Basics of Linux

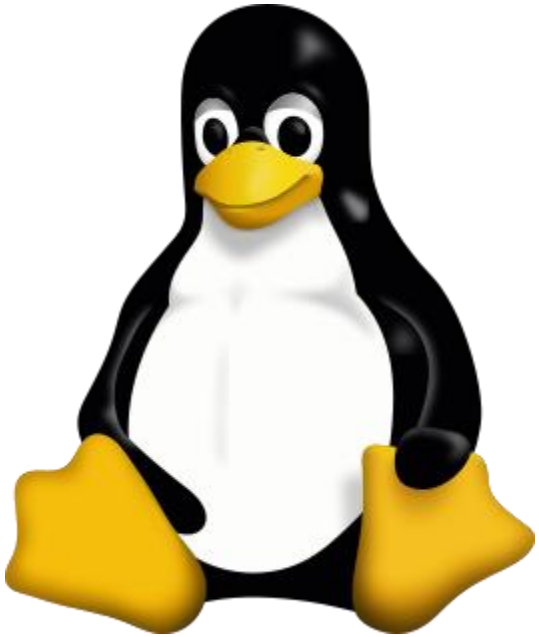
Vishram L. Terse, Ph.D.
Scientific Officer D,
Hematopathology Department, ACTREC
26th September, 2024

Outline

- What is Linux ?
- Who Cares ?
- Organization of files
- Installation of programs
- Use in bioinformatics
- Most frequently used commands

What is Linux ?

- Linux is a common name for a family of open source OS based on Linux kernel
- Evolved from a kernel created by Linus Torvalds, owns the Linux trademark
- Started as a hobby to create an alternative, free and open source version of MINIX (Unix)
- First released in 1991 (Linux kernel + GNU utilities)
- GNU => Richard Stallman, provides utilities (eg: shell)



Tux (Linux kernel mascot)

https://en.wikipedia.org/wiki/Tux_(mascot), https://www.redhat.com/en/topics/linux, https://www.linux.org/threads/what-is-linux.4106/

Linux is one of the biggest examples of open-source success

- Open Source Software (OSS) : allows programmers to develop software and then release it with no licensing fees attached

- Use the software, modify / incorporate it into his or her own system without having to pay

- Advantages:

1. Reliability and transparency : thousands of independent programmers testing and fixing bugs of the software

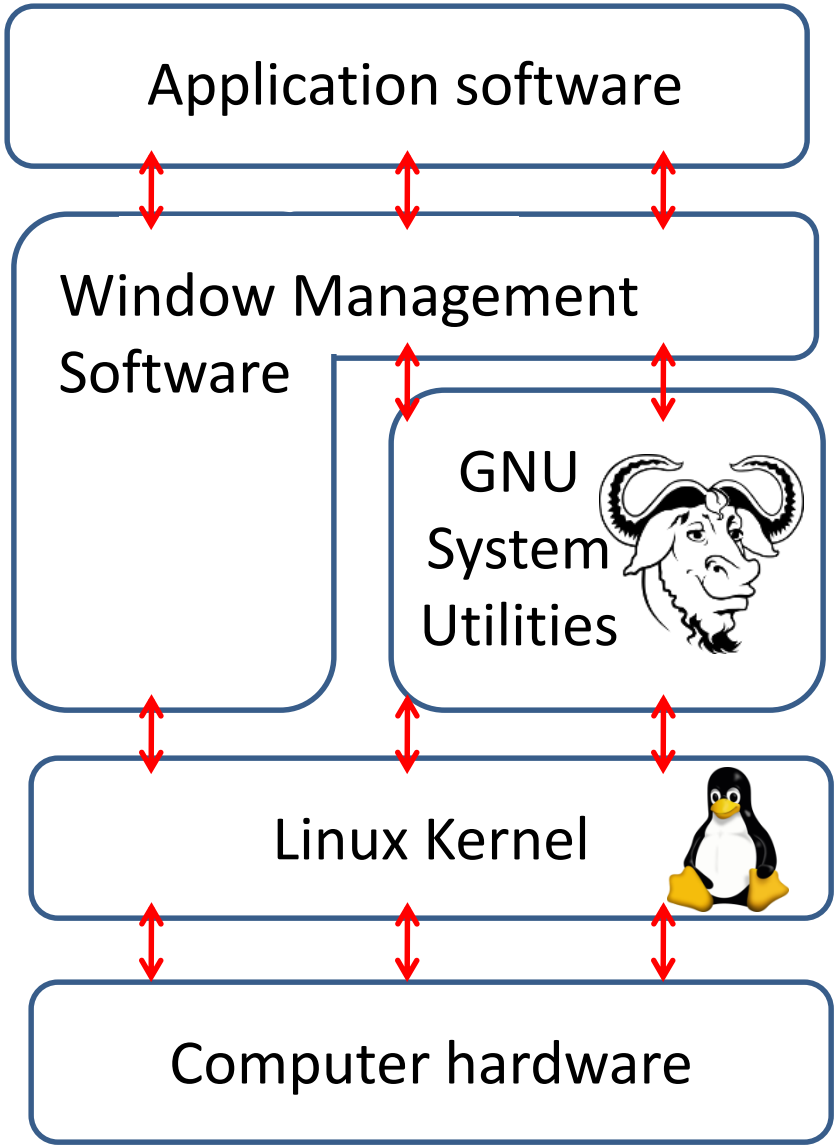
2. Tool to promote a company's image / product

3. Flexibility: Allows building custom interfaces/ add new abilities, promotes innovation

https://en.wikipedia.org/wiki/Open-source_software

Core of a Linux system is the kernel and the utilities

- Kernel controls all the hardware and software
- It is freely available on the internet
- Utilities are required for controlling files / programs ….
- Graphical desktop / Window management gives Windows like interface
- Unlike other OSs , different styles of interfaces are present (KDE, GNOME …)



https://www.gnu.org/graphics/agnuhead.en.html, Blum, R., & Bresnahan, C. (2015). *Linux command line and scripting bible* (3rd ed.). Wiley.

GitHub is a platform to share and manage code

Linus Torvalds

Linux kernel

Git (Version control)

GitHub

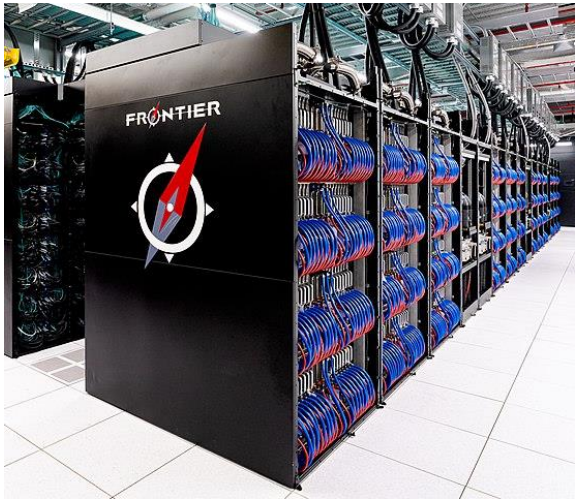
- Linux kernel is freely available on GitHub: https://github.com/torvalds/linux

https://en.wikipedia.org/wiki/Linus_Torvalds

Who cares ?

Smartphones with android (43 %)

Most of internet servers



Top 500 fastest supercomputers

- Not as widely used on Desktop computers

https://stock.adobe.com/search?k=phone+logo,https://commons.wikimedia.org/wiki/File:Android_robot.svg, https://www.alamy.com/stock-photo/internet-server.html?sortBy=relevant, https://www.redhat.com/en/topics/linux, , https://en.wikipedia.org/wiki/Frontier_(supercomputer)#/media/File:Frontier_Supercomputer_(2).jpg

Linux for everyday use ?

- Support for popular web browsers
- Video , music and image editing softwares
- Limited viruses and malware threats
- Free and open source
- Software developers (inbuilt support for text editors/ version control / languages)

- Libre Office : open-source office suite, ~Microsoft Office (Word, Excel, PowerPoint)
- Limited gaming support

Choice of distribution depends on requirements

- A Linux distribution (distro) is an installable operating system
- 2 major types
  Community (free, developed and maintained by community)
  Enterprise (subscription from a vendor)

- Can be downloaded from respective sources
- Windows subsystem for linux (run linux as a program)

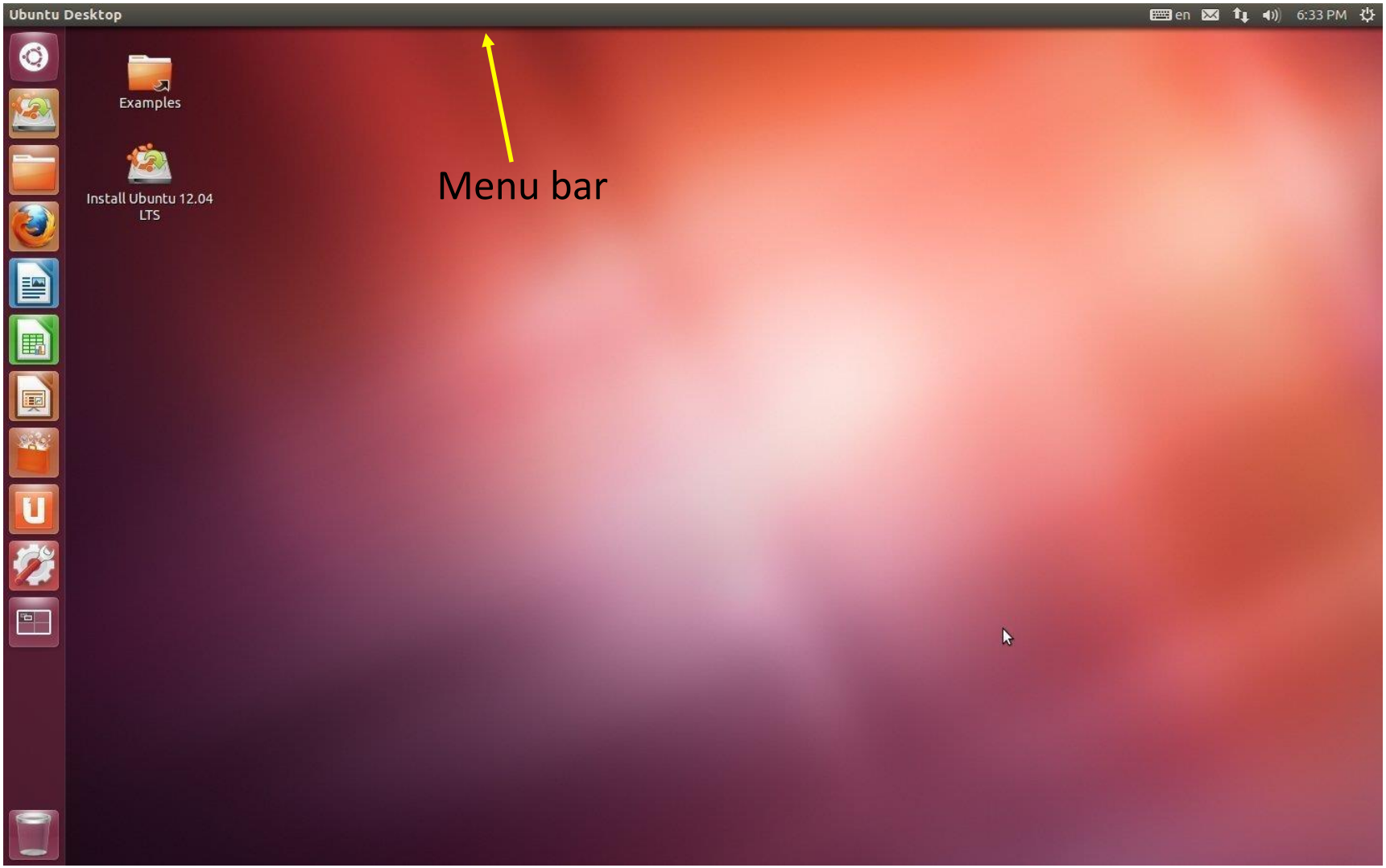
https://www.redhat.com/en/topics/linux/whats-the-best-linux-distro-for-you, https://design.ubuntu.com/resources,
https://www.redhat.com/en/about/brand/standards/product-logos

Command prompt allows interaction with shell

- Shell is an interactive program
- Core of the shell is the command prompt, the interactive part
- Shell contains a set of commands that allow you to copy files, compute values, rename files ….
- A group of shell commands into files, executed as a program is a shell script

Blum, R., & Bresnahan, C. (2015). *Linux command line and scripting bible* (3rd ed.). Wiley.

Desktop environment of a Ubuntu system



Examples of GUIs: **GNOME**, **KDE**, **XFCE, Unity**

https://www.amazon.in/Linux-OS-Preloaded-Desktop-Prettier/dp/B0039BZZ7A

CLI is the place to enter your shell commands

- The default prompt symbol for the bash shell is the dollar sign ($). This symbol indicates that the shell is waiting for you to enter text

- The current user ID name, “ilmnuser” is shown in the prompt. Also, the name of the system is shown, “vl00151”

- After pressing “Enter” wait till $ sign appears again

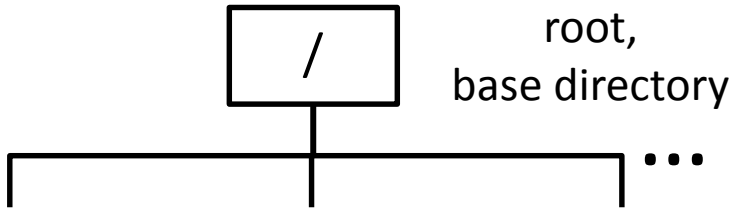
Bash is the default shell in all linux distros

- A variety of shells are available on linux systems
- Used for writing scripts and managing processes
- Bash (Bourne again shell) is a GNU utility, replacement for bourne shell from unix
- Others include :
  1. tcsh: incorporates elements from the C programming language into shell scripts
  2. korn: advanced programming features like associative arrays and floating-point arithmetic
  3. zsh : advanced programming features (bash+ tcsh + korn)
- How to identify ? `$ echo $SHELL` (press Enter) => `/bin/bash`
- MacOS: zsh as default, command structure differs `~ % echo $SHELL` (press Enter) => `/bin/zsh`

Blum, R., & Bresnahan, C. (2015). *Linux command line and scripting bible* (3rd ed.). Wiley.

Navigating the linux filesystem

- File and directories unlike Windows are stored in a virtual directory



binary directory, GNU user-level utilities

Eg: /bin/bash

system configuration files directory

home directory, user directories

Eg: /home/username/

- Linux uses a forward slash (/) instead of a backward slash (\) to denote directories in file paths
- You can traverse directories using “cd” command, Eg : cd /usr/bin (absolute directory ref.)
- Use “pwd” (present working directory) to display current directory location

File permissions provide a crucial level of security

- Users can provide permissions to objects (files/folders/links..) present in their account.

- 3 kinds of permission : control who can read, write, or execute files (rwx)

- 3 permission groups:
  - User : owner of the object
  - Group : multiple users who share a common set of objects
  - Others : Everyone else

- Enter the command : `$ ls -l`

Blum, R., & Bresnahan, C. (2015). *Linux command line and scripting bible* (3rd ed.). Wiley.

Linux for bioinformatics

- Linux : open-source nature, flexibility, and cost-efficiency
- Supports a vast array of open-source bioinformatics tools and frameworks which are cli based
- Handles large datasets efficiently (e.g., genomic data, RNA sequencing)
- Graphical interfaces /web servers (e.g., Galaxy) are limited by the options, cli provides flexibility of analysis
- Large community-driven forums provides bioinformatics tools that are continuously developed and improved
- Windows command prompt / powershell (recently introduced) are paid software and not as versatile
- MacOS is not available as a software

Linux allows for workflow automation

- Bioinformatics analysis need to be reproducible, efficient and scalable
- Workflow engines: Execution of programs with chained input and output

	BASH	nextflow	snakemake
Language	Bash	Groovy	Python
Ease of use	✓	~	✓
Scalability	✗	✓	~
Parallelization	✗	✓	~

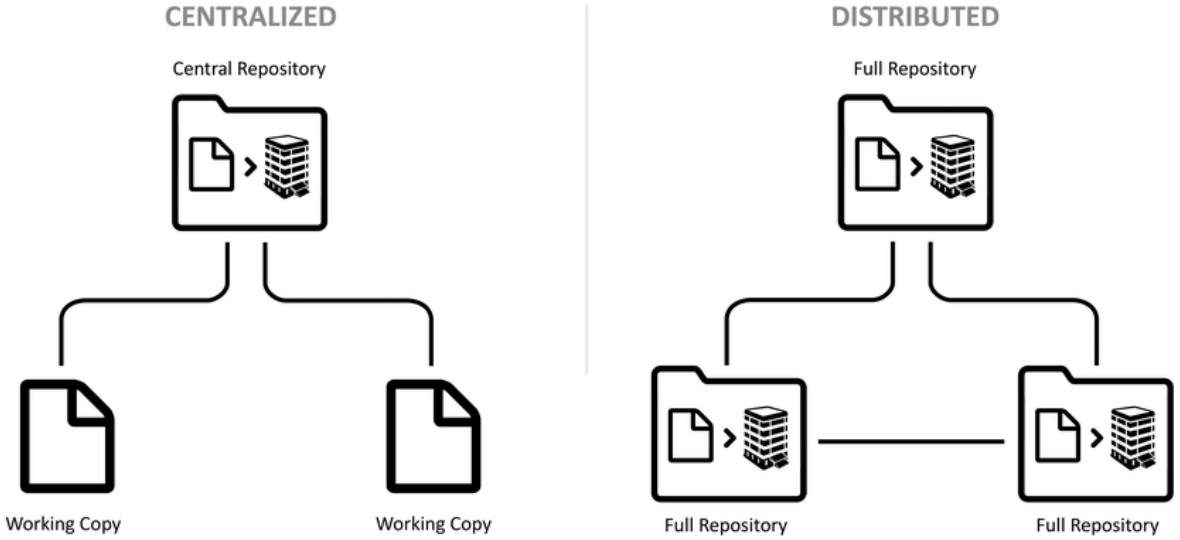
- Others include Bpipe, Make, Cromwell , WDL …

2021, PeerJ.; 9: e10897, 2018, J Mol Diagn. 20(1):4,

Version control programs help track changes to a project

- System to record changes to files over time, recall specific

- Large projects need to:
  - revert back selected files / entire project
  - Compare changes over time
  - Identify the contributor / last modification



CVS, Subversion, and Perforce

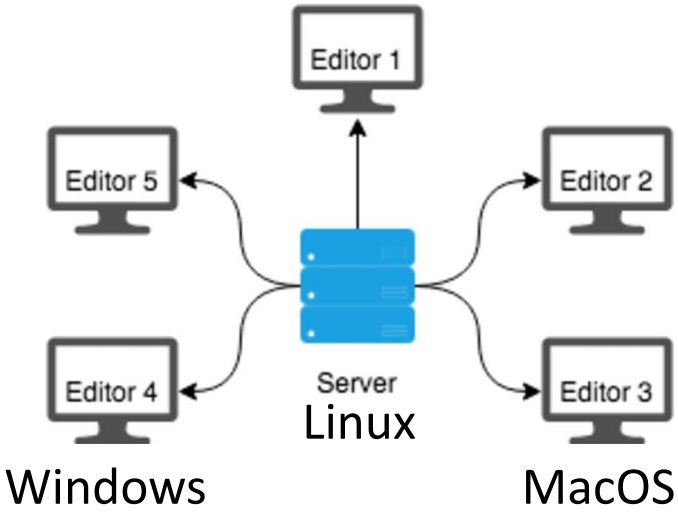
git , Mercurial and Darcs

- GitHub is an online platform that uses Git

https://git-scm.com/book/en/v2/Getting-Started-About-Version-Control, 2017, Collaborative Algorithmic-based Building Information Modelling

Accessing linux servers from Windows / MacOS

- Analysis on Linux servers: faster processing, more memory, and parallel computation
- Obtain a username , hostname and password from sysadmin
- MacOS : unix based, open a terminal, type “ssh username@hostname” => password
- Linux: same as MacOS
- Windows: PuTTY, Cygwin, **MobaXterm**. Freely available, provide a terminal and GUI



https://www.owc.com/blog/how-do-servers-communicate-with-other-computers

Basic shell scripting

- Step 1: Create a file on a text editor (Notepad, vi)

```
#!/usr/bin/bash
# This is a demo script

ls -l   # list the files
whoami  # username
~
~
~
```

- Specify the shell on the 1st line
- Text preceded by # are comments, except on the first line when followed by a !
- Type the description of the script
- Enter the commands (executed sequentially)
- Make the script as executable

```
$ chmod +x source.sh
```

- Execute the script

```
$ ./source.sh
```

https://www.owc.com/blog/how-do-servers-communicate-with-other-computers

Software installation on linux systems

- Package manager: CLI or graphical tool used to automate installation/ upgrade/ removal of software packages

- Connects to online repositories, detection of dependencies and their resolution

1. APT (Advanced Package Tool): CLI based, default package manager for Ubuntu
   Eg: `$ sudo apt-get install package-name`

2. Conda : Command line tool for package & *environment management* (Windows, macOS, and Linux), Anaconda (GUI). Eg: `$ conda install bioconda::trimmomatic`

3. Yum, Pacman, dpkg …

4. Installation from source code: download the source package, follow the instructions, access the executable

https://www.geeksforgeeks.org/understanding-package-managers-and-systemctl/, https://docs.conda.io/projects/conda/en/latest/user-guide/getting-started.html

Solutions to frequently encountered problems

- Details about usage of any command: `$ man ls` => manual pages
- Software not found : verify installation, `$ sudo apt-get install package-name`
  Locate the binary of the software, adjust PATH variable
- System slows down : check the running processes `$ top`
  terminate processes : `$ kill <process id>`
- Permission denied : lack of authorization , check permission `$ ls -l`
  Change permissions `$ chown`
- File not found : Check the file path , search a file using `$ find`

Blum, R., & Bresnahan, C. (2015). *Linux command line and scripting bible* (3rd ed.). Wiley.

Basic linux commands

- **ls**: List Files
  Shows files and directories in the current directory.
  Eg: ls -l for detailed listing

- **cd** - Change Directory. Navigate between directories.
  Eg: cd /home/user/Documents

- **pwd** - Print Working Directory. Displays the current directory path.
  Eg: /home/user

- **cp** - Copy Files. Copies files from one location to another.
  Eg: cp file1.txt /home/user/

- **mv** - Move/Rename Files. Moves or renames files & directories.
  Eg: mv oldname.txt newname.txt

- **rm** - Remove Files. Deletes files or directories.
  Eg: rm file1.txt

- Please find the cheat sheet here : GitHub => MolHemat => Workshop_2024

Summary

- Linux is a versatile operating system, suited for bioinformatics
- In order to use linux efficiently: understand its structure
- Workflow automation tools => reproducible results, scalable analysis & resource management
- Bash scripting is a simple yet powerful method to parse output

Acknowledgement

- Hematopathology department, ACTREC

- Organisers

Thank you !!