结直肠癌差异甲基化区域

艾米森生命科技有限公司
生物信息部

2019-05-27

Contents

Chapter 1

问题描述

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！

Chapter 2

解决方案

2.1 思路

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

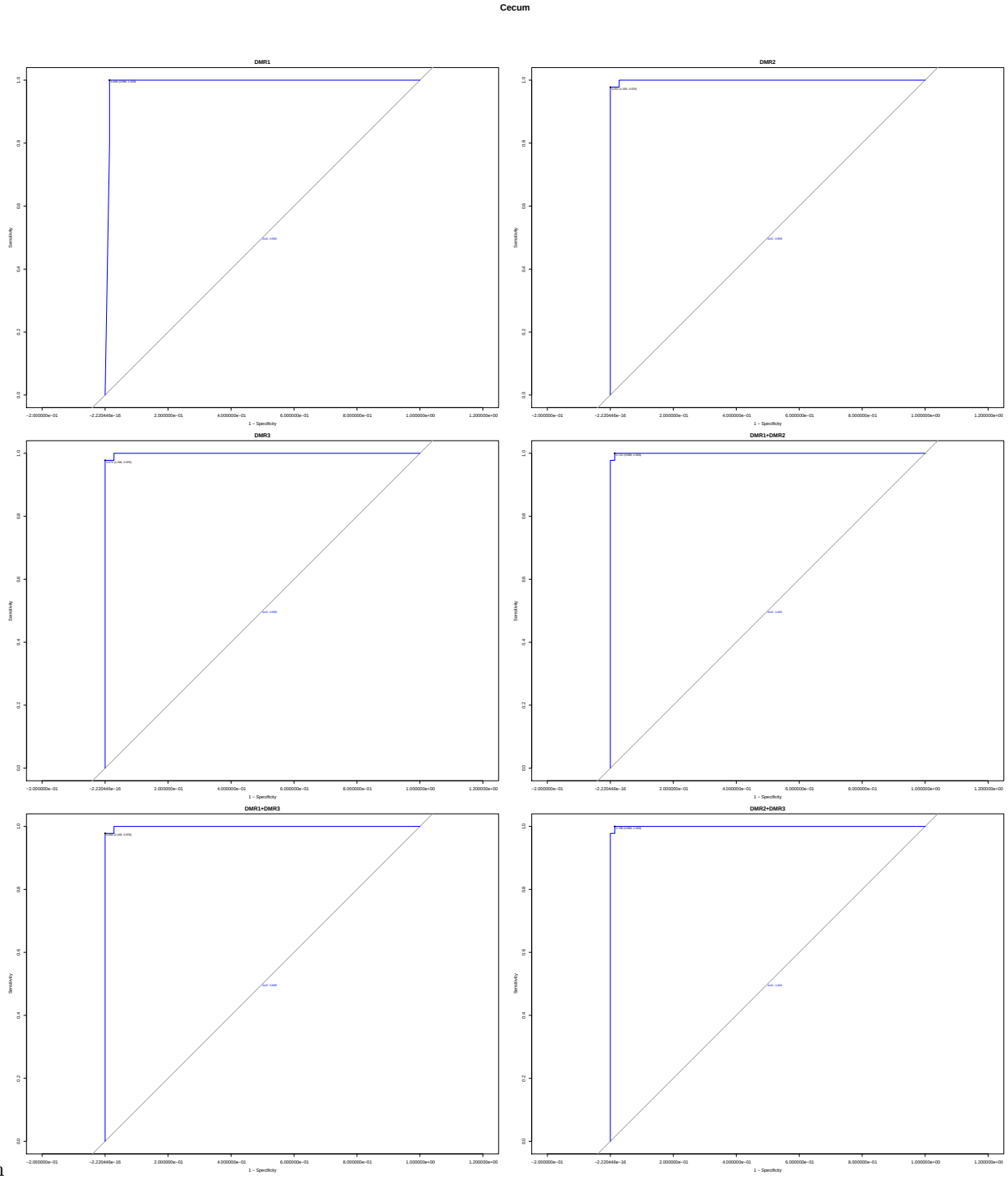
Table 2.1: 结直肠各部位样本数

部位	样本类别	样本数量
Cecum	cancer	71
Cecum	normal	7
Sigmoid colon	cancer	67
Sigmoid colon	normal	7
Ascending colon	cancer	60
Ascending colon	normal	6
Colon, NOS	cancer	49
Colon, NOS	normal	12
Rectum, NOS	cancer	49
Rectum, NOS	normal	6
Rectosigmoid junction	cancer	46
Rectosigmoid junction	normal	2
Descending colon	cancer	14
Descending colon	normal	2
Hepatic flexure of colon	cancer	13
Hepatic flexure of colon	normal	3
Transverse colon	cancer	14
Splenic flexure of colon	cancer	5
–	cancer	3
Connective, subcutaneous and other soft tissues of abdomen	cancer	2

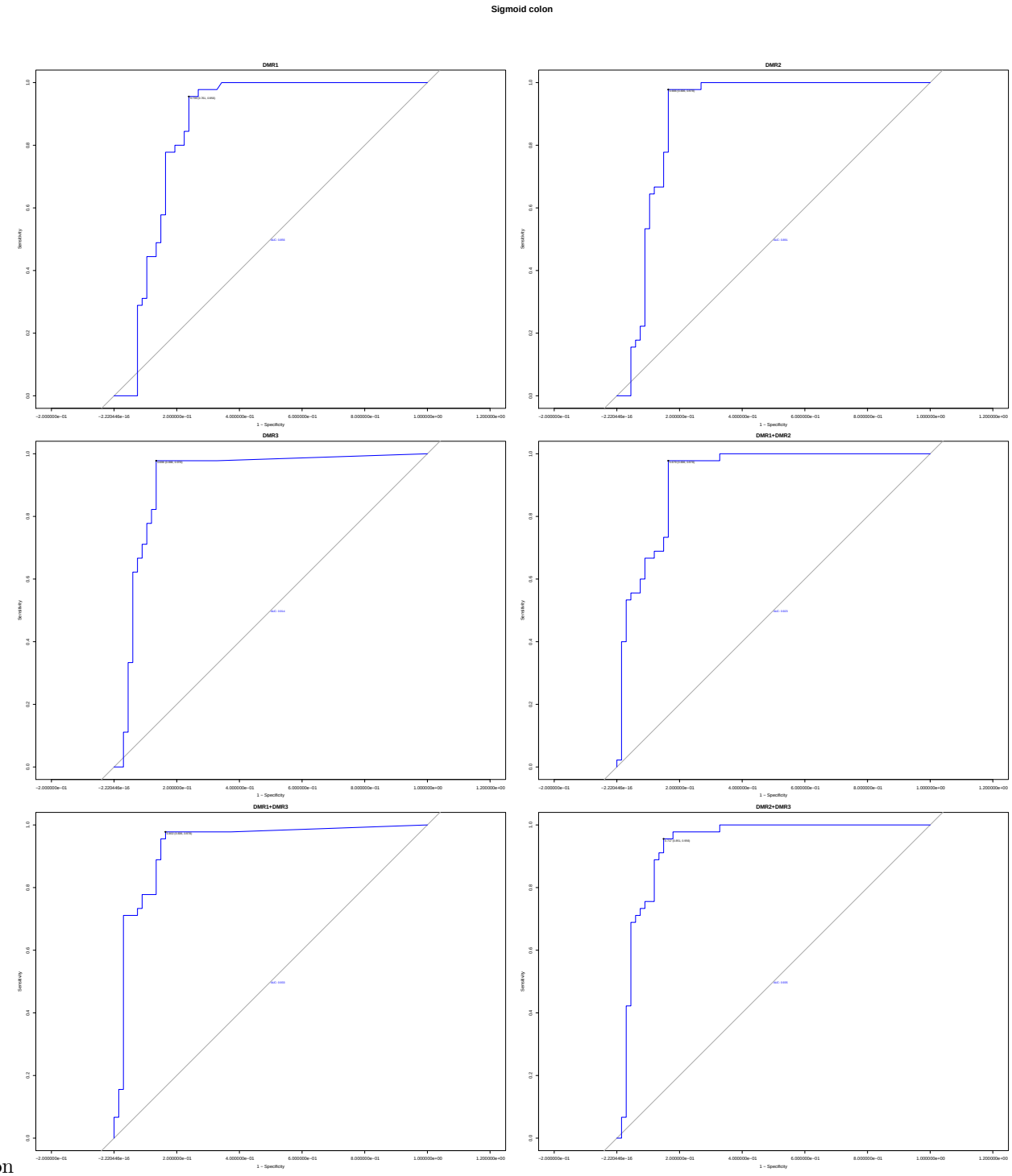
从上表可见，大多数部位的癌旁样本非常少。为此，我们将所有 45 例癌旁样本当作一个整体，与其它各个部位的癌症样本进行比较分析，做出 ROC 曲线。根据你提供的序列，我们找到对应的三个 DMR 区域的探针构成如下:

- DMR1：cg13096260; cg18719750; cg24732574; cg08979737; cg25070637
- DMR2：cg08979737; cg25070637; cg14538332; cg16935295
- DMR3：cg14538332; cg16935295

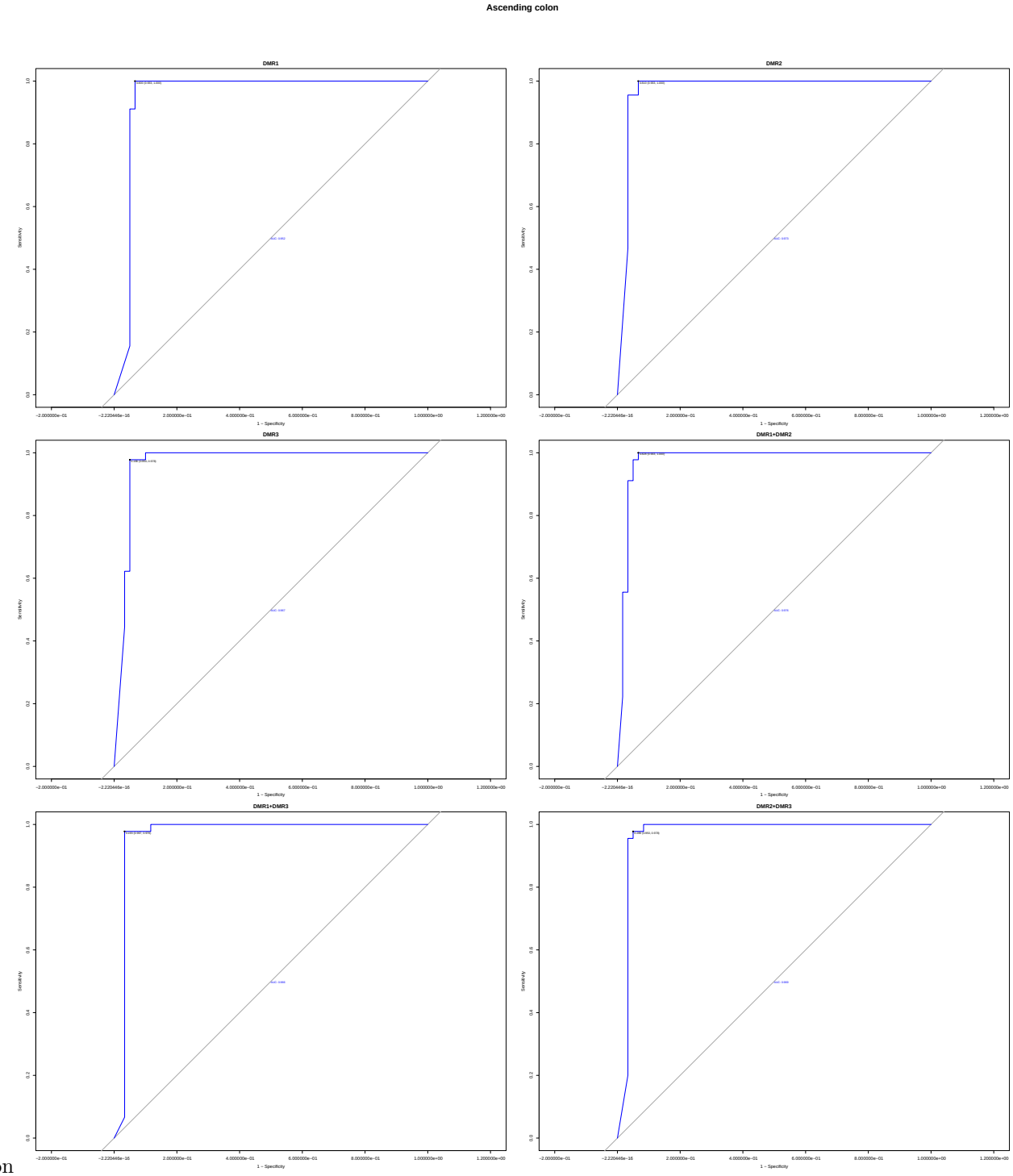
对于一个样本来说，我们采用该样本在这个 DMR 上的所有探针的甲基化水平的均值，作为该样本在这个 DMR 上的甲基化水平。如果两个或两个以上 DMR 组合使用，我们分别求出该样本在不同 DMR 上的甲基化水平，然后用 random forest 模型进行建模，获得该模型的 ROC 曲线。



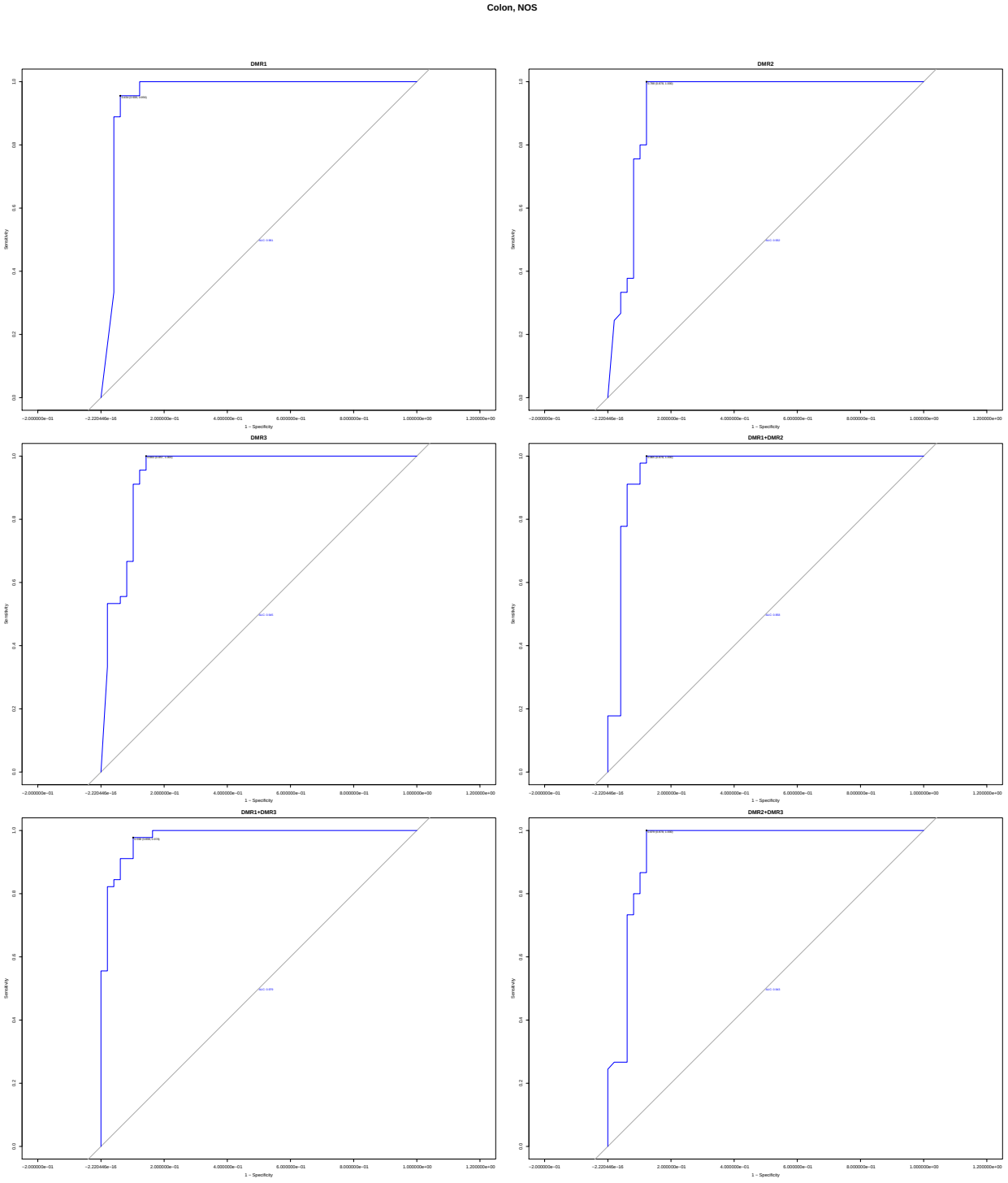
Cecum



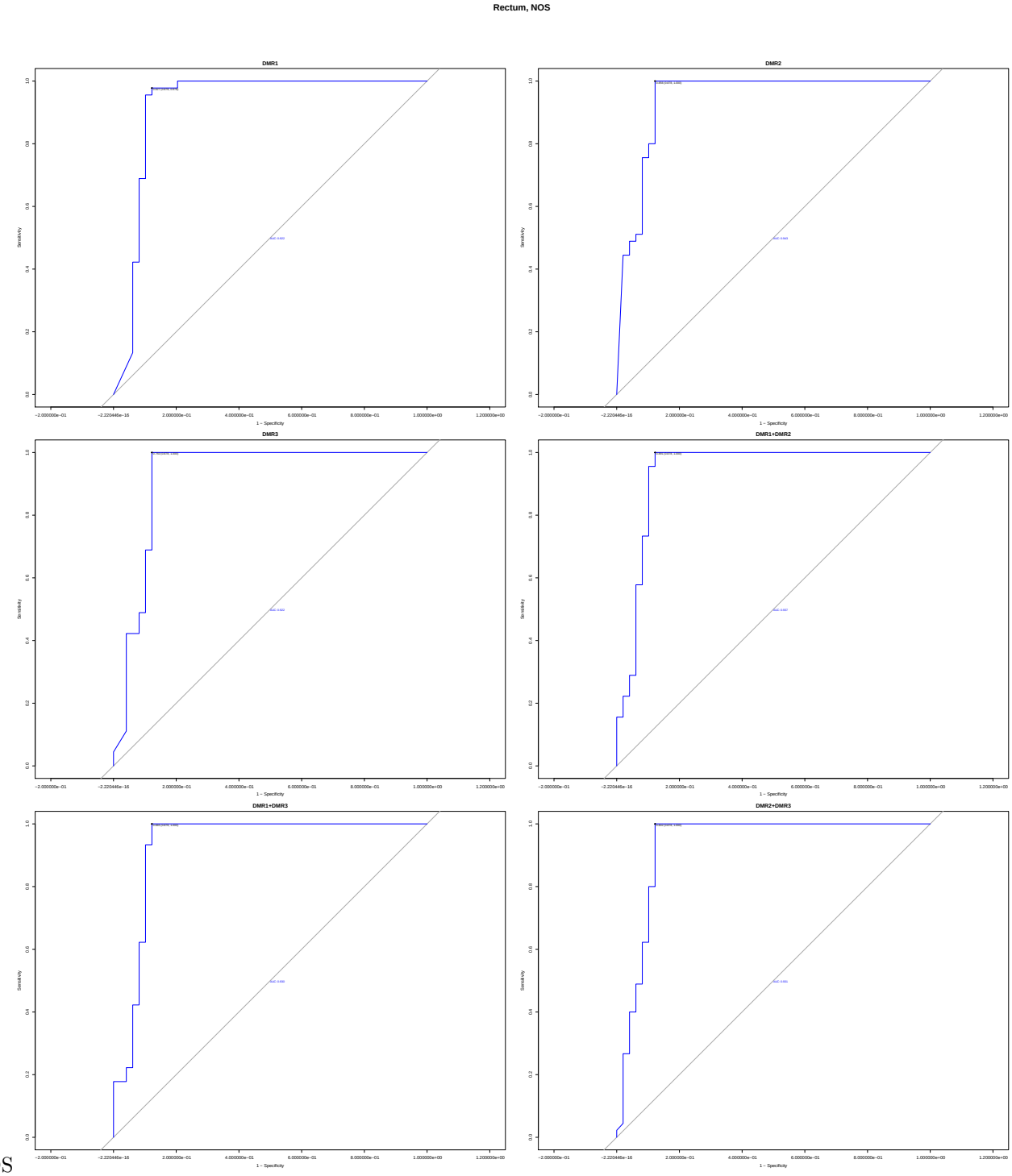
Sigmoid colon



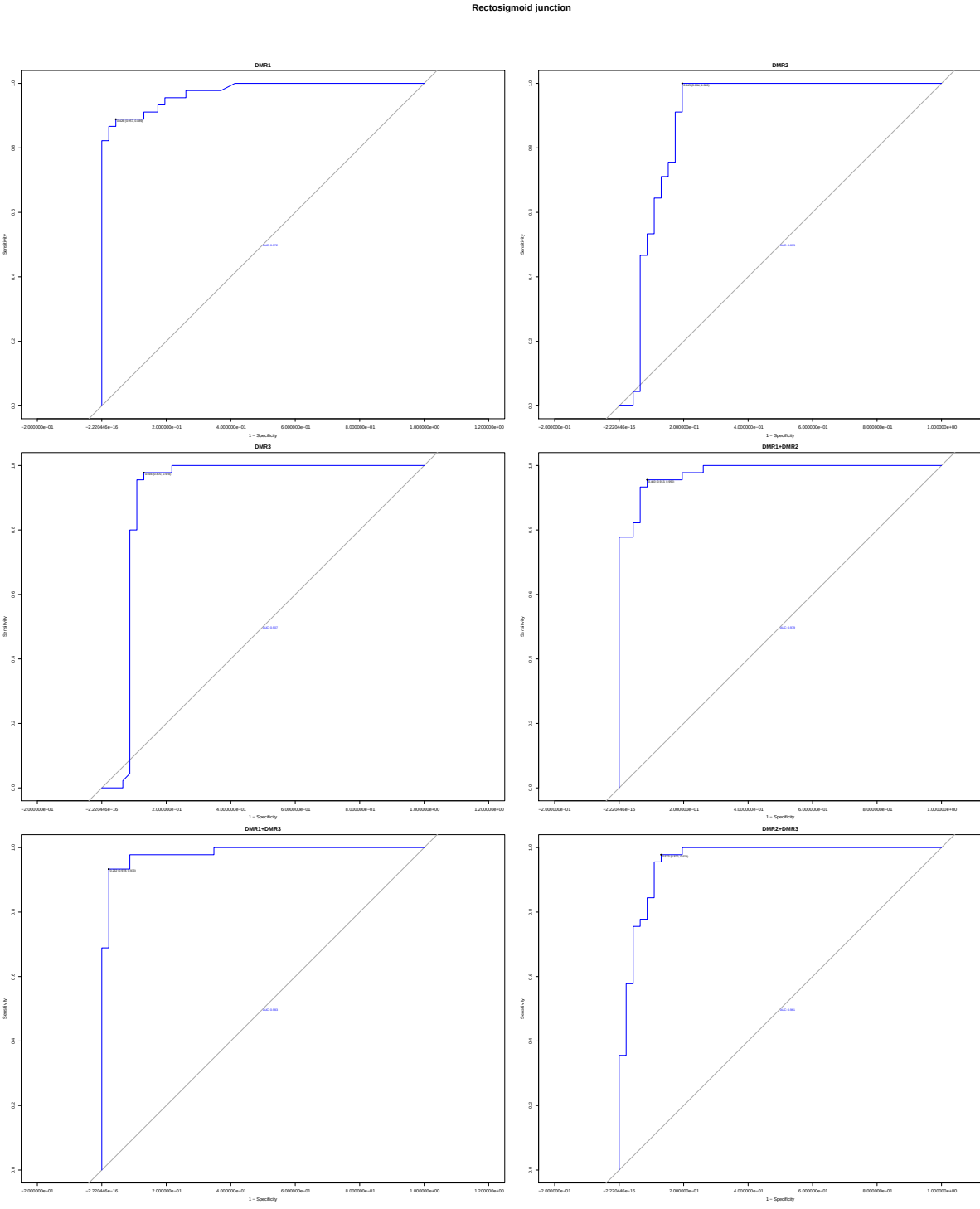
Ascending colon


Colon, NOS
DMR1
DMR2
DMR3
DMR1+DMR2
DMR1+DMR3
DMR2+DMR3
Sensitivity
1 − Specificity

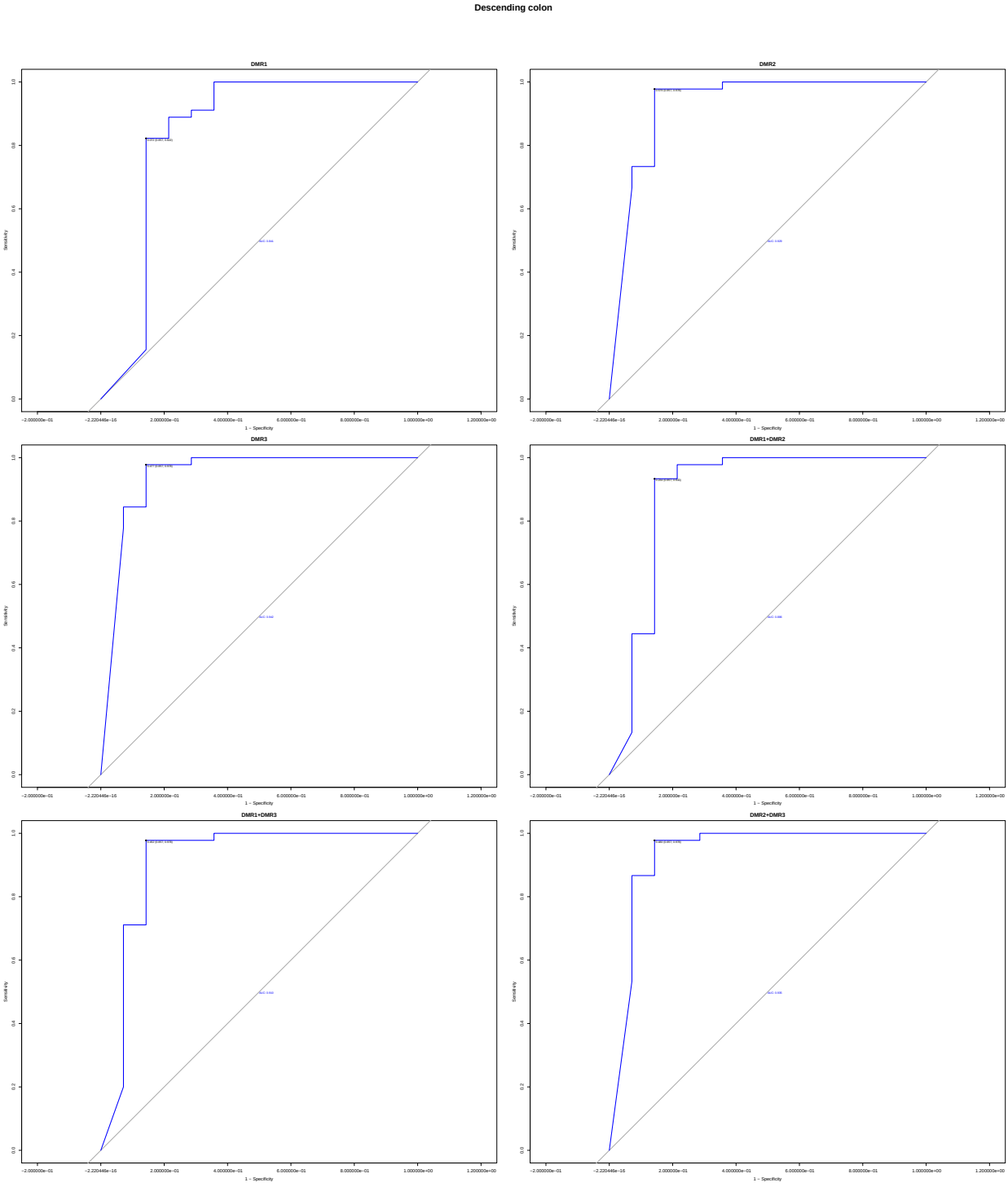
Colon, NOS


Rectum, NOS
DMR1
DMR2
DMR3
DMR1+DMR2
DMR1+DMR3
DMR2+DMR3
Sensitivity
1 − Specificity

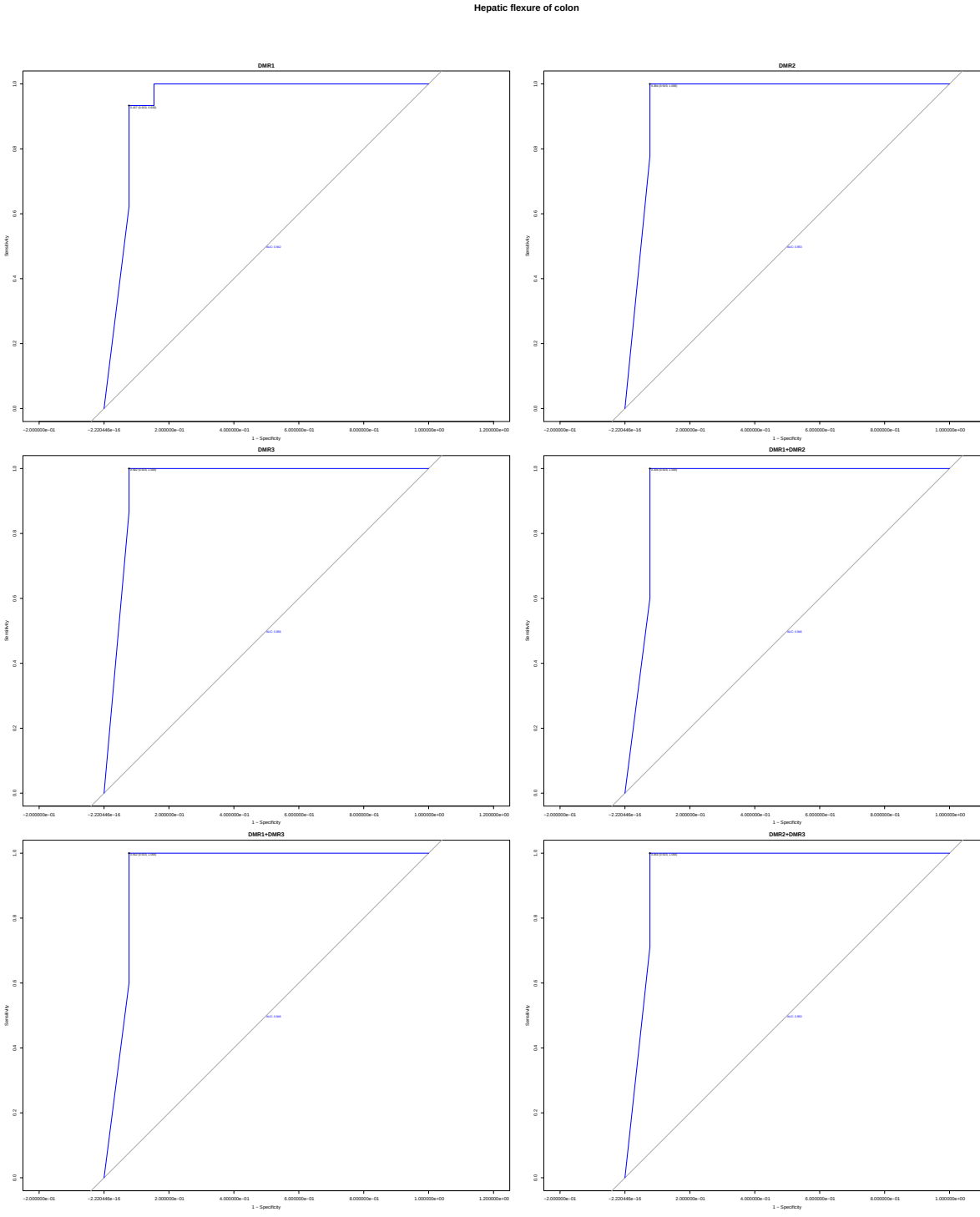
Rectum, NOS



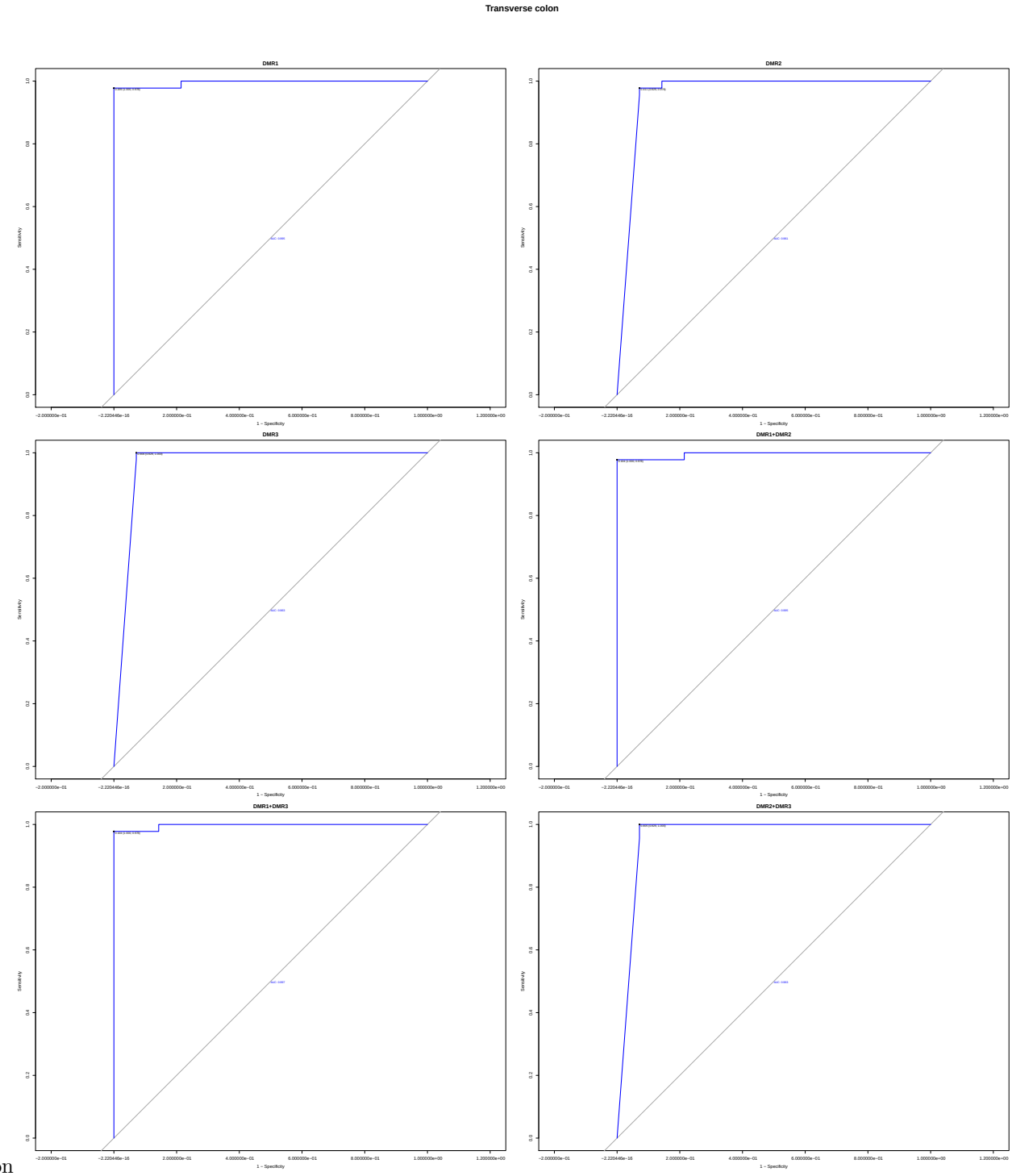
Rectosigmoid junction



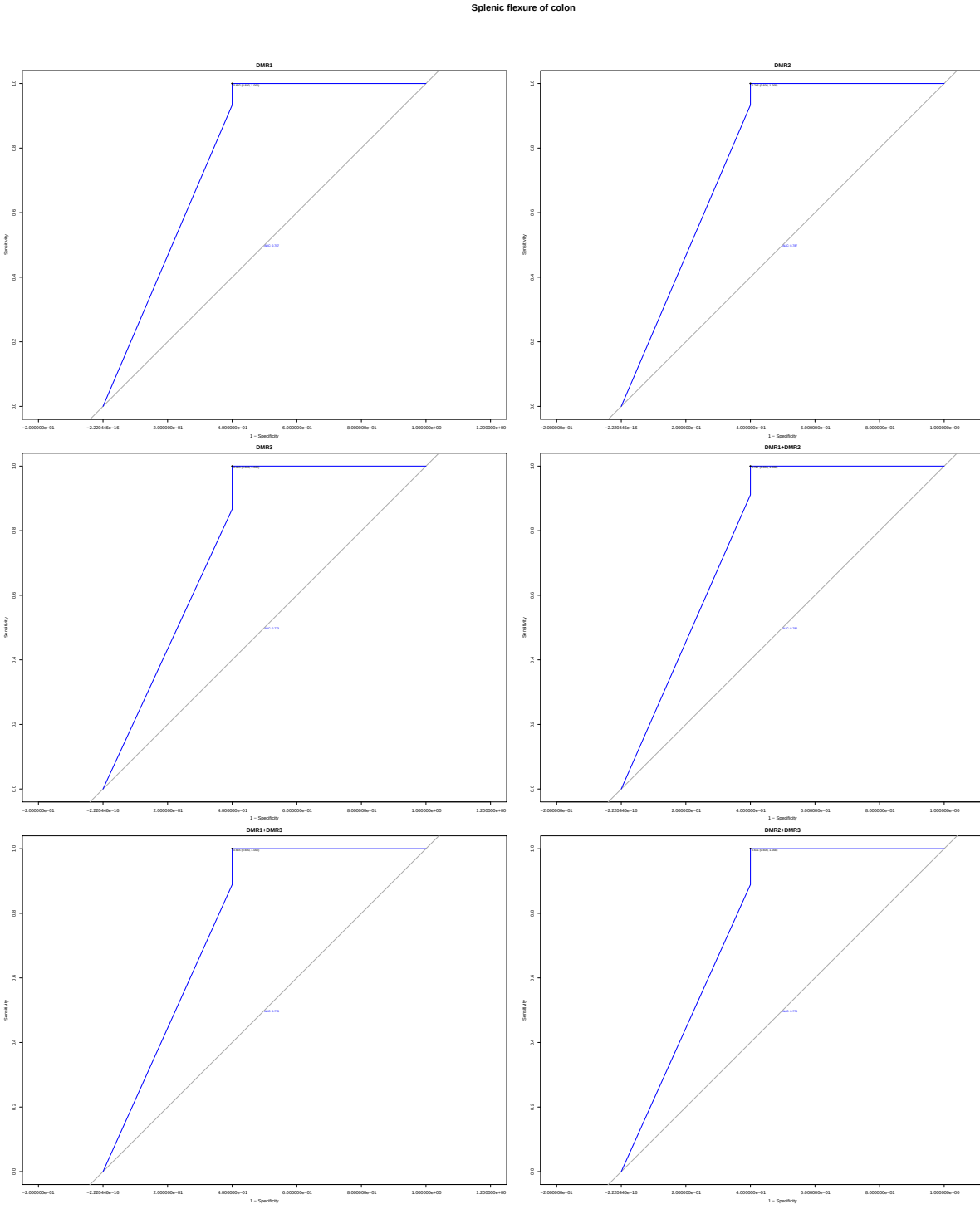
Descending colon



Hepatic flexure of colon

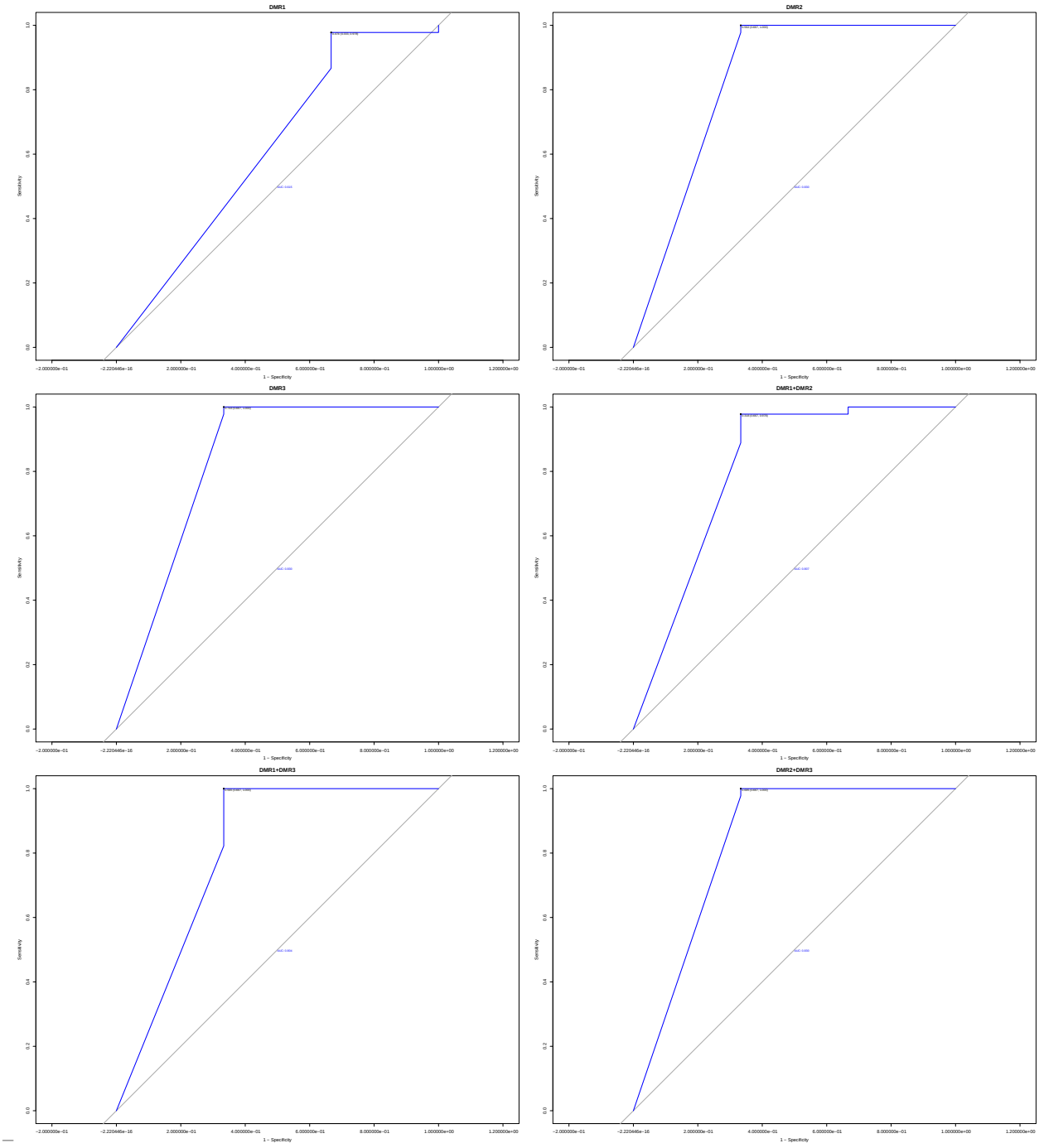


Transverse colon

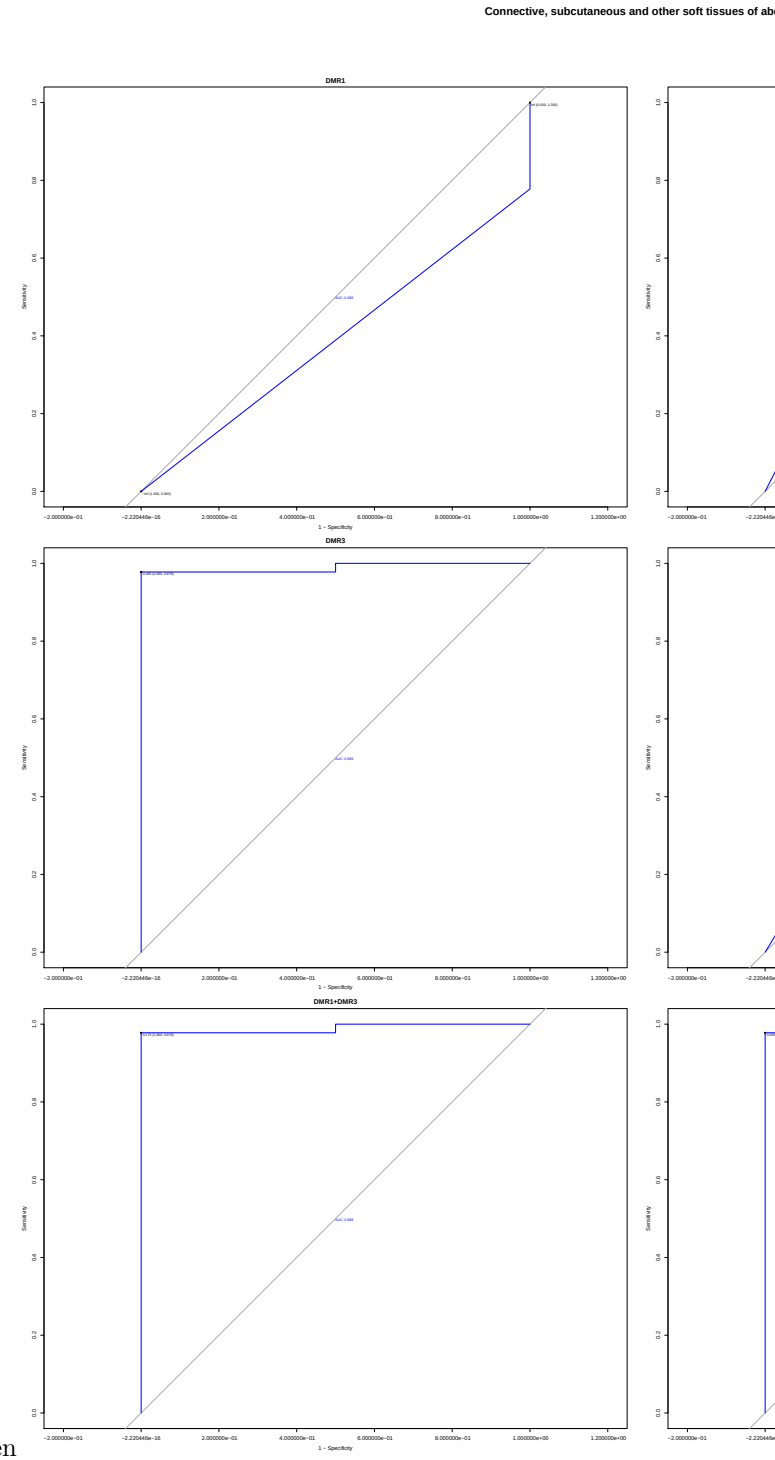


Splenic flexure of colon

--


DMR1
DMR2
DMR3
DMR1+DMR2
DMR1+DMR3
DMR2+DMR3
Sensitivity
1 − Specificity

—



Connective, subcutaneous and other soft tissues of abdomen