我的建议

1, 很有必要深入了解450芯片的数据集的细节, 1), 分析的目标区的性质, 2), 数据的内涵和表示方式, 3), TCGA数据的450芯片数据的请款,和可用数据的总结.

(希望同时将有450 芯片数据的肿瘤类型的突变谱, CNV谱的数据下载, 进行同类的分析)

2,, 一个类型的肿瘤的数据进行全面的分析,建立流程, 流程确立后,再将分析扩展到其他类型的肿瘤

基本信息

1, 启动子高甲基化是导致抑癌基因的肿瘤特征性转录失活的重要机制, 与突变和缺失介导的遗传性改变共同构成了抑癌基因失活的机制.

2, 虽然约60%的启动子为CpG岛(CpG rich)区, DNA甲基化高甲基化状态也可以发生在付CpG岛的启动子区, 450芯片上应该包括了绝大多数,如果不是全部的蛋白编码基因的启动子区.

|  |  |  |
| --- | --- | --- |
| 肿瘤特征性DNA甲基化异常 | | |
| 局部高甲基化(启动子, 增强子) | 抑癌基因(单拷贝序列, 启动子区, CGI 和非CGI 区) | 1, 无突变的; 2, 有突变(在不同的等位基因上, |
| 全局性去甲基化 | 癌基因 | 极为少见 |
|  | 重复序列 | 主要的 |

分析目的, 确定驱动性DNA甲基化异常和突变间对肿瘤发生,发展和临床行为的贡献的量化分析.

1, 抑癌基因为目标的 DNA甲基化/突变频率的标胶, ( 癌基因的异常主要取决于突变,CNV, 启动子区甲基化的降低所介导的基因转录活化, 较为少见, 如,testis 基因在其他组织肿瘤中活化)

2, 蛋白编码基因全局性甲基化状态的改变(编码区等)

讨论备忘录：

1. 补充关于数据处理的细节文档，明确数据的有效性。
2. 先集中分析某一类癌症（例如乳腺癌），在基因组水平对每个基因启动子区甲基化水平分析，计算每个样本，比较癌症 I 期 与正常期的甲基化水平，对每个样本进行超甲基化或低甲基化检验（计算 p 值）. 根据计算结果筛选癌症早期发生超甲基化或低甲基化的基因。
3. 统计编码区的 DNA 甲基化水平的变化，并与基因突变数据进行比较。
4. 目前已经下载了十余种癌症的数据，对一类癌症的分析完成以后，再对其他癌症类型进行分析。比较不同癌症类型的结果。
5. 初步下周四上午再讨论一次。