

# 1. Linux

袁荣卿 朱运鹏 for linux

钟砺涵 吴根艺 for markdown

谢真乐 for vscode

1.1 作业

1.2 awk简析

1.3 双系统的文件交换

1.4 Git的安装

# 1.1 作业

- 第1题 1. 解释1.gtf文件中第4、5列代表什么, exon长度应该是\$5-\$4+1还是\$5-\$4?

```
IV    ensembl gene    1802    2953    .    +    .    gene_id "YDL248W"; gene_version "1"; gene_name "COS7"; gene_source "ensembl"; gene_biotype "protein_coding";
```

第4列指这个序列开始的坐标,  
第5列指这个序列结束的坐标。  
exon长度应该是\$5-\$4+1。

# 1.1 作业

- 第2题 列出1.gtf文件中 XI 号染色体上的后 10 个 CDS（按照每个CDS终止位置的基因组坐标进行sort）。

```
cat 1.gtf | awk '$1 == "XI" && $3 == "CDS"' | sort -n -k 5 | tail -10
```

pipe

- 第3题 统计 IV 号染色体上各类 feature（1.gtf文件的第3列，有些注释文件中还应同时考虑第2列）的数目，并按升序排列。

```
grep -v '^#' 1.gtf | awk '{if($1 == "IV") print $3}' | sort | uniq -c | sort -n -k 1
```

正则表达式

<https://deerchao.cn/tutorials/regex/regex.htm>

## 1.2 awk简析

- “awk是一种处理文本文件的语言，是一个强大的文本分析工具”
- awk指令的语法：

```
awk [选项参数] 'script' var=value file(s)
```

或

```
awk [选项参数] -f scriptfile var=value file(s)
```

- awk脚本的基本结构：

```
BEGIN{循环前代码块}; 条件语句 {逐行执行代码块}; END{循环后代码块}
```

- 重要的选项参数与内建变量：

1. -F [分隔符 or 正则表达式]，为文本文件的读取指定分隔符，默认为空白符（空格、制表键）。例如：指定为','后适用于csv文件。
2. -f [awk脚本]，指定一个脚本文件，代替脚本输入。使用脚本文件的好处是可以更好地组织代码，适用于对文件进行复杂处理时。
3. -v [变量=值]，变量赋值。
4. \$n，表示某行的第n列字符串。
5. FS，字段分隔符，修改其值可以代替-F选项。
6. OFS，输出字段分隔符，默认等于FS。

# 几点说明

- awk的输出除print外还有printf（格式化输出），用法与C语言基本一致。另外awk语言也包含与C用法基本一致的条件与循环语句。
- awk内置了很多实用函数，比如算数函数的atan2、sqrt等，字符串函数的gsub、split等。用户还可以通过function语法来自定义函数。
- awk的变量创建时不需要指定类型，其变量类型只有数与字符串，及对应的数组。
- 使用vim编辑器编写awk脚本(\*.awk)，用以复杂处理文本文件。
- 更多信息见[Linux awk 命令 | 菜鸟教程 \(runoob.com\)](https://www.runoob.com/linux/linux-command/linux-awk-command.html)

# 1.3 双系统间的文件交换

- ~~Linux的\$/share与Windows的~~

- ~~~\Desktop\bioinfo\_Tsinghua\_share间无法实现文件交换？~~

- 手动转移的方法：

在Windows的shell内执行指令 `docker cp [src] [dest]`

1. `src` = Linux内的文件绝对路径, `dest` = Windows内的目标文件夹绝对路径 --> 将Linux内文件复制到Windows目标文件夹
2. `src` = Windows内的文件绝对路径, `dest` = Linux内的目标文件夹绝对路径 --> 将Windows内文件复制到Linux目标文件夹

# 1.4 Git的安装

- 最简单的方式：

1. 以root身份进入容器：

```
PS C:\Users\zhuyp> docker exec -it -u root bioinfo_tsinghua bash
```

2. 使用apt-get指令从网络获取git：

```
root@bioinfo_docker:/home/test# apt-get install git.
```

- 注意到程序的下载地址是：

<http://mirrors4.tuna.tsinghua.edu.cn/ubuntu bionic-updates>

然而下载的git版本是2.17.1（官方最新版为2.35.1）



- 从源码安装：

1. 从git官方下载源码包：[Index of /pub/software/scm/git/ \(kernel.org\)](https://kernel.org/pub/software/scm/git/)，后缀为.tar.gz和.tar.xz均可，将其转移入Linux中并使用tar指令解压。

2. 安装编译源码所需的包与工具（apt-get install ...）：

dh-autoreconf libcurl4-gnutls-dev libexpat1-dev gettext  
libz-dev libssl-dev asciidoc xmlto docbook2x install-info

3. 最后在源码包解压后的文件夹内（如git-2.35.1）输入以下指令：

- 1) make prefix=/usr/local all doc info

- 2) make install prefix=/usr/local install-doc install-html install-info

**缺点**：复杂、耗时长、占用空间大

**优点**：软件版本不受制于镜像的软件源

# 2. Markdown

Lihan Zhong

Genyi Wu



# 什么是markdown

- 轻量级文本编辑语言
  - Just using <tag> to mark down the text
- 与Word不同的地方
  - Word (\*.doc & \*.docx) is a **binary format** file
  - Markdown is a **plain text format** file

# 如果用记事本打开同样的word和markdown文件



```
Test.md - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)
# 一级标题
## 二级标题
### 三级标题
#### 四级标题
##### 五级标题

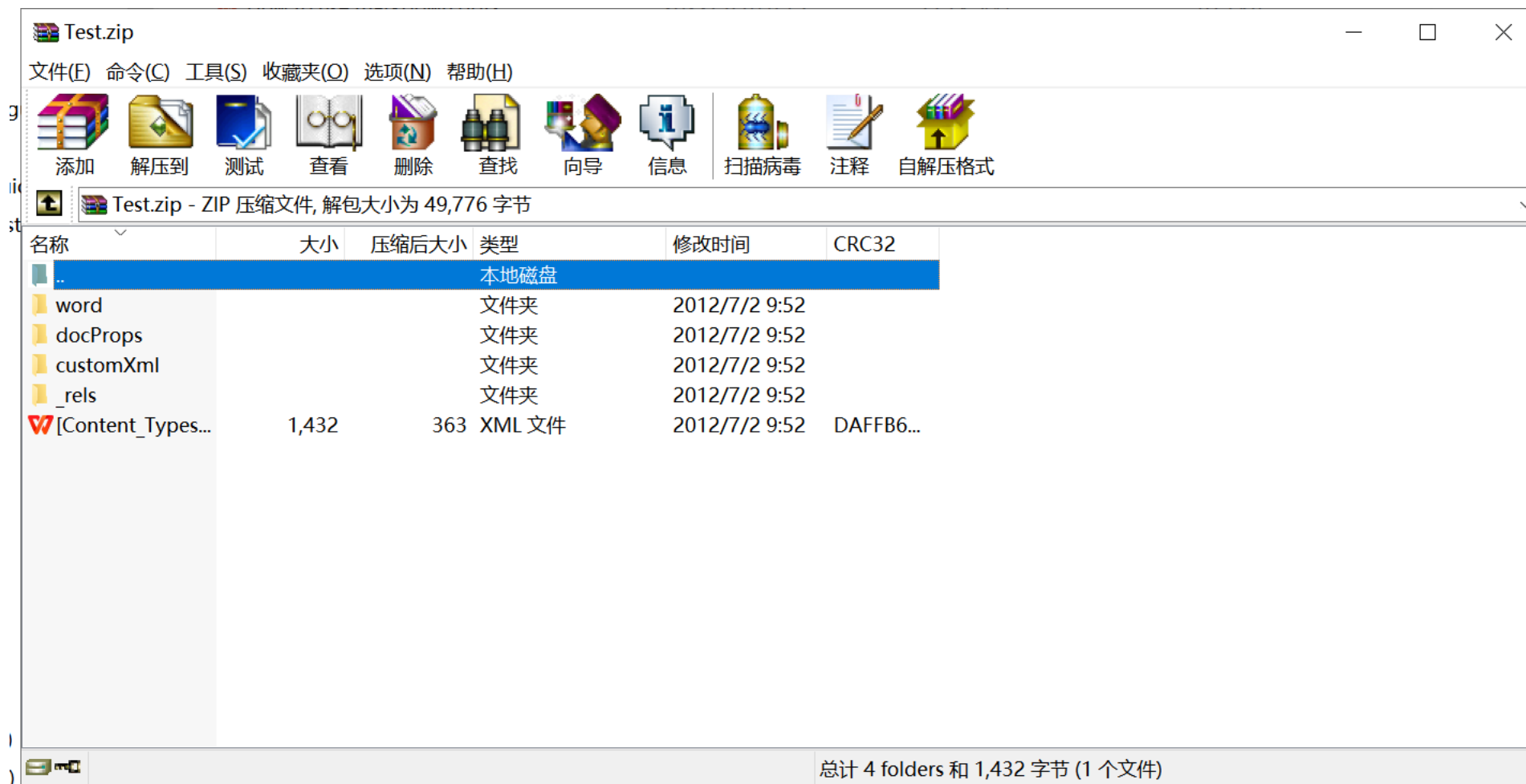
正文

***\加粗\****、<span style="background-color:yellow">高亮</span>、**斜体**、<font color="red">标红</font>、<u>下划线</u>
```

第 7 行, 第 10 列 100% Windows (CRLF) UTF-8

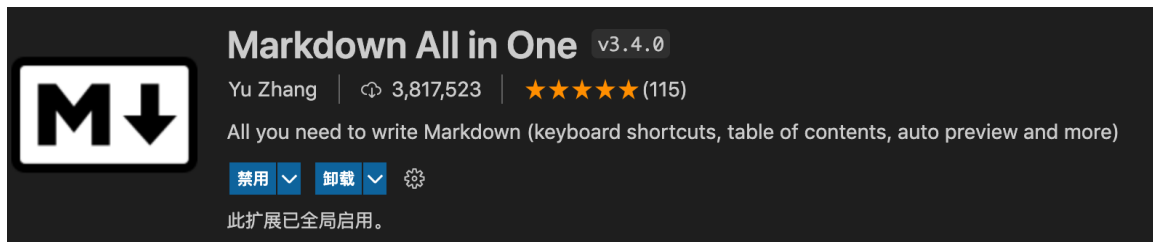
# Word本质上是一个.zip文件

编辑困难！

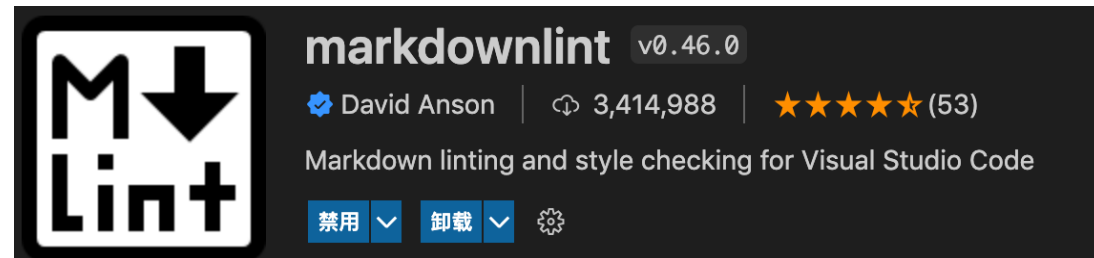


# 什么是markdown

- Markdown的优点：
  - 语法简单：只需要加入对应的<tag>即可。兼容HTML的tag，功能非常强大
  - 移植性好：跨平台，无论是win还是mac，格式完全相同
  - 易编辑：纯文本，易于**被程序**读取、编辑
  - 导出方便：可以轻松导出为PDF、HTML、LaTex、EPub、Word格式
- Markdown的缺点
  - 图片问题：图片必须要有一个“路径”
  - 语法兼容：目前有了一些高级的markdown编辑格式，可能会出现版本不兼容问题
  - 细节控制：肯定没有Word那么好
    - 图片的位置精细调节
    - 字体大小的精细调节
    - .....



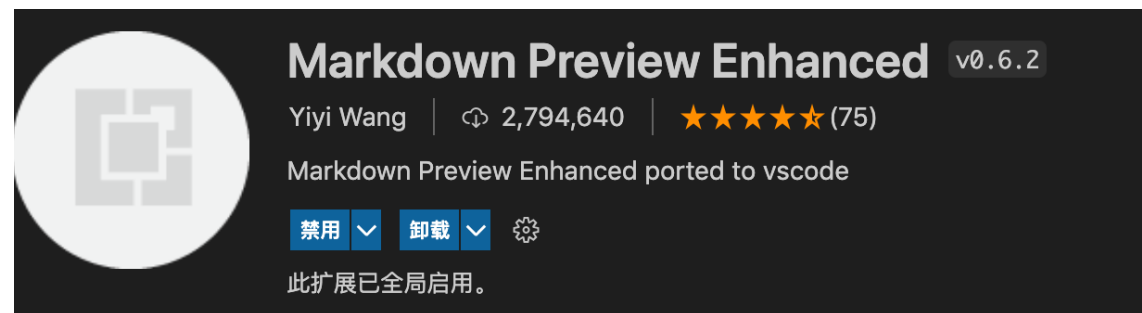
集成了很多MD功能的插件



针对MD的语法风格检查



可以将MD文件转为PDF文件



即时预览功能

Element	Markdown Syntax
<a href="#">Heading</a>	# H1 ## H2 ### H3
<a href="#">Bold</a>	<b>**bold text**</b>
<a href="#">Italic</a>	<i>*italicized text*</i>
<a href="#">Blockquote</a>	> blockquote
<a href="#">Ordered List</a>	1. First item 2. Second item 3. Third item
<a href="#">Unordered List</a>	- First item - Second item - Third item ( * - + )
<a href="#">Code</a>	`code`
<a href="#">Horizontal Rule</a>	---
<a href="#">Link</a>	[title](https://www.example.com)
<a href="#">Image</a>	![alt text](image.jpg)

<a href="#">Fenced Code Block</a>	<pre>```\n{\n  "firstName": "John",\n  "lastName": "Smith",\n  "age": 25\n}\n```</pre>
<a href="#">Footnote</a>	<p>Here's a sentence with a footnote. [^1]</p> <p>[^1]: This is the footnote.</p>

<https://www.markdownguide.org/cheat-sheet/>



Element	Markdown Syntax
<a href="#">Table</a>	Syntax   Description     -----   -----     Header   Title     Paragraph   Text

<a href="#">Definition List</a>	term : definition
---------------------------------	----------------------

<a href="#">Strikethrough</a>	~~The world is flat.~~
-------------------------------	------------------------

<a href="#">Task List</a>	- [x] Write the press release - [ ] Update the website
---------------------------	---

<a href="#">Highlight</a>	==very important words==.
---------------------------	---------------------------

<a href="#">Subscript</a>	H~2~O
<a href="#">Superscript</a>	X^2^
<a href="#">Emoji</a>	:joy: :angry:

Syntax	Description
Header	Title
Paragraph	Text

<b><i>term</i></b>
definition

~~The world is flat.~~

- ☒ Write the press release
- ☐ Update the website

very important words

H<sub>2</sub>O  
X<sup>2</sup>



# Tips

homework1.md 9+ •

homework1.md > # HM1-吴根艺-2019012253 > ## 3

24

gene\_id "YKR105C"; gene\_version "1"; transcript\_id "YKR105C";  
transcript\_version "1"; exon\_number "1"; gene\_name "VBA5";  
gene\_source "ensembl"; gene\_biotype "protein\_coding";  
transcript\_name "VBA5"; transcript\_source "ensembl";  
transcript\_biotype "protein\_coding"; protein\_id "YKR105C";  
protein\_version "1";  
XI ensembl CDS 661442 663286 . + 0  
gene\_id "YKR106W"; gene\_version "1"; transcript\_id "YKR106W";  
transcript\_version "1"; exon\_number "1"; gene\_name "GEX2";  
gene\_source "ensembl"; gene\_biotype "protein\_coding";  
transcript\_name "GEX2"; transcript\_source "ensembl";  
transcript\_biotype "protein\_coding"; protein\_id "YKR106W";  
protein\_version "1";

25

26 - 运行代码

27

28 grep CDS 1.gtf | awk '\$1=="XI" | sort -n -k 5| tail

29

Preview homework1.md ×

• 结果:

XI	ensembl	CDS	631152	632798	.	+	0	gene_id "YKR097W"; gene
XI	ensembl	CDS	633029	635179	.	-	0	gene_id "YKR098C"; gene
XI	ensembl	CDS	635851	638283	.	+	0	gene_id "YKR099W"; gene
XI	ensembl	CDS	638904	639968	.	-	0	gene_id "YKR100C"; gene
XI	ensembl	CDS	640540	642501	.	+	0	gene_id "YKR101W"; gene
XI	ensembl	CDS	646356	649862	.	+	0	gene_id "YKR102W"; gene
XI	ensembl	CDS	653080	656733	.	+	0	gene_id "YKR103W"; gene
XI	ensembl	CDS	656836	657753	.	+	0	gene_id "YKR104W"; gene
XI	ensembl	CDS	658719	660464	.	-	0	gene_id "YKR105C"; gene
XI	ensembl	CDS	661442	663286	.	+	0	gene_id "YKR106W"; gene

• 运行代码

grep CDS 1.gtf | awk '\$1=="XI" | sort -n -k 5| tail

# Tips

转化成不同的格式

Open in Browser	
HTML	▶
Chrome (Puppeteer)	▶
PDF (prince)	
eBook	▶
Pandoc	
Save as Markdown	
Image Helper	
Sync Source	
Preview Theme	▶

**Image Helper**

Link


enter image URL here, then press 'Enter' to insert.

Copy image to workspace /assets folder

Click me to browse image file

Upload

Click me to browse image file

use  to upload images

选择不同的预览背景

# Typora：免费的轻量级Markdown编辑器

- 一秒启动
- 编辑简单
- 支持表格、代码、数学公式
- 所见即所得



# Typora：方便的编辑工具

- 可以简单地用**快捷键**，做好**标签**



# Typora：方便的编辑工具

- 可以简单地用**快捷键**，做好标签
- 轻松导出各种格式




# Typora：方便的编辑工具

- 可以简单地用**快捷键**，做好标签
- 轻松导出各种格式
- 复制粘贴可以快速生成图像地址

正文

**\*加粗\***、高亮、*斜体*、标红、下划线、~~删除线~~、<sup>上标</sup>、<sub>下标</sub>、正文

 ![image-20220310083242737](C:\Users\Lenovo\AppData\Roaming\Typora\typora-user-images\image-20220310083242737.png)



# 什么是markdown

- Markdown的优点：
  - 语法简单：只需要加入对应的<tag>即可。兼容HTML的tag，功能非常强大
  - 移植性好：跨平台，无论是win还是mac，格式完全相同
  - 易编辑：纯文本，易于**被程序**读取、编辑
  - 导出方便：可以轻松导出为PDF、HTML、LaTex、EPub、Word格式
- Markdown的缺点 **（瑕不掩瑜）**
  - 图片问题：图片必须要有一个“路径”
  - 语法兼容：目前有了一些高级的markdown编辑格式，可能会出现版本不兼容问题
  - 细节控制：肯定没有Word那么好
    - 图片的位置精细调节
    - 字体大小的精细调节
    - .....





# Typora：方便的编辑工具

$$E = \begin{Bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{Bmatrix} \quad (1)$$

- 可以简单地用**快捷键**，做好标签
- 轻松导出各种格式
- 复制粘贴可以快速生成图像地址
- 支持**数学公式**

$$\begin{cases} 3x + 5y + z \\ 7x - 2y + 4z \\ -6x + 3y + 2z \end{cases} \quad (2)$$

$$f(x) = \int_a^b \frac{1}{\sqrt{x^2 + 1}} \times \sqrt[3]{x^3 + 1} dx \quad (3)$$

- 公式指令集总结1：<https://zhuanlan.zhihu.com/p/261750408>
- 公式指令集总结2：<https://zhuanlan.zhihu.com/p/165342174>

# Typora：所见即所得

## 一级标题

---

## 二级标题

---

### 三级标题

#### 四级标题

##### 五级标题

正文

**\*加粗\***、**高亮**、*斜体*、**标红**、下划线、~~删除线~~、<sup>上标</sup>、<sub>下标</sub>、正文



# 3. Vscode与Docker与Git

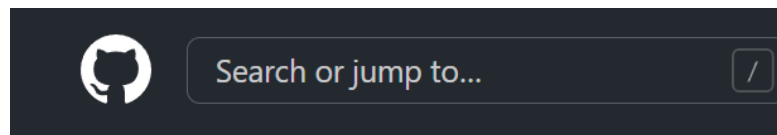
Zhenle Xie

3.1 利用Vscode向github同步代码

3.2 利用Vscode访问Docker镜像

# 3.1 利用Vscode向github同步代码




- 首先是创建一个库



## Recent Repositories



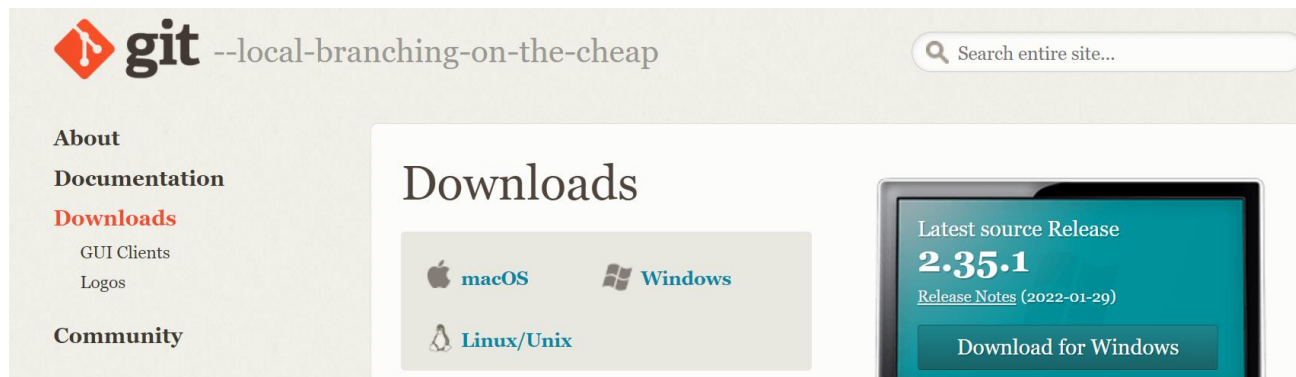
Find a repository...

-  233xzl/biohw
-  233xzl/Python
-  233xzl/hello-world

## Recent activity

When you take actions across GitHub, we'll provide links to that activity here.

- 下载并安装git



- 配置git并将在线库同步到本地

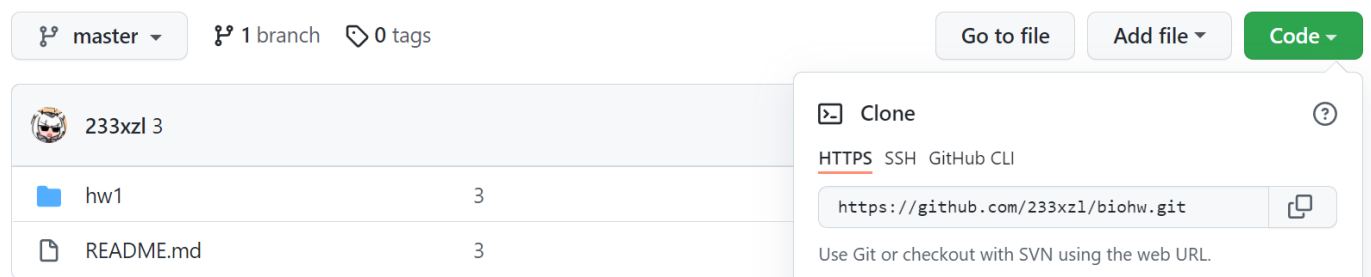


- 配置用户名和邮箱

```
xzl@DESKTOP-LF6J56B MINGW64 ~/Desktop/bioinfo_tsinghua_share
$ git config --global user.name "233xzl"

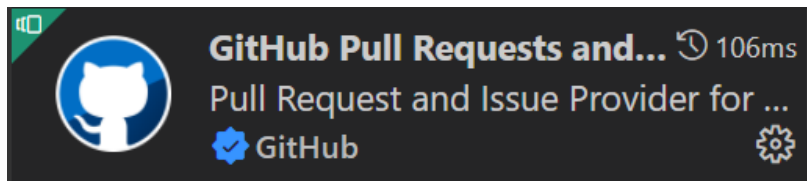
xzl@DESKTOP-LF6J56B MINGW64 ~/Desktop/bioinfo_tsinghua_share
$ git config --global user.email "747086546@qq.com"
```

- 同步在线库

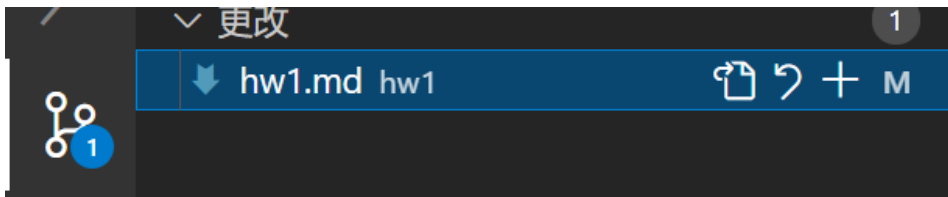


```
xzl@DESKTOP-LF6J56B MINGW64 ~/Desktop/bioinfo_tsinghua_share
$ git clone https://github.com/233xzl/biohw.git
```

- 安装github扩展

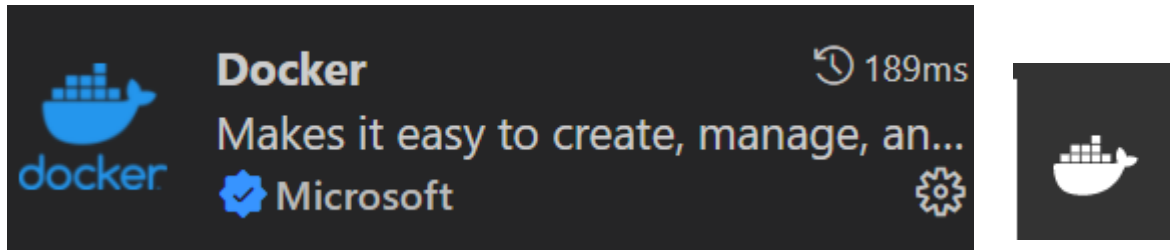


- 尝试提交代码



## 3.2 利用Vscode访问Docker镜像

- 安装一个扩展



- 搞定！



# 仍存在的问题

- 无法将Vscode的工作目录直接设置在docker容器内
- 因此也无法在docker内直接用Vscode完成代码的提交等操作