

# TCGA、GEO 及 SEER 公共数据库挖掘与应用学习班

玮瑜主办 2018 年 5 月 5-6 日（周末两天） 上海

## 课程背景

随着大数据时代的到来，各种生物类公共数据库井喷，其中就包括癌症领域熟为人知的癌症基因图谱 The Cancer Genome Atlas (TCGA) 数据库和 Gene Expression Omnibus (GEO) 基因表达数据库和 (SEER) 数据库，临床科研人员有没有一种方法可以不做实验不查病史，直接调用现有数据发表 SCI 论文呢？癌症公共数据库即提供了这样的可能。本次学习班讲授癌症领域熟为人知的癌症基因图谱 The Cancer Genome Atlas (TCGA) 数据库和基因表达数据库及癌症监测、流行病学和结果 (SEER) 数据库。TCGA 由 NCI 牵头，作为美国攻克癌计划的一个大项目，系统提供了癌症多组学测序和芯片数据，包括 Gene expression, DNA methylation, Copy Number Variation, Mutation 等结果，同时也附有相应各测序样本的完整临床资料。TCGA 为肿瘤基础医学和转化医学研究者提供了海量的基因组数据和与其关联的临床数据，这为挖掘有意义的基因组变化和发现影响肿瘤起始、发展、分化、转移等生物学机制提供了海量数据基础。GEO 数据库是当今最大、最全面的公共基因表达数据资源。不仅可以上传自己的数据，而且还可以免费下载数据库中和自己研究方向类似甚至相同的数据来进行分析，为自己的研究提供一些启示或者验证。而美国的 SEER 数据库由美国国立癌症研究所 (National Cancer Institute, NCI) 于 1973 年所建立，其后每年定期更新，是北美最具代表性的大型临床肿瘤登记注册数据库之一，收集了各个癌种的临床病理信息和预后数据，并向全世界开放，为临床医师的循证实践及临床肿瘤学研究提供了宝贵的第一手资料。

然而传统的基础医学和临床医学研究者缺乏信息学基础来处理这些大规模癌症数据，因而在面对这些极其有价值的公共数据时，往往心有余而力不足。作为医学信息领域研究者，我们需要将信息学和统计学知识运用到癌症公共数据分析的研究当中，作为连接大数据与肿瘤研究者之间的一个纽带，帮助研究者去更好地挖掘利用这些数据。

## 课程内容、目标与特色

本次提供了一次系统了解 TCGA、GEO 和 SEER 数据产生，糅合、分析及挖掘的课程，使基础医学和临床医学研究者能更好地挖掘这些公共数据，以便为自身科研项目服务。本次涵盖拟解决的问题包括：1、免费获取并安装 genespring

和 cytoscape 软件；2、将系统了解芯片分析相关知识，手把手教会使用 genespring，轻松完成各种 GEO 数据类型的芯片分析；3、结合文章实例和实战经验，熟悉数据挖掘文章等的基本套路，详细学习系统生物学的相关知识，手把手指导，学会使用 cytoscape，掌握例如构建蛋白互作网络和调控网络等芯片下游分析等实用技能；4、了解 TCGA 数据库知识，无需编程分析就能使用各种类型数据（mRNA, miRNA, 蛋白, 甲基化, 拷贝数等）。5、SEER 数据库的基本情况、获取方式、数据库结构及基本统计方法。6、近年来国内外基于 TCGA 和 SEER 数据挖掘的经典案例，重点对基于 TCGA 和 SEER 数据库的数据挖掘和课题设计进行讨论

## 授课老师

本次邀请两位主讲老师。一位来自浙大医学院生物信息老师，另一位是三甲医院临床医生。两位授课老师将分别从 TCGA 数据下载，数据整合，清洗及 SCI 文章常规思路对这些数据进行分析，并从临床角度出发，把测序数据与临床资料整合，为临床医生提供科研思路，为自身科研服务。

## 课程安排

### 第一天上午：GEO 数据库介绍和 genespring 软件分析

1. 安装 genespring
2. 介绍芯片分析相关知识
3. GEO 数据介绍和使用
4. 文章套路讲解和深度解读
5. Affymetrix 芯片上机操作
6. Illumina 芯片上机操作
7. Agilent 芯片上机操作

### 第一天下午：TCGA 介绍及生物网络调控分析

1. 介绍 TCGA 数据库相关知识
2. TCGA 实用工具的演示
3. 如何做成发表文章中的图表
4. 安装 cytoscape
5. 学习系统生物学的相关知识
6. 构建 miRNA-mRNA 互作网络上机操作
7. 蛋白质-蛋白质互作网络上机操作

## 第二天上午：SEER 数据挖掘

1. SEER 数据库入门（软件安装、数据获取、数据检索）
2. 几篇基于 SEER 数据库发表文章的深度解析
3. 上机操作实战：SEER 数据库注册、数据获得、处理和分析

## 第二天下午：基于 TCGA 数据和 SEER 数据的转化医学课题探索

1. 如何利用 TCGA 进行数据挖掘及临床转化课题设计
2. 几篇基于 TCGA 数据挖掘文章的深度解析（偏向临床）
3. 整合 TCGA 和 SEER 数据进行数据挖掘及临床转化

### 主办单位：

玮瑜科研平台（上海玮瑜生物科技有限公司）

### 承办单位：

上海玮瑜生物科技有限公司

### 时间地点：

2018 年 5 月 5-6 日 4 日报到 田林宾馆（徐汇区田林路 1 号）

### 住宿酒店：

田林宾馆：标准间 380 元/间 合住 190 元/间

### 注册费用：

2800 元/人。授课期间发放纸质邀请函（盖章）和发票。按交费先后顺序确定座位号。会务期间提供午餐，晚餐自理。

### 付款方式：

#### A：银行转账

账户名称：上海玮瑜生物科技有限公司 账户号：97340154740007035

开户行：上海浦东发展银行大华支行

#### B：支付宝转账

收款人：wybiot@163.com

支付宝户名：上海玮瑜生物科技有限公司

#### C：刷卡或现金

支持公务卡

汇款时写上您的姓名，如果朋友代付一定要注明您本人的名字，便于好查询。

**疑问咨询：**

联系人：谢老师 13611825136      报名邮箱：sci1976@126.com

**报名方式（任选一种即可）**



方式一：你可以扫描左侧二维码填写后提交

我们将在 24 小时内和您联系

方式二：请将您的姓名 手机号码 参加班级发送到 sci1976@126.com, 并编辑“参会”二字发信息到 13611825136 即可