**基于DNA甲基化及突变的食管癌侵袭转移的分子病理特征**

**研究方案**

**1. 样本量**

160对食管癌（I，II，III，IV各40例）及癌旁对照+5种食管癌细胞系

**2. 主要研究目的**

2.1 探索食管癌远端骨转移的分子病理特征，鉴定可用于预测的DNA甲基化生物标记物。

2.2 探索食管癌远端脑转移的分子病理特征，鉴定可用于预测的DNA甲基化生物标记物。

2.3 通过上述样本的5mC图谱鉴定食管癌分期 (I, II, III, IV) 相关的DNA甲基化生物标记物。

2.4 收集12个食管癌患者（癌组织和癌旁组织，共24例样本），进行多组学分析。

2.4.1 对3个食管患者（各1对组织：癌组织和癌旁组织，计6个组织样本）：进行多组学分析。包括：全基因组甲基化测序（WGBS）、DNA甲基化850K芯片、MBD-seq、外显子测序（WES）、RNA-seq，m6A-RNA甲基化、miRNA-seq。样本分别提取DNA和RNA，DNA做WGBS、DNA甲基化850K芯片、MBD-seq、WES；RNA同步做m6A-RNA甲基化、RNA-seq、miRNA-seq。同时注意保留一定的样本用于未来新型组学的检测。

2.4.2 为了丰富信息学分析，含序号2.4.1的3个样本在内的12个食管患者（癌组织和癌旁组织，共24例样本）抽取DNA后，做DNA甲基化850K芯片，评估WGBS和850K芯片的correlation。

【说明1】肿瘤样本的癌细胞占比估算，可利用WES估计，参考文献[1]或AbsCN-seq[2]或[3]。

【说明2】WGBS：建立食管癌和正常组织的甲基化图谱。测序数据量（Raw data）为平均测序数据量90 G，每个样品的测序数据量不少于承诺平均数据量的98%，且clean data最低不低于5×的测序深度。【数据传输到SRA】

【说明3】未来我们想做新的组学，可能还需要序号2.4.1的3个食管患者（计6个组织样本），妥善保管好6个样本的组织、DNA、RNA和蛋白。

2.5 收集160对食管癌癌组织和配对癌旁组织。利用食管癌公共甲基化大数据，筛选出100个左右诊断、预后、复发、转移相关的潜在位点。根据2.4步骤中国人图谱对100个位点进行filtering，通过该课题收集的160对样本进行验证后（甲基化检测深度depth>1000×），最终确定约50个甲基化位点，然后进行各临床表型和甲基化的关联分析。

【说明1】样本数的选择：考虑到检测失败率约15%，最终获得成功的样本数约为136对。

【说明2】结果出来后决定抗体的购买情况。

2.6 选择5~10种食管癌细胞系，利用5-AZA处理细胞系，完成去甲基化的目的。然后收集处理前后的DNA、RNA和蛋白，对处理前后的RNA进行RNA-seq检测，明确DNA甲基化是否调控上述50个基因的表达，如有需要利用western检测蛋白质变化。

**食管癌细胞株（5种：xx1 + xx2 + xx3 + xx4 + xx5）+ 1个正常细胞株（HEEC）。**

【说明1】5-AZA处理的主要目的是证明哪些候选基因受甲基化转录调控。对于基因的转录调控，DNA甲基化和组蛋白乙酰化是两种重要手段，是否受乙酰化调控可以用TSA处理进行研究。2008年cancer research有篇文章对乳腺癌中哪些基因受甲基化和乙酰化的调控进行了研究，实验细节可部分参考[4]。

【说明2】乳腺癌细胞系和食管癌细胞系可能会有不同。两类细胞的增殖速度是否相当？5-AZA实验最好加入时间梯度和浓度梯度，通过Real-time PCR检测几个基因的表达情况，可以判断时间梯度和浓度梯度的effect。最后确定一个最佳的时间和浓度。

2.7 分析5mC甲基化谱式与食管癌**生存率**的关系，分析甲基化与各临床指标的关系如突变谱式、肿瘤亚型、转移的方向，利用多变量生存分析，评估各个feature的相对权重及生存预测贡献度。

2.8 收集20例食管癌患者及20例正常对照的血浆游离DNA (cfDNA)，利用RRBS建立食管癌cfDNA的特异甲基化图谱。利用多种模型包括甲基化熵、甲基化多态性、甲基化单倍型载荷、甲基化均值水平、甲基化方差等方式对上述甲基化基因与各种临床表型进行关联分析，从而挖掘出与不同临床表型相关的甲基化feature。【说明】步骤2.8要求比较高，是cell-free DNA甲基化，先收集样本，暂时不建议开展。

**3. 靶向DNA甲基化测序技术流程【全程50-60天时间】**

3.1 建立多种预测模型，feature selection，确定50个候选区域，确定候选区域的bed文件。

3.2 提交bed文件，进行引物评估，2周左右完成评估，送出引物合成。

3.3 引物单位点优化。

3.4 Panel优化（完成步骤2.1~2.4预计40天时间）。

3.5 样本建库。

3.6 测序。

3.7 测序结果拆分（完成步骤2.5~2.7预计10天时间）。

**4. 注意事项**

4.1 需要提前将临床表型、临床资料整理完毕，理论上在送出测序之前临床资料就需要整理好，这样可以避免送一些临床资料非常不全的样本。

4.2 测序结果的文件名（S0001.fastq.gz）和临床样本的文件名或者ID（S0001）如何对应需要在送出样本测序的同时尽快分享。测序文件的名字和临床样本ID不一致，没有对应文件，无法完成对应及后续分析。

4.3 对于甲基化研究1）样本的亚型、 2）是否复发、3）预后OS，等是非常有意思的研究内容，这几项信息最好要齐全，此外包括年龄、性别、BMI吸烟、种族等都需要详细记录。

**5. 参考文献**

[1] Luo Z, Fan X, Su Y, Huang YS. Accurity: accurate tumor purity and ploidy inference from tumor-normal WGS data by jointly modelling somatic copy number alterations and heterozygous germline single-nucleotide-variants. Bioinformatics. 2018;34(12):2004-11.

[2] Bao L, Pu M, Messer K. AbsCN-seq: a statistical method to estimate tumor purity, ploidy and absolute copy numbers from next-generation sequencing data. Bioinformatics. 2014;30(8):1056-63.

[3] Li Y, Xie X. Deconvolving tumor purity and ploidy by integrating copy number alterations and loss of heterozygosity. Bioinformatics. 2014;30(15):2121-9.

[4] Heller G, Schmidt WM, Ziegler B, Holzer S, Mullauer L, Bilban M, Zielinski CC, Drach J, Zochbauer-Muller S. Genome-wide transcriptional response to 5-aza-2'-deoxycytidine and trichostatin a in multiple myeloma cells. Cancer Res. 2008;68(1):44-54.