



转录组生信软件安装和测试手册

软件安装视频



○ <https://www.bilibili.com/video/BV1Qk4y117ZY> (P4,P5)

易生信软件安装基本指南

1790播放 · 0弹幕 · 2020-08-24 18:40:25



生信宝典和宏基因组 发消息

欢迎大家关注公众号《生信宝典》和《宏基因...

为TA充电

+ 关注 8482

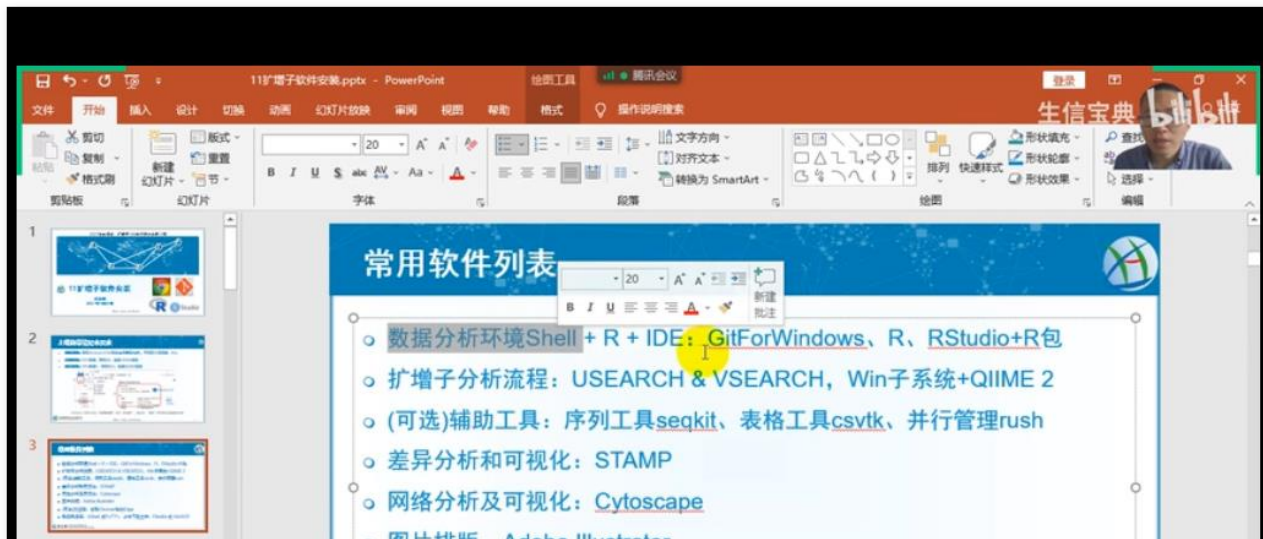
弹幕列表 :

展开

视频选集 (1/9)

自动连播

- P1 扩增子和宏基因组软件安装 (更新版 2... 32:29
- P2 Window下安装Ubuntu安装和Qiime2 - 1 04:44
- P3 Window下安装Ubuntu安装和Qiime2 - 2 21:37
- P4 转录组Windows软件安装20210726 33:44
- P5 转录组Windows本地环境配置 (2021072... 03:45
- P6 易生信软件安装基本介绍 (从这开始的... 02:45
- P7 Windows下安装R/Rstudio/Git (从这开始... 07:36
- P8 网络图绘制cytoscape软件安装 (从这开... 06:24
- P9 扩增子usearch/vsearch软件安装 (从这开...02:10



易汉博基因科技(北京)有限公司
EHBIO Gene Technology (Beijing) co., LTD




易生信, 毕生缘; 培训版权所有。

上课自带笔记本要求



- **操作系统：**推荐Windows10 64位版，不推荐32位系统、Mac、Win7
- **推荐配置：**CPU双核、内存4G+、硬盘30GB空间以上



打开“计算机/我的电脑”()，再点击鼠标右键，选择“属性”
或
打开“资源管理器”()，点击“此电脑”，再点击鼠标右键，选择“属性”
或
快捷键 win 键() + E，点击“此电脑”，再点击鼠标右键，选择“属性”

Windows 10用户点击“资源管理器”中的“此电脑”，再点击--“属性”即可显示电脑基本信息



软件安装清单

- R + IDE: R、Rstudio (按顺序安装)
- Windows + Ubuntu: transcriptome环境 (可选)
- Chrome浏览器: 远程访问Rstudio
- 网络分析及可视化: java 11, Cytoscape 3.9 (按顺序安装)
- GSEA富集分析: gsea-4.0
- 图片排版: Adobe Illustrator
- IGV: 高通量数据可视化工具
- 上传下载文件: Filezilla 或 Xftp (可选)
- 登录远程服务器: Xshell (windows也可选其它工具) (可选)

注: 均选择默认安装, 路径不能有中文。Xshell选家庭和学校版本。
绿色软件双击即可自行安装



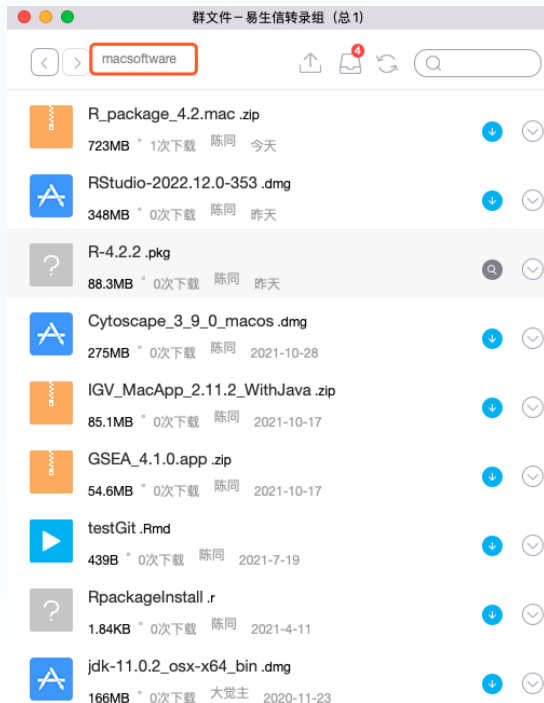
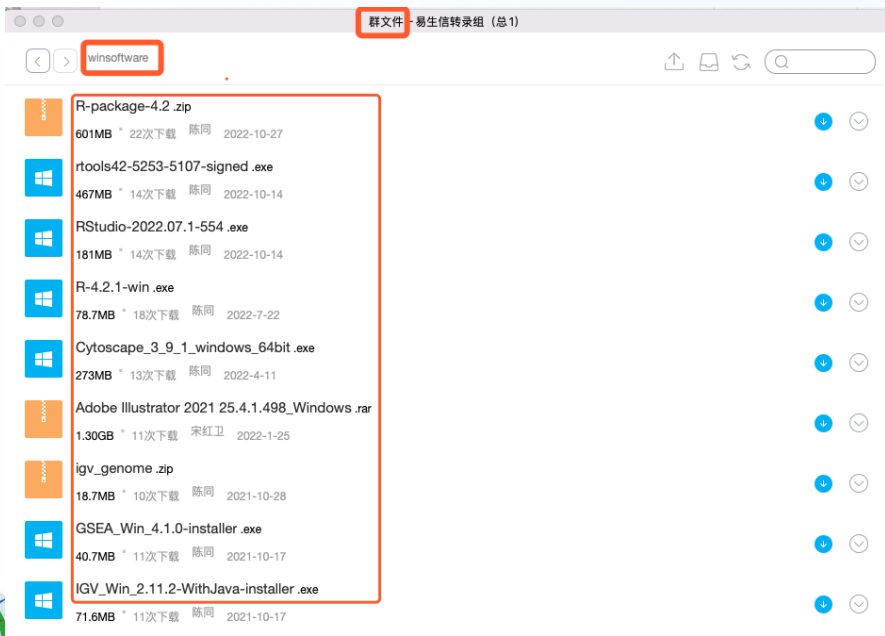
QQ 群下载需要的数据

- IGV_data.zip : IGV导入的 bam 和 wig 文件，下载后解压
- ServerData.zip: 服务器运行需要的流程代码（可以不下载，服务器已部署好）
- 转录组教案：课程 PPT，下载解压
- 转录组本地代码和数据.zip：本地需要运行的 Rmd 代码和数据，下载后解压，放在C盘根目录下
- 如果解压后出现_MacOSX类似的文件夹，忽略或直接删除即可，是 Mac 的特有文件，windows 电脑用不到。

	转录组本地代码和数据-更新版.zip
451MB	2次下载 陈同 今天
	转录组教案-更新.zip
67.5MB	2次下载 陈同 今天
	transcriptomeSoftInstall.sh
14KB	39次下载 陈同 2022-1-13
	serverData.zip
355MB	72次下载 陈同 2022-1-13
	转录组服务器环境配置.wmv
429MB	31次下载 陈同 2021-7-26
	02-转录组Windows本地环境配置.wmv
23MB	30次下载 陈同 2021-7-26
	01-转录组Windows软件安装20210726.wmv
191MB	28次下载 陈同 2021-7-26
	IGV_data.zip
2.05GB	45次下载 大觉主 2020-11-23

QQ 群下载需要安装的软件 (4.2,4.3皆可)

- Windows 电脑 — 下载后一路双击安装；如果是压缩包则解压。
- Mac 电脑



宏基因组



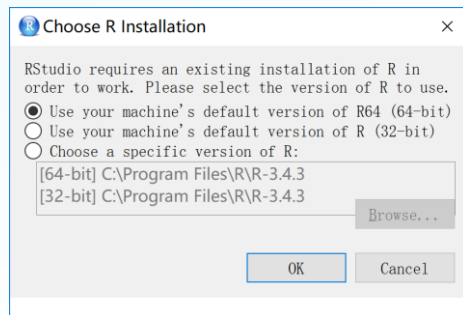
1. R解析程序

- R语言是目前生物学、经济学等领域最流行的统计分析语言
- 在文件夹下或点击链接下载R-4.2.1-win.exe安装程序
- R语言清华镜像站：[链接](#)
- 下载最新版：Download R for Windows —— base ——[Download R 4.2.1 for Windows](#)
- 双击安装程序，直接按默认参数安装即可（注意：选择组件步可去掉32-bit，节约空间并减少选择）
- Mac文件夹中有Mac系统安装包R-4. 2. 1. pkg



2. R/Shell编程环境——Rstudio

- 下载页面：[点击下载](#)
- 选择适合自己系统的版本，下载安装程序，如 [RStudio 2022 - Windows Vista/7/8/10](#)，或在文件夹下或点击链接下载安装程序 RStudio-2022
- 如果用户名为中文，则右键使用管理员身份安装
- 完成后打开时，会选择R版本
系统允许下建议选第一项，OK



3. 安装R包——批量解压4.2.zip

Windows/Mac安装R包都按这个操作，mac需要从QQ群mac目录下下载R包压缩文件

Windows/Mac安装R包都按这个操作，mac需要从QQ群mac目录下下载R包压缩文件

从QQ群下载Rpackage4.1.zip
解压
获得 485个子文件夹

进入到 第3步指示的目录下，把 第4步获得的所有子文件夹拷贝到 第5步进入的目录下。

已完成 1%
正在将 50,413 个项目从 4.1 复制到 4.1
已完成 1%
速度: 1.90 MB/秒
名称: eaply.Bimap-method.html
剩余时间: 大约 10 分钟
剩余项目: 49,407 (1.69 GB)

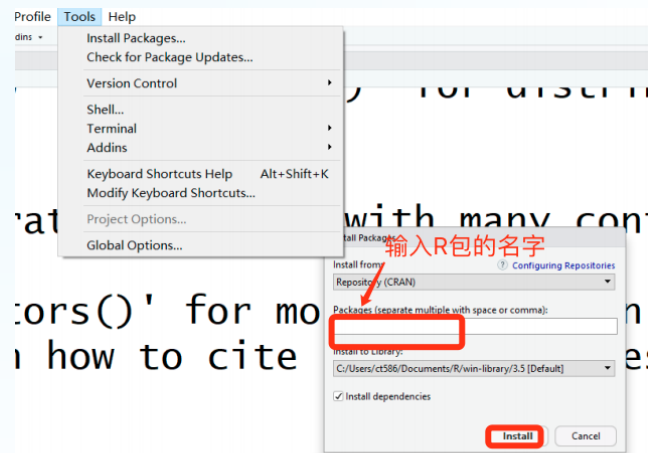
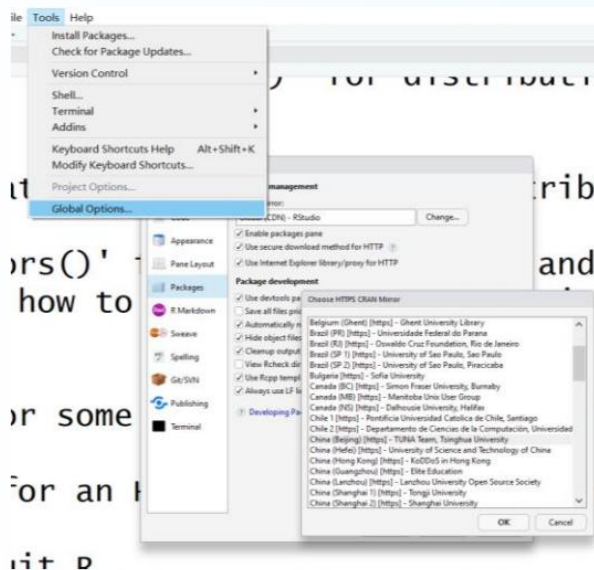
Tools
1. Install Packages...
2. Check for Package Updates...
Version Control
Shell...
Terminal
Jobs
Addins
Memory
Keyboard Shortcuts Help
Modify Keyboard Shortcuts...
Show Command Palette
Project Options...
Global Options...

Install Packages
Install from: Repository (CRAN)
Packages (separate multiple with space or comma):
Install to Library: C:/Users/qihy/Documents/R/win-library/4.1 [Default]
3. 按顺序操作 1,2,3三个步骤，目的就是拿到 3所在方框的路径。这个路径就是我们接下来要拷贝R包进入的路径。

基因组

4. 重头安装R包——CRAN (只看不操作)

- 按左侧图示配置最近的CRAN镜像，方便快速下载；
- 按右侧图示输入待安装包的名字，安装CRAN包。



宏基因组

4. 重头安装R包——Bioconductor/Github (只看不操作)

- #安装CRAN包(斜体部分为包的名字, 需要替换)
- site= "<https://mirrors.tuna.tsinghua.edu.cn/CRAN>"
- install.packages("devtools", repo=site)
- #安装Bioconductor的包(斜体部分为包的名字, 需要替换)
- BiocManager::install(c("AnnotationDbi", "GO.db", "preprocessCore", "impute"))
- #安装Github的R包(斜体部分为包的名字, 需要替换)
- devtools::install_github("JustinaZ/pcaReduce")

宏基因组

生信宝典

易生信



R包——测试和常用问题解决

- 替换之后，重新启动 Rstudio，在左下角 > 后面输入 `library(ggplot2)` 测试，如下图则是安装成功。
- 如果出现 `Error in library("ggplot2") : 不存在叫 'ggplot2' 这个名字的程辑包`，则说明包未安装成功。可能是前面操作解压覆盖没有成功，如位置不正确，可人为检查并修改。
- 如果是自己需要用的新包，不存在于易生信提供的压缩包中，则需按后面的方式重头安装。

Console /var/www/html/ehbio_doc/train/ ↗

```
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.
```

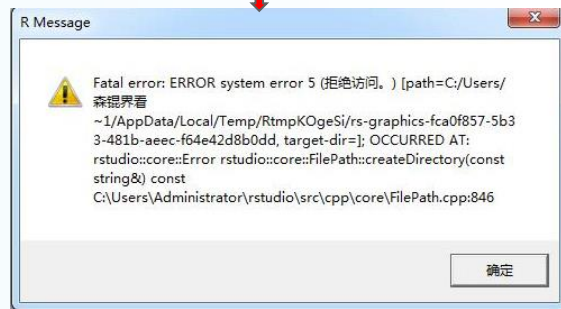
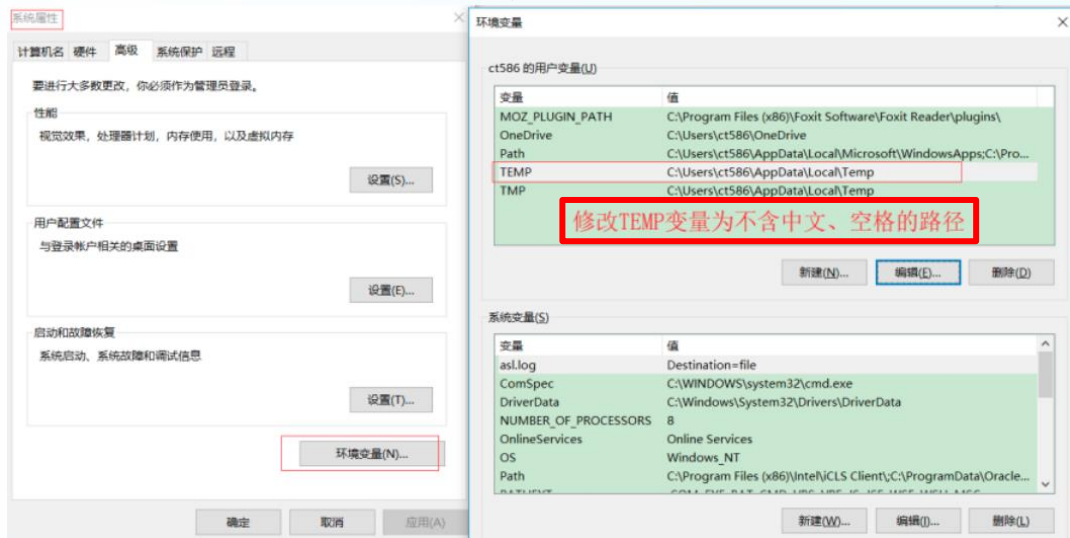
```
[Workspace loaded from /var/www/html/ehbio_doc/train/.RData]
```

```
> library(ggplot2)  
>
```

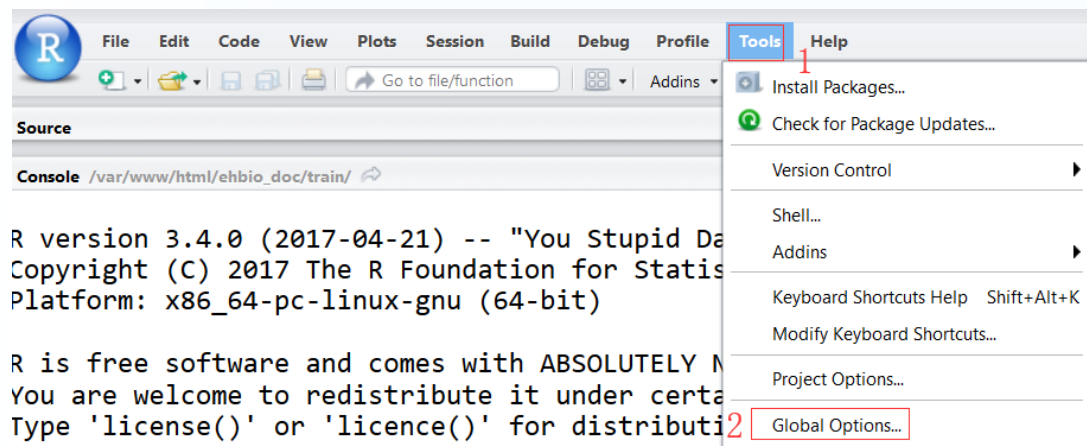


启动Rstudio时若遇到中文用户名乱码

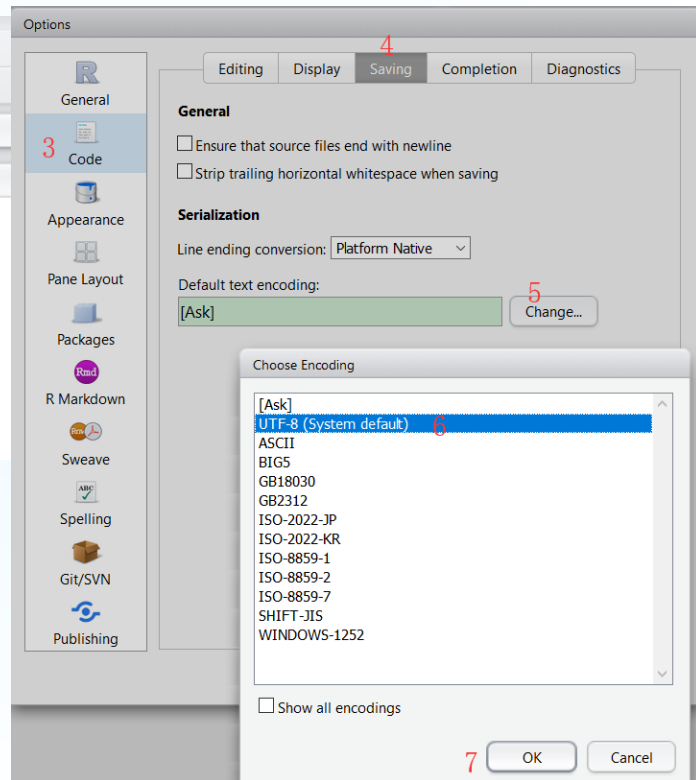
- 如果您碰到如下错误，是因为用户名中存在中文，导致乱码不能识别，请新建一个**英文用户**，再行安装。
- 也可以尝试下面的解决方案：
- 进入“我的电脑”，点击右键进入“属性”，进入“高级系统设置”



Rstudio中字符乱码 – 改为UTF-8进行字符编码



设置后重新打开之前乱码的文件或重启Rstudio
再打开之前的文件



5. 谷歌Chrome浏览器

- 远程访问 Rstudio server (192.178.1.107:8787, 此IP不可用)
- 兼容性最好的浏览器
- 在文件夹下或点击链接下载ChromeStandaloneSetup64.exe安装包
- 按默认参数完成安装即可

宏基因组

生信宝典

易生信



6. Java 11



- Cytoscape、gesa-4.0和IGV都依赖于Java（所以务必保证先安装上java，其他三个软件才能正常运行）
- 软件下载地址
- 在文件夹下或点击链接下载安装程序jre-8u181-windows-x64.exe或jdk-11.0.7_windows-x64_bin.exe
- 按默认参数完成安装即可



7. Cytoscape



- Cytoscape是一款图形化显示网络软件，生物学中常用于分析转录因子与基因或蛋白与蛋白之间互作关系、GO和KEGG富集分析
- [软件下载地址](#)
- 在文件夹下或点击链接下载安装程序
Cytoscape_3_9_0_windows_64bit.exe
- 按默认参数完成安装即可
- [Cytoscape使用视频教程](#)

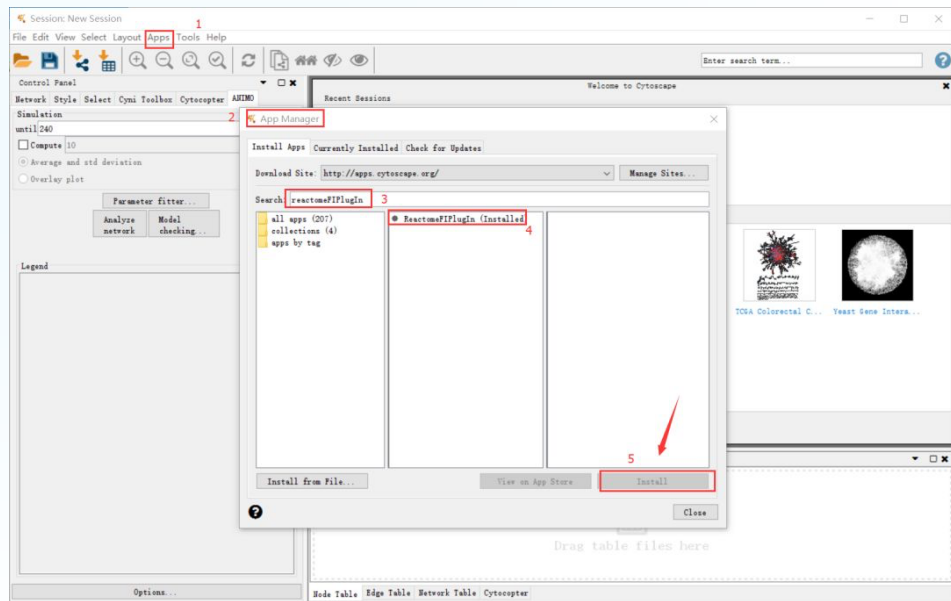


Cytoscape插件安装



- 安装好Cytoscape后打开软件，按照图示方式安装以下几个插件
- 需要安装的插件：

- yFiles Layout Algorithms;
- CyKEGGParser;



8. javaGSEA



- GSEA富集分析本地软件；软件依赖java8/java11运行（请确保前面已安装好java）
- 软件下载地址
- 在文件夹下或点击链接下载安装程序gsea-4.1，按默认参数完成安装即可
- 关于GSEA富集分析基础知识查看公众号推送：《GSEA富集分析——界面操作》



9. Adobe Illustrator



- 即我们常说的“AI”，是一款非常好的矢量图形处理工具、图片排版工具
- 是Adobe公司开发的一款收费软件，在此为大家提供一个破解免安装版的包仅供学习使用
- 在文件夹下或点击链接下载安装包Illustrator 32.zip，解压后运行IllustratorPortable.exe程序即可使用
- 使用视频教程



9. Adobe Illustrator



- 当双击IllustratorPortable.exe无法打开或者遇到报错情况如下图时
删掉软件包中的Data文件夹，再双击IllustratorPortable.exe即可打开
(Data文件夹是每次运行AI时自动产生的文件，每次打开前删除即可)



App	2018/8/5 9:34	文件夹
Data	2018/8/7 9:47	文件夹
IllustratorPortable.exe	2015/7/27 4:14	应用程序
IllustratorPortable.ini	2015/7/28 10:08	配置设置
说明.txt	2015/7/27 3:42	文本文档
未来软件园_百度搜索	2014/10/30 11:31	Internet 快捷方式
未来软件园-首页	2006/7/20 22:26	Internet 快捷方式
下载安装说明.txt	2015/3/18 16:58	文本文档



10. IGV



- 整合基因组浏览器(IGV)是一种高性能的可视化工具，用来交互式地探索大型综合基因组数据。
- 需要在java环境下运行，Java安装见前。
- [IGV下载地址](#)
- [IGV使用教程](#)
- [浏览器基本使用见](#)
- 可在文件夹下或点击链接下载安装包IGV_Win_2.7.2



IGV 使用方法



名称	修改日期	类型
jre1.8.0_131	2018/7/19 13:54	文件夹
lib	2018/7/19 14:04	文件夹
.DS_Store	2018/7/19 13:54	DS_Store
igv.bat	2018/7/19 13:54	Windows 批处理文件

igv.bat - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

::Get the current batch file's short path

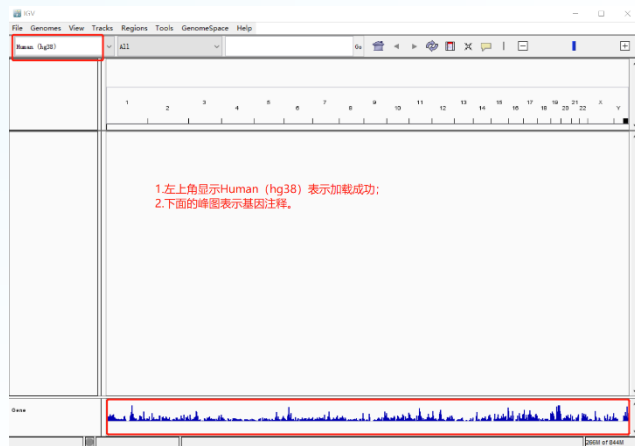
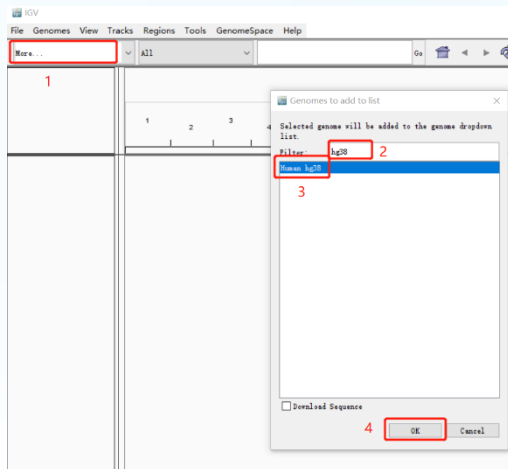
for %%x in (%0) do set BatchPath=%%~dpsx

for %%x in (%BatchPath%) do set BatchPath=%%~dpsx

jre1.8.0_131\bin\javaw -Xmx2G -Dproduction=true -Djava.net.preferIPv4Stack=true -Dsun.java2d.noddraw=true -jar %BatchPath%lib\igv.jar %*

1. 双击igv.bat启动。
2. 若因内存问题启动失败，右键igv.bat选择编辑，修改标出的2G为1G或500M等。

请提前加载好hg38基因组信息（见右边两张图）



1. 左上角显示Human (hg38) 表示加载成功;
2. 下面的峰图表示基因注释。



11. Xshell & Xftp/FileZilla (不需要)



- Xshell是终端模拟软件，Xftp是文件传输助手
- FileZilla可以断点续传进行上传、下载（需要服务器支持）
- 在文件夹下或点击链接下载安装程序Xshell5.exe、Xftp5.exe和FileZilla_3.35.1_win64.exe
- 安装Xftp6时用户名和单位名均填写：ehbio
- 按默认参数完成安装即可，但是需要按照图示选择许可类型。

请选择一个许可类型。如果你是注册用户或评估Xftp5，请选择商业。

☐ 商业（包括评估）

☒ 免费为家庭/学校



Xshell安装和使用



- Xshell是Windows下功能强大的终端。Linux和Mac可以直接ssh访问远程服务器。
- [最新版下载地址](#)
- 安装：注意选择Home/School类型可免费使用
- 运行：首次需要配置服务器信息，文件-新建-名称YSX-主机IP-确定-输入用户名并勾选记住-输入密码并勾选记录-确定

- Rstudio中的Terminal也可以访问服务器
- `ssh username@serverIP`

如 `ssh amplicon@192.168.1.130`

```
Xshell 5 (Build 1005)
Copyright (c) 2002-2016 NetSarang Computer, Inc. All rights reserved.

Type 'help' to learn how to use Xshell prompt.
[c:\~]$

Connecting to 201.113.224.110:4000...
Connection established.
To escape to local shell, press 'Ctrl+Alt+J'.

Welcome to Ubuntu 16.04.1 LTS (GNU/Linux 4.4.0-34-generic x86_64)

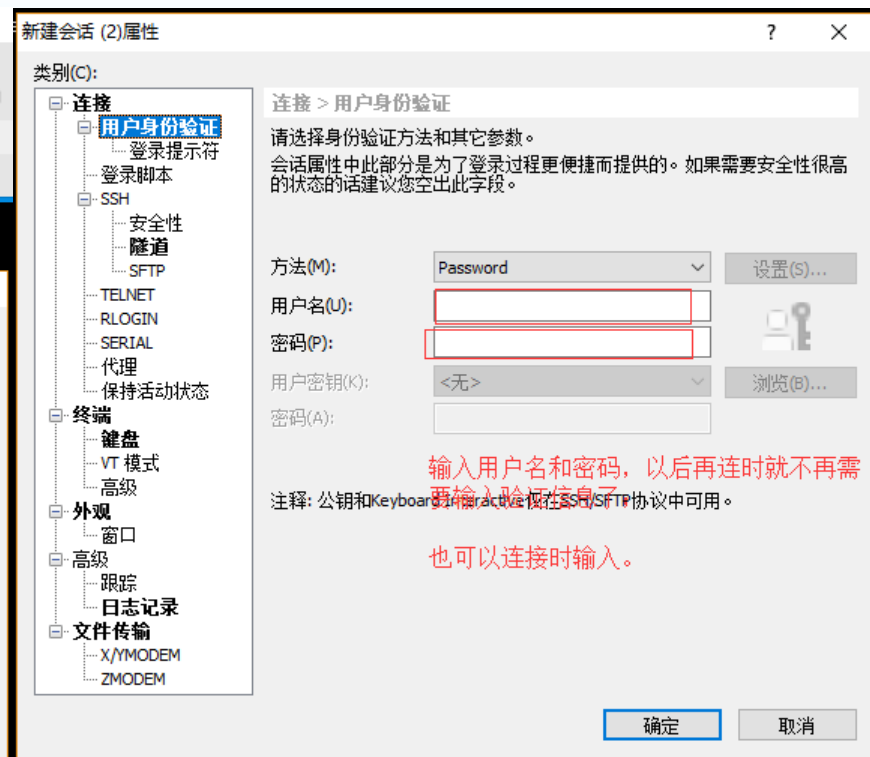
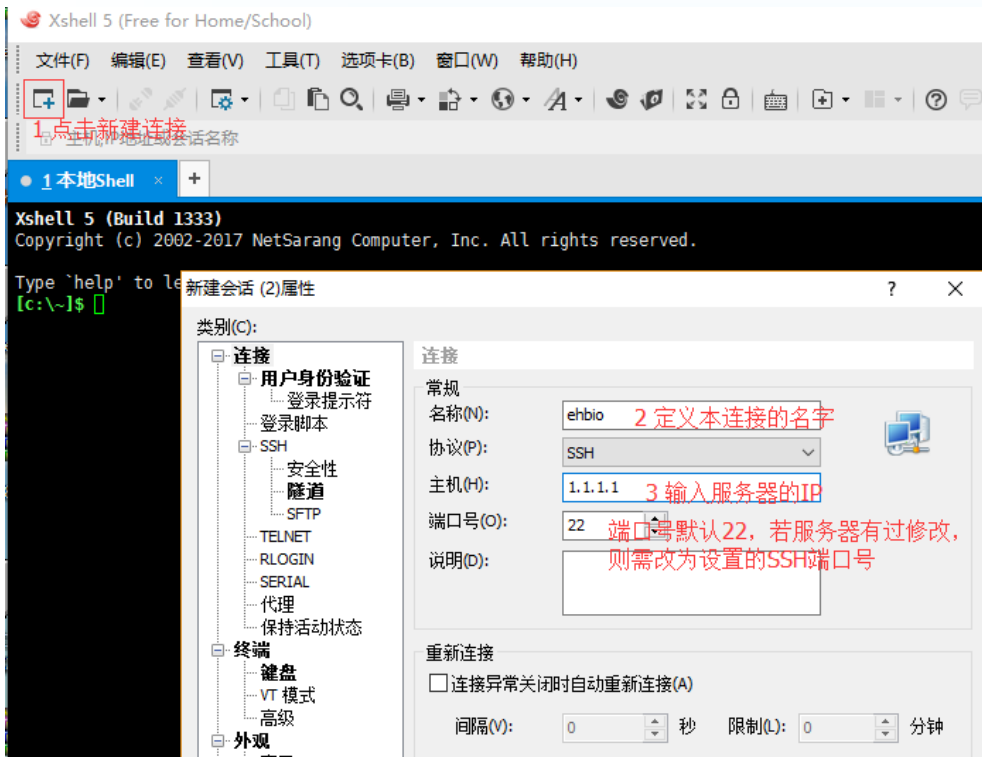
 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

12 packages can be updated.
7 updates are security updates.

Last login: Tue Aug 30 08:33:05 2016 from 124.10.132.133
[yongxin@biocloud:~]$
```



Xshell使用



Xshell使用



新建会话 (2)属性



类别(C):

- 连接
 - 用户身份验证
 - 登录提示符
 - SSH
 - 登录脚本
 - 安全性
 - 隧道
 - SFTP
 - TELNET
 - RLOGIN
 - SERIAL
 - 代理
 - 保持活动状态
- 终端
 - 键盘
 - VT 模式
 - 高级
- 外观
 - 窗口
- 高级
 - 跟踪
 - 日志记录
- 文件传输
 - X/YMODEM
 - ZMODEM

外观

调整配色方案, 字体大小等

预览:

Normal Bold Underline Reversed Cursor
black red green yellow blue magenta cyan white
black red green yellow blue magenta cyan white

字体和颜色

配色方案(O):

XTerm

编辑(E)...

字体名(F):

New White
New Black
ANSI Colors on Black
ANSI Colors on White
White on Black
Black on White
Pastel on Black
Pastel on White
XTerm

字体大小(S):

"加粗"文本(B):

字体质量(Q):

光标

光标颜色(B):

■

☐ 使用闪烁的光标(U)

文本颜色(T):

■

速度(P):

millisec

外观(A):

☒ 块

☐ 竖线

☐ 下划线

确定

取消

新建会话 (2)属性



类别(C):

- 连接
 - 用户身份验证
 - 登录提示符
 - SSH
 - 登录脚本
 - 安全性
 - 隧道
 - SFTP
 - TELNET
 - RLOGIN
 - SERIAL
 - 代理
 - 保持活动状态
- 终端
 - 键盘
 - VT 模式
 - 高级
- 外观
 - 窗口
- 高级
 - 跟踪
 - 日志记录
- 文件传输
 - X/YMODEM
 - ZMODEM

终端

仿真

终端类型(T):

xterm

☐ 连接后使用以下终端大小(U)

列(O):

80

行(R):

24

滚动缓冲区

缓冲区大小(S):

1024

行

☒ 将清除的屏幕推到滚动缓冲区(P)。

转换

选择远程主机的编码。已选择默认语言时Xshell使用Windows的默认语言。

编码(E):

Unicode (UTF-8)

☐ 韩中日语言中的不确定字符处理为宽字符(A)

确定

取消

默认都是UTF-8编码, 若出现乱码, 可来此查看是否已配置为UTF-8编码



FileZilla的安装和使用（不需要）



- [官网](#) Download FileZilla Client 下载适合自己系统的安装包
- 默认安装我们提供的FileZilla_3.35.1_win64-setup_bundled.exe
- 配置 Configure

打开软件-左上角图标 ‘打开站点管理器’ - 新站点 - 命名为 YSX - 主机 填写 192.168.1.130, 端口 22; 协议 SFTP, 登陆类型 普通, 用户名为姓名全拼, 密码默认为ysx001。首次连接选择总是信任, 确定。



FileZilla上传下载使用



主机(H): sftp://192.168.1.10 用户名(U): yishengxin 密码(W): 端口(P): 22 快速连接(Q) ▾

上传下载前的登录设置，注意sftp是必须的，端口号根据需要修改

本地站点: D:\

文件名 文件大小

..	
11940958FOCA	4
VitaminD switches BAF complexes to protect b...	6,827
业务宣传片.wmv	19,866
基因组分析测序.docx	166,33
FSCapture.rar	4,264
poster.pdf	14,618
EHBIO	

设置 隶属于“编辑”菜单栏

选择页面(P): 请选择目标文件存在时的默认操作。

文件存在时的默认操作

下载(D): 继续文件传输

上传(U): 继续文件传输

要是选择“如果较新就覆盖”，您的系统时间需要与服务器同步。若时间不同（例如不同的时区），请在站点管理器中选择一个时区。

☐ 允许续传 ASCII 文件(L)
如果服务器使用不同格式的行结尾，继续传输 ASCII 格式文件可能导致问题。

连接

- FTP
 - 主动模式
 - 被动模式
- FTP 代理服务器
- SFTP
- 通用代理
- 传输
 - 文件类型
 - 对已存在文件的操作
- 界面
 - 主题
 - 日期/时间格式
 - 文件大小格式
 - 文件列表
- 语言
- 文件编辑

○ 上传高通量数据见 [链接/](#)

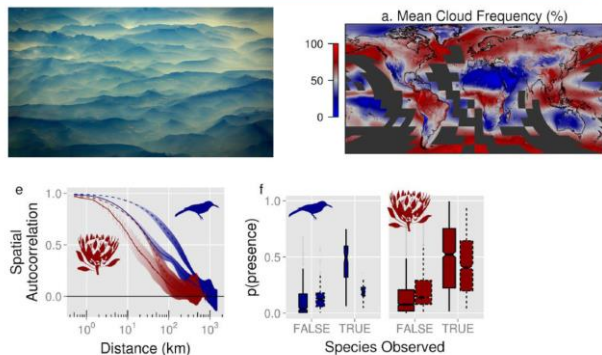


免费视频资料领取



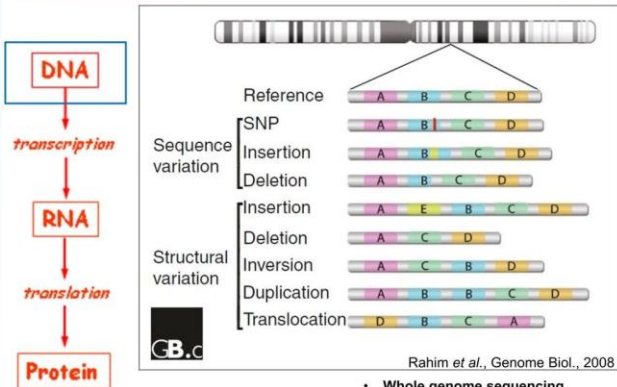
- 生物信息培训免费资料: <http://www.ehbio.com/Training/index.php/Home/Index/Articles.html>
- 组学讲解经典视频, 免费观看
- 视频从中心法则出发, 从不同层面介绍组学分析的手段, 并解释了不同方法都能解决什么具体的生物学问题。本视频内容全面, 简单易懂, 很适合想知道组学能做什么和生物数据分析没有思路的朋友, 能为你在学习、科研项目以及分析实战的过程中提供很多的指导。

Big Data in ecology



Wilson and Jetz, PLoS Biol., 2016

Central dogma of molecular biology



- Whole genome sequencing
- Genome resequencing





扫码关注生信宝典，学习更多生信知识



扫码关注宏基因组，获取专业学习资料

易生信，没有难学的生信知识



易汉博基因科技(北京)有限公司
EHBIO Gene Technology (Beijing) co., LTD