机器感知与智能教育部重点实验室学术报告

Ab initio identification of transcription start sites in the Rhesus macaque genome by histone modification and RNA-Seq

Yi Liu, Ph.D. Assistant Research Staff,
CAS Key Laboratory of Molecular Developmental Biology
Institute of Genetics and Developmental Biology, Chinese Academy of Sciences

Abstract

Rhesus macaque is a widely used primate model organism. Its genome annotations are however still largely comparative computational predictions derived mainly from human genes, which precludes studies on the macaque-specific genes, gene isoforms or their regulations. Here we took advantage of histone H3 lysine 4 trimethylation (H3K4me3)'s ability to mark transcription start sites (TSSs) and the recently developed ChIP-Seq and RNA-Seq technology to survey the transcript structures. We generated 14,013,757 sequence tags by H3K4me3 ChIP-Seq and obtained 17,322,358 paired end reads for mRNA, and 10,698,419 short reads for sRNA from the macaque brain. By integrating these data with genomic sequence features and extending and improving a state-of-the-art TSS prediction algorithm, we ab initio predicted and verified 17,933 of previously electronically annotated TSSs at 500-bp resolution. We also predicted approximately 10,000 novel TSSs. These provide an important rich resource for close examination of the species-specific transcript structures and transcription regulations in the Rhesus macaque genome. Our approach exemplifies a relatively inexpensive way to generate a reasonably reliable TSS map for a large genome. It may serve as a guiding example for similar genome annotation efforts targeted at other model organisms.

简介
短尾猴是一种重要的灵长类模式动物，然而它的基因组注释仍主要是通过物种对比的计算方法从人类基因组的注释搬移过来的，这限制了研究短尾猴的特异基因，基因多态性和特殊的调控方式。我们利用组蛋白H3 4'赖氨酸的三甲基化修饰的特征能够标识转录起始位点(TSSs)的特性，并依靠近年来发展的ChIP-Seq和RNA-Seq技术来研究短尾猴的转录结构。我们在猴脑组织中，通过H3K4me3 ChIP-Seq实验得到了14,013,757个cDNA序列片段，还通过mRNA-Seq得到17,322,358个序列片段对，sRNA-Seq得到10,698,419个序列片段。通过整合这些数据、利用基因组序列特征、改进和扩展了一个当前流行的TSS预测算法，我们在500bp的分辨率下，重新预测并验证了17,933个先前标注的TSS。我们还预测和验证了约10,000个新的TSS。这些信息提供了一个重要的新的基因组资源，使研究者能够仔细观察短尾猴基因组物种特异的转录结构和转录调控特征。我们的工作展示了一个费用较低的方法，能够对较大的基因组生成一个相对可靠的TSS位点预测。该方法同样可以用于其它物种的类似的基因组标注工作中。

Time: 15:00, Nov. 16, 2010. Location: 2736, Science Building 2, Peking University