

nf-core-workflow

Meng Luo

目录

CONTENTS

1 **Creating a pipeline**

2 **工作内容及体会**

3 **职业发展规划**

➤ Creating a pipeline

```
/ Name: scrnaworkflow
```

```
tion: scRNA Pipeline
```

```
Meng Luo
```

```
Creating new nf-core pipeline: 'nf-core/scrnaworkflow'
```

```
Initialising pipeline git repository
```

```
Done. Remember to add a remote and push to GitHub:
```

```
cd /mnt/ilustre/users/meng.luo/Pipeline/RNA/Pipeline/nf_core/nf-core-scrnaworkflow
```

```
git remote add origin git@github.com:USERNAME/REPO_NAME.git
```

```
git push --all origin
```

```
This will also push your newly created dev branch and the TEMPLATE branch
```

```
!!!!!! IMPORTANT !!!!!!
```

```
If you are interested in adding your pipeline to the nf-core community,  
PLEASE COME AND TALK TO US IN THE NF-CORE SLACK BEFORE WRITING ANY CODE!
```

```
Please read: https://nf-co.re/developers/adding\_pipelines#join-the-community
```

```
Workflow Name: █
```

<https://nf-co.re/tools/#creating-a-new-workflow>

- 基础流程包含fastqc 和multiqc两个程序（可以持续添加你想加的流程）
- 需要重新修改整个流程，接口需要修改不能直接运行
- 下载例子：
- nextflow run
/mnt/ilustre/users/meng.luo/Pipeline/RNA/Pipeline/nf_core/circrna-test/ --input
/mnt/ilustre/users/meng.luo/Pipeline/RNA/Pipeline/nf_core/test-datasets-
circrna/samples.csv --input_type fastq --fasta
/mnt/ilustre/users/meng.luo/Pipeline/RNA/Pipeline/nf_core/test-datasets-
circrna/reference/chrl.fa --gtf
/mnt/ilustre/users/meng.luo/Pipeline/RNA/Pipeline/nf_core/test-datasets-
circrna/reference/chrl.gtf --mature
/mnt/ilustre/users/meng.luo/Pipeline/RNA/Pipeline/nf_core/test-datasets-
circrna/reference/mature.fa --tool circexplorer2 --phenotype
/mnt/ilustre/users/meng.luo/Pipeline/RNA/Pipeline/nf_core/test-datasets-
circrna/phenotype.csv --trim_fastq false --module
circrna_discovery,mirna_prediction,differential_expression --outdir
circrna_result --bsj_reads 2 --species cel

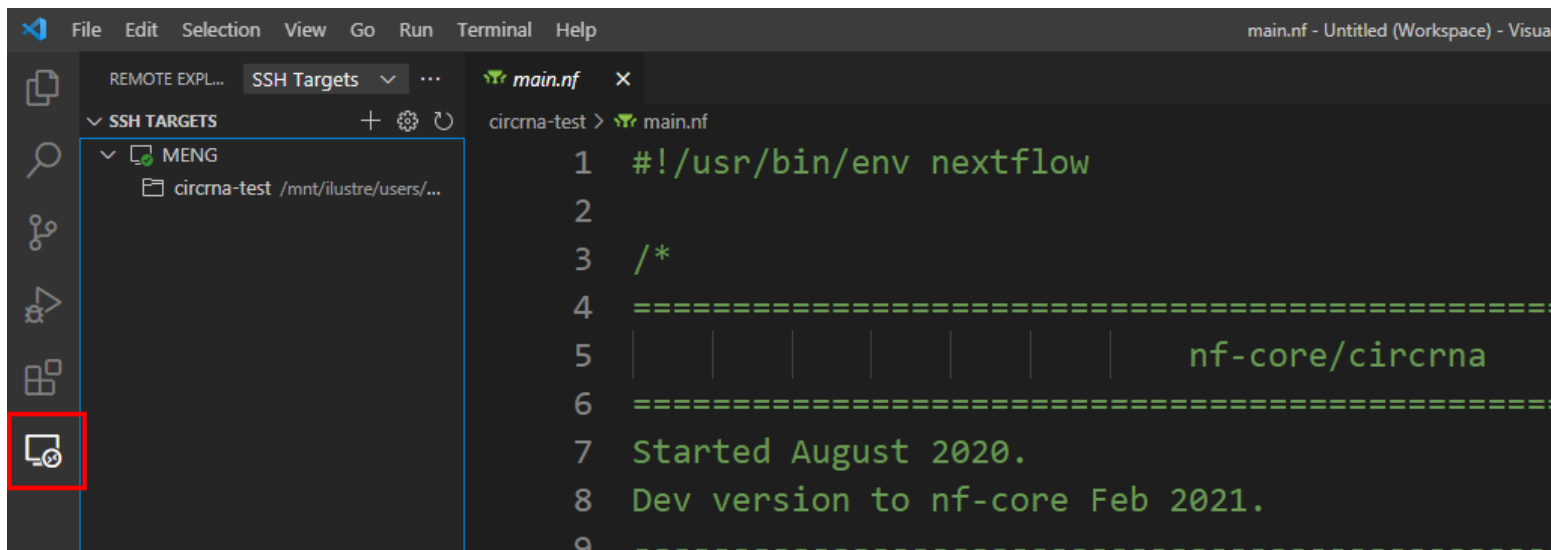
➤ Creating a pipeline

- `-bg` 想nohup一样
- 单纯的nextflow 加上clean可以清除一些cache
- `file`（可以直接有文件，链接过来的文件）和`val`（就是文件名）的变量关系要区分

➤ VSCODE remote-ssh

安装remote development

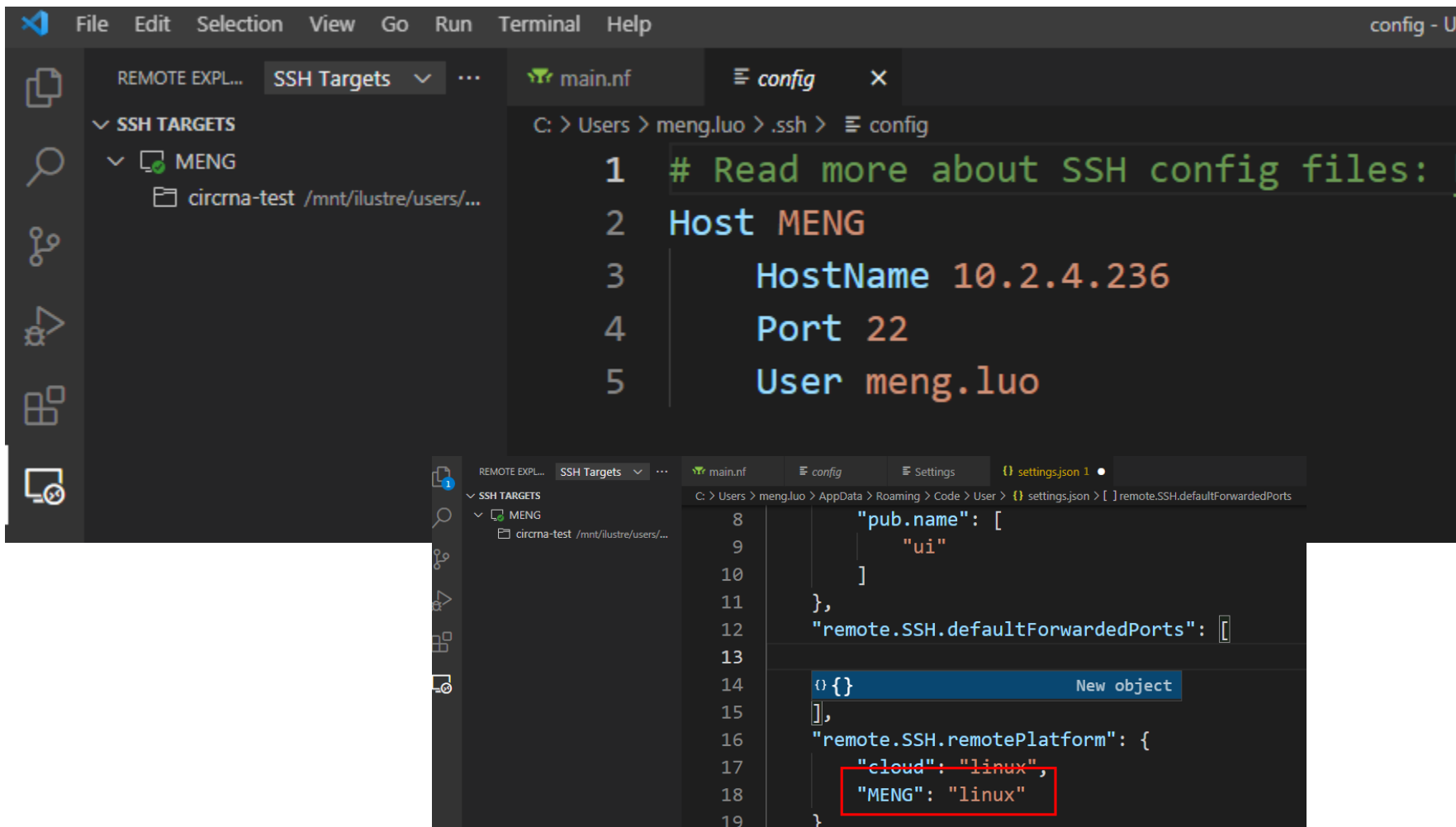
安装完Remote-Development后，左侧会多一个按钮



➤ Creating a pipeline

按下shift+ctrl+p, 输入Remote-SSH: Connect to Host

此时, 添加一个新的SSH Host



The screenshot displays the Visual Studio Code interface with two panels. The top panel shows the 'SSH Targets' sidebar on the left, listing a target named 'MENG' with the path 'circma-test /mnt/ilustre/users/...'. The main editor area shows the 'config' file for the 'MENG' target, containing the following configuration:

```
1 # Read more about SSH config files: [link]
2 Host MENG
3     HostName 10.2.4.236
4     Port 22
5     User meng.luo
```

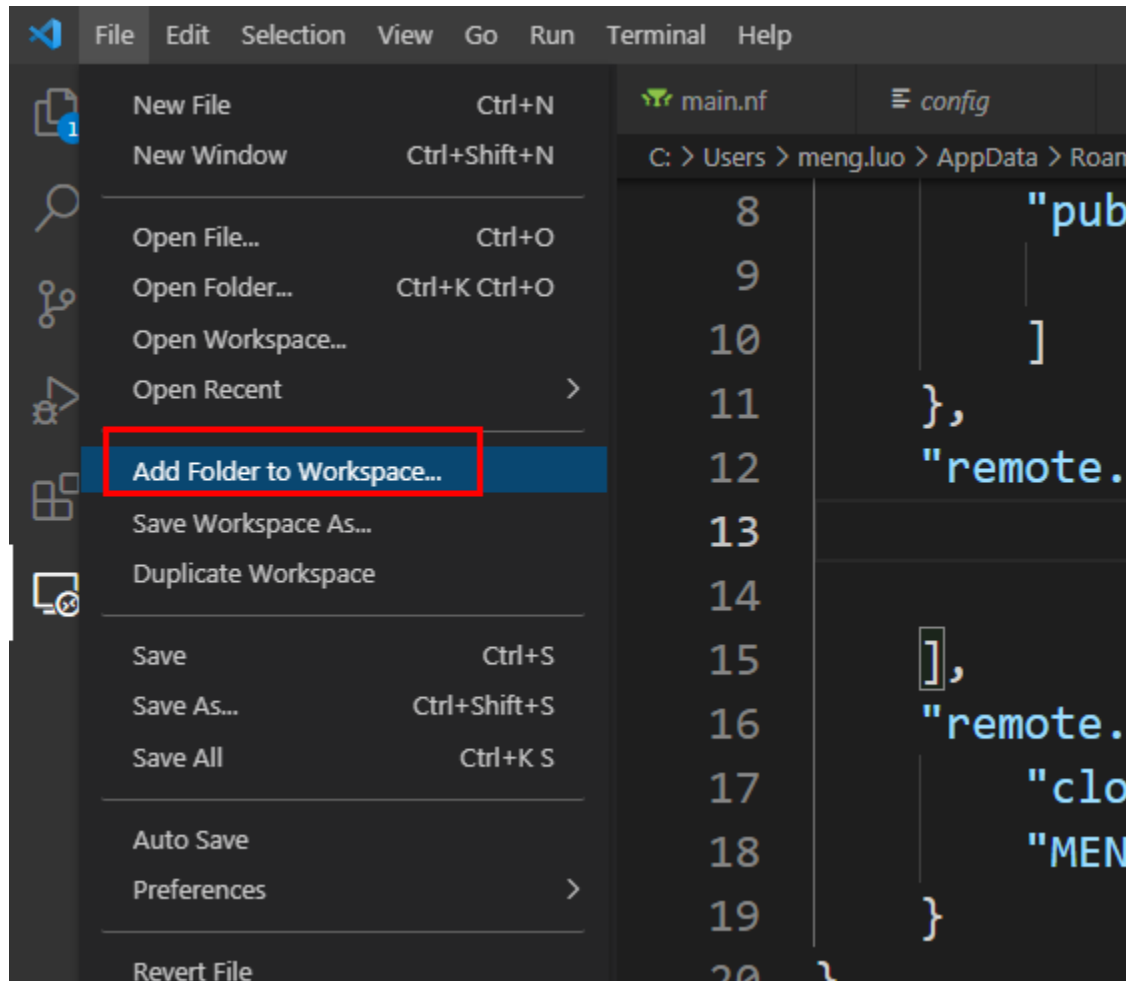
The bottom panel shows the 'Settings' view, specifically the 'remote.SSH.defaultForwardedPorts' setting. The configuration is as follows:

```
8 "pub.name": [
9     "ui"
10 ],
11 },
12 "remote.SSH.defaultForwardedPorts": [
13     {
14         "cloud": "linux",
15         "MENG": "linux"
16     }
17 ],
18 "remote.SSH.remotePlatform": {
19     "cloud": "linux",
20     "MENG": "linux"
21 }
```

A red box highlights the 'MENG' entry in the 'remote.SSH.remotePlatform' object, indicating the platform configuration for the 'MENG' target.

➤ Creating a pipeline

- 添加多个文件夹到同一个操作窗口，设置不占内存的方法，打开shift+ctrl+p json，添加："search.followSymlinks": false,



➤ groovy 编写

groovy编写得语言可以使用在任何nextflow中进行编写

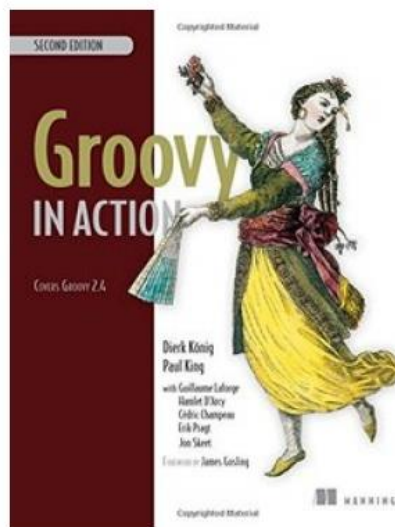
<http://groovy-lang.org/groovy-dev-kit.html#Collections-Lists>

<https://www.nextflow.io/docs/latest/script.html>

书籍下载:

<https://bbooks.info/viewmore/groovy-in-action-second-edition>

[Groovy in Action, Second Edition](#) Sented by Shon



Groovy in Action, Second Edition

Groovy in Action, Second Edition

[Download book \(pdf - 18.67 MB\)](#)

This link for educational purpose only. Please remove file from your computer after familiarization.

<https://www.nextflow.io/blog/2021/5-more-tips-for-nextflow-user-on-hpc.html>

tuple list 直接给定义命名为新的变量才行，不能只给一个，**tuple**只能给一个[],不能在**map**中给多个[],不然无法分隔出变量。

input的**tuple**变量，要给定义一个变量从**input**的里面给出对应的变量，一一对应，才能相应的把对应的变量提出来。

file 只是提出来文件名称，但不是文件绝对路径

nextflow 只要上一步中有一个样本运行成功，下一个流程中也会跟着运行分析这个样本所有相关的分析。

shell的情况下是” 是单引号

script: 是双引号 “”“”

➤ nextflow 更新使用情况

<https://www.nextflow.io/blog/2021/5-more-tips-for-nextflow-user-on-hpc.html>

➤ Creating a pipeline

2 的生命科学技术服务专家

构建的要求(书写格式很重要, 变量名):

PRINT HELP

VALIDATE PARAMETERS (check)

PRINT PARAMETER SUMMARY

Stage Parameters

process

Auxiliary functions

println (打印不出来, 有的变量是不能打印出来的, 在process中不用打印, 记住输出的变量是对的就行, 多尝试跑跑验证)

注意单引号及双引号的使用, 变量中的使用形式是不一样的, 就在script和input及output中使用的是不一样的。也channel 只能出现一次。每个channel 可以赋值1, 2, 3去重复使用在下一个process中。

publishDir 直接copy就行, 不管是文件或文件夹都可以

input file 的变量参数, 就是直接会把文件软链接到分析的当前路径来

➤ Creating a pipeline

基于nf-core

写完的pipeline使用如下进行优化提升搭配

<https://nf-co.re/tools/#pipeline-schema>

nf-core schema build nf-core-testpipeline

只是一个框架，直接使用就行
需要添加python环境

-bg 就像nohup，在后台运行

-q 不打印信息，但是warning还是会提醒，所以.
如果有问题，加上上面的-bg中间中断的话，还是流程运行不能完成。

-log nf-core-circrnaworkflow.log

感谢您的欣赏



地址/Addr: 上海市浦东新区国际医学园区康新公路3399号3号楼

电话/Tel: 021-51875086

服务热线: 400 660 1216

网址/Web: www.majorbio.com

传真/Fax: 021-51875086-8002