

Présentation du e-workshop

Dr A. Jamet

Médecin Microbiologiste

@Necker

Dr A. Bridier-Nahmias

Maitre de conférence

@IAME

Les bioinformaticiens vus par les biologistes

0100100010101001!



```
/dev/block/platform/sdhci-tegra.3/by-name/system          241888  197864   44624  82% /system  
/dev/block/platform/sdhci-tegra.3/by-name/userdata        30182852  1241724  28941128  4% /data  
/dev/block/platform/sdhci-tegra.3/by-name/cache           170320      4260   166060  3% /cache  
/dev/block/platform/sdhci-tegra.3/by-name/pdsb            1976      106   1870  5% /pds  
/dev/fuse          30182852  1241724  28941128  4% /mnt/sdcard  
/ $ ll  
drwxr-xr-X  14 root    root      0 Mar 18 14:14 .  
drwxr-xr-X  14 root    root      0 Mar 18 14:14 ..  
drwxr-xr-X  3 root    root      0 Mar 18 14:14 acct  
drwxrwx---  4 system  cache    4096 Mar 30 14:20 cache  
dr-x-----  2 root    root      0 Mar 18 14:14 config  
lrwxrwxrwx  1 root    root      17 Mar 18 14:14 d => /sys/kernel/debug  
drwxrwxr-x  18 system  system   4096 Mar 16 17:59 data  
-rw-r--r--  1 root    root     318 Dec 31 1969 default.prop  
drwxr-xr-X  11 root    root    2280 Mar 30 18:50 dev  
drwxrwxrwx  1 root    root      11 Mar 18 14:14 etc -> /system/etc  
-rwxr-x---  1 root    root     94344 Dec 31 1969 init  
-rwxr-x---  1 root    root     2210 Dec 31 1969 init.goldfish.rc  
-rwxr-x---  1 root    root     9299 Dec 31 1969 init.olympus.rc  
-rwxr-x---  1 root    root     15379 Dec 31 1969 init.rc  
-rwxr-x---  1 root    root     9299 Dec 31 1969 init.stingray.rc  
drwxrwxr-x  6 root    system   2048 Feb 14 12:55 mnt  
drwxr-xr-X  9 root    root      0 Mar 18 14:14 mnt  
dr-xr-xr-X  152 root   root     0 Dec 31 1969 proc  
drwxr----- 2 root    root      0 Feb  7 15:23 root  
drwxr-xr-X  2 root    root      0 Dec 31 1969 sbin  
drwxrwxrwx  1 root    root      11 Mar 18 14:14 sdcard -> /mnt/sdcard  
drwxr-xr-X  12 root   root     0 Mar 18 14:14 sys  
drwxr-xr-X  13 root   root     4096 Mar 32 13:16 system  
-rwxr--r--  1 root   root     221 Dec 31 1969 ueventd.goldfish.rc  
-rwxr--r--  1 root   root     629 Dec 31 1969 ueventd.olympus.rc  
-rwxr--r--  1 root   root     3707 Dec 31 1969 ueventd.rc  
-rwxr--r--  1 root   root     629 Dec 31 1969 ueventd.stingray.rc  
drwxrwxrwx  1 root   root     14 Mar 18 14:14 vendor -> /system/vendor  
/ $
```

A. Jamet

Les biologistes vus par les bioinformaticiens



Conseils

- Sachez être un bon biologiste expérimental en vous aidant de la bioinformatique
- Très peu de possibilité de ne faire que de la bioinformatique
- « Un vrai plus » de connaître les 2 aujourd’hui
- « Une obligation » demain...

La microbiologie clinique @Necker

- L'hôpital Necker
 - 400 lits de pédiatrie (médecine, chirurgie, réa, urgences)
 - 200 lits pour le pôle Adulte (hémato, néphro, SMIT)
- Le laboratoire
 - Centre de référence Mucoviscidose



A. Jamet



Le NGS en microbiologie @Necker

- Premiers pas avec un 454 en 2013 (Roche)
- Acquisition d'un MiniSeq en 2017 (Illumina)

2013



Anne Jamet @DrAnneJamet · 26 Sep 2017
MiniSeq is in the lab !

2017



A. Jamet



Choix du séquenceur



| Sequencing System | iSeq™ | MiniSeq™ | MiSeq® | NextSeq® | HiSeq® | HiSeq® X | NovaSeq® |
|-----------------------------------|---------|----------|--------|----------|---------------------|---------------------------------------|--------------------------|
| | | | | 4000 | Five/Ten | | 6000 |
| Output per run | 1.2 Gb | 7.5 Gb | 15 Gb | 120 Gb | 1.5 Tb | 1.8 Tb | 1 Tb - 6 Tb ¹ |
| Instrument price | \$19.9K | \$49.5K | \$99K | \$275K | \$900K | \$6M ² /\$10M ² | \$985K |
| Installed base³ | NA | ~600 | ~6,000 | ~2,400 | ~2,300 ⁴ | | ~285 |

De 20 000 euros à 1 000 000 euros

1. Output per run for the S1, S2 and S4 flow cells equal 1 Tb, 2 Tb and 6 Tb, respectively assuming two flow cells per run
2. Based on purchase of 5 and 10 units for HiSeq X Five and HiSeq X Ten, respectively
3. Based on end of fiscal year 2017
4. Combined HiSeq family

5

illumina®

Workflow séquençage

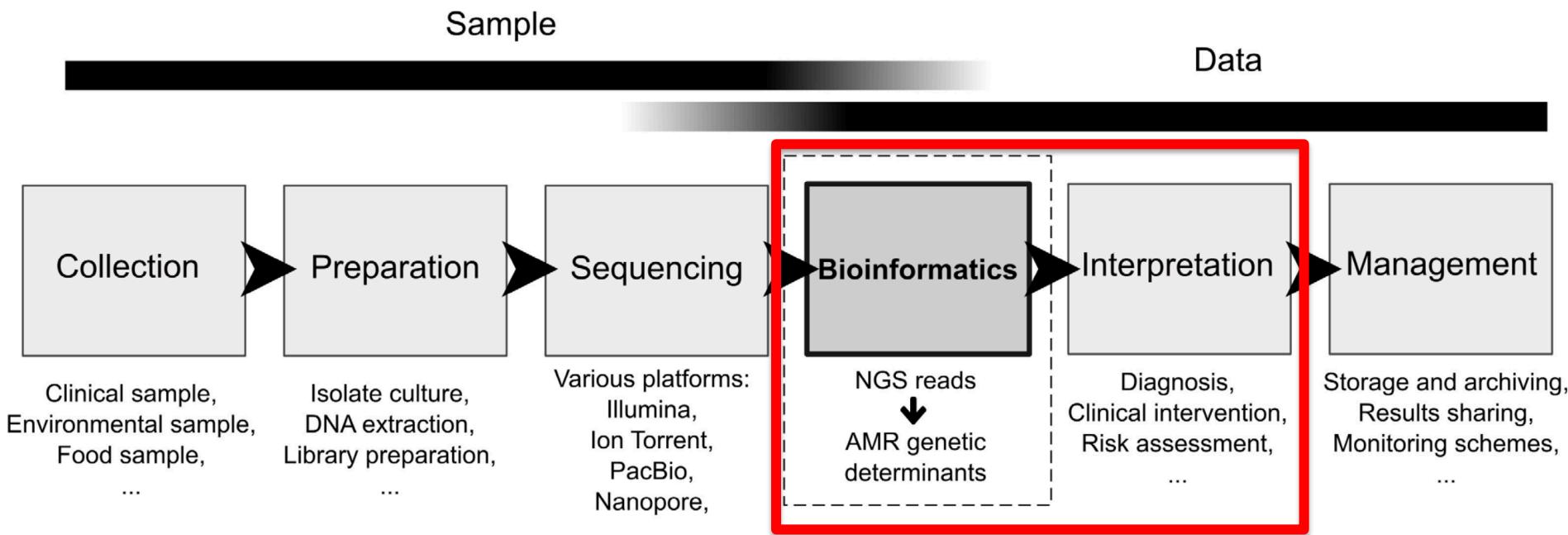


Figure 1. Overview of the different steps involved in the use of Next-Generation Sequencing technologies for the detection and monitoring of antimicrobial resistance. The benchmark strategy discussed in the current article focuses on the bioinformatics steps, the pipeline converting the output of the sequencing experiment into a list of identified antimicrobial resistance genetic determinants (dashed rectangle).

Workflow commun pré-analytique

1- Extraire les ADN

2- Préparer les librairies

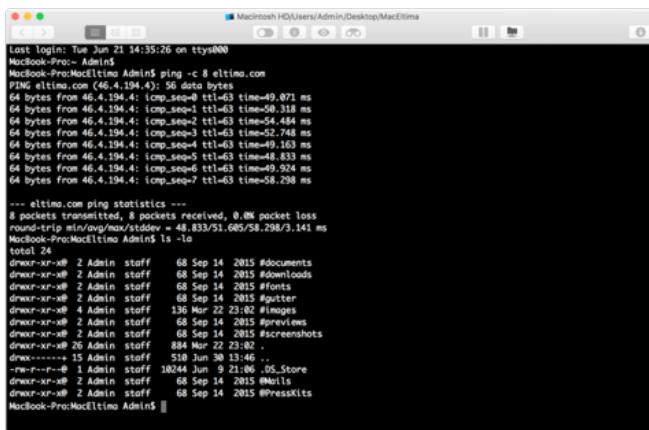
- > multiplexage d'échantillons
- > ex: *S. aureus* X3 + *M. tuberculosis* X2
- En pratique 6 à 30 souches selon taille genome

3- Séquençage sur machine

4- Récupération des **fichiers bruts “.fastq”**

Objectifs du TP de bioinformatique

A partir des données brutes issues d'un séquenceur haut débit, être capable d'évaluer la **qualité** des séquences, **d'assembler** un génome, de le **comparer** à d'autres génomes, de **déetecter des gènes d'intérêt** pour le typage, la virulence et la résistance.



```
Macintosh HD/Users/Admin/Desktop/MacElUltimo
Last Login: Tue Jun 21 14:35:26 on ttys000
MacBook-Pro:~ Admin$ MacBook-Pro:MacElUltimo Admin$ ping -c 8 eltimo.com
PING eltimo.com (46.4.194.4): 56 data bytes
64 bytes from 46.4.194.4: icmp_seq=0 ttl=63 time=49.071 ms
64 bytes from 46.4.194.4: icmp_seq=1 ttl=63 time=50.318 ms
64 bytes from 46.4.194.4: icmp_seq=2 ttl=63 time=50.318 ms
64 bytes from 46.4.194.4: icmp_seq=3 ttl=63 time=52.748 ms
64 bytes from 46.4.194.4: icmp_seq=4 ttl=63 time=49.163 ms
64 bytes from 46.4.194.4: icmp_seq=5 ttl=63 time=48.833 ms
64 bytes from 46.4.194.4: icmp_seq=6 ttl=63 time=49.924 ms
64 bytes from 46.4.194.4: icmp_seq=7 ttl=63 time=58.298 ms
...
8 packets transmitted, 8 packets received, 0.0% packet loss
round-trip min/avg/max/stddev = 48.833/51.685/54.298/3.141 ms
MacBook-Pro:MacElUltimo Admin$ ls -l
total 24
drwxr-xr-x  2 Admin  staff  68 Sep 14  2015 #documents
drwxr-xr-x  2 Admin  staff  68 Sep 14  2015 #downloads
drwxr-xr-x  2 Admin  staff  68 Sep 14  2015 #fonts
drwxr-xr-x  2 Admin  staff  68 Sep 14  2015 #gutter
drwxr-xr-x  4 Admin  staff  136 Mar 22 23:02 #images
drwxr-xr-x  2 Admin  staff  68 Sep 14  2015 #reviews
drwxr-xr-x  2 Admin  staff  68 Sep 14  2015 #screenshots
drwxr-xr-x  2 Admin  staff  68 Sep 14  2015 #scripts
drwxr-xr-x  15 Admin  staff  510 Jun  9  21:06 .05.Store
-rw-r--r--  1 Admin  staff  10244 Jun  9  21:06 .05.Store
drwxr-xr-x  2 Admin  staff  68 Sep 14  2015 #tools
drwxr-xr-x  2 Admin  staff  68 Sep 14  2015 #presskits
MacBook-Pro:MacElUltimo Admin$
```

Avec des outils en ligne...de commande



Center for Genomic Epidemiology

Galaxy PROJECT

PATRIC 3.6.11 ORGANISMS DATA WORKSPACES SERVICES HELP

BACTERIAL BIOINFORMATICS RESOURCE CENTER

Avec des outils en ligne

Vous devrez présenter dans ce compte-rendu

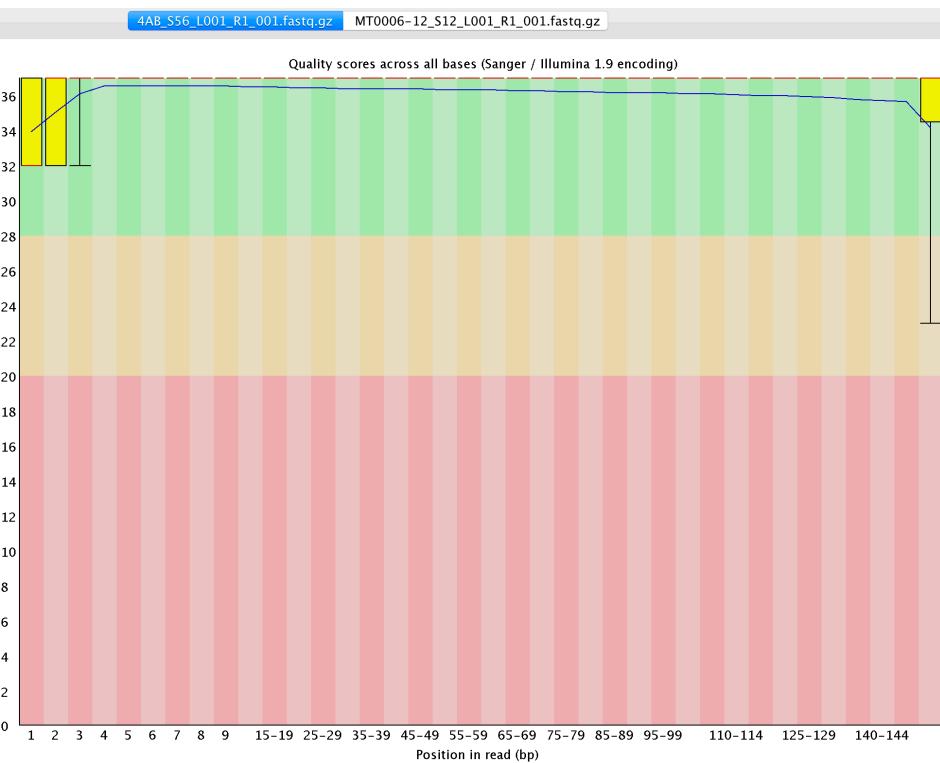
1. les méthodes utilisées ;
2. les résultats obtenus ;
3. les conclusions biologiques
(qu'a-t-on appris sur la séquence analysée ?)
4. les conclusions méthodologiques
(quelles sont les limitations de la méthode employée ?)
5. faire une synthèse globale

Faites des captures d'écran, des tableaux...

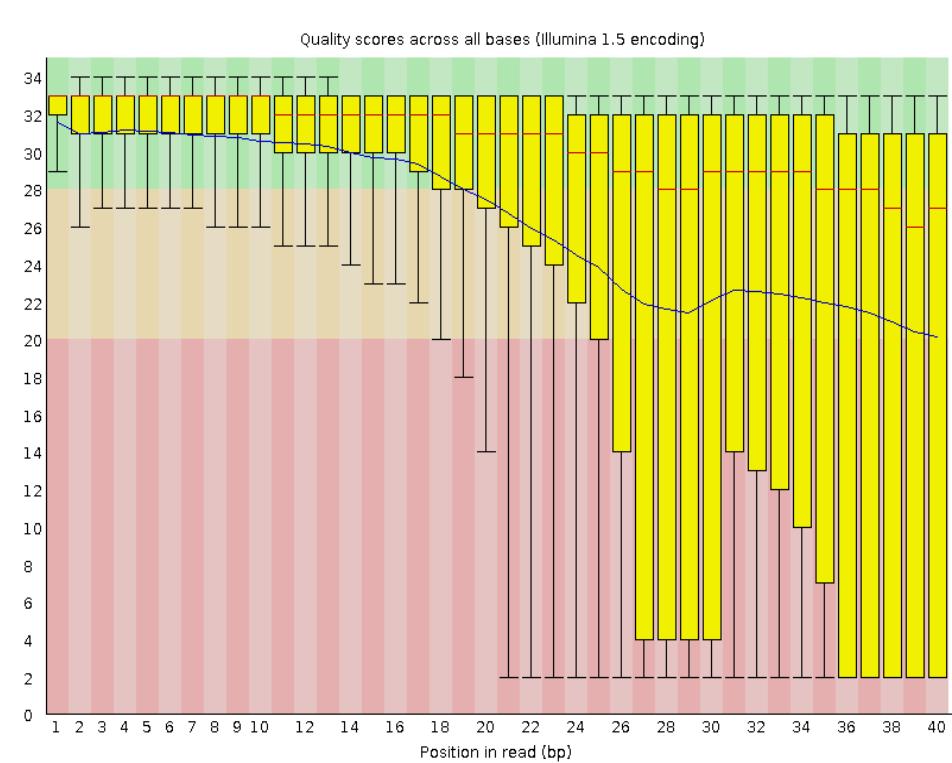
Evaluer les “reads”

FastQC

GOOD



BAD



QC & Trim

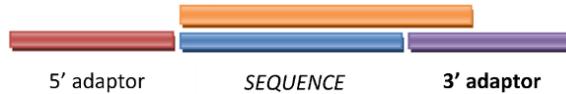
What do we want to trim?

Trimomatic

Low quality or adaptor sequence

Quality is variable and often decreases towards end of read.

Poor quality = high error rate

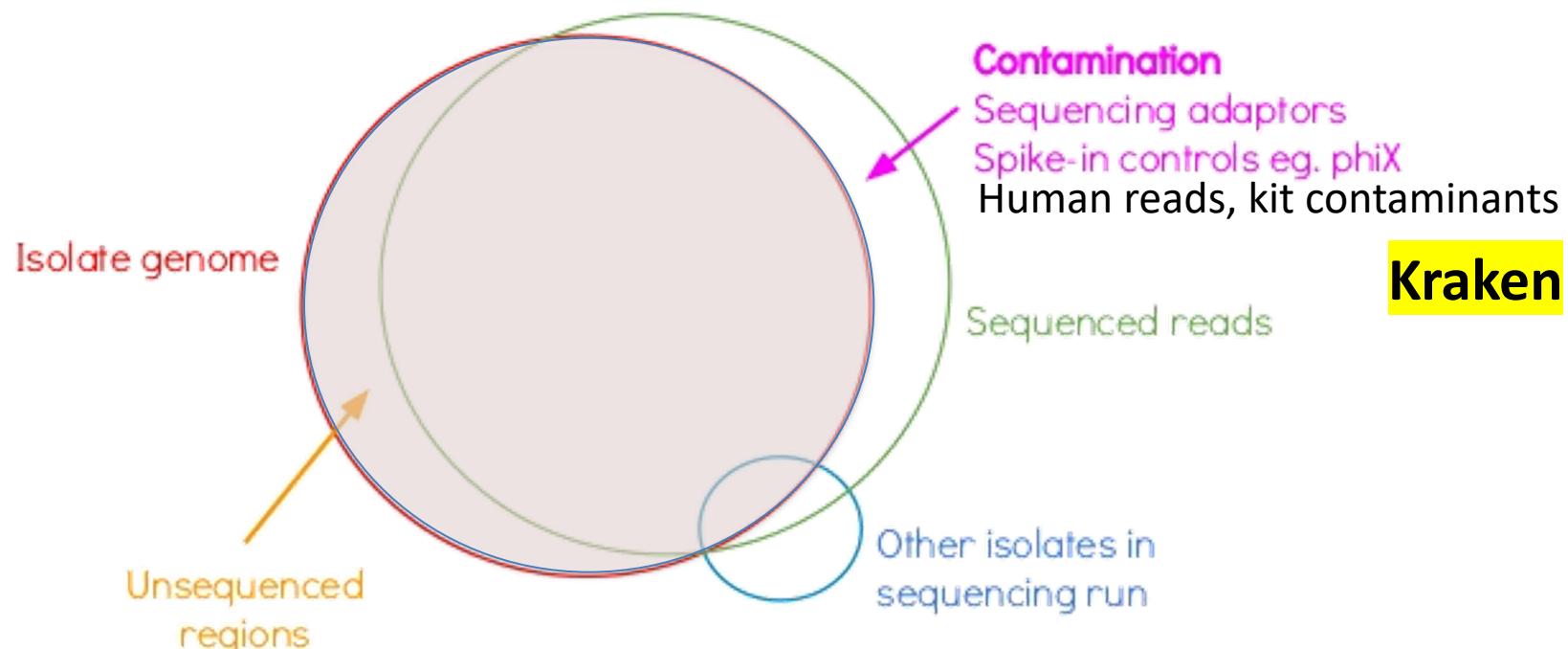


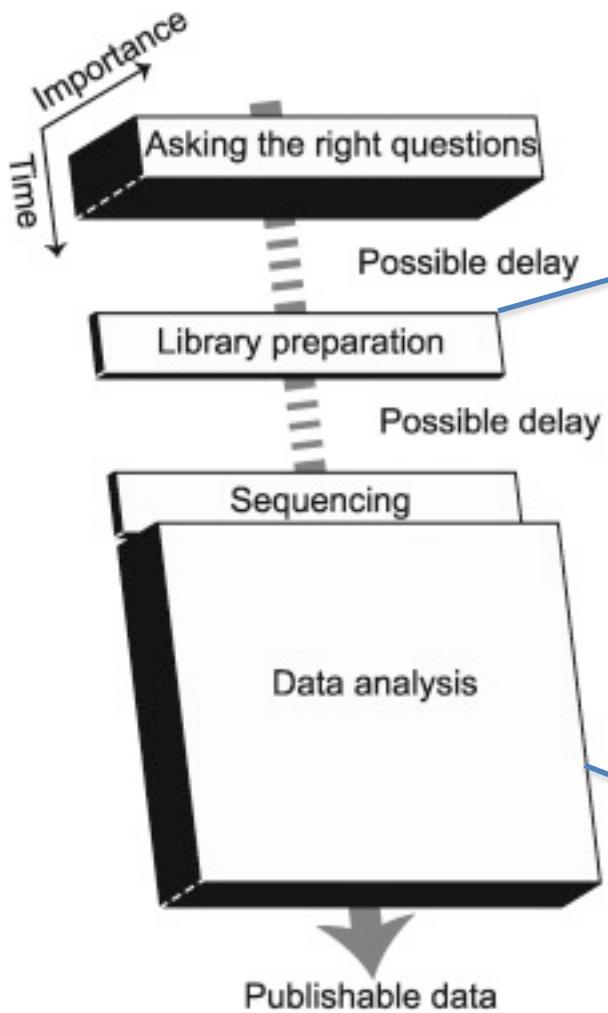
Illumina sequences typically include some of adaptor sequence

Why trim?

Low quality sequence / adaptors makes alignment harder + slows analyses down.

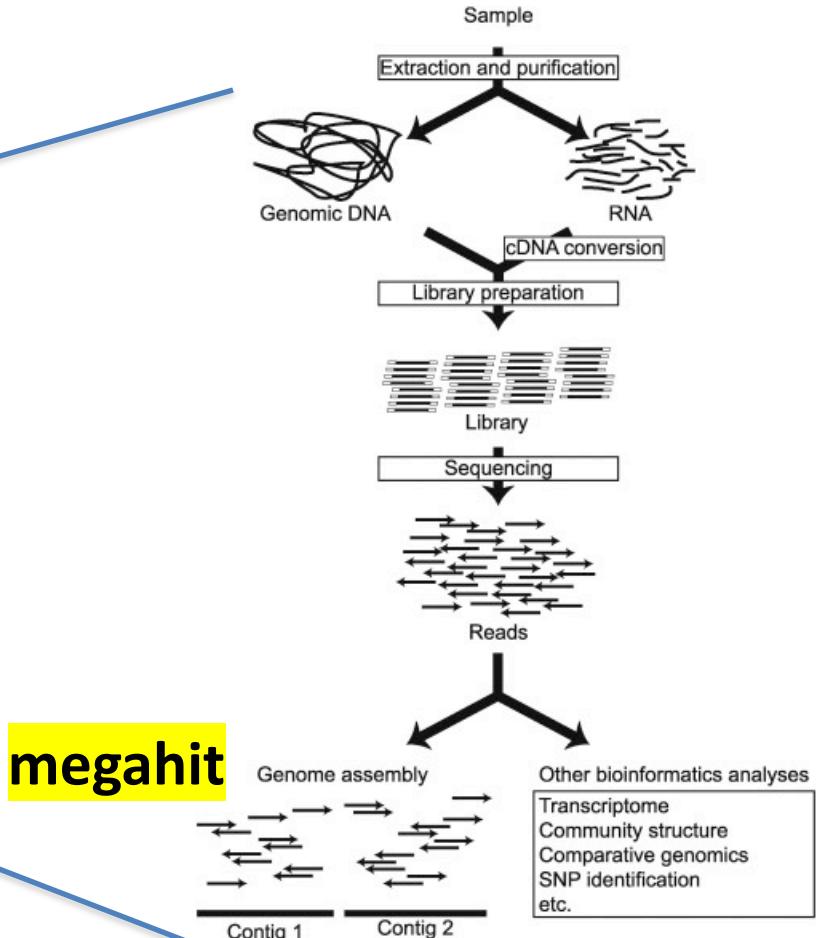
What data do we really have?





Conceptual workflow of a complete NGS based project with the relative importance and time spent for each step.

A. Jamet



10.1016/j.mimet.2016.02.016

Computer Hardware

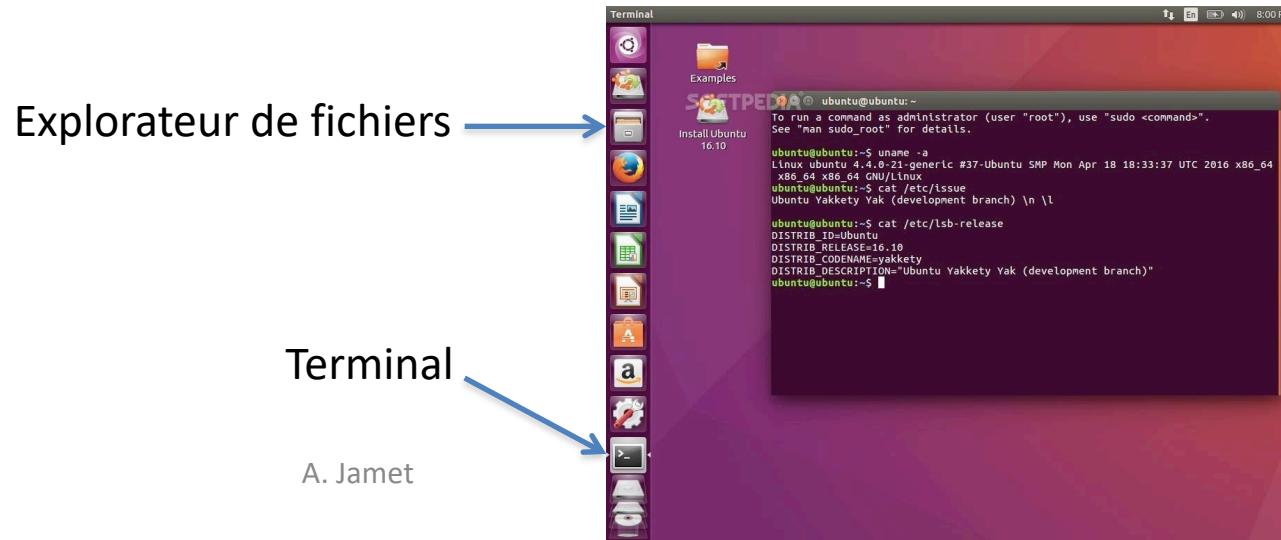
- In general, you will need:
 - 4 Gb of RAM Minimum (-> 16 Gb)
 - 500 Gb of disk space (-> 2 To)
 - better to have at least 4 cores (-> 8 CPU)

Votre “cahier de manip” électronique

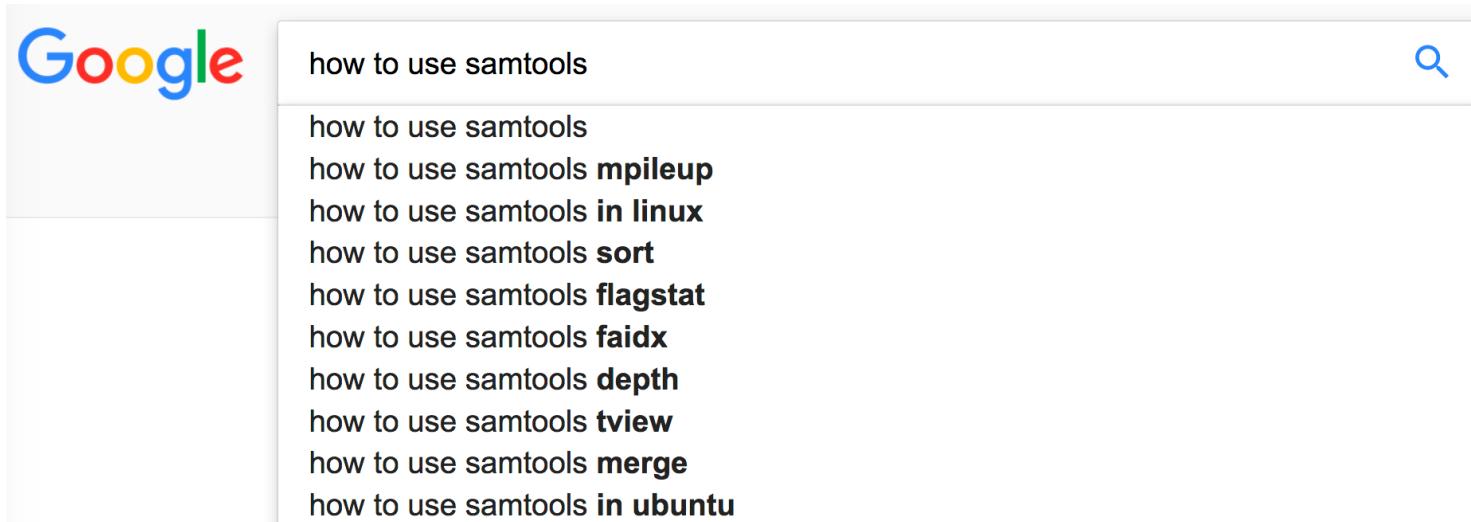
- JAMAIS de séquences ou de lignes de commandes dans word
- Créer un « simple » fichier texte tout de suite
 - C'est votre cahier de manip !!!!
 - écrire toutes les infos
 - Les dates
 - Les adresses des sites internet de téléchargement
 - Copier/coller toutes les commandes utilisées
 - Copier/coller les résultats des commandes ou décrire ces résultats si trop longs

Votre “cahier de manip” électronique

- allez sur le bureau en tapant ce qui suit dans le terminal puis "enter":
cd puis espace puis le chemin du bureau
cd '/home/anne/Desktop'
- tapez ensuite ce qui suit mais avec votre nom (!) puis "enter":
touch votrenom_notebook.txt
- **votre fichier texte se trouve maintenant sur le bureau, ouvrez le et copier/coller toutes vos commandes ainsi que vos résultats.**
- A récupérer à la fin du module et à reformater chez vous pour faire un compte rendu propre



Google is your bioinformatician friend !



Marco Galardini and Nouri Ben Zakour liked

 **Philipp Bayer**  @PhilippBayer · 14h

My ideal bioinformatics course starts with 1h on 'how to google properly'. Yet it's impossible for me to summarise how I do it at all

Laura Williams @MicroWavesSci

Realized one key lesson for Genomics lab course is that looking up commands on the Internet is half of bioinformatics, and not cheating.

2

8

31
A. Jamet