

Yakoubene Selma - Sellam Meriem - Souaré Aïcha

S'ÉCHAPPER DU

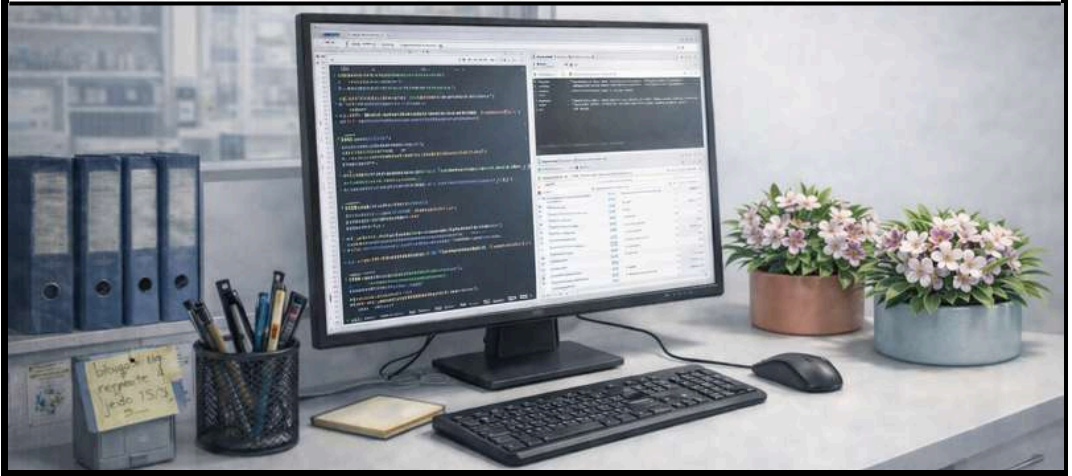
LABORATOIRES MAUDIT



Lundi matin, 9h.
Je me rends au laboratoire pour poursuivre mes analyses.



Tout m'est familier : les écrans, le bourdonnement des machines, les lignes de code ouvertes sur mon poste...



Puis...

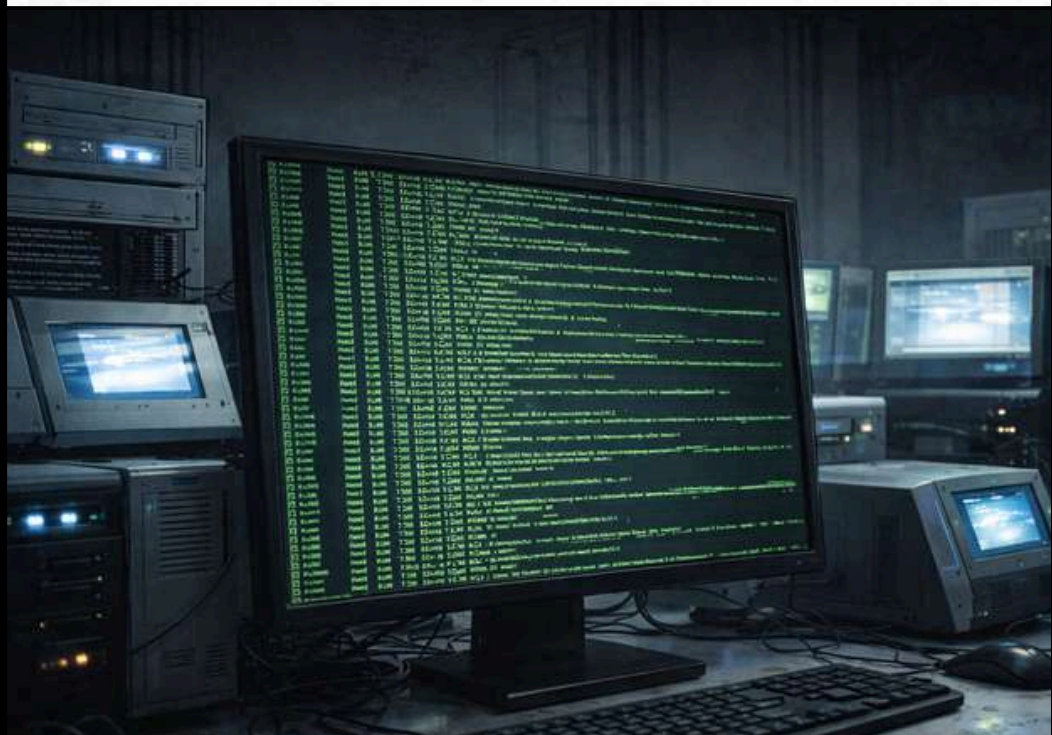
Plus rien.

Je me réveille comme sorti d'un trou noir.

Désorienté et étourdi, je mets quelques secondes à reprendre mes esprits.

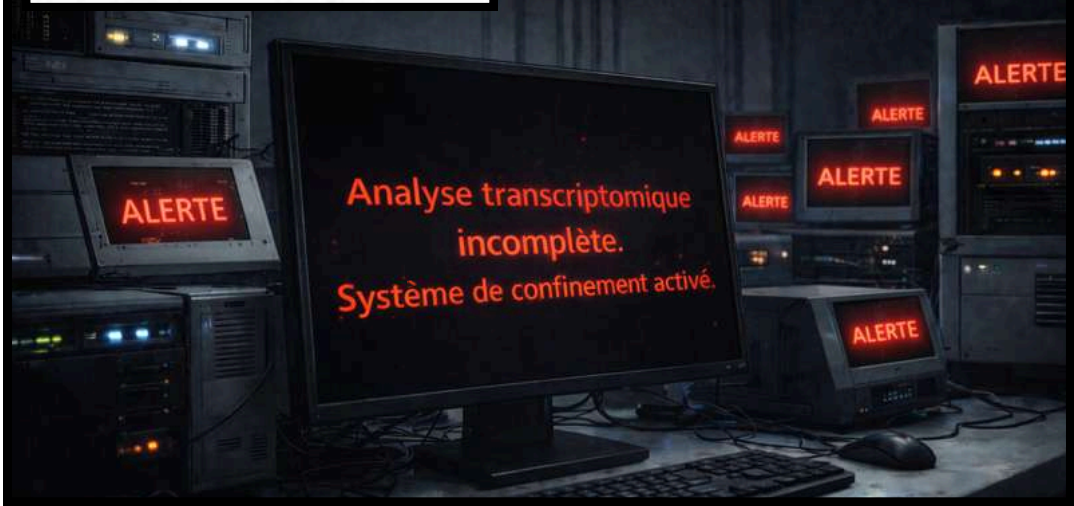


Lorsque je lève les yeux, je comprends immédiatement que je ne suis pas dans mon environnement habituel.



Les écrans sont tous allumés, des données défilent sans interruption, et des machines de séquençage bourdonnent autour de moi.

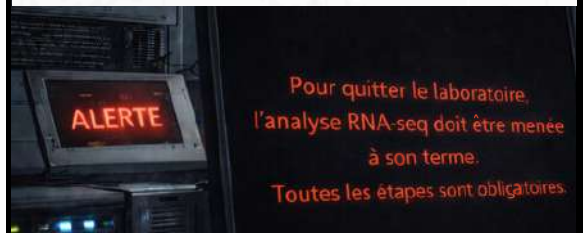
D'un coup, mon écran affiche :



Une horloge numérique apparaît au-dessus de l'ordinateur.

60:00

Un second message s'ajoute :



Une voix synthétique s'élève dans la pièce :

Chaque décision impactera la durée de l'analyse.
Des choix inefficaces entraineront des retards.
Si le temps est écoulé, vous mourrez.



Le compte à rebours démarre.



59:59

Je comprends que mon temps
est compté.
Je n'ai pas d'autre option.

Pour comprendre ce qui m'est arrivé et échapper à la mort, je vais
devoir utiliser mes compétences en bio-informatique...



Préparation
pour la NGS

Analyse
primaire

Mapping

RNA-Seq

FastQC

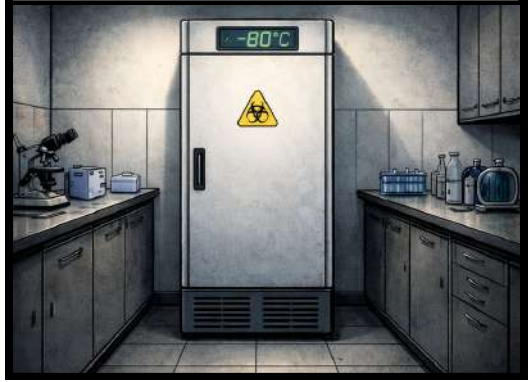
Reads

...et faire les bons choix.

Une porte s'ouvre soudainement
derrière moi.



J'avance prudemment et entre
dans une pièce où je découvre
un immense congélateur.



Je saisis la poignée et ouvre la
portière.

Devant moi, trois tubes contenant
des liquides aux couleurs
différentes reposent dans le froid.



Je me demande comment connaître leur contenu, lorsque je remarque
qu'à l'intérieur de la portière se trouve un écran sur lequel s'affiche un
message :



Je lui explique que j'aimerais savoir ce que contient chacun des tubes, ce à quoi Mestiv'IA me répond :

Voici le contenant de chacun des trois tubes :

Sample MO	Sample DA	Sample HU
Fragment d'ADN court Absence de contaminants Liquide rouge	Fragment d'ARN long Absence de contaminants Liquide vert	Fragment d'ARN long Activité ARNase constaté Liquide jaune

Lequel devrais-je choisir pour poursuivre mon analyse ?



Sample MO
→ Aller à la page 9



Sample DA
→ Aller à la page 11



Sample HU
→ Aller à la page 10

Je choisis le sample MO et subitement, Mestiv'IA m'indique le message suivant :



Échantillon non conforme

Je prends mon classeur de préparation pour la NGS pour mieux comprendre.



Préparation pour la NGS

SAMPLE MO

Fragment d'ADN court

Absence de contaminants

Liquide rouge

✓ Le sample ne présente pas de contaminants :

→ C'est une information rassurante pour la NGS.

⊖ Le sample est composé de fragments courts d'ADN :

→ En RNA-seq on ne veut pas de fragments d'ADN courts car cela signifie que l'ADN est déjà dégradé, ce qui empêche la reconstruction des transcrits.

L'ordinateur avait pourtant prévenu : « RNA-seq », et moi, j'ai choisi un échantillon d'ADN. C'était forcément le mauvais choix !
Lequel parmi les deux choix restants est le bon ?

Sample DA

Fragment d'ARN long

Absence de contaminants

Liquide vert

→ Lancer

Sample HU

Fragment d'ARN long

Activité ARNase constatée

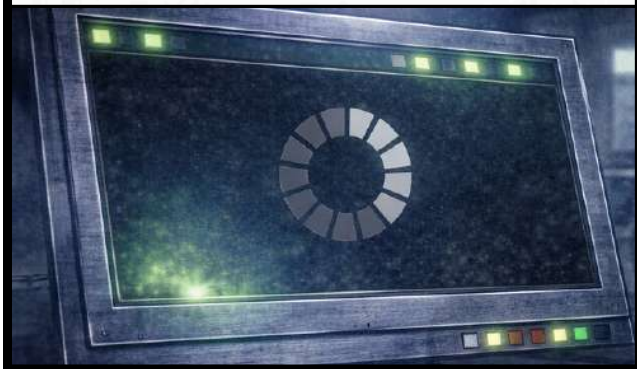
Liquide jaune

→ Lancer

< Page 11

Page 10 >

Je choisis le sample HU et subitement,
Mestiv'IA se met à bugger.



Perturbé, j'attrape mon
classeur de préparation
pour la NGS



SAMPLE HU

Fragment d'ARN long

Activité ARNase
constatée

Liquide jaune

✓ Le sample HU est composé
de longs fragments d'ARN :

→ C'est l'idéal pour la
poursuite de la NGS.

⊖ Une activité ARNase est
constatée :

→ L'ARNase détruit l'ARN en
quelques minutes.

→ Les fragments vont se
raccourcir.

→ Cela peut compromettre la
qualité du séquençage.

Je comprends alors que mon sample est bon à l'instant, mais sera
inutilisable d'ici quelques minutes...

Lequel parmi les deux choix restants est le bon ?

Sample MO

Fragment d'ADN court

Absence de contaminants

Liquide rouge

→ Lancer

Sample DA

Fragment d'ARN long

Absence de contaminants

Liquide vert

→ Lancer

J'ai pris la bonne décision !
Les voyants de l'écran passent au vert.

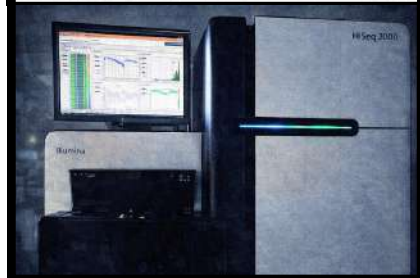


Dix minutes se sont déjà
écoulées...



Rassuré par la qualité de mon échantillon,
je quitte la pièce et me dirige vers la salle
suivante.
J'entre dans la salle du NGS.

Je m'approche d'une
machine de séquençage.



Sur l'écran, trois configurations différentes me sont proposées :



Quelle est la proposition la plus adaptée à l'analyse que je souhaite mener ?

PH - Page 13

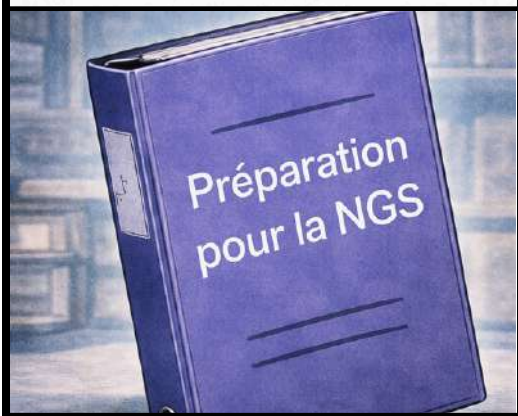
MA - Page 12

NK - Page 14

Je sélectionne la configuration MA et rien ne s'affiche.



Un doute me traverse. J'ouvre mon classeur de préparation NGS.



Configuration MA

Technologie : long reads

Mode : paire-end

Profondeur de séquençage :
très élevée

Temps de run : long

✗ Configuration MA

Technologie : long reads

→ surdimensionné pour une RNA-Seq d'expression, est plus adapté pour reconstruire des transcrits complets.

Mode : paired-end

→ permet un alignement plus précis et facilite la détection ainsi que la correction des erreurs.

Profondeur : très élevée

→ génère un volume excessif de données, complique l'analyse sans gain réel.

Temps de run : long

→ perte de temps alors que ce niveau de détail n'est pas nécessaire.

En lisant le classeur, je comprends que cette configuration est totalement inadaptée à mon analyse. Une dernière chance s'offre à moi :

PH - Page 13

NK - Page 14

Le séquençage NGS s'est déroulé sans incident grâce aux réglages effectués.

Le message suivant apparaît :



Je m'avance vers la salle suivante... puis je m'arrête net.
Devant moi, trois portes apparaissent.



Quelle porte dois-je ouvrir ?

Entrer par la porte NB

L'analyse primaire ne me semble pas nécessaire. Les échantillons étaient de bonne qualité avant le séquençage, et je veux sortir au plus vite.

→ Page 15

Entrer par la porte EY

J'ai déjà perdu trop de temps. Autant passer directement à l'interprétation des résultats.

→ Page 16

Entrer par la porte NE

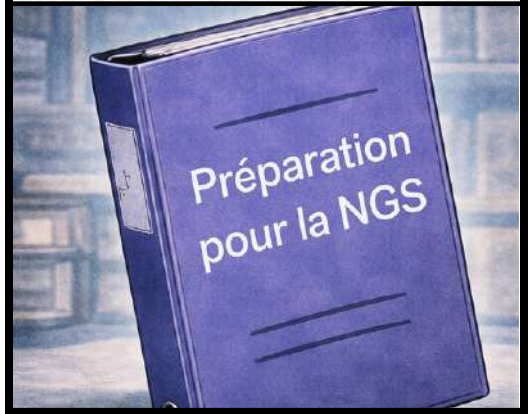
Même avec de bons échantillons, rien ne garantit que le séquençage se soit déroulé parfaitement.

→ Page 17

Je sélectionne la configuration NK et rien ne s'affiche.



Un doute me traverse. J'ouvre mon classeur de préparation NGS.



Configuration NK

Technologie : short reads

Mode : single-end

Profondeur de séquençage : faible

Temps de run : très court

✗ Configuration NK

Technologie : short reads

→ longueur des fragments du sample est différente de la longueur des reads produits par la technologie

Mode : single-end

→ mapping moins précis, difficulté à distinguer correctement les transcrits.

Profondeur : faible

→ sous-échantillonnage du transcriptome, quantification peu fiable.

Temps de run : très court → rapidité au détriment de la qualité des résultats.

Même si je souhaite que l'analyse soit rapide, ce n'est pas la meilleure configuration pour mon analyse. Une dernière chance s'offre à moi :

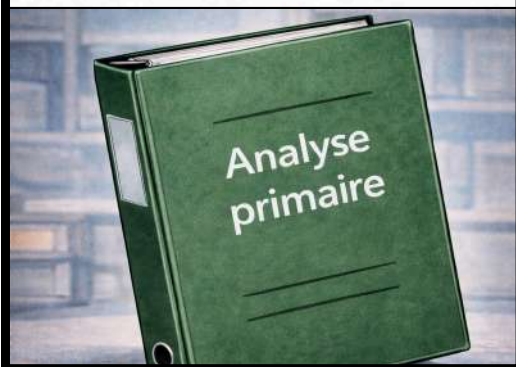
PH - Page 13

MA - Page 12

J'entre dans la salle des résultats et je découvre un géant écran affichant plusieurs graphiques que je n'arrive pas à analyser...



Je décide de prendre le classeur d'analyse primaire.



FastQ est un fichier brut produit par le séquenceur après la NGS.

Son analyse brute peut donner l'illusion que les données sont prêtes à être exploitées, alors qu'elles peuvent encore contenir des erreurs invisibles qui ne se révèlent qu'à l'étape

Mince, la fin de la page est déchirée !

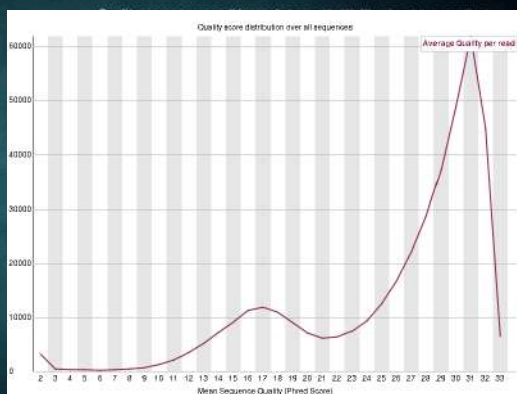
Je me rends compte que je ne suis pas dans la bonne salle. Je dois vite retourner en arrière avant qu'il ne soit trop tard



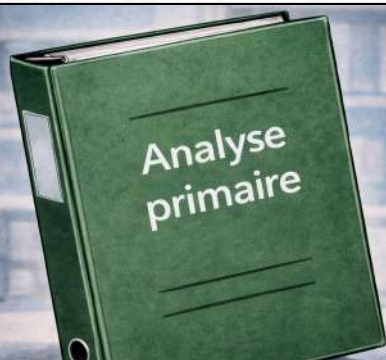
J'entre dans la salle d'interprétation des résultats NGS où je découvre un géant écran affichant des éléments auxquels je ne m'attendais pas...

SMT627479.1 *Daphne odora chlorodress*, complete genome

```
TTAAATTAATTTTAAATATTTTAAATATCTTTTATTTTGTTCATGCGCGACACGAGGAGTTGAA
CCCGGAGTGGTGGATTCCAAATCCACTGCTTAACTCCACTTTGGCTACATCGCGCCCTCTTTTATTTA
TAATATTTTSTTCGCTTTTATTTAAATTTAAATTAAGTAAATTAAGGACTTCAAAATAGATTTGGTAAGGA
CTGATCCATTTATAGAAATTTATATATATATATGATACACAAACAAAGATATGATACTAAACAAACCAAT
TAAATGAGCCCTTTCTTTTATTTTATTTTATTTATGAAAGTAAAGTCAAGATAAAGTAAAGTAAAGTAAAGT
AATAAAGAAAGTATTTAAATTAATTAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT
TTCTTTTCAAAAAATAGTATACACTAGACAAACCTTTATCTATTGTAGATGGAGCTTCAATAGCAGC
TAGATCTAGAGGAAATTTAGAGCTTACGTTCTATGATACCTTCAATAGCAGCTTCAATAGCAGCTTCAAT
ATATGAGCAGGATTAATTTACACGACCTTGAATCAACTACGAGTGTGGATTTGAATTTGAATCTTTTAT
GTTTGAAGGCAATAGTGTGATACCTTGAAGGCAATGAGGCAATGAGGCAATGAGGCAATGAGGCAATGAGGCA
GAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCA
GCTACAAATTTGATAGTTTCTCTCTCTGAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTA
TTCTCTGATCAAGTATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGG
AGCTACGCTAGCATGTAAGTATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTG
AAAGTACGAGGATTTCTGAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAG
CGCGGTAGCATGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTG
AGGTCCGCTACGAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTA
CGAGGATTTGATTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTG
CAGAGTAAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGG
AATACGCTCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAAT
GTAGGATCTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTA
AGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGT
ATCATGAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCAATGAGGTGTAAGGCA
```



Je décide d'ouvrir le classeur d'analyse primaire



FastQ est un fichier brut produit par le séquenceur après la NGS.

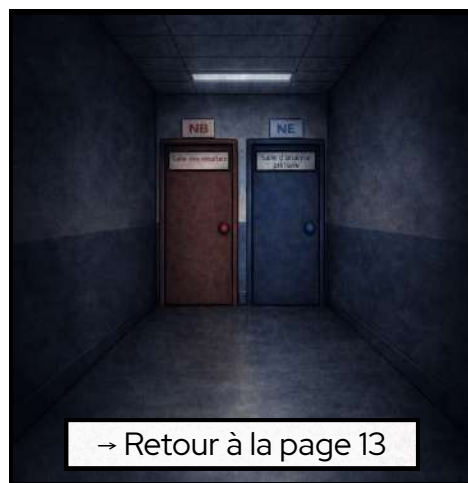
Il contient, pour chaque read :

- la séquence nucléotidique
- le score de qualité (Phred) de chaque base

Pour obtenir des résultats à interpréter, il faut utiliser l'outil FastQC, disponible dans la salle

Mince, la fin de la page est déchirée !

Je me rends compte que je ne suis pas dans la bonne salle. Je dois vite retourner en arrière avant qu'il ne soit trop tard

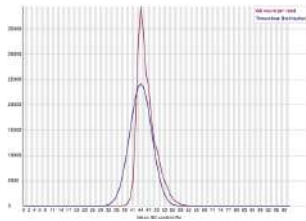


→ Retour à la page 13

Super ! J'entre dans la salle dédiée à l'analyse primaire et lance le contrôle qualité à l'aide de FastQC. Trois résultats différents s'affichent :

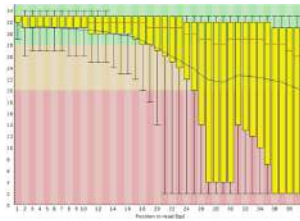
Résultat S

Per sequence GC content

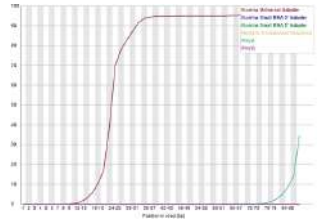


→ 44 % GC

Per base sequence quality

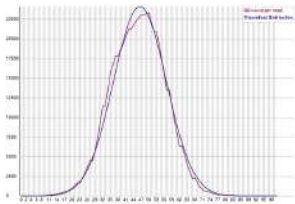


Adapter content



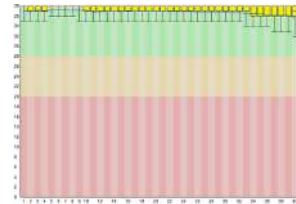
Résultat O

Per sequence GC content

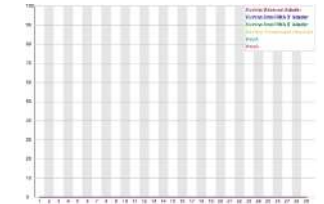


→ 45 % GC

Per base sequence quality

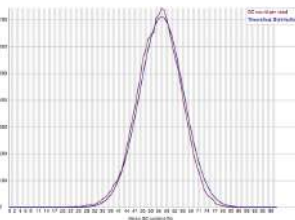


Adapter content



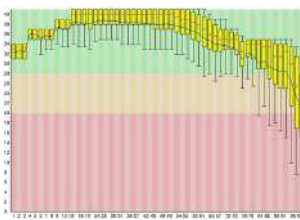
Résultat E

Per sequence GC content

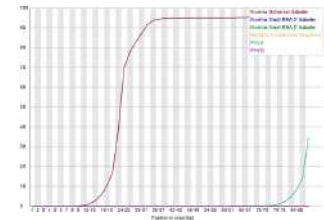


→ 52 % GC

Per base sequence quality



Adapter content



Quel résultat dois-je sélectionner ?

Résultat S

→ Page 19

Résultat O

→ Page 20

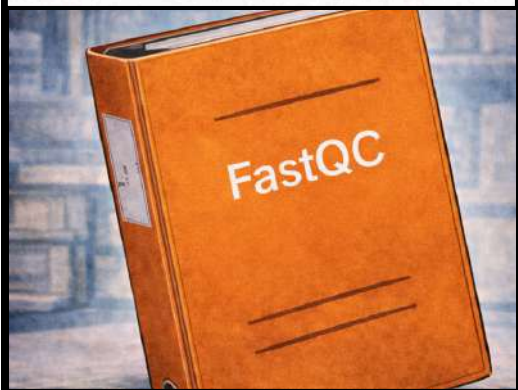
Résultat E

→ Page 18

Je sélectionne le résultat S et mon écran affiche le message suivant :

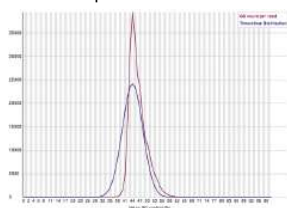
FAIL

Ne comprenant pas l'apparition de ce message, j'ouvre le manuel FastQC



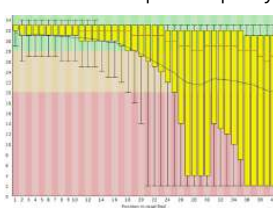
Résultat S

Per sequence GC content

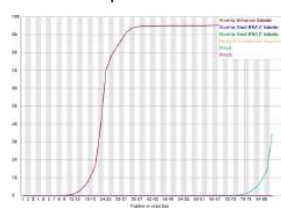


→ 44 % GC

Per base sequence quality



Adapter content



Habituellement, la distribution GC suit une distribution suivant la loi normale \mathcal{N} .

Le score qualité par base est considéré comme acceptable à partir de Q20.

Le résultat indique qu'à partir de 22-25 bases, il y a énormément d'adaptateurs.

Mince, j'ai raté mon analyse...
Je me demande si j'ai choisi des samples de qualité ? ou si la configuration NGS était la bonne...

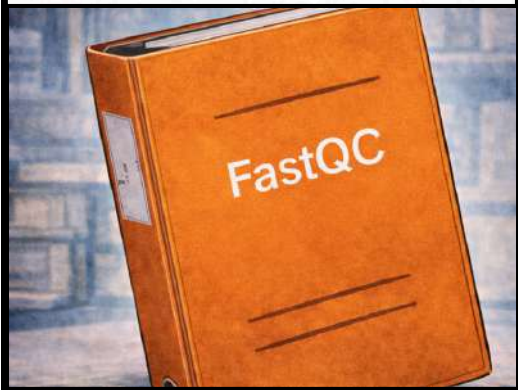
Je n'ai plus le temps d'y réfléchir, ma priorité est de trouver le bon résultat.

← Retour à la page 17

Je sélectionne le résultat E et mon écran affiche le message suivant :

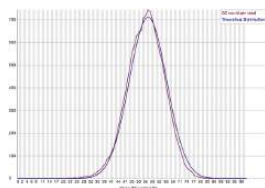
WARNING

Ne comprenant pas l'apparition de ce message, j'ouvre le manuel FastQC



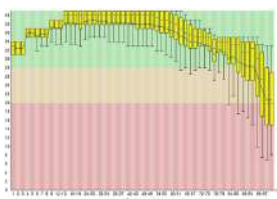
Résultat E

Per sequence GC content



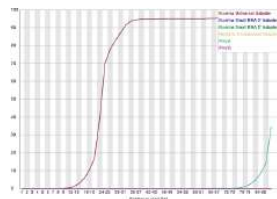
→ 44 % GC

Per base sequence quality



Le score qualité par base est considéré comme acceptable à partir de Q20.

Adapter content



Le résultat indique qu'à partir de 22-25 bases, il y a énormément d'adaptateurs.

Habituellement, la distribution GC suit une distribution suivant la loi normale \mathcal{N} .

Le logiciel FastQC me propose une autre question :

Voulez-vous appliquer un trimming ?

Oui

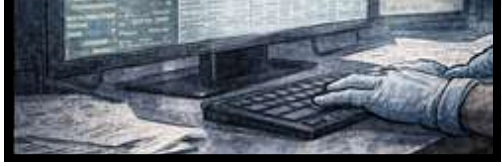
→ Page 20

Non

→ Page 17

L'analyse FastQC de l'échantillon est enfin terminée. Les indicateurs sont au vert : la qualité est suffisante pour continuer.

Je prends une grande inspiration et décide de passer à l'étape suivante : le mapping des reads.

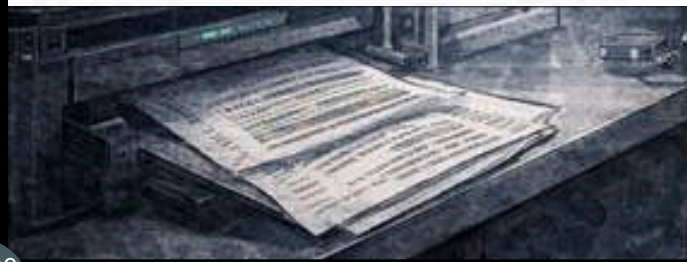


Je m'installe devant la console principale et lance le logiciel d'alignement. Les fichiers FASTQ défilent à l'écran tandis que chaque read est comparé à la séquence de référence.



Le processus paraît simple, presque automatique, mais je sais qu'une seule mauvaise décision peut fausser toute l'analyse...

Quelques secondes plus tard, l'imprimante se met à vibrer et crache plusieurs feuilles. En les ramassant, je reconnais immédiatement ce que j'ai sous les yeux : des alignements de reads sur la séquence de référence.



Pourtant,
quelque chose
me trouble...

Je pose les trois feuilles côte à côte et, malheureusement, elles se ressemblent presque parfaitement...

Feuille INGS

TCAAT

CGGCCAA

AATAACC

TCAATCGGCCAAAATAACC

→ Aller à la page 22

Feuille INGE

TCAATCGGCCA

CGGCCAAAATA

ATAACG

TCAATCGGCCAAAATAACC

→ Aller à la page 23

Feuille DORA

TCAATCGGCCAA

CGGC

AATAACC

GCCAAAATAACC

TCAATCGGCCAAAATAACC

→ Aller à la page 24

Laquelle dois-je choisir ?

Je choisis la feuille INGS et subitement, un voix s'élève dans la pièce :

**VOTRE CHOIX
EST
INCORRECT**

Dégouté, j'attrape le classeur des reads pour comprendre



Feuille INGS

TCAAT

CGGCCAA

AATAACC

TCAATCGGCCAAAATAACC

Mapping

L'alignement des reads, ou mapping, consiste à retrouver l'endroit précis du génome de référence où chaque fragment doit se placer.

Pour y parvenir, il faut vérifier :

- ✓ Le sens du read (5'→3')
- ✓ La complémentarité des bases nucléotidiques (A-T / C-G)
- ✗ S'il y a bien un chevauchement entre les différents reads

Mince, à cause de mon manque d'attention je me suis trompé. Il faut vite que je retourne chercher le bon read pour avancer.

Je ne suis plus très éloignée de la fin de l'analyse mais je perds trop de temps...

← **Retour à la page 21**

Je choisis la feuille INGE et subitement, une voix s'élève dans la pièce :

**VOTRE CHOIX
EST
INCORRECT**

Dégoûté, j'attrape le classeur des reads pour comprendre



Feuille INGE

TCAATCGGCCA
CGGCCAAAATA
ATAACG

TCAATCGGCCAAAATAACC

Mapping

L'alignement des reads, ou mapping, consiste à retrouver l'endroit précis du génome de référence où chaque fragment doit se placer.

Pour y parvenir, il faut vérifier :

- ✓ Le sens du read (5'→3')
- ✓ S'il y a bien un chevauchement entre les différents reads
- ✗ La complémentarité des bases nucléotidiques (A-T / C-G)

Je réalise que je me suis trompée. Il faut vite que je retourne chercher le bon read pour avancer.

Je ne suis plus très éloigné de la fin de l'analyse mais je perds trop de temps...

← **Retour à la page 21**

J'ai réussi à obtenir le bon alignement !
Une immense vague de soulagement
m'envahit.

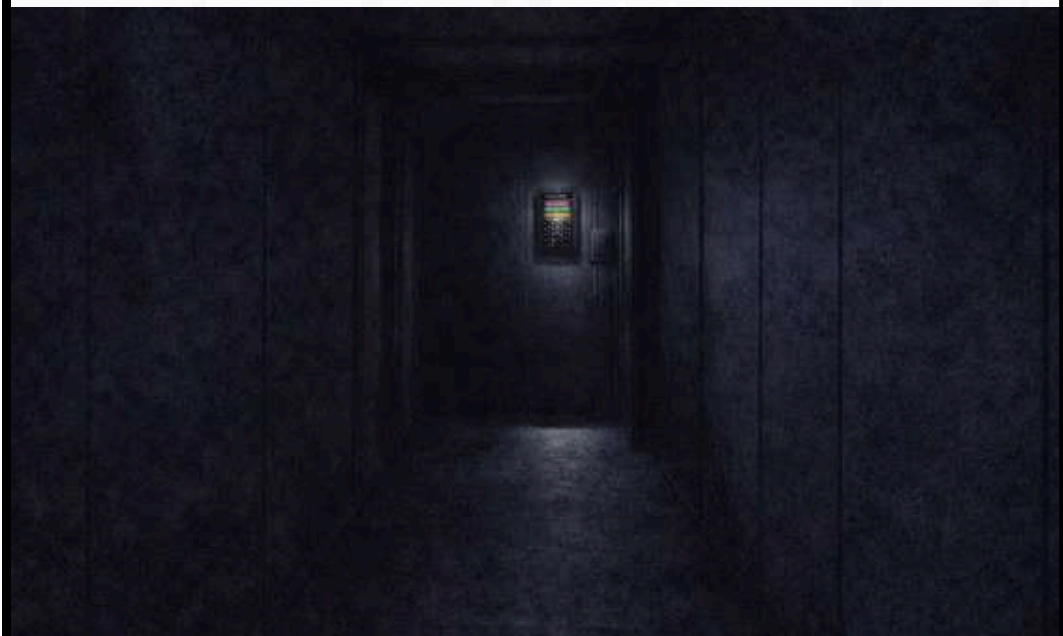
Je laisse échapper un cri de joie !

Cette fois, j'en suis sûr, je vais enfin
pouvoir sortir d'ici.



Mais mon enthousiasme retombe aussitôt...

En quittant la salle du mapping, j'entre dans une pièce totalement plongée dans l'obscurité. Le silence est pesant. Le seul élément éclairé dans la pièce est une grande serrure métallique fixée sur une porte.



Moi qui pensais que c'était terminé...

En m'approchant, j'aperçois sur la porte des gravures mais je ne sais pas si ça va m'aider. Quel est le code ?



44 88
62
6622
33
444664777
PAGE 27

7777444
664
336
88
82668
PAGE 26

32
744
6633
666
36667772
PAGE 28

Je me suis trompé de code. J'ai un dernier essai. Lequel je prends entre les deux ?



44 88
62
6622
33
444664777
PAGE 27

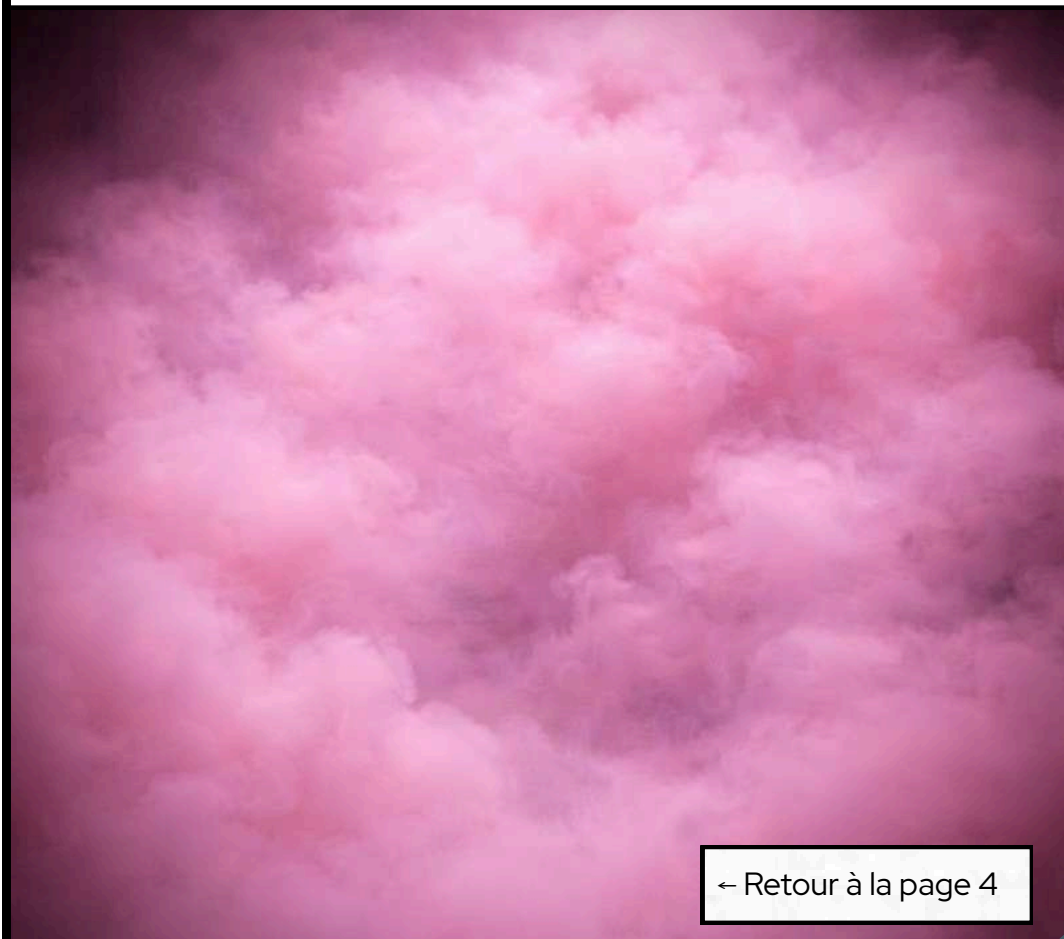
32
744
6633
666
3667772
PAGE 28

La porte se bloque.

CODE ERRONÉ
X



Un épais nuage de fumée rose envahit le couloir. Je perds connaissance...



← Retour à la page 4

J'ai choisi le bon code.
La serrure s'ouvre dans un cliquetis
métallique.



Je fais quelques pas... puis je m'arrête net.



Je reconnais la rue, c'est celle de mon laboratoire...

Peu à peu, mes esprits
me reviennent.

Je repense à mon bureau...



Aux changements récents...



Aux nouvelles plantes posées près de
mon poste de travail...



L'odeur ! Je me souviens maintenant. En entrant ce matin, j'avais été
surpris par cette agréable odeur florale. Douce. Presque apaisante.



Puis... plus rien.



Tout s'explique !

Le Daphné odora.



Cette plante, aussi appelée Bois-joli, est toxique de ses feuilles à ses fruits. Elle émet naturellement, par ses fleurs et ses tissus végétaux, de petites molécules appelées composés organiques volatils. Ces molécules s'évaporent facilement dans l'air et sont responsables de l'odeur florale agréable... mais peuvent aussi provoquer des maux de tête, des vertiges, une confusion mentale, et parfois même des hallucinations.

Je comprends alors que...
Pendant tout ce temps...
J'étais simplement dans mon propre bureau.

Remerciements

Nous souhaitons adresser nos plus sincères remerciements à Monsieur Mestivier pour son accompagnement, sa pédagogie et son exigence scientifique tout au long de nos trois années de formation.

Ses enseignements ont constitué le socle théorique et méthodologique qui nous a permis de concevoir cet ouvrage avec rigueur et cohérence.

Nous exprimons aussi notre gratitude envers l'ensemble de l'équipe pédagogique de la filière ISBS.

Nous remercions nos camarades de promotion, et plus particulièrement Koussou Camara et Margaux Dupré, pour les échanges et les regards critiques qui ont nourri notre réflexion et enrichi ce travail.

Enfin nous nous remercions nous même, car ce livre interactif est le fruit d'un travail collectif.

S'échapper du laboratoire maudit

Vous êtes bio-informaticien.

Ce lundi matin ressemble à tous les autres. Vous passez prendre votre café, vous vous installez à votre poste... puis plus rien. À votre réveil, vous vous retrouvez enfermé dans un laboratoire inconnu. Les écrans affichent des données, les machines de séquençage tournent encore, et un compte à rebours vient de s'enclencher.

Pour sortir, une seule solution : mener à bien une analyse RNA-seq complète. Chaque salle que vous traversez correspond à une étape réelle du pipeline bio-informatique : choix du sample, paramétrage du NGS, contrôle qualité, mapping, interprétation... À chaque décision, vous devrez faire appel à vos connaissances scientifiques. Mais il faut se dépêcher, car le compte à rebours tourne...

Ce livre interactif vous plonge au cœur d'un escape game scientifique où vos compétences en bio-informatique seront votre seule issue. Saurez-vous faire les bons choix avant la fin du compte à rebours ?

Selma, Meriem et Aïcha, étudiantes de la promotion ISBS 2023-2026, ont conçu cet ouvrage dans le cadre de la spécialité bio-informatique. Il a été réalisé avec l'objectif d'accompagner leurs camarades ingénieurs dans leurs révisions de la matière, à travers une approche à la fois ludique et immersive.