北京中科云畅应用技术研究院
中国科学院北京基因组研究所基因组科学与信息重点实验室

CCAT-01-1709

关于举办“生物大数据分析及 TCGA 数据库挖掘”高级培训班的通知

各有关单位:

随着基因组学的发展，利用高通量测序等技术，全球生物研究机构建立了众多基因信息数据库，如描述人类基因组中所编码的全部功能性序列元件的 ENCODE 项目（DNA 元件百科全书）、全球最大的癌症基因信息数据库 TCGA 等。这些数据库包含了海量的基因信息数据，为了对这些海量数据进行有效的分析和深入的研究，人们结合了计算机科学技术开发了多种生物数据挖掘工具，获得了很多新的成果。生物数据挖掘技术已经成为当前生物信息领域的重要突破口。为进一步推动我国生物信息学特别是生物数据挖掘技术的发展，提高从业人员的技术水平，北京中科云畅应用技术研究院举办“生物信息学最新技术”暨“生物大数据分析及 TCGA 数据库挖掘”高级培训班，并由北京中科润开生物科技有限公司具体承办，具体事宜通知如下：

一、培训目标及特点：

本培训邀请的主讲人均是有理论和实际研究经验的人员。课程力求"有点有面，实用高效"，理论授课与上机实践有机结合，使学员既能够学会思考和设计科学实验，也能够独立使用生物信息学基本方法和常用工具，力求学有所用，学有所值。

1、讲师拥有丰富的数据分析和软件研发经验。

2、主题明确，针对性强。理论与实践结合，针对专题内容进行深入讲解。

3、精心挑选上机软件，实用性强，方便学员课下独立操作。

学员通过与专家直接交流，能够分享到这些顶尖学术机构的研究经验和实验设计的思路。学员通过集中专题学习后能够扩展思路，在研究技术方面领悟更多。

二、培训对象：

大中专院校生物信息、生物计算、生命科学、医学、化学、农学、计算机科学、数学类专业的课程负责人、一线教师、教研室骨干人员、教学管理人员；科研单位从事生物、生命科学、微生物研究的相关人员；生物、医药、化学及相关企业的领导与技术骨干。

三、时间地点：

2017 年 9 月 22 日——9 月 26 日　　北京

（时间安排：第 1 天报到，授课 4 天）

四、培训内容

基因组学的进展及应用	DNA 测序技术-转录组分析的进化 第一代测序技术：Sanger 测序原理 第二代测序技术：Illumina，454, Ion Torrent 原理 第三代测序技术：PacBio, Hellicos 原理 第四代测序技术: Oxford NanoPore 原理 其他技术 Hybridization based methods (NabSys) 生物信息学基本技能： 基于 virtual box 的 linux 虚拟机的安装及使用（实验） 云服务器登陆及文件传输（实验） 常用 LINUX 命令（实验） 简单 shell 脚本的编写（实验）
高通量测序应用	基因组测序和组装 第二代测序技术应用 ChIRP-Seq, GRO-Seq, Ribo-Seq)/ARTseq, RIP-Seq, HITS-CLIP, CLIP-Seq, PAR-CLIP, iCLIP, NET-Seq, TRAP-Seq, CLASH-Seq, PARE-Seq, GMUCT, TIF-Seq, PEAT 等方法的原理解析 第三代测序的应用 植物、动物全基因组测序研究实例
转录组分析	Experimental procedure for transcriptomic analysis Introduction Number of duplications, Sequencing coverage Transcriptomic analysis using NGS (RNA-Seq) Transcriptomic analysis using PacBio (Iso-Seq) IsoSeq Experimental design Data analysis （part 1）：data pre-processing evaluation of data quality 数据分析 Data format，fasta，fastq，quality value，gff3 Data cleanup Quality filter, trimmer, clipper Data analysis （part 2）：reference free analyses，无参转录组分析 Gene discovery Trinity de novo transcriptome assembly Analysis of Differential Expressed Gene （DEGs） Abundance estimation using RSEM Differential expression analysis using EdgeR Explore the results (cummerbund) MA plot, Volcano plot, False Discovery Rate (FDR) hierarchical two-way clustering, pairwise sample-distance,gene expression profiles.

	Data analysis （part 3）：reference based analyses，有参转录组分析 Gene discovery Mapping reads to the reference (tophat) Assemble mapped reads (cufflinks) Merge sample-specific assemblies (cuffmerge) Analysis of Differentially Expressed Gene （DEGs） Identify DEGs (cuffdiff) Explore the results (cummerbund) Data analysis （part 4）：from gene list to gene function，基因功能注释 File format for annotation information: GFF3 Annotation Enrichment analysis using DAVID
表观基因组数据	数据的产生、处理与分析 ChIP-Seq 与 DNA 甲基化测序 主要分析方法和原理 -数据质量检查和质量控制 -数据前处理 Peak calling 算法介绍 DNA Accessibility 解析 3D Chromatin Structure 解析 UCSC Genome Browser 工具介绍与应用
ENCODE 项目	ENCODE 项目背景、目标、实验种类介绍 数据和数据仓库的概况 工具介绍: ChromHMM, Segway, chromImpute 等. 数据资源: regulomeDB, HaploReg 等. 其他应用与项目如 Epigenome Roadmap 等
TCGA 项目及其他	TCGA 项目背景、目标、实验种类介绍 数据和数据仓库的概况 临床数据介绍 SNV, CNV 等识别 多维基因组数据分析 cBioPortal 工具的介绍与应用 CCLE、CGP 项目的介绍与应用

五、主讲专家：

主讲专家来自中科院等科研机构的高级专家，拥有丰富的科研及工程技术经验，长期从事生物信息领域项目研究，具有资深的技术底蕴和专业背景。

六、颁发证书：学员经培训考试合格后可以获得：

1．由北京中科云畅应用技术研究院颁发的生物信息培训班结业证书。

2．注：请学员提供电子版白底彩照、身份证复印件一张。

七、报名办法及费用：

每人￥4300 元（含报名费、培训费、资料费、考试费相关费用）食宿统一安排，费用自理。请各有关部门统一组织本地区行政、企事业单位报名参加培训，各单位也可直接报名参加，报名回执表请邮件发送至会务处。

八、联系方式：

联 系 人：张爱国

移动电话：136 8311 1214

联系电话/传真： 010-58032788　　报名邮箱：zky_im@vip.126.com

北京中科云畅应用技术研究院

二零一七年八月十八日

北京中科润开生物科技有限公司

二零一七年八月十八日