

毕业设计(论文)任务书

- 一、学生姓名： 黄凯辉 学号： 41467024
- 二、题目：癌症样本纯度与染色体倍性的算法的设计与开发
- 三、题目来源：自拟
- 四、结业方式：设计
- 五、主要内容（背景）：

肿瘤纯度和倍性对肿瘤样本的下一代序列分析有重要作用，并可能改变其结果的生物学和临床解释。目前存在几种专门用于评估来自癌症基因组图谱肿瘤 - 正常细胞全基因组测序（WGS）数据的肿瘤纯度或倍性的计算方法，但仍然缺乏一种能够在广泛测序范围内，肿瘤纯度水平和肿瘤内异质性水平下，准确，快速且完全自动化的方法。

本实验旨在设计一种计算方法，用来准确推断肿瘤的纯度和染色体倍性。

六、主要（技术）要求：

- 1.以 rust 语言为算法实现的主要编写语言
- 2.使用 HashMap 为主要的多个库文件
- 3.使用一些特殊文件格式诸如 bam,vcf 等对数据进行存储和操作

七、日程安排：

- 第 1 周至第 3 周：查阅文献，初步设计算法
- 第 4 周：撰写开题报告
- 第 5 周至第 9 周：完成肿瘤的纯度和染色体倍性计算方法的设计并编写代码
- 第 10 周：中期检查，完善程序
- 第 11 周至第 13 周：设计实现数据可视化系统，开始撰写论文
- 第 14 周：查重，完成整个系统的设计与实现，完成论文撰写

第 15 至 16 周：评阅及答辩

八、主要参考文献和书目：

- [1]Alkodsi,A. et al. (2015) Comparative analysis of methods for identifying somatic copy number alterations from deep sequencing data. Brief.Bioinf., 16,242-254.
- [2]Andor,N. et al. (2014) EXPANDS: expanding ploidy and allele frequency on nested subpopulations.Bioinformatics, 30, 50-60.
- [3]Aran,D. et al. (2015) Systematic pan-cancer analysis of tumour purity. Nat. Commun., 6, 8971.
- [4]Benjamini,Y. and Speed,T.P. (2012) Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res., 40, e72.
- [5]Beroukhim,R. et al. (2010) The landscape of somatic copy-number alteration across human cancers. Nature, 463, 899-905.
- [6]Li,Y. and Xie,X. (2014) Deconvolving tumor purity and ploidy by integrating copy number alterations and loss of heterozygosity. Bioinformatics, 30, 2121-2129.

指导教师签字： 年 月 日

学 生 签 字： 年 月 日

系（所）负责人章： 年 月 日