1) 建立起完善的描述测序数据分析流程以及获取大量CpG分布及甲基化状态的实验数据。

A. 从低等生物到高等生物收集不同物种以及不同组织的BS-Seq测序数据。收集癌细胞BS-Seq测序数据，同时收集对应细胞的一些临床癌症及其癌旁组织样品。对这些样品我们将提取DNA构建BS-Seq测序文库并进行测序，获取自己的测序数据。

B. 建立完整的数据分析pipeline，对收集的大量测序数据进行分析。包括数据质量评估与过滤、序列比对、CpG分布数据提取、甲基化状态数据提取。此外我们还将获取DNA不同特征区域的CpG分布及其对应的甲基化状态数据。

2)

通过对这些分布数据的分析，我们将系统地建立描述CpG位点以及甲基化对称性以及对称性破缺程度的数学模型。