

宁波大学 2019 年博士研究生招生考试初试试题(A 卷)

(答案必须写在考点提供的答题纸上)

科目代码: 3824 总分值: 100 科目名称: 生物信息学

一、名词解释 (每题 6 分, 共 42 分)

1. E 值
2. 模体(Motif)
3. 基因本体(Gene Ontology)
4. 隐马尔可夫模型(HMM)
5. 蛋白质组(Proteome)
6. KEGG
7. 基因组

二、选择题 (每题 2 分, 共 20 分)

1. GenBank 是()。
A. 在线人类孟德尔遗传数据 B. 国际核酸数据库 C. 人类基因组计划 D. 水稻基因组计划
2. DDBJ 的含义是()。
A. 美国国家生物信息中心 B. 欧洲分子生物学实验室 C. 日本 DNA 数据库 D. 中国基因组研究中心
3. NCBI 中人类无冗余基因数据库是()。
A. UniGene B. UniPro C. UniRef D. URF
4. 差异表达主要采用的统计方法是()。
A. 卡方检验 B. 超几何分布检验 C. T 检验 D. 正态性分布检验
5. ChIP-seq 不能用于以下哪种情况的检测()。
A. 转录因子 DNA 结合位点 B. 基因突变 C. H3K4me3 组蛋白修饰 D. H3K27me3 组蛋白修饰
6. 下列哪种测序方式可以用于 DNA 甲基化检测?()
A. RNA-seq B. BS-seq C. CLIP-seq D. ChIP-seq
7. contig 的含义是()。
A. 基序 B. 跨叠克隆群 C. 碱基对 D. 结构域
8. BLAST 程序中, 哪个方法是不存在的?()
A. BLASTP B. BLASTN C. BLASTX D. BLASTQ

宁波大学 2019 年博士研究生招生考试初试试题(A 卷)

(答案必须写在考点提供的答题纸上)

科目代码: 3824 总分值: 100 科目名称: 生物信息学

9. TCGA 的含义是()。

A. 基因组研究中心 B. 核苷酸序列数据库 C. 疾病基因组数据库 D. 肿瘤基因组图谱

10. Read 的含义是()。

A. 测序的片段 B. 测序片段的读长 C. 碱基 D. 模体

三、简答题(每题 8 分, 共 24 分)

1. 简述生物信息学的发展阶段。

2. 序列的相似性与同源性有什么区别与联系?

3. 人类基因组计划的结果显示, 蛋白编码基因占整个基因组的不到 5%, 剩下的都是非编码区域, 请您说出常见的非编码 RNA 类型, 并列举几个常用的非编码 RNA 数据库及其主要内容。

四、论述题 (14 分)

CRISPR(clustered regularly interspaced short palindromic repeats, 规律成簇间隔短回文重复序列), 最早在 K12 大肠杆菌中的碱性磷酸酶基因附近发现。细菌的 CRISPR 序列来源于先前入侵的病毒 DNA 片段, 在后续类似病毒入侵时, 可用于识别和破坏外来 DNA, 是细菌用来识别和摧毁抗噬菌体和其他病原体入侵的防御系统。CRISPR 对外来 DNA 的编辑需要依赖一种叫做 CRISPR 相关蛋白的 DNA 内切酶, 常用的 Cas 蛋白为 Cas9, 因此, 这种技术也被称为 CRISPR/Cas9 基因编辑系统。CRISPR/Cas9 系统通过设计基因的单个向导 RNA (guide RNA, gRNA), 可以精确地切割和操纵细胞中该基因的 DNA, 是一种有效的基因组编辑工具, 对细胞系或生物过程中基因功能的鉴定起着重要的作用, 在疾病和癌症的治疗上具有巨大的潜力。

2018 年, 来自上海科技大学、加州大学旧金山分校、中山大学等地的研究人员在 PNAS 杂志上发表了题为 “A genome-wide CRISPR screen identifies FAM49B as a key regulator of actin dynamics and T cell activation” 的文章。作者利用全基因组 CRISPR 筛选技术, 系统性地研究了 T 细胞激活的分子机制, 绘制了人类 T 细胞功能的调控图谱。

已知 CD69 表达升高是 T 细胞活化的标志。在这个筛选中, 包含 250,000 个基于整个基因组蛋白编码基因设计的 gRNA 的库被慢病毒封装, 转染进 T 细胞。然后, T 细胞经抗 T 细胞受体 V β 8 的抗体刺激 13 个小时后, 基于 T 细胞表面 CD69 的表达水平被分为 CD69 高表达 (CD69^{high}) T 细胞和 CD69 低表达 (CD69^{low}) T 细胞, 提取这两类细胞群的 DNA 进行 gRNA 的 PCR 扩增和测序, 获取 gRNA

宁波大学 2019 年博士研究生招生考试初试试题(A 卷)

(答案必须写在考点提供的答题纸上)

科目代码: 3824 总分值: 100 科目名称: 生物信息学

的数量。通过比较两类细胞各个 gRNA 的数量,从而判断对应靶基因是 T 细胞受体信号通路的激活因子还是抑制因子。

根据筛选结果,研究人员除了发现大量已知的调控基因外,还包含了一批尚未见报道的参与 T 细胞信号转导的基因,比如,研究人员最终证实,功能未知基因 FAM49B 在 T 细胞中的敲除会使 T 细胞过度激活,细胞骨架蛋白 actin 聚集增强,是 T 细胞受体信号通路的抑制因子。

阅读以上材料,请回答下列几个问题。

(1) 什么是全基因组 CRISPR 筛选技术? (4 分)

(2) 如何根据 gRNA 在 CD69^{low} T 细胞和 CD69^{high} T 细胞中数量的变化,判断对应靶基因是 T 细胞受体信号通路的激活因子还是抑制因子? (5 分)

(3) CRISPR 筛选技术的应用主要有哪些? (5 分)