**R在生物信息学中的运用示范**

由于本人电脑是Windows系统，2G内存，所有操作均在此平台上完成。

一、R简介及操作

1. R的简介

与起源于贝尔实验室的S语言类似，R也是一种为统计计算和绘图而生的语言和环境，它是一套开源的数据分析解决方案，由一个庞大且活跃的全球性研究型社区维护。但是，市面上也有许多其他流行的统计和制图软件，如Microsoft Excel、SAS、IBM SPSS、Stata以及Minitab。为何偏偏要选择R？

R有着非常多值得推荐的特性。

*  多数商业统计软件价格不菲，投入成千上万美元都是可能的。而R是免费的！如果你是位教师或一名学生，好处显而易见。
*  R是一个全面的统计研究平台，提供了各式各样的数据分析技术。几乎任何类型的数据分析工作皆可在R中完成。
*  R拥有顶尖水准的制图功能。如果希望复杂数据可视化，那么R拥有最全面且最强大的一系列可用功能。
*  R是一个可进行交互式数据分析和探索的强大平台。任意一个分析步骤的结果均可被轻松保存、操作，并作为进一步分析的输入。
*  从多个数据源获取并将数据转化为可用的形式，可能是一个富有挑战性的议题。R可以轻松地从各种类型的数据源导入数据，包括文本文件、数据库管理系统、统计软件，乃至专门的数据仓库。它同样可以将数据输出并写入到这些系统中。
*  R是一个无与伦比的平台，在其上可使用一种简单而直接的方式编写新的统计方法。它易于扩展，并为快速编程实现新方法提供了一套十分自然的语言。
*  R囊括了在其他软件中尚不可用的、先进的统计计算例程。事实上，新方法的更新速度是以周来计算的。如果你是一位SAS用户，想象一下每隔几天就获得一个新SAS过程的情景。
*  如果你不想学习一门新的语言，有各式各样的GUI（Graphical User Interface，图形用户界面）工具通过菜单和对话框提供了与R语言同等的功能。
*  R可运行于多种平台之上，包括Windows、UNIX和Mac OS X。这基本上意味着它可以运行于你所能拥有的任何计算机上。

1. R的下载与安装

R总体分为主程序（含基础包）和拓展包两部分，前者从官方网站https://www.r-project.org

下载并安装，后者则是需要某个包时再去安装。

2.1主程序的下载安装

一般下载页面提供三种常用操作系统上的安装程序（Windows、Linux以及MacOS X）

这里主要介绍Windows平台。登录R官网https://www.r-project.org——点击download R——选择一个本国镜像网站（一般选第一个）——点击Download R for Windows——进入 R for Windows界面——初次安装选择安装基本包base——点击下载。

2.2 拓展包的下载与安装

三种方式：

A 在线安装。打开R——选择菜单栏 Packages ——Install package（s）——出现“CRAN mirror”界面——选择中国境内的CRAN镜像——选择后点击“OK”——出现Packages界面——选择自己需要的包——点击“OK”，之后会下载自动安装。

可用命令来装：

options(CRAN=”http://cran.r-project.org”) # 指定镜像网站

install.packages(“包的名称”)

B 下载到本地安装。打开R——选择Packages——Install packages from local zip……——找到下好的包——点击安装

此种方法不推荐，因为Windows平台只能识别 “.zip”格式的包，但实际上大部分注释包都没有这个格式下载，因此遇到这种情况我们使用第三种方法。

C 使用第三方提供的脚本实现在线安装（生物信息我们常用Bioconductor，之后都以这个为例）使用命令安装：

source("https://bioconductor.org/biocLite.R")

biocLite()

biocLite("org.Mm.eg.db")

其中org.Mm.eg.db就是一个示例的文件名，将它替换成你需要下载的包的名称即可。