同的m5C修饰模式的患者进行分类，根据Delta 图我们判断选择k=2。最终利用无监督聚类确定了两种不同的修饰模式，包括模式A的266例、模式B的449例，我们将这些模式分别称为cluster A-B。对两种主要m5C修饰亚型的预后分析显示p=0.001，认为两种m5C修饰模式之间存在明显的生存差异。

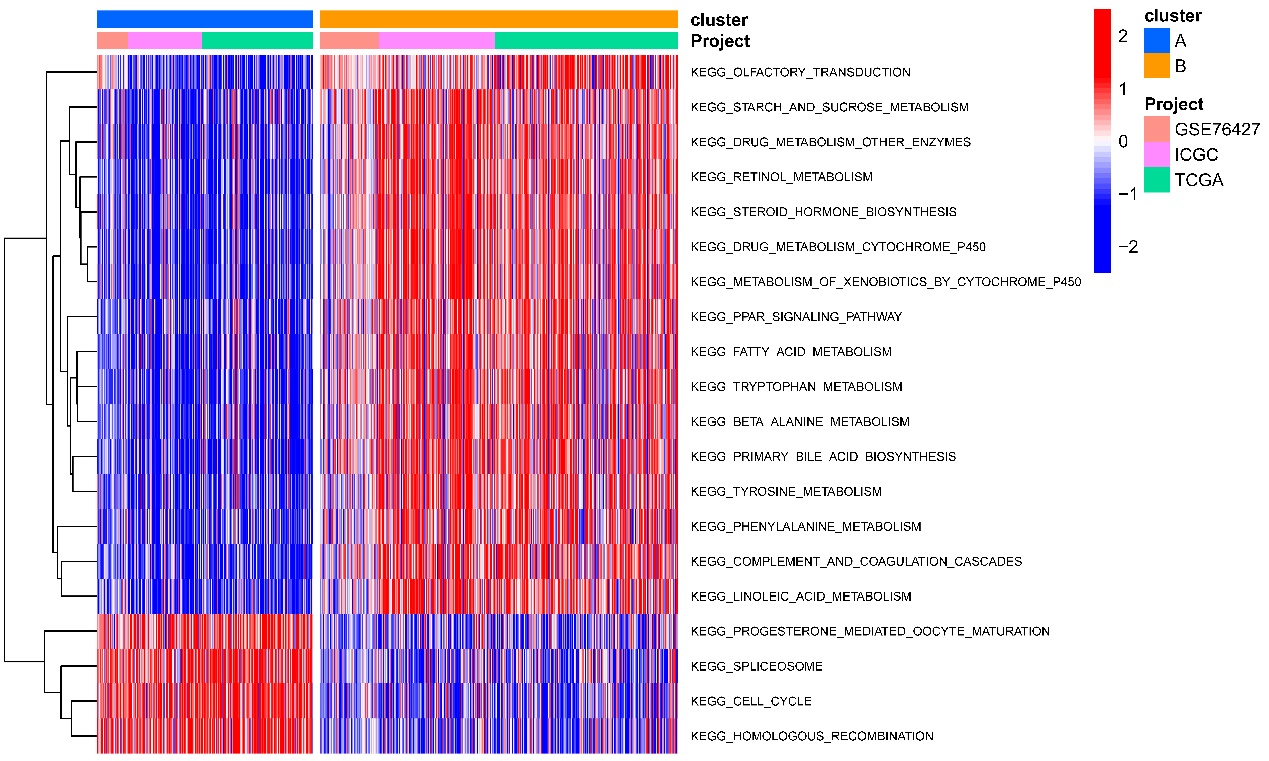




同时我们分析15个m5C调节器在不同簇的表达状况，可以发现A簇中显著高调了除NOP2和 NSUN6所以的基因

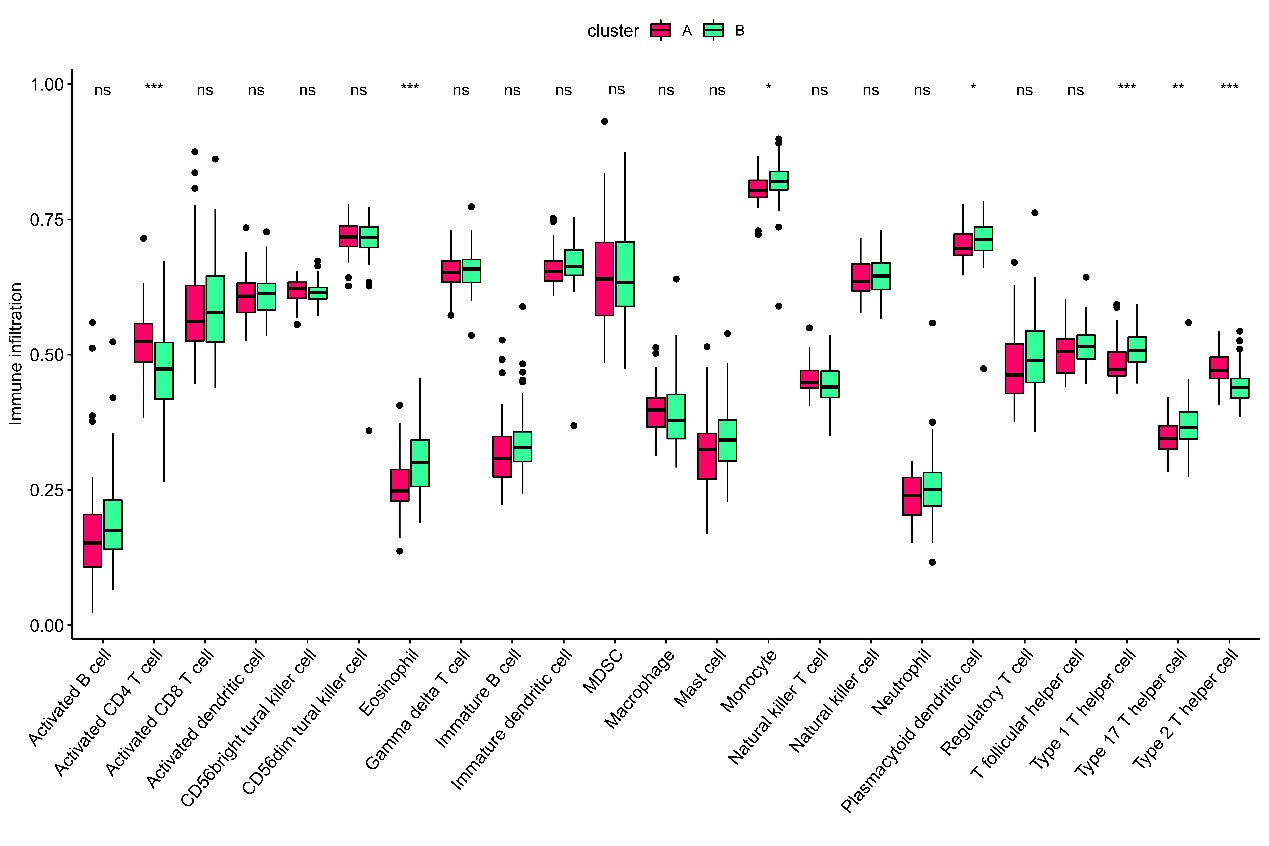


为了探索这些不同的m6A修饰模式之间的生物学行为，我们进行了GSVA富集分析。cluster-A明显富集于的途径各种代谢相关机制，而cluster-A明显富集于细胞周期（可根据附件B-A.diff.txt中的选择认为较好叙述的重新绘制图）

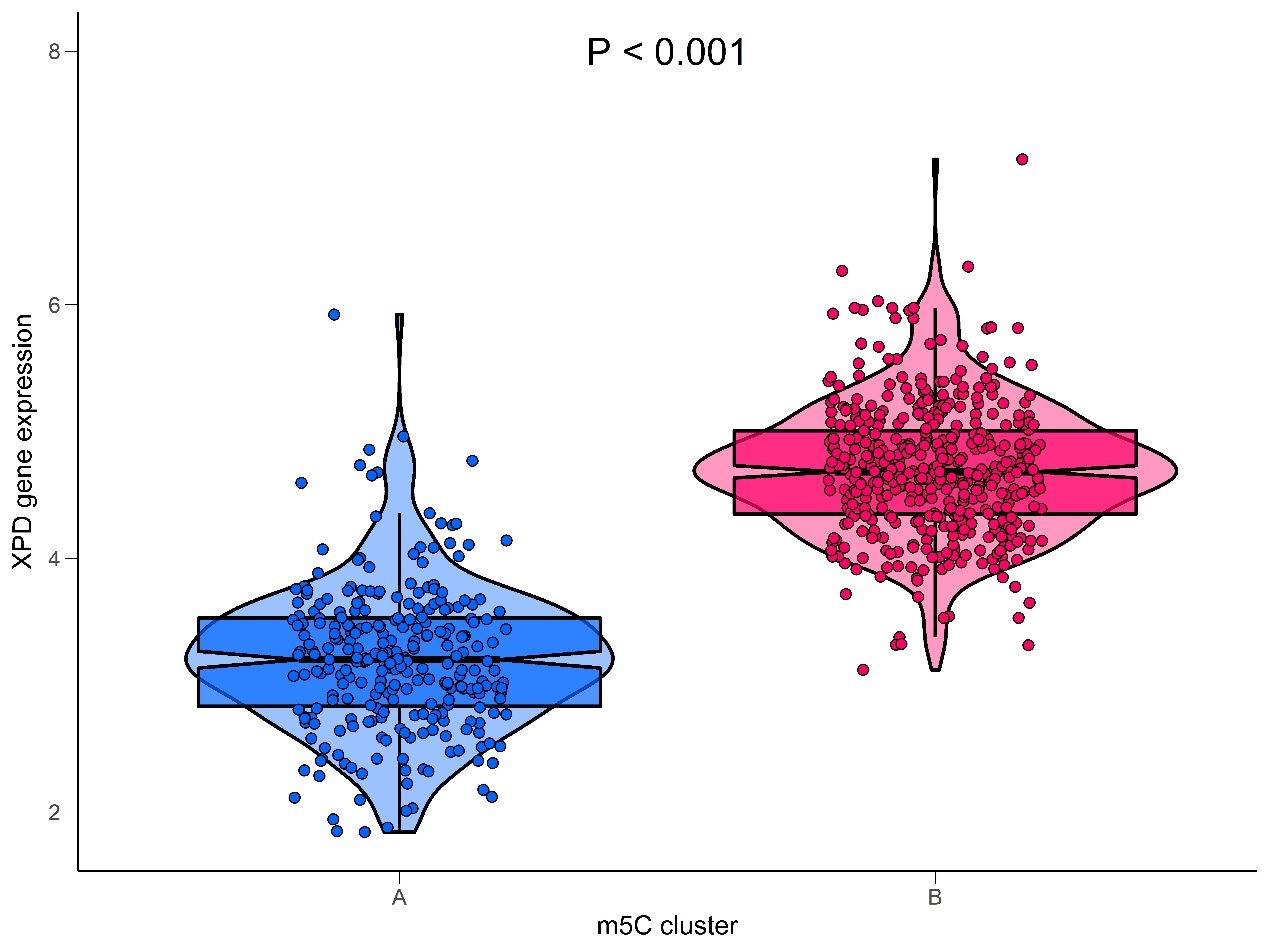


我们运用ssGSEA的方法对HCC样本进行分析，选取23个免疫细胞进行分析，通过 ssGSEA 分析计算的富集分数用于表示每个样品中代谢物相对丰度，仅有Activated CD4 T cell、Eosinophil、Monocyte、Plasmacytoid dendritic cell、Type 1 T helper cell、Type 17 T helper cell、

Type 2 T helper cell具有统计学意义。



接着我们分析不同簇间XPD的表达，可以发现A簇的表达明显低于B簇，我们发现XPD与NOP2呈正相关的关系。



随后为了进一步了解XPD与m5C调控因子的相关性，我们发现XPD与NOP2呈高度正相关,与DNMT1呈负相关，与NSUN2等系列家族基因呈弱的负相关。

