

Reproducibility

- data
- software
- computer

- data
- *software*
- computer

Can change over time (months or years)

Data

Freeze d.	GENCODE	Release d.	Genome v.	UCSC version
5.2018	29	10.2018	GRCh38	29 (GRCh38)
6.2014	21	10.2014	GRCh38	-
7.2013	19	12.2013	GRCh37	19 (GRCh37)
7.2011	10	12.2011	GRCh37	10
12.2010	7	4.2011	GRCh37	7
7.2009	3c	9.2009	NCBI36/GRCh37	3c (NCBI36)

You can even build your own reference genome(s)

- unkown species
- unkown genes
- ...

Your lab can generate data over time

and

you must re-analyze it multiple times or combining it in many ways

- data
- *software*
- computer

They transform data

Each transformation has an input and an output

The output of the previous transformation is the input for the next one

There are many tools for many purposes and our duty is to combine them in a proper way

- Operating System (OS)
- Applications
- Bioinformatics Applications

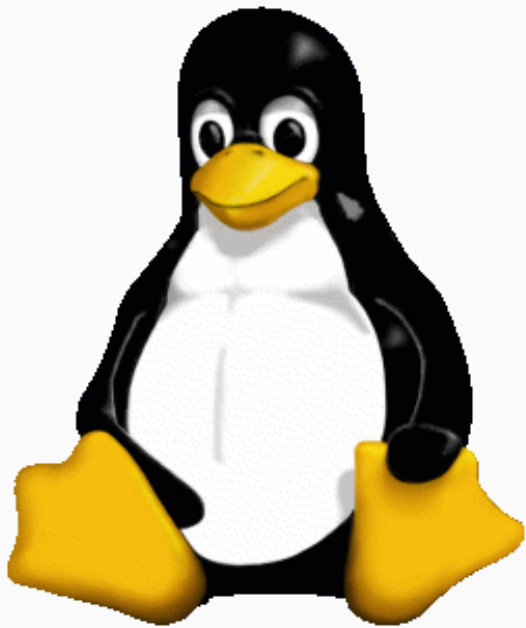




Figure 1: Linux Distribution

- Linux is open source

Its core code can be inspected and you can contribute !

[torvalds/linux](https://github.com/torvalds/linux)

- Linux is open source

It is a very very stable system

Our cluster uptime is about Years

Source	Date	Unix	Windows	Unkw
W3Tech	Feb 15	36.85	32.3	30.5
W3Tech	Feb 19	69.3	30.7	0

W3Tech

- Standalone or Remote

You can install Linux main OS or access a remote machine by network

[illegible]

Figure 2: Terminal

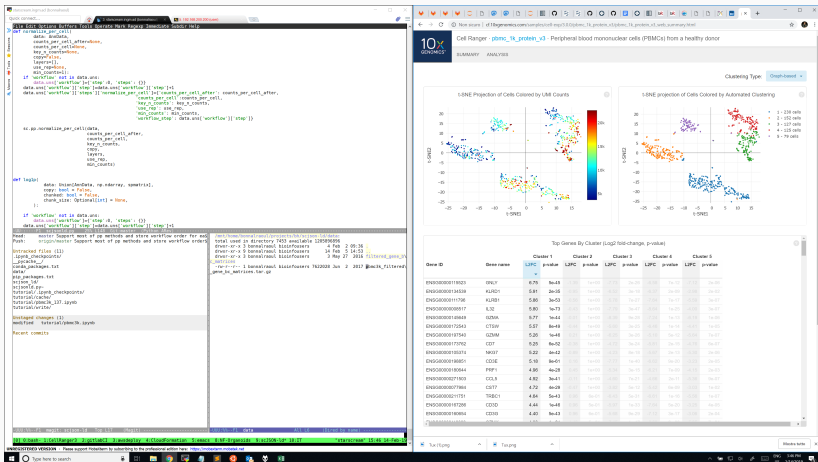


Figure 3: TerminalWeb

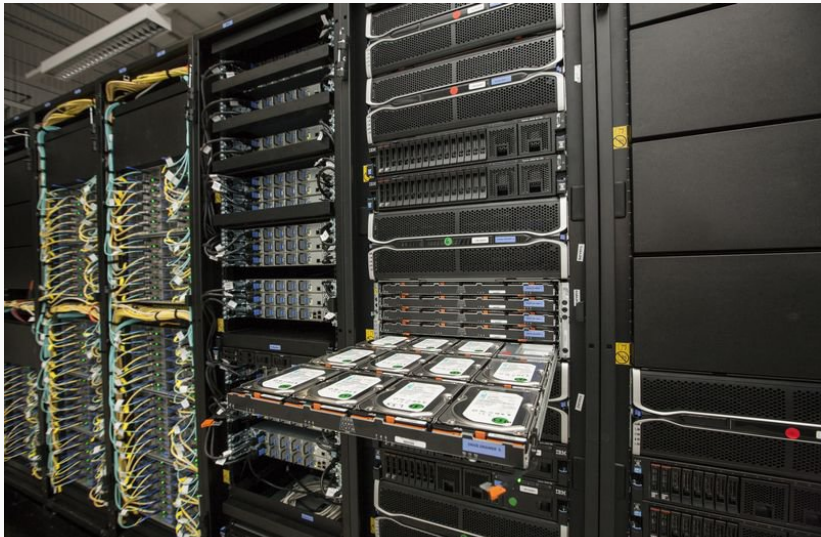
- Multi user
 - Disk
 - Memory
 - CPU

Most of the time you share resources with others

- Multi user
 - Disk (HDD: long term storage)
 - Memory (RAM: volatile)
 - CPU (engine)

Most of the time you share resources with others

- Cluster of thousands of machines



- Open Source
- Standalone or Remote
- Multi User
- Cluster of thousands of machines
- Stable
- Huge communities

Whatever action you are doing through a computer you are going to use a software somewhere

In any case we have 2 best friends:

- An Interface
- An Editor

- Graphical User Interface
- The ugly terminal

- Graphical User Interface (ICONS !!!! Mouse !!!!)
- The ugly terminal

Interfaces

- Graphical User Interface
- The ugly terminal

[illegible]

The winner is ?

Interfaces

[illegible]

Figure 5: The Ugly Terminal

You can instruct the machine to do whatever you want using the keyboard typing commands.

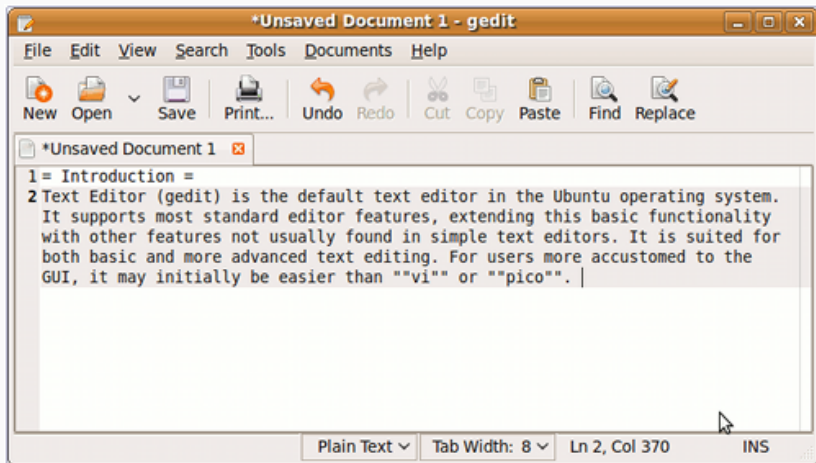


Figure 6: gedit

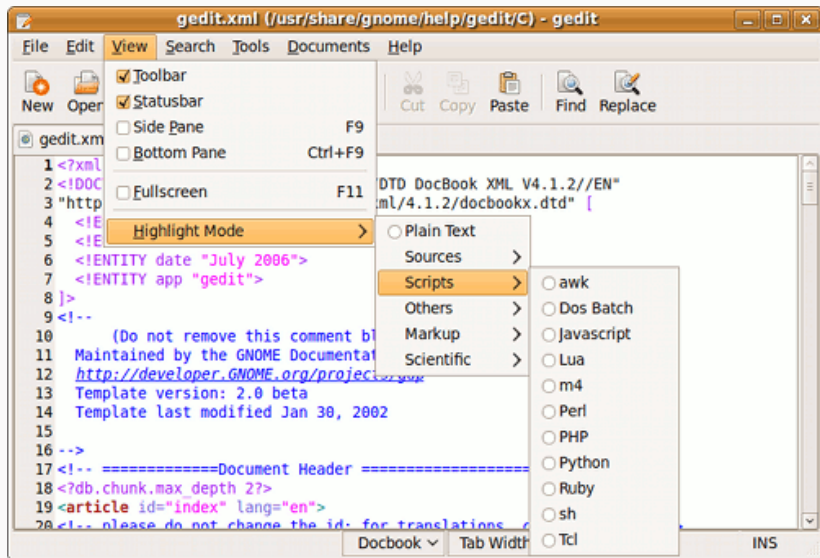


Figure 7: gedit

Terminal

With the terminal you can even write documents, text files.

NOT WORD FILES!!!

```
File Edit Options Buffers Tools Python Virtual Envs Elpy YASnippet Help
1#!/usr/bin/env python
2# -*- coding: utf-8 -*-
3
4"""Tests for 'scutils' package."""
5
6
7import unittest
8
9from scutils import scutils as scu
10import scanpy as sc
11
12
13class TestScutils(unittest.TestCase):
14    """Tests for 'scutils' package."""
15
16    def setUp(self):
17        """Set up test fixtures, if any."""
18        self.data = sc.read_10x_mtx(
19            "tests/data/filtered_gene_bc_matrices/hg19", var_names="gene_symbols", cache=True)
20
21    def tearDown(self):
22        """Tear down test fixtures, if any."""
23
24    def test_min_cells(self):
25        """Test the minimum number of cells that must express a gene."""
26        self.assertEqual(scu.min_cells(self.data), 3)
27
28
29if __name__ == '__main__':
30    unittest.main()
```

```
UUU:----F1 test_scutils.py All L10 Git-master (Python || Elpy FlyC) -----
1*: mode: compilation; default-directory: "~/projects/core/scutils/" -*-
2Compilation started at Tue Feb 12 13:03:21
3
4python -m unittest tests.test_scutile
```

It is a shell which interprets your commands and acts appropriately

- you will love it

The Bash

It is a shell which interprets your command and act appropriately

- you will hate it

The Bash

```
ls -la
```

The Bash

```
ls -l /root
```


Write a file called “myfile.txt” with `gedit` and read it with the terminal

```
cat myfile.txt
```

The Bash

Copy a file

```
cp myfile.txt myfile.backup.txt
```

now do a `ls`

The Bash

Move a file aka Rename

```
mv myfile.txt MyFile.txt
```

now do a ls

The Bash

Connect to your own machine

```
ssh user@localhost
```

if you get a message like this:

```
The authenticity of host 'localhost (:::1)' can't be established.  
ECDSA key fingerprint is SHA256:Zf595J6gKYjVYVD3R/E1AnQmsLu7QB  
Are you sure you want to continue connecting (yes/no)?
```

Type yes

The Bash

Connect to a server

```
ssh user@192.168.200.213
```

now list all the connected users

```
who
```

There are many users all w/the same name

Using applications

```
STAR --help
```

Installing programs

The functionalities of a computer can be extended installing software

Installing programs

1. Ubuntu/Debian packages **list** or apt search
2. **Conda** packages. **search** or conda search
3. **Linuxbrew**. **List** or brew search
4. From specialized archives: e.g. pip
5. Compile the program

- Lots of packages ready `apt-get install` ne
- Very robust
- Additional packages with PPAs

`add-apt-repository ppa:user/ppa-name`

- Not very useful for science

Advantages

- Safe and efficient
- Easy to rebuild from source

- Sometimes outdated (Ubuntu LTS are released every 2 years)

- Must be downloaded and installed (with bash)
- Relies on channels
 - *base* with installation
 - *bioconda*: bioinformatics programs
 - *conda-forge*: misc programs
- Not very robust/stable

Advantages

- Programs are usually the latest version

Cons

- Needs update to config files (`~/.bashrc`)
- Weird interactions with system programs/libraries
- Messes up with `$PATH` (e.g. shebang)

Linuxbrew packages

- Must be downloaded and installed
- Only one source of packages
- Similar to conda
 - fewer packages

On our system is already installed by the procedure is quite easy if you start from scratch.

You can start creating a conda environemnt.

A conda environment let's you install and use software *softly separated* by the original system

```
conda create -n unimiPhD
```

Then you are ready to install the application for RNA-seq

```
conda install -y -c bioconda \  
    fastqc=0.11.8 \  
    trimmomatic=0.38 \  
    samtools=1.9 \  
    star=2.6.1b
```

- data
- software
- *computer*

- Local
- Remote
- Physical
- Virtual
- Cloud