资助项目进展报告

项目基本信息

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 项目名称： | DNMT1通过与RNA聚合酶polII相互作用而导致抑癌基因启动子区域异常甲基化的机制研究 | | | | |
| 项目批准号： | 81472637 | 申请代码1： | H1603 | | |
| 批准金额： | 72.0000万元 | 负责人： | 李志广 | 依托单位： | 大连医科大学 |
| 资助类别： | 面上项目 | 附注说明： | 常规面上项目 |  |  |

重要研究进展

填写说明：主要阐述本年度研究工作中取得的重要进展，发表论文和取得专利等成果，请利用“成果在线”导入报告正文后面的基金委项目研究成果目录。

(**字数限制4000字**)

在过去一年中，我们的课题都在紧张有序的开展中。课题的开展主要分为两部分，实验部分与数据处理部分。实验部分包括细胞培养，各种测序文库的构建及质量评估。数据处理部分包括写作完成用于比对whole genome bisulfite sequencing (WGBS)测序reads到参考基因组的mapping程序，建立能够进行differential methylation regions (DMR) calling的统计模型并完成相关程序写作，熟悉对DMR调控基因的生物学功能进行分析的流程，从NCBI Gene Expression Omnibus (GEO)数据库下载甲基化组及基因表达谱的数据并进行初步分析等。

1. 实验部分
2. RNA-seq测序文库的构建
3. BS-seq测序文库的构建
4. 细胞培养
5. 数据处理部分
6. 比对whole genome bisulfite sequencing (WGBS)测序reads到参考基因组的mapping

我们完善并设计了WGBS数据mapping的流程：



我们建立了使用Segemehl进行WGBS数据比对Mapping的分析流程。我们获得的测序数据会先进行FastQC数据质量分析，根据分析结果我们来评估计测序数据的质量，并根据这些质量报告对我们的测序建库实验进行改进，完善我们的测序建库实验流程。随后根据FastQC的结果我们会选择性地对测序数据进行Adapter去除，然后我们会进入测序比对mappding处理环节。这里我们使用Segemehl进行程序比对：首先我们会根据参考序列分别建立C->T以及G->A转变的索引文件；随后Segemehl程序会使用这些索引文件以及参考序列文件本身进行WGBS的数据Mapping。在此过程中程序会自动对正负序列上的甲基化相关的碱基变换做相应的处理。当比对程序结束后，我们将会对比对结果进行处理，删除重复的比对、不合理的比对（大约去处10%左右的数据量），然后进入后续的数据分析（MTBR分析、IMPRINTING分析等等）。

1. 建立能够进行Differential Methylation Regions (DMR) calling的统计模型并完成相关程序写作

DMR是分析甲基化组数据的一种重要方法。DMR的精确定位是进行数据分析的非常重要的前提。目前比较流行的方法主要有基于机器学习的Hidden Markov Model (HMM)方法以及基于数据特征建立各种统计学模型的方法。在这个部分，我们使用后一种方法建立了一种DMR精准检测的方法。该方法基于以下两个对WGBS数据的统计分布的合理假设：a）每个CpG位点的甲基化水平满足 β二项分布b）相邻CpG位点的甲基化水平之间有着很强的空间关联性。我们通过对每个CpG位点的模型拟合得出相应的分布参数，然后在使用Getis-Ord局域统计方法进行DMR的检测。我们新发展的这种方法有着非常好的检测阳性率和灵敏性。此外该方法也能非常精确地确定DMR的边界。目前，基于该方法的程序编写工作已经完成，大量的对真实数据以及模拟数据的检测验证正在进行中。

1. 从NCBI Gene Expression Omnibus (GEO)数据库下载甲基化组及基因表达谱数据

GEO数据库中的甲基化组数据，对于正常组织细胞以及肿瘤细胞进行甲基化组的基

目前我们对甲基化组数据分析的pipeline正在稳步建立和完善中。与此同时我们也开始整理各种甲基化组的数据以用于后续的分析工作。甲基化相关的表观遗传学特性与癌症的表征有着重要的关系。Bing Ren et. al.发表了关于乳腺癌的甲基化组相关的数据，基于这些数据他们提出了Partial Methylation Domain(PMD)的概念。我们正在对他们数据进行下载。初步的数据分析工作也在进行之中。我们后续的分析工作将着眼于在正常上皮组织细胞(HMEC)与乳腺癌细胞(HCC1954)之间的DMR相关的数据分析与挖掘方面。

HMEC_methylevel.pdfHMEC_coverage.pdf

存在问题及解决办法

填写说明：简要说明项目实施中遇到的问题以及采取或拟采取的措施，研究计划调整情况及原因。(字数限制2000字)

其他需要说明的情况

填写说明：前面几部分没有涵盖而需要说明的问题。若有经费管理办法规定范围内的预算科目间调整，也需在此予以说明。(字数限制2000字)