

Présentation du e-workshop

Dr A. Jamet

Médecin Microbiologiste

@Necker

Dr A. Bridier-Nahmias

Maitre de conférence

@IAMÉ

Les bioinformaticiens vus par les biologistes



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



Today



A. Jamet



Tomorrow ?

Sanger center
Hinxton, UK

Le NGS en microbiologie @Necker

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



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



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

- 5
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

illumina®

Workflow séquençage

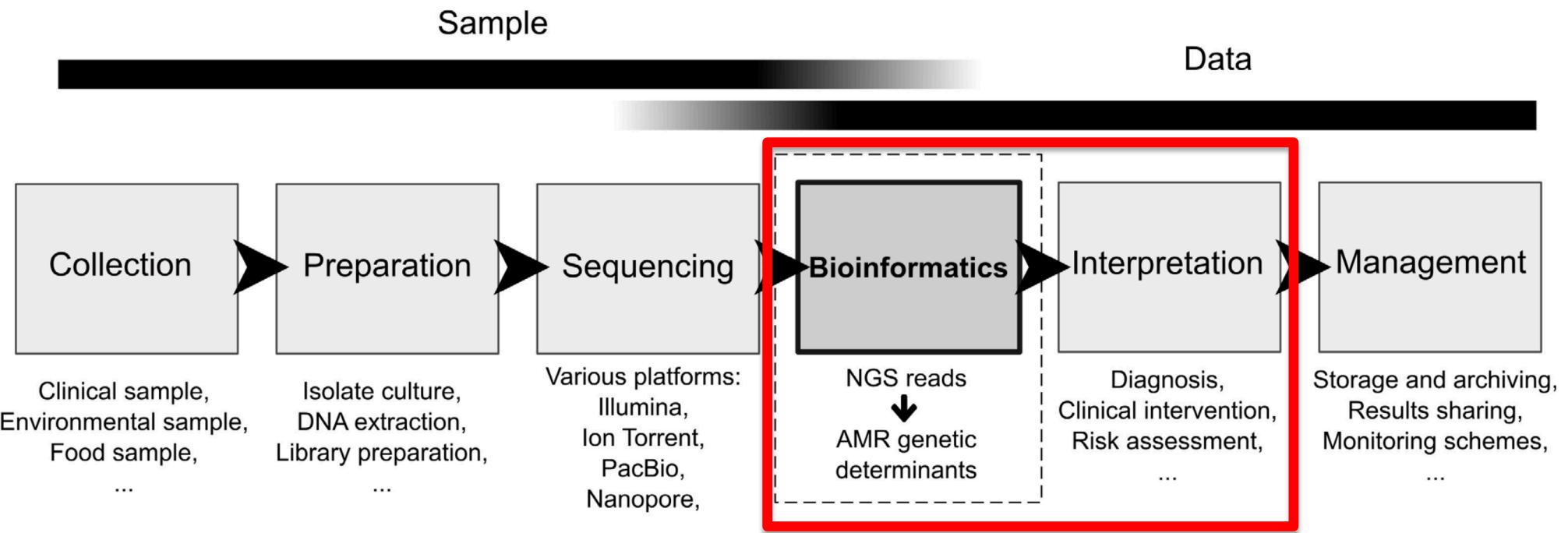


Figure 1. Overview of the different steps involved in the use of Next-Generation Sequencing technologies for the detection and monitoring of antimicrobial resistance. The benchmark strategy discussed in the current article focuses on the bioinformatics steps, the pipeline converting the output of the 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

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.

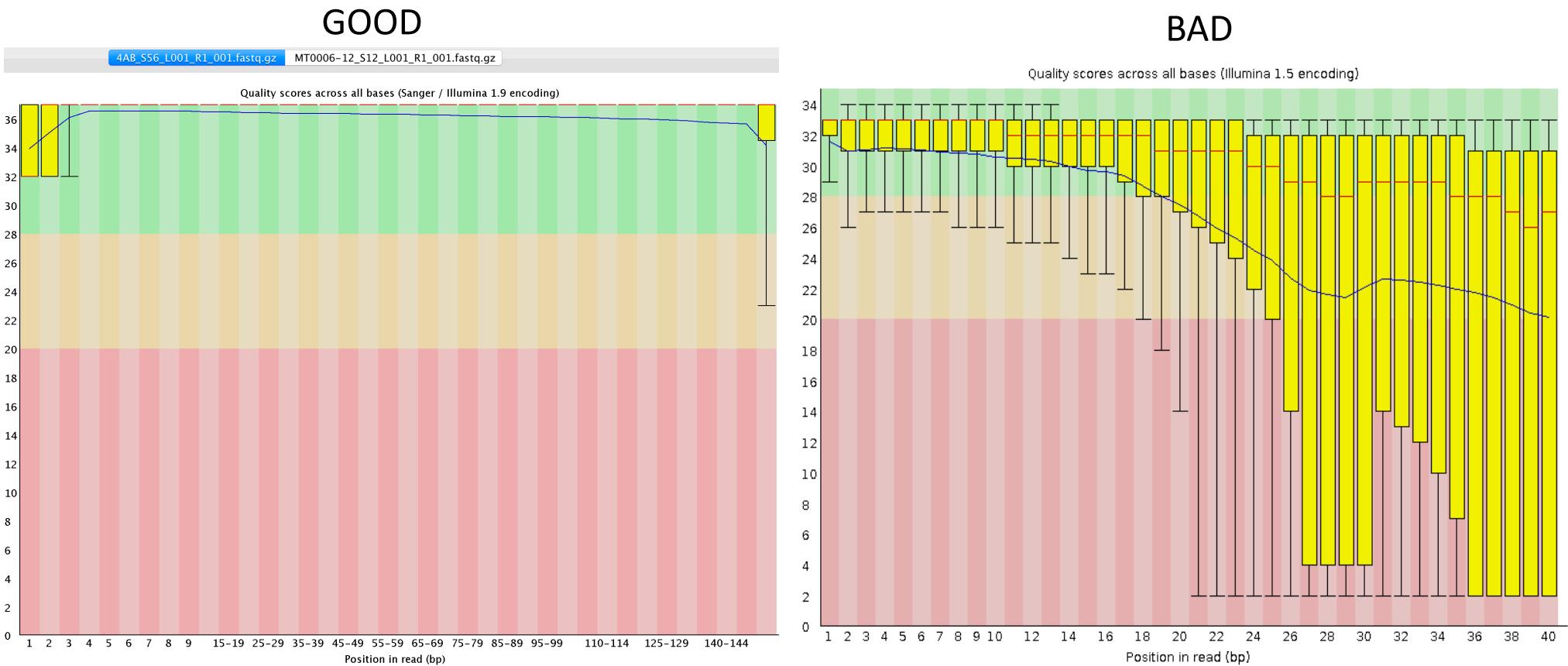
Vous devrez présenter dans ce compte-rendu

1. Vous devez créer un compte PATRIC <https://www.patricbrc.org/>
2. Uploader les FastQ R1 et R2
3. Quelle est l'espèce séquencée ?
4. Donnez un génome proche
5. Assembler ce génome
6. Annoter ce génome
7. Comparer vos résultats avec un génome de référence
(GC%, taille du génome, nb de gènes...)
9. Définir le MLST, les gènes de résistance et de virulence...
<https://cge.cbs.dtu.dk/services/>
8. faire une synthèse globale

Précisez les méthodes utilisées et les conclusions biologiques
Faites des captures d'écran, des tableaux...

Evaluer les “reads”

FastQC



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

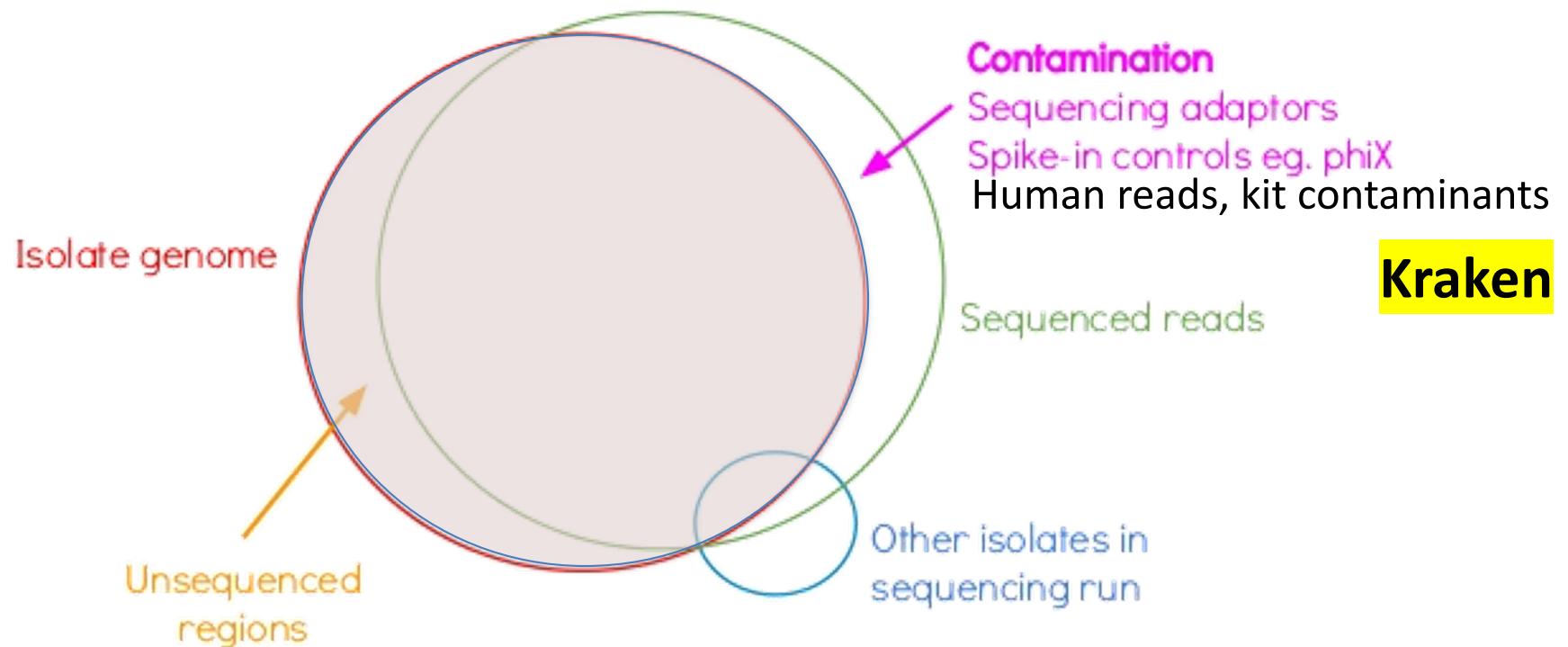


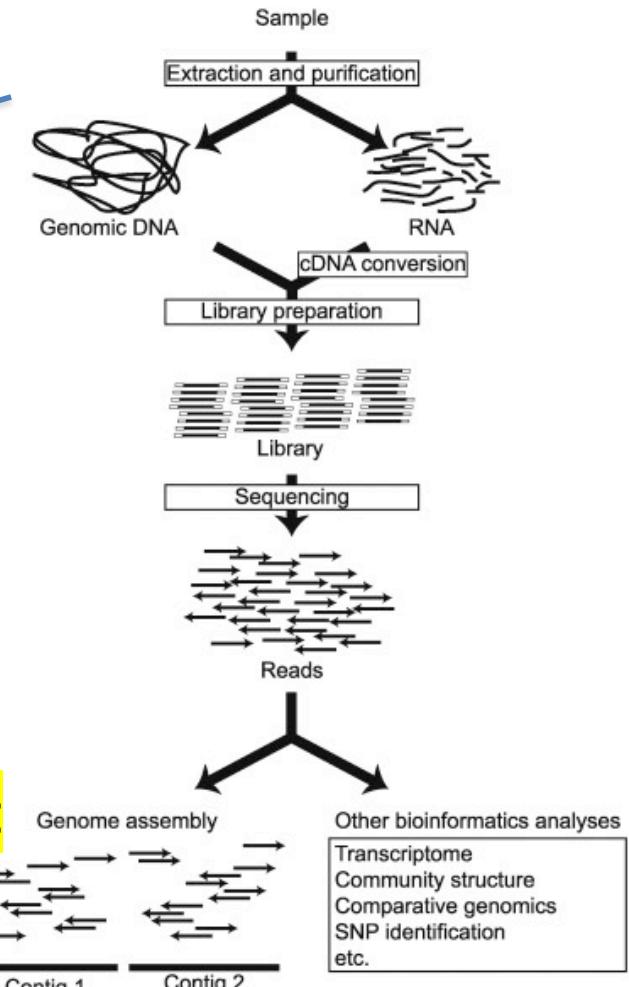
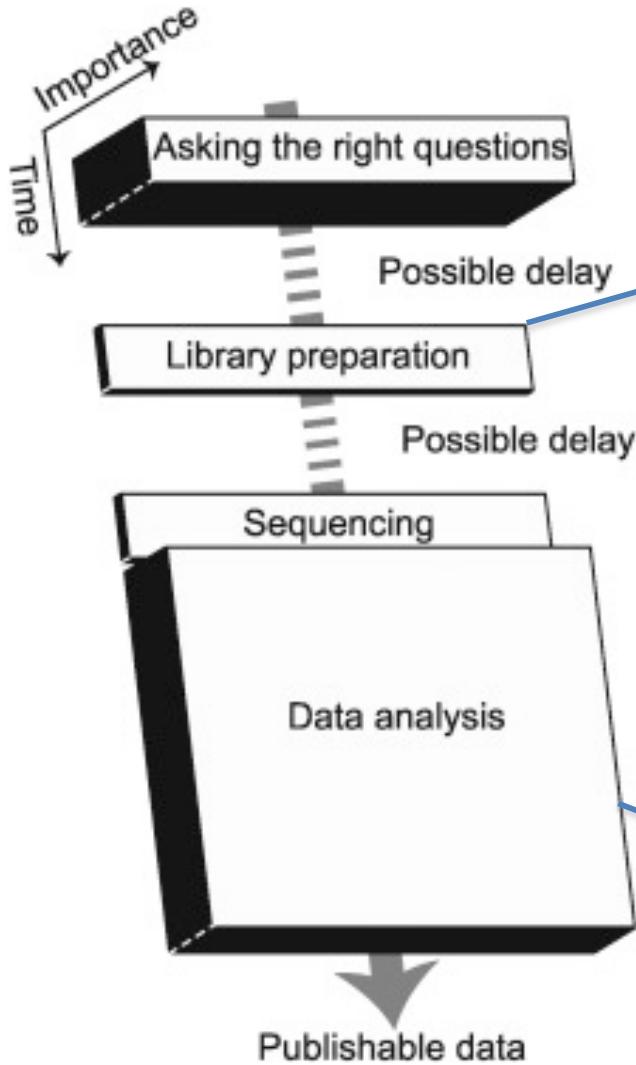
Illumina sequences typically include some of adaptor sequence

Why trim?

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

What data do we really have?





megahit

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

10.1016/j.mimet.2016.02.016

Computer Hardware

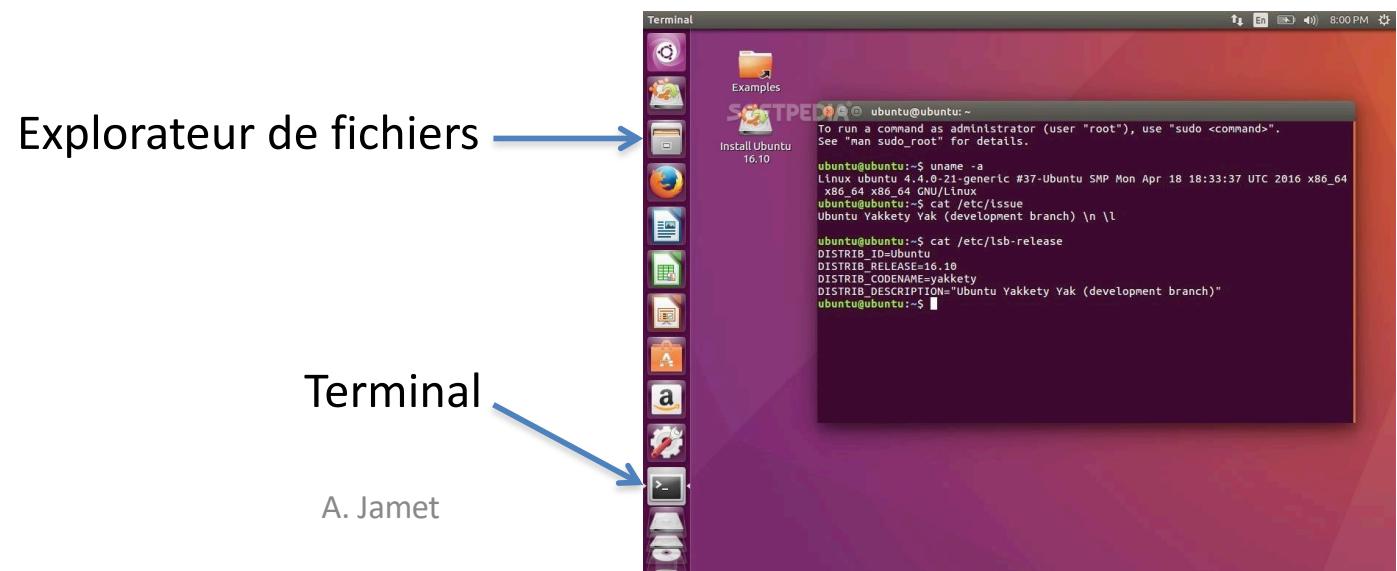
- In general, you will need:
 - 4 Gb of RAM Minimum (-> 16 Gb)
 - 500 Gb of disk space (-> 2 To)
 - better to have at least 4 cores (-> 8 CPU)

Votre “cahier de manip” électronique

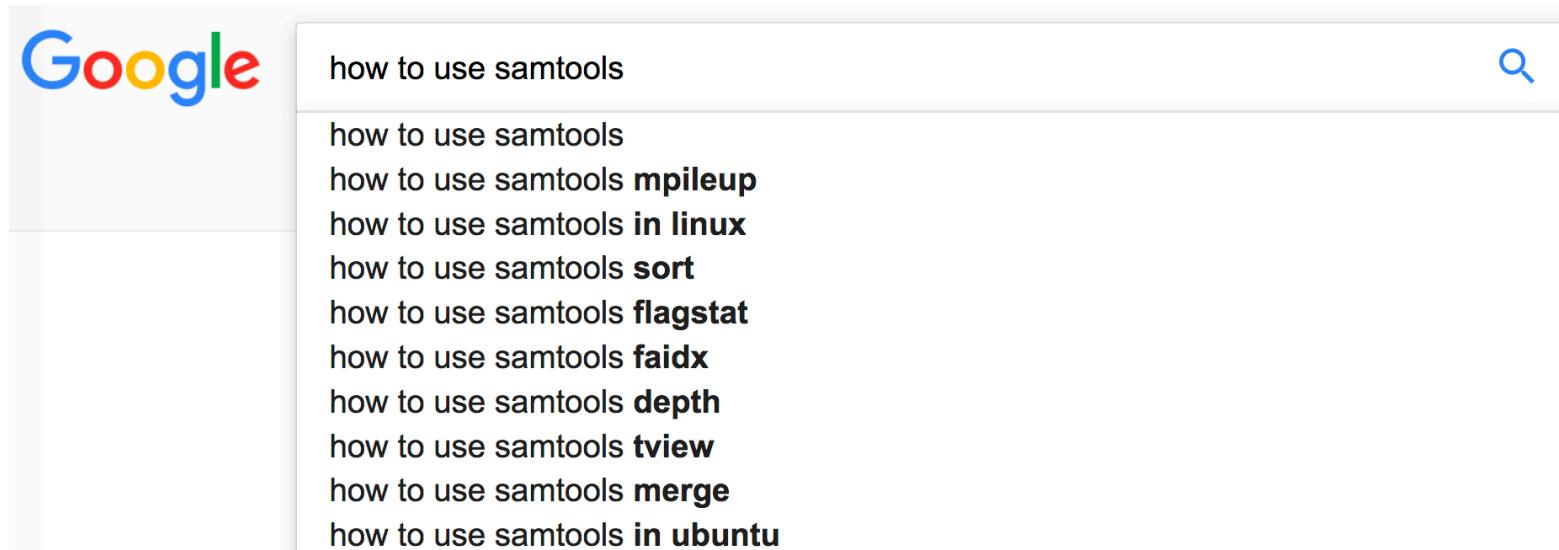
- JAMAIS de séquences ou de lignes de commandes dans word
- Créer un « simple » fichier texte tout de suite
 - C'est votre cahier de manip !!!!
 - écrire toutes les infos
 - Les dates
 - Les adresses des sites internet de téléchargement
 - Copier/coller toutes les commandes utilisées
 - Copier/coller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

