20190522 结直肠癌补充生信分析结果

吴老师，

SDC2 三个DMR的序列和位置已经发给您，烦请补充以下结果：

首先，将390多例癌症样本按照部位进行分类（如直肠、乙状结肠、升结肠等等）

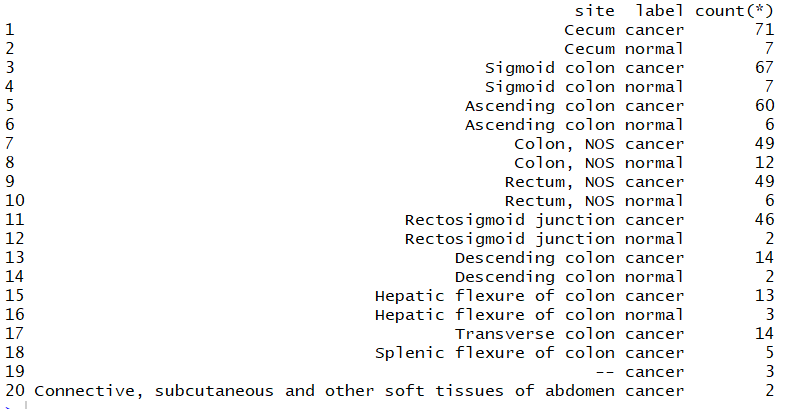
对于每个部位的癌症，做6条ROC曲线，分别是单个DMR的ROC和两个DMR

组合的ROC，各三个。

谢谢！

Response：

TCGA结直肠癌数据集共包括387给病例，总计438例样本，其中癌症样本393，癌旁45. 查阅临床信息，整理各个部位的样本分布列表如下：



从上表可见，大多数部位的癌旁样本非常少。为此，我们将所有45例癌旁样本当作一个整体，与其它各个部位的癌症样本进行比较分析，做出ROC曲线。

根据你提供的序列，我们找到对应的三个DMR区域的探针构成如下：

DMR1： cg13096260; cg18719750; cg24732574; cg08979737; cg25070637

DMR2： cg08979737; cg25070637; cg14538332; cg16935295

DMR3： cg14538332; cg16935295

对于一个样本来说，我们采用该样本在这个DMR上的所有探针的甲基化水平的均值，作为该样本在这个DMR上的甲基化水平。如果两个或两个以上DMR组合使用，我们分别求出该样本在不同DMR上的甲基化水平，然后用random forest模型进行建模，获得该模型的ROC曲线。