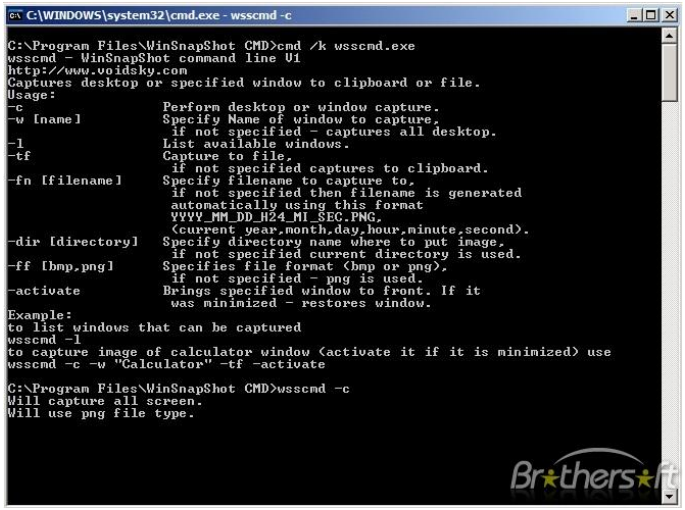
01. The shell

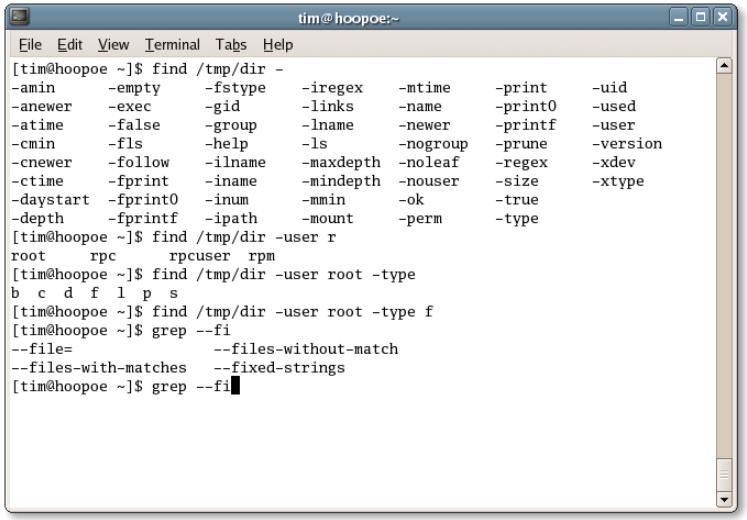
Erik Dassi & Davide Bressan

Genomics Technologies Lab

The Shell

- A command-line interpreter, the historical user interface to computer systems, also called *terminal*
- Allows to perform actions by inserting text commands
- Many types available in Linux (sh, Bash, csh, etc)





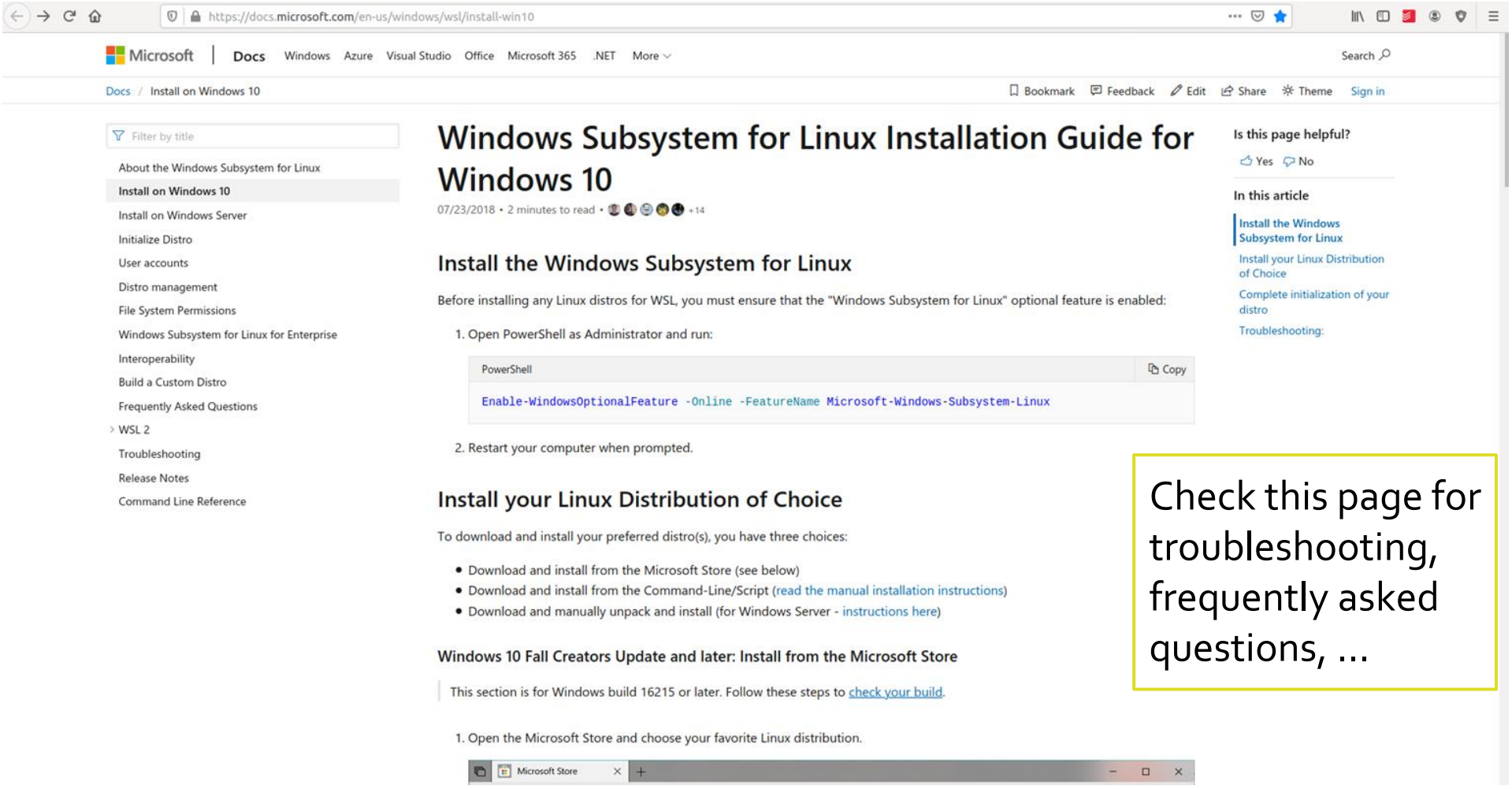
What can I do with a shell ?

- List directories and files
- Move, rename, delete and view files
- Launch any kind of software, even the ones having a graphical interface
- View various information on the status of your system
- Perform complex text manipulations (with the help of some specific software)
- … almost anything that might come to your mind (*no coffee*, though….)

> Run a set of commands as a script: a *pipeline* <

Install WSL for Windows

https://docs.microsoft.com/en-us/windows/wsl/install-win10



Check this page for troubleshooting, frequently asked questions, ...

Install WSL for Windows

Filter by title

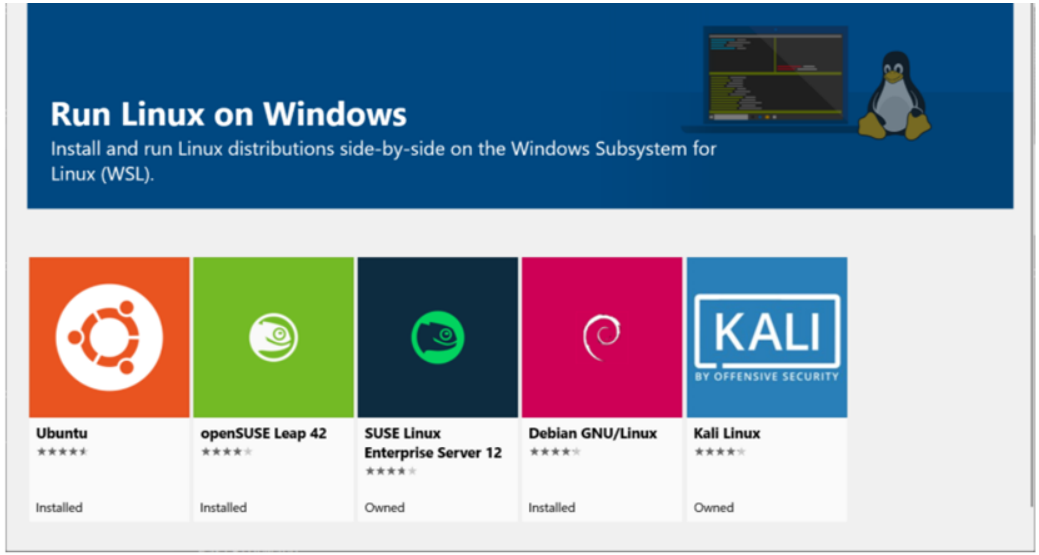
WSL Documentation
- Overview
- Quickstart
  - Install WSL & update to WSL 2
  - Install on Windows Server
  - Create a user account & password
- Tutorials
- How-to

Frequently Asked Questions

WSL 2 FAQ

Troubleshooting

- Release Notes



The following links will open the Microsoft store page for each distribution:

- Ubuntu 16.04 LTS
- Ubuntu 18.04 LTS
- Ubuntu 20.04 LTS
- openSUSE Leap 15.1
- SUSE Linux Enterprise Server 12 SP5
- SUSE Linux Enterprise Server 15 SP1
- Kali Linux
- Debian GNU/Linux
- Fedora Remix for WSL

Is this page helpful?

Yes No

In this article

Simplified Installation for Windows Insiders

Manual Installation Steps

Step 1 - Enable the Windows Subsystem for Linux

Step 2 - Check requirements for running WSL 2

Step 3 - Enable Virtual Machine feature

Step 4 - Download the Linux kernel update package

Step 5 - Set WSL 2 as your default version

Step 6 - Install your Linux distribution of choice

Install Windows Terminal (optional)

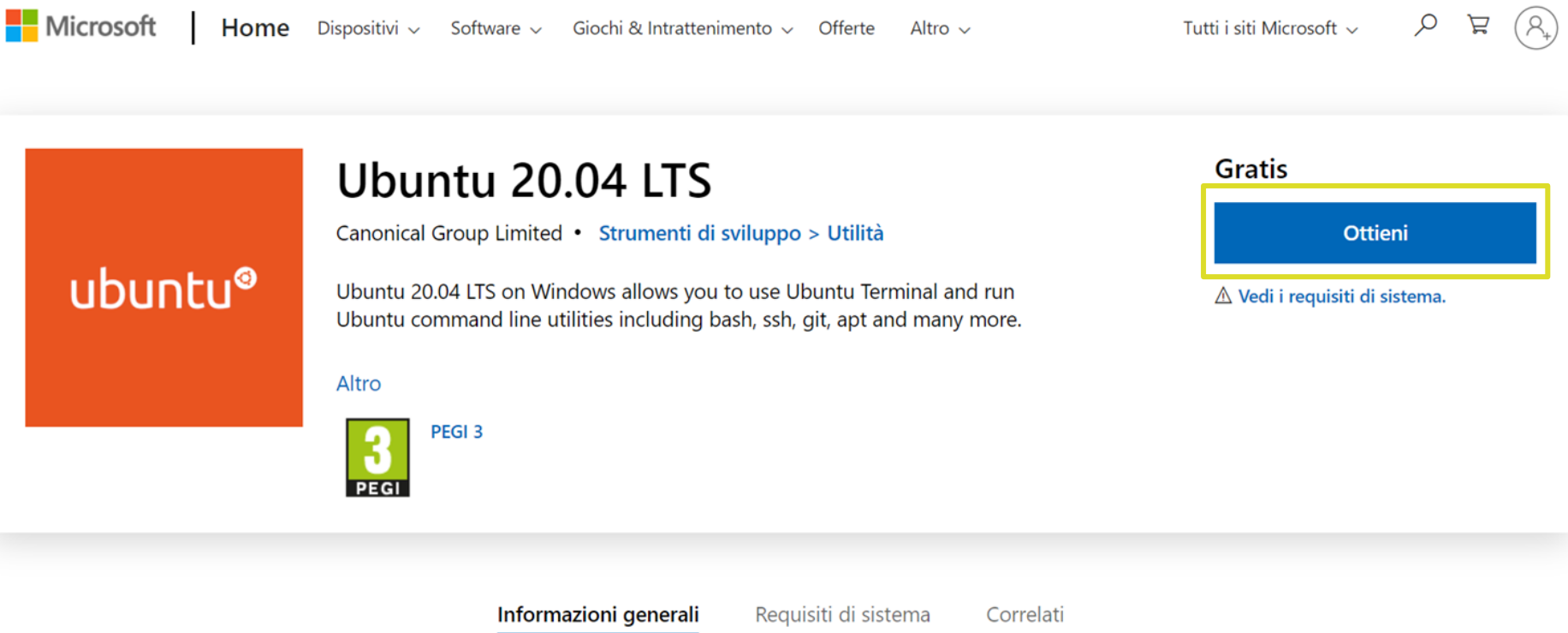
Set your distribution version to WSL 1 or WSL 2

Troubleshooting installation

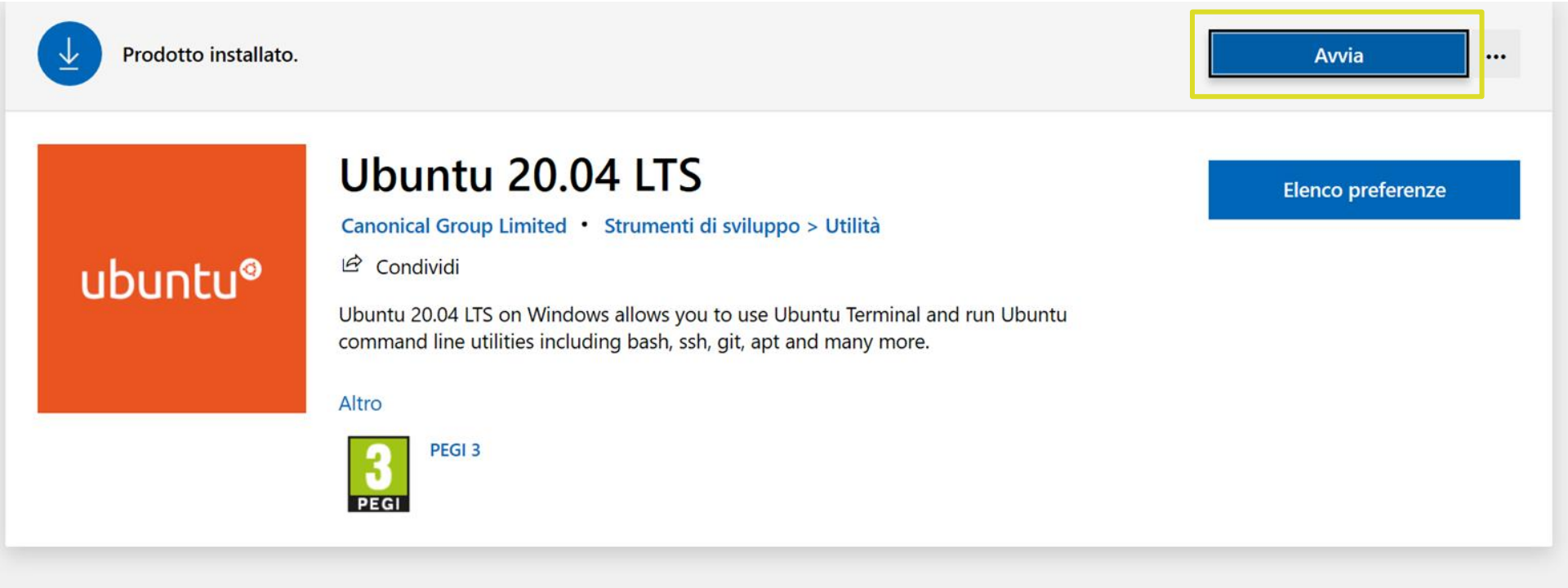
UNIVERSITÀ DI TRENTO

CIBIO

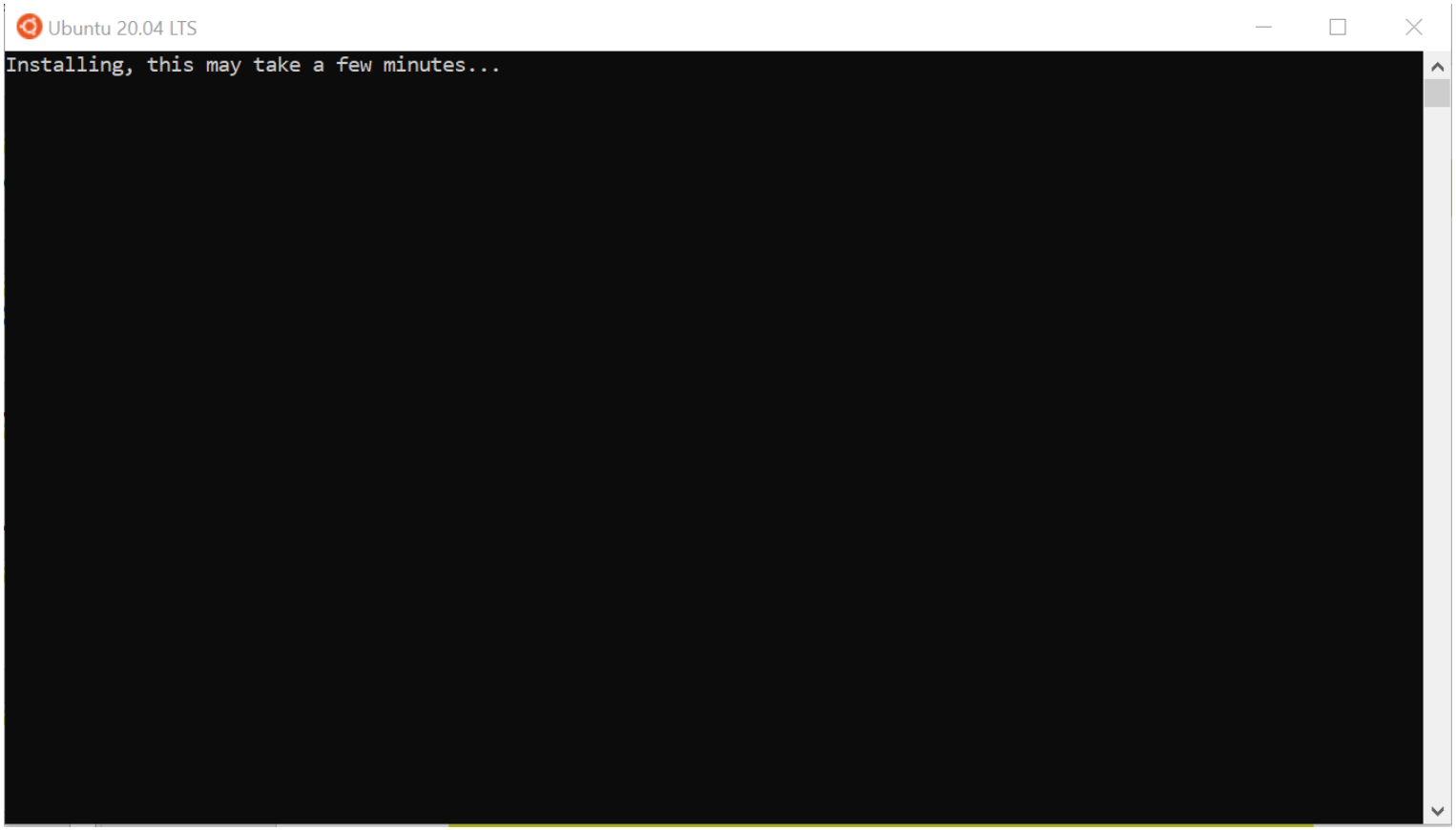
Install WSL for Windows



Install WSL for Windows

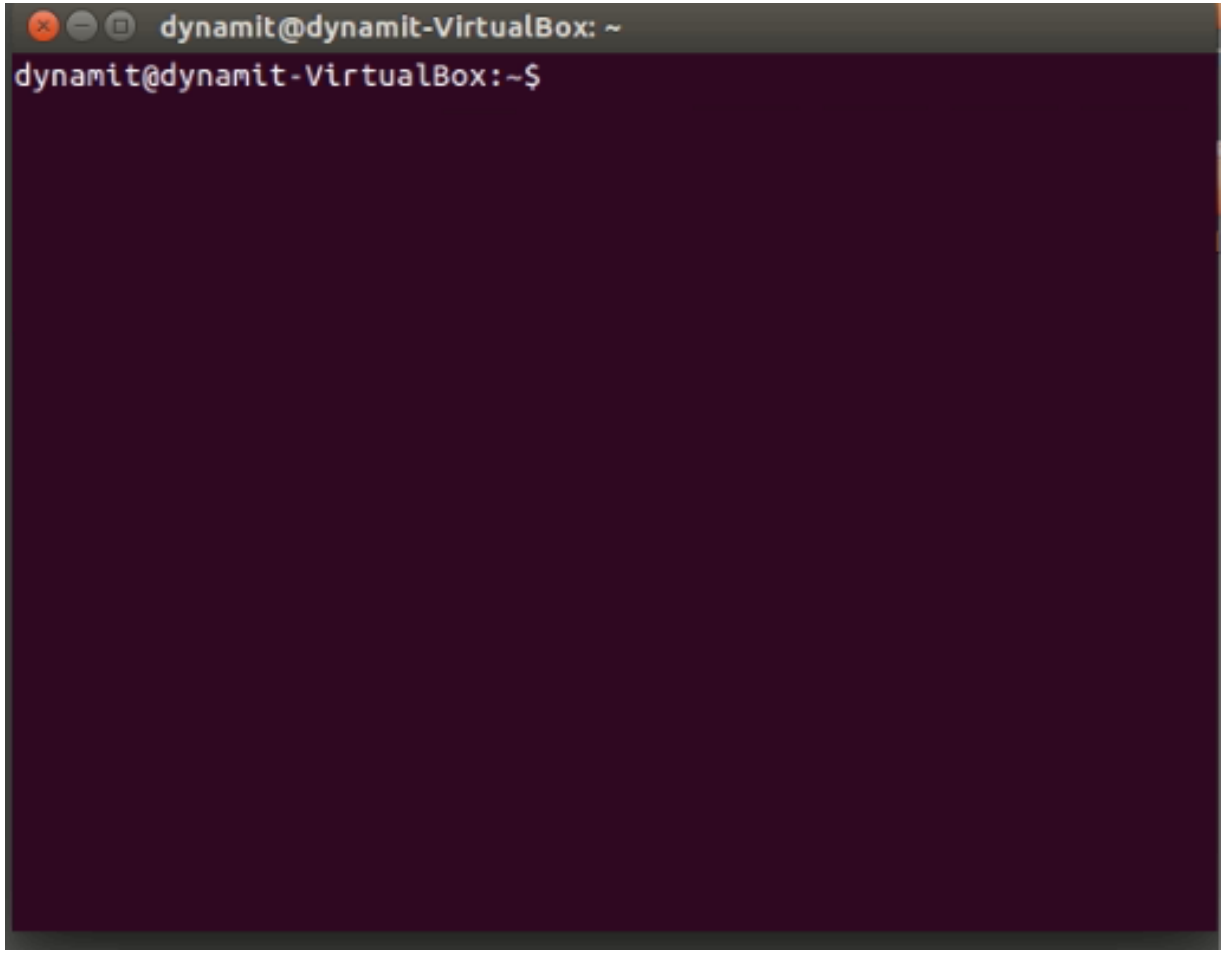


Install WSL for Windows



Choose a username and password

Linux



Code conventions

- Text in red surrounded by < and > signs, e.g.:

  echo <text>

- Means that you have to substitute that with your text/command/folder name, e.g:

  echo «Genomics Technologies ‘23-’24»

Hello, World!

- Print something to the shell with the command:

  echo <text>

  echo Hello World

- Print current time and date

  echo $(date)

- Print some text to a file (do it!)

  echo <the text you want> > output.txt

Windows WSL: similar to C:\Users\<user>\AppData\Local\Packages\CanonicalGroupLimite.Ubuntu20.04onWindow_79rhkp1fndgsc\LocalState\rootfs\home\<user>\output.txt

macOS: /Users/<user>/output.txt

Linux: /home/<user>/output.txt

Let's have a look around

- Check the current location:

  pwd *print name of current/working directory*

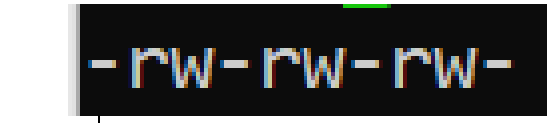
- List directories and files in the current location:

  ls *lists everything that is in current directory*

  ls –lh *lists files details and human-readable sizes*

  ls –lh *.txt *list only files with txt extension*

File permissions:



- = file
d = directory

r = read
w = write
x = execute
- = no permission

Let's have a look around

- Move to another directory:

cd <DirectoryName>	*move to that directory*
cd ..	*move up one level*
cd	*move to the home directory*

- Create a new directory:

mkdir <DirectoryName>	*create directory*
mkdir mywork	*create that directory in home directory*
mkdir mywork/GTExercises	*create that directory in Desktop*

Moving stuff around

- Copy a file to another folder:

  cp <source file> <destination file>

  cp output.txt mywork/ GTExercises /output2.txt *copy to our directory in Desktop*

- Move a file to another folder:

  mv <source file> <destination file>

  mv output.txt mywork/GTExercises/output.txt *moves the file to our directory*

  mv mywork/GTExercises/output.txt . *moves back the file*

  mv output.txt output3.txt *renames the file*

Help!

- Display the user manual of the command:

  man <command>

  man less

- Display a usage message:

  command --help

  less --help

  head --help

  tail --help

Viewing stuff

- Display a file contents (one screen at a time):

  less <filename>

  less output3.txt

  man less > less.txt

  less less.txt

- Display the beginning or the end of a file:

  head <filename> **tail** <filename>

  head less.txt

  head –n 200 less.txt

  tail –n 200 less.txt

Pipes

- Pipes allow to forward the output of a command to the input of the next one, thus creating a chain or pipe

- A pipe is created by «**|**», multiple pipes can be chained in the same command

- Allow to spare commands and be more efficient

- Basic examples:

  man less | **head**

Try more!

- Download the **mygenelist.txt** file from moodle

- Copy the file in the mywork/GTExercises directory

- Move to that directory

  cd mywork/GTExercises

Sorting, counting

sort mygenelist.txt — *sort lines of text files*

wc mygenelist.txt — *print number of lines, words and character for each file*
wc –l mygenelist.txt
wc –w mygenelist.txt
wc –c mygenelist.txt

uniq mygenelist.txt — *report or omit repeated lines*

Cat, cut & paste

cat mygenelist.txt — *create, view and concatenate files and print output*

cat output2.txt mygenelist.txt

cut — *print selected parts of lines*

cut -f 1 mygenelist.txt — *print column 1 of mygenelist.txt*

paste output2.txt mygenelist.txt — *merge lines of files*

Text manipulation tools

- **grep** is a command looking for lines matching a pattern in a file

  i.e. All lines containing «*motorcycle*»
  i.e. All lines **not** containing «*wasp*»

- can use simple patterns (i.e. a word) but also regular expressions (i.e. all words beginning with *co* and ending with *er* - computer, controller, etc.)

- A basic example:

  man more > more_manual.txt

  grep less more_manual.txt

Grep examples

- Select non-matching lines:

  grep –v less more_manual.txt

- Reading patterns from a file:

  grep –f patterns.txt more_manual.txt

Grep examples / 2

- Counting lines corresponding to matches:

  grep –c less more_manual.txt

- Ignoring case (**C = c**, while normally **C <> c**):

  grep LiNe more_manual.txt
  grep –i LiNe more_manual.txt

- Matching whole words only:

  grep in more_manual.txt
  grep –w in more_manual.txt

Grep exercise

- Search in the **awk** program manual for lines including the program name «*AWK*» and also the word «*language*»

- Save the matching lines to a file

- Count the number of matching lines

- Try to use pipes!

A solution

- Pipe-extreme solution:

man awk | **grep** AWK | **grep** language > exercise.txt
man awk | **grep** AWK | **grep -c** language

A solution

- Pipe-extreme solution:

  man awk | **grep** AWK | **grep** language > output.txt
  man awk | **grep** AWK | **grep -c** language

- Simpler solutions not using pipes can work and are ok as well, although they are less efficient:

  man awk > awk_manual.txt
  grep AWK awk_manual.txt > tmp.txt
  grep –c language tmp.txt
  grep language tmp.txt > exercise.txt

GTF file format

The GTF (General Transfer Format) / GFF (General Feature Format) is a text file format, tab-separated and consists of one line per feature, each containing 9 columns of data, plus optional track definition lines, and "empty" columns should be denoted with a '.'

1. **seqname** - name of the chromosome or scaffold; chromosome names can be given with or without the 'chr' prefix.
2. **source** - name of the program that generated this feature, or the data source (database or project name)
3. **feature** - feature type name, e.g. Gene, Variation, Similarity
4. **start** - Start position of the feature, with sequence numbering starting at 1.
5. **end** - End position of the feature, with sequence numbering starting at 1.
6. **score** - A floating point value.
7. **strand** - defined as + (forward) or - (reverse).
8. **frame** - One of '0', '1' or '2'. '0' indicates that the first base of the feature is the first base of a codon, '1' that the second base is the first base of a codon, and so on..
9. **attribute** - A semicolon-separated list of tag-value pairs, providing additional information about each feature.

GTF file example

```
chr1    HAVANA  gene    11869   14409   .       +       .       gene_id "ENSG00000223972.5"; gene_type "transcribed_unprocessed_pseudogene"; gene_name "DDX11L1"; level 2; hgnc_id "HGNC:
chr1    HAVANA  transcript      11869   14409   .       +       .       gene_id "ENSG00000223972.5"; transcript_id "ENST00000456328.2"; gene_type "transcribed_unprocessed_pseudogene"; g
chr1    HAVANA  exon    11869   12227   .       +       .       gene_id "ENSG00000223972.5"; transcript_id "ENST00000456328.2"; gene_type "transcribed_unprocessed_pseudogene"; gene_name
chr1    HAVANA  transcript      12010   13670   .       +       .       gene_id "ENSG00000223972.5"; transcript_id "ENST00000450305.2"; gene_type "transcribed_unprocessed_pseudogene"; g
chr1    HAVANA  exon    12010   12057   .       +       .       gene_id "ENSG00000223972.5"; transcript_id "ENST00000450305.2"; gene_type "transcribed_unprocessed_pseudogene"; gene_name
chr1    HAVANA  exon    12179   12227   .       +       .       gene_id "ENSG00000223972.5"; transcript_id "ENST00000450305.2"; gene_type "transcribed_unprocessed_pseudogene"; gene_name
chr1    HAVANA  exon    12613   12721   .       +       .       gene_id "ENSG00000223972.5"; transcript_id "ENST00000456328.2"; gene_type "transcribed_unprocessed_pseudogene"; gene_name
chr1    HAVANA  exon    12613   12697   .       +       .       gene_id "ENSG00000223972.5"; transcript_id "ENST00000450305.2"; gene_type "transcribed_unprocessed_pseudogene"; gene_name
chr1    HAVANA  exon    12975   13052   .       +       .       gene_id "ENSG00000223972.5"; transcript_id "ENST00000450305.2"; gene_type "transcribed_unprocessed_pseudogene"; gene_name
chr1    HAVANA  exon    13221   14409   .       +       .       gene_id "ENSG00000223972.5"; transcript_id "ENST00000456328.2"; gene_type "transcribed_unprocessed_pseudogene"; gene_name
chr1    HAVANA  exon    13221   13374   .       +       .       gene_id "ENSG00000223972.5"; transcript_id "ENST00000450305.2"; gene_type "transcribed_unprocessed_pseudogene"; gene_name
chr1    HAVANA  exon    13453   13670   .       +       .       gene_id "ENSG00000223972.5"; transcript_id "ENST00000450305.2"; gene_type "transcribed_unprocessed_pseudogene"; gene_name
chr1    HAVANA  gene    14404   29570   .       -       .       gene_id "ENSG00000227232.5"; gene_type "unprocessed_pseudogene"; gene_name "WASH7P"; level 2; hgnc_id "HGNC:38034"; havan
chr1    HAVANA  transcript      14404   29570   .       -       .       gene_id "ENSG00000227232.5"; transcript_id "ENST00000488147.1"; gene_type "unprocessed_pseudogene"; gene_name "WA
chr1    HAVANA  exon    14404   14501   .       -       .       gene_id "ENSG00000227232.5"; transcript_id "ENST00000488147.1"; gene_type "unprocessed_pseudogene"; gene_name "WASH7P"; t
chr1    HAVANA  exon    15005   15038   .       -       .       gene_id "ENSG00000227232.5"; transcript_id "ENST00000488147.1"; gene_type "unprocessed_pseudogene"; gene_name "WASH7P"; t
chr1    HAVANA  exon    15796   15947   .       -       .       gene_id "ENSG00000227232.5"; transcript_id "ENST00000488147.1"; gene_type "unprocessed_pseudogene"; gene_name "WASH7P"; t
chr1    HAVANA  exon    16607   16765   .       -       .       gene_id "ENSG00000227232.5"; transcript_id "ENST00000488147.1"; gene_type "unprocessed_pseudogene"; gene_name "WASH7P"; t
chr1    HAVANA  exon    16858   17055   .       -       .       gene_id "ENSG00000227232.5"; transcript_id "ENST00000488147.1"; gene_type "unprocessed_pseudogene"; gene_name "WASH7P"; t
chr1    HAVANA  exon    17233   17368   .       -       .       gene_id "ENSG00000227232.5"; transcript_id "ENST00000488147.1"; gene_type "unprocessed_pseudogene"; gene_name "WASH7P"; t
chr1    ENSEMBL gene    17369   17436   .       -       .       gene_id "ENSG00000278267.1"; gene_type "miRNA"; gene_name "MIR6859-1"; level 3; hgnc_id "HGNC:50039";
chr1    ENSEMBL transcript      17369   17436   .       -       .       gene_id "ENSG00000278267.1"; transcript_id "ENST00000619216.1"; gene_type "miRNA"; gene_name "MIR6859-1"; transcr
chr1    ENSEMBL exon    17369   17436   .       -       .       gene_id "ENSG00000278267.1"; transcript_id "ENST00000619216.1"; gene_type "miRNA"; gene_name "MIR6859-1"; transcript_type
chr1    HAVANA  exon    17606   17742   .       -       .       gene_id "ENSG00000227232.5"; transcript_id "ENST00000488147.1"; gene_type "unprocessed_pseudogene"; gene_name "WASH7P"; t
chr1    HAVANA  exon    17915   18061   .       -       .       gene_id "ENSG00000227232.5"; transcript_id "ENST00000488147.1"; gene_type "unprocessed_pseudogene"; gene_name "WASH7P"; t
chr1    HAVANA  exon    18268   18366   .       -       .       gene_id "ENSG00000227232.5"; transcript_id "ENST00000488147.1"; gene_type "unprocessed_pseudogene"; gene_name "WASH7P"; t
chr1    HAVANA  exon    24738   24891   .       -       .       gene_id "ENSG00000227232.5"; transcript_id "ENST00000488147.1"; gene_type "unprocessed_pseudogene"; gene_name "WASH7P"; t
chr1    HAVANA  exon    29534   29570   .       -       .       gene_id "ENSG00000227232.5"; transcript_id "ENST00000488147.1"; gene_type "unprocessed_pseudogene"; gene_name "WASH7P"; t
chr1    HAVANA  gene    29554   31109   .       +       .       gene_id "ENSG00000243485.5"; gene_type "lncRNA"; gene_name "MIR1302-2HG"; level 2; hgnc_id "HGNC:52482"; tag "ncRNA_host"
chr1    HAVANA  transcript      29554   31097   .       +       .       gene_id "ENSG00000243485.5"; transcript_id "ENST00000473358.1"; gene_type "lncRNA"; gene_name "MIR1302-2HG"; tran
chr1    HAVANA  exon    29554   30039   .       +       .       gene_id "ENSG00000243485.5"; transcript_id "ENST00000473358.1"; gene_type "lncRNA"; gene_name "MIR1302-2HG"; transcript_t
chr1    HAVANA  transcript      30267   31109   .       +       .       gene_id "ENSG00000243485.5"; transcript_id "ENST00000469289.1"; gene_type "lncRNA"; gene_name "MIR1302-2HG"; tran
chr1    HAVANA  exon    30267   30667   .       +       .       gene_id "ENSG00000243485.5"; transcript_id "ENST00000469289.1"; gene_type "lncRNA"; gene_name "MIR1302-2HG"; transcript_t
chr1    ENSEMBL gene    30366   30503   .       +       .       gene_id "ENSG00000284332.1"; gene_type "miRNA"; gene_name "MIR1302-2"; level 3; hgnc_id "HGNC:35294";
chr1    ENSEMBL transcript      30366   30503   .       +       .       gene_id "ENSG00000284332.1"; transcript_id "ENST00000607096.1"; gene_type "miRNA"; gene_name "MIR1302-2"; transcr
chr1    ENSEMBL exon    30366   30503   .       +       .       gene_id "ENSG00000284332.1"; transcript_id "ENST00000607096.1"; gene_type "miRNA"; gene_name "MIR1302-2"; transcript_type
chr1    HAVANA  exon    30564   30667   .       +       .       gene_id "ENSG00000243485.5"; transcript_id "ENST00000473358.1"; gene_type "lncRNA"; gene_name "MIR1302-2HG"; transcript_t
chr1    HAVANA  exon    30976   31097   .       +       .       gene_id "ENSG00000243485.5"; transcript_id "ENST00000473358.1"; gene_type "lncRNA"; gene_name "MIR1302-2HG"; transcript_t
chr1    HAVANA  exon    30976   31109   .       +       .       gene_id "ENSG00000243485.5"; transcript_id "ENST00000469289.1"; gene_type "lncRNA"; gene_name "MIR1302-2HG"; transcript_t
chr1    HAVANA  gene    34554   36081   .       -       .       gene_id "ENSG00000237613.2"; gene_type "lncRNA"; gene_name "FAM138A"; level 2; hgnc_id "HGNC:32334"; havana_gene "OTTHUMG
chr1    HAVANA  transcript      34554   36081   .       -       .       gene_id "ENSG00000237613.2"; transcript_id "ENST00000417324.1"; gene_type "lncRNA"; gene_name "FAM138A"; transcri
chr1    HAVANA  exon    34554   35174   .       -       .       gene_id "ENSG00000237613.2"; transcript_id "ENST00000417324.1"; gene_type "lncRNA"; gene_name "FAM138A"; transcript_type
chr1    HAVANA  transcript      35245   36073   .       -       .       gene_id "ENSG00000237613.2"; transcript_id "ENST00000461467.1"; gene_type "lncRNA"; gene_name "FAM138A"; transcri
chr1    HAVANA  exon    35245   35481   .       -       .       gene_id "ENSG00000237613.2"; transcript_id "ENST00000461467.1"; gene_type "lncRNA"; gene_name "FAM138A"; transcript_type
chr1    HAVANA  exon    35277   35481   .       -       .       gene_id "ENSG00000237613.2"; transcript_id "ENST00000417324.1"; gene_type "lncRNA"; gene_name "FAM138A"; transcript_type
chr1    HAVANA  exon    35721   36081   .       -       .       gene_id "ENSG00000237613.2"; transcript_id "ENST00000417324.1"; gene_type "lncRNA"; gene_name "FAM138A"; transcript_type
chr1    HAVANA  exon    35721   36073   .       -       .       gene_id "ENSG00000237613.2"; transcript_id "ENST00000461467.1"; gene_type "lncRNA"; gene_name "FAM138A"; transcript_type
chr1    HAVANA  gene    52473   53312   .       +       .       gene_id "ENSG00000268020.3"; gene_type "unprocessed_pseudogene"; gene_name "OR4G4P"; level 2; hgnc_id "HGNC:14822"; havan
chr1    HAVANA  transcript      52473   53312   .       +       .       gene_id "ENSG00000268020.3"; transcript_id "ENST00000606857.1"; gene_type "unprocessed_pseudogene"; gene_name "OR
chr1    HAVANA  exon    52473   53312   .       +       .       gene_id "ENSG00000268020.3"; transcript_id "ENST00000606857.1"; gene_type "unprocessed_pseudogene"; gene_name "OR4G4P"; t
chr1    HAVANA  gene    57598   64116   .       +       .       gene_id "ENSG00000240361.2"; gene_type "transcribed_unprocessed_pseudogene"; gene_name "OR4G11P"; level 2; hgnc_id "HGNC:
```

Let's practice!

- Download the **hs_annotation.gtf** file from moodle
- Move it to your *GTExercises* directory
- Check where you are:

  pwd

- Move in your directory

  cd mywork/GTExercises

Let's practice!

1. Count how many lines are present in hs_annotation.gtf

Let's practice!

1. Count how many lines are present in hs_annotation.gtf

   wc -l hs_annotation.gtf

Let's practice!

1. Count how many lines are present in hs_annotation.gtf

   wc -l hs_annotation.gtf

2. Count how many genes are present in hs_annotation.gtf

Let's practice!

1. Count how many lines are present in hs_annotation.gtf

   wc -l hs_annotation.gtf

2. Count how many genes are present in hs_annotation.gtf

   cut -f 3 hs_annotation.gtf | **grep** –c gene
   cut -f 3 hs_annotation.gtf | **grep** gene | **wc** -l
   grep -w gene hs_annotation.gtf | **wc** -l

Let's practice!

1. Count how many lines are present in hs_annotation.gtf

   wc -l hs_annotation.gtf

2. Count how many genes are present in hs_annotation.gtf

   cut -f 3 hs_annotation.gtf | **grep** –c gene
   cut -f 3 hs_annotation.gtf | **grep** gene | **wc** -l
   grep -w gene hs_annotation.gtf | **wc** -l

3. Count how many "protein_coding" genes are present in hs_annotation.gtf

Let's practice!

1. Count how many lines are present in hs_annotation.gtf

   wc -l hs_annotation.gtf

2. Count how many genes are present in hs_annotation.gtf

   cut -f 3 hs_annotation.gtf | **grep** –c gene
   cut -f 3 hs_annotation.gtf | **grep** gene | **wc** -l
   grep -w gene hs_annotation.gtf | **wc** -l

3. Count how many "protein_coding" genes are present in hs_annotation.gtf

   grep -w gene hs_annotation.gtf | **grep** –c protein_coding
   grep -w gene hs_annotation.gtf | **grep** protein_coding| **wc** -l

   grep -w gene hs_annotation.gtf > hs_annotation_genes.gtf
   grep protein_coding hs_annotation_genes.gtf | **wc** -l

Exercises

1. Count how many genes of type “lncRNA” are present in hs_annotation.gtf, get their gene names and sort them alphabetically

2. Count how many gene types are present in hs_annotation.gtf

Questions ?

The meaning of ~~life~~ ./ («dot-slash»)

- Builtin commands VS custom commands

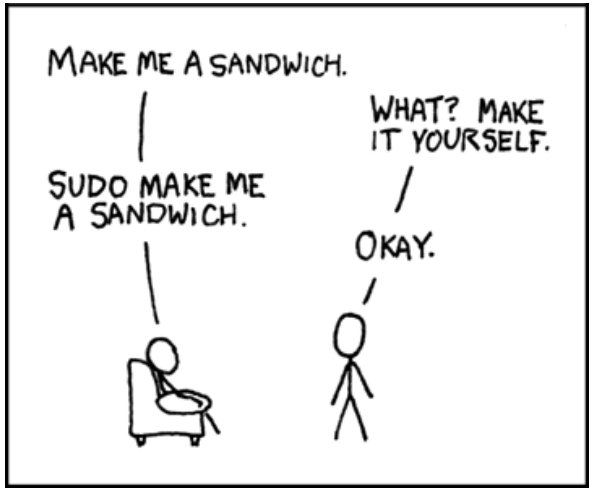
less output.txt
man less

./myscript.sh
./install.sh

. -> refers to the current directory

/ -> directory path separator

- Details here: **http://www.linfo.org/dot_slash.html**

Need Power ?

- Some (*quite a lot, indeed…*) actions can not be run by the normal user
- Needs super-user (SU) status to install specific tools or to change system settings
- If you know the SU password, you can gain temporary SU privileges (*dangerous!*)
- Just type: **sudo** «*command to execute*»