

	<p style="text-align: center;"><b>ANNEE UNIVERSITAIRE 2024-2025</b> <b>SESSION 1 DE PRINTEMPS</b></p> <p><b>PARCOURS / ETAPE : Master Bioinformatique</b>  <b>Code UE : 4TBI808U/4TBA810U UE Nouvelles Générations de Séquençage</b>  <b>Epreuve : Cours</b>  <b>Date : 09/04/2025      Heure : 14h00-15h30      Durée : 1h30</b>  <b>Documents : non autorisés</b>  <b>Epreuve de S. Karkar et P. Sirand-Pugnet</b></p>	<p style="text-align: center;"><b>Collège Sciences et technologies</b></p> <p style="text-align: center;"><b>Masters</b></p>
---	--	--

**Remarques :**

- Les deux sujets sont indépendants et à traiter sur des copies séparées.
- Il sera tenu compte de la qualité de rédaction et de présentation

**Sujet 1 (10 points):**

Le projet de science participative\* « Le French Gut » est porté par INRAE et vise à explorer le potentiel du microbiote intestinal, définir ce qu'est un microbiote de personnes en bonne santé et ses changements dans les maladies chroniques (<https://lefrenchgut.fr>). Pour ce projet, les selles de 100 000 volontaires majeurs, résidant en France métropolitaine, seront collectées et analysées pour déterminer leur composition en microorganismes.

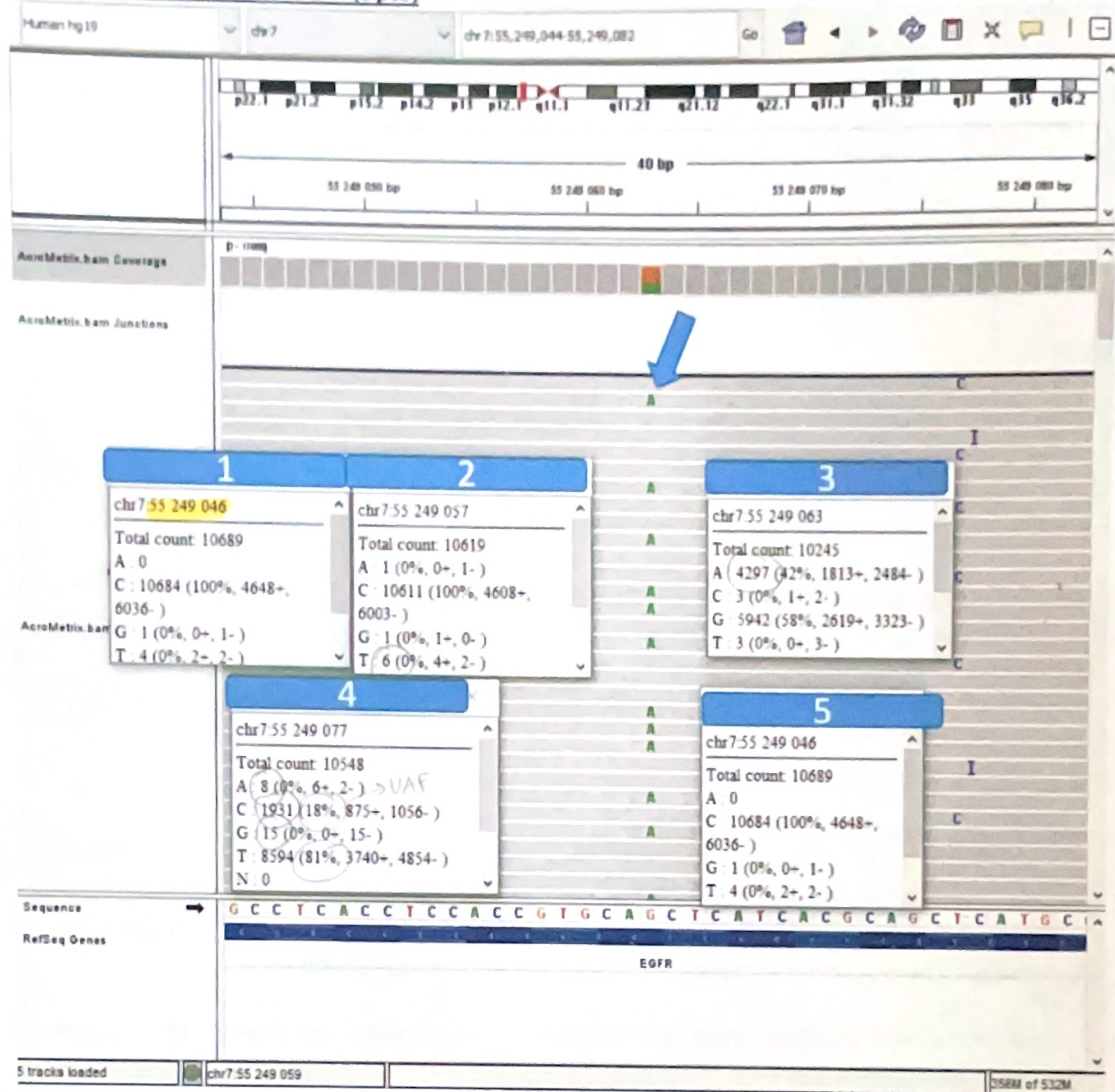
(\* vous aussi, vous pouvez participer !)

1. Définissez le microbiote intestinal (1 point)
2. Quelle stratégie proposez-vous pour réaliser un simple inventaire des micro-organismes procaryotes et eucaryotes dans des échantillons de selles ? Représentez cette stratégie sous la forme d'un schéma annoté présentant les différentes étapes. (2 points)
3. Quelle technologie de séquençage vous semble la plus adaptée pour réaliser cet inventaire ? Justifiez votre réponse. (1 point)
4. Détaillez sur un schéma cette technologie de séquençage. (3 points)
5. Dans une perspective plus large que de faire un simple inventaire des espèces microbiennes présentes, le projet vise également à mieux comprendre les interactions métaboliques entre les microorganismes du microbiote intestinal et entre ces microorganismes et les cellules humaines. En quoi cette perspective peut éventuellement modifier votre choix de la technologie de séquençage ? Justifiez votre réponse. (2 points)
6. Quelles étapes spécifiques seraient nécessaires pour aboutir à une modélisation de l'ensemble de ces interactions métaboliques ? Quelles données expérimentales, ne relevant pas des technologies de séquençage haut-débit pourraient être intéressantes pour affiner de telles modélisations ? (1 point)



## Sujet 2 (10 points):

### Exercice 1 - Variants et IGV (3pts)



- Quelle représente la valeur "55 249 046" pour le cas 1 ?
- Choisissez quelle cas putatif 1, 2, 3, 4 ou 5 décrit la mutation indiquée avec la flèche
- Pour la mutation indiquée par la flèche :
  - Indiquer la VAF (Variant Allele Frequency, en %) =
  - Indiquer l'allèle de référence =
  - Indiquer le chromosome =
  - Indiquer la version (build) du génome de référence =
- Choisir le cas putatif 1, 2, 3, 4 ou 5 pour le variant ayant la 2<sup>nd</sup> VAF la plus élevée.
- Pour les cas 1 à 5, quels sont les variants qui atteignent ou dépassent un nombre minimal de support reads de 5 (c'est-à-dire trouvé sur au moins 5 reads) ?



## Exercice 2 - QC (3pts)

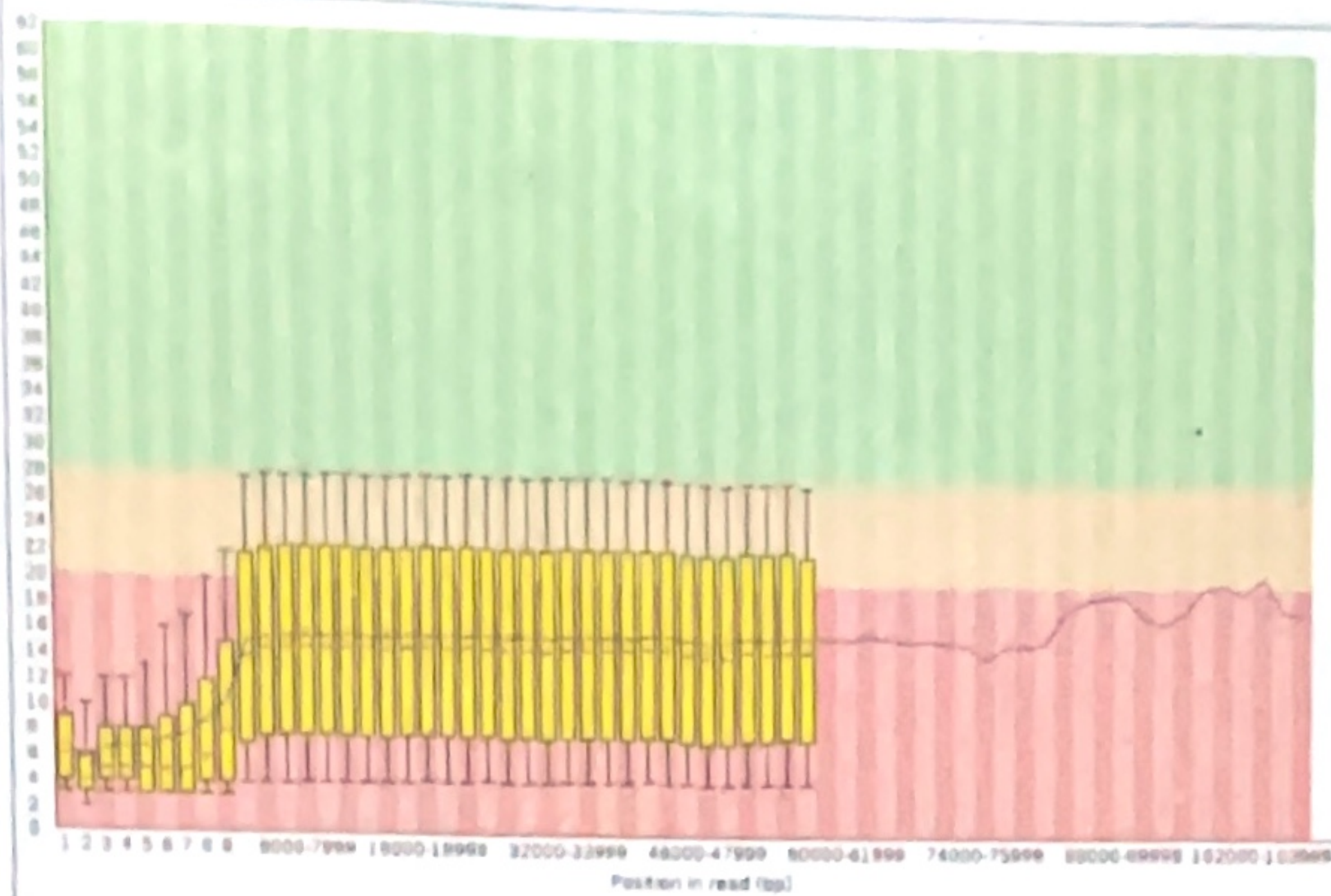


Fig. 1

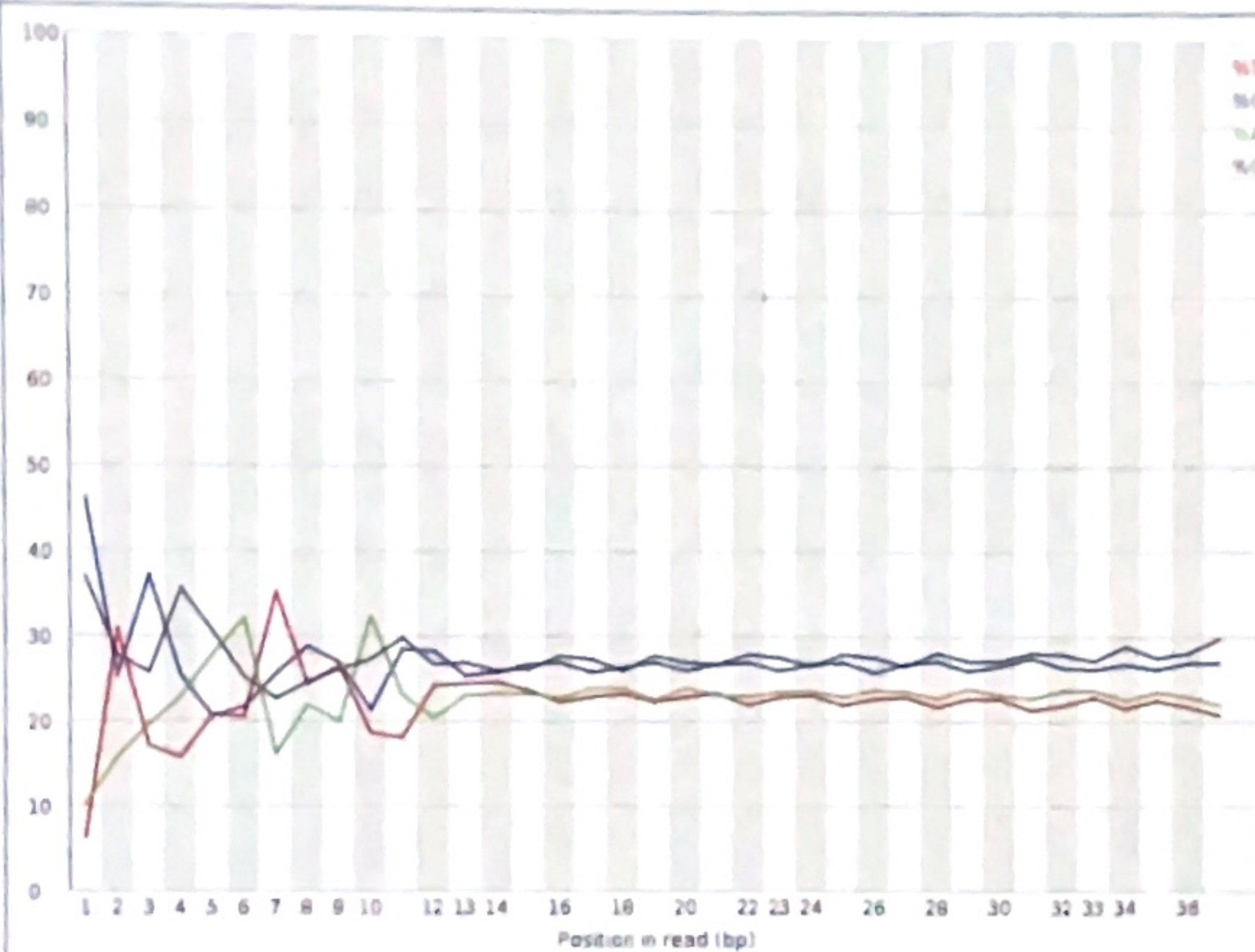


Fig 2

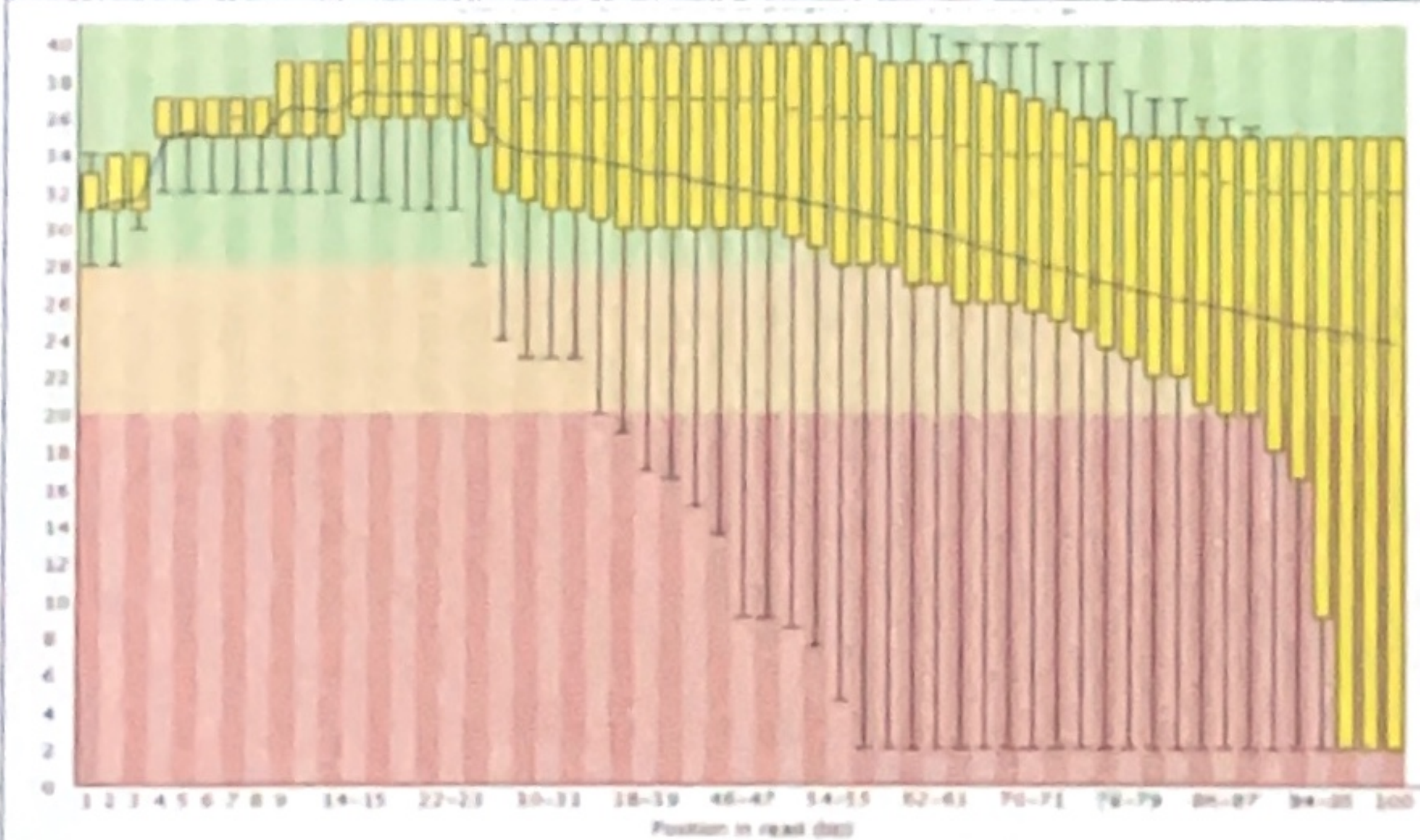


Fig 3

Rédigez une légende pour chaque figure (1,2,3) en complétant le modèle suivant :

"La figure [1/2/3] est un extrait d'un contrôle de qualité pour des reads [courts/longs] et résume [la qualité/ le contenu] des bases au long des reads. Le trimming du [début/fin] des reads [améliorera / n'améliorera pas] les analyses postérieures telle que le mapping et le variants calling"



### Exercice 3 - Format de fichier (3pts)

Voici un extrait de l'historique de galaxy

29: VarScan mpileup on data 28	  
28: samtools mpileup on data 18 pileup	  
21: Samtools stats on data 19	  
20: Samtools flagstat on data 19	  
19: BAM-to-SAM on data 18: converted SAM	   
18: Map with BWA on data 15 (mapped reads in BAM format)	   
17: FastQC on data 15: RawData	  
16: FastQC on data 15: Webpage	  
15: FASTQ Quality Trimmer on data 12	  
14: FastQC on data 12: RawData	  
13: FastQC on data 12: Webpage	  
12: FASTQ Groomer on data 11	  
11: <a href="https://swift.rc.nectar.org.au:8888/v1/AUTH_a3929895f9e94089ad042c9900e1ee82/VariantDet_BASIC/NA12878.GAIIx.exome_chr22.1E6reads.76bp.fastq">https://swift.rc.nectar.org.au:8888/v1/AUTH_a3929895f9e94089ad042c9900e1ee82/VariantDet_BASIC/NA12878.GAIIx.exome_chr22.1E6reads.76bp.fastq</a>	  

Rédigez une légende pour les éléments d'historique 12, 15, 18, 19, 28 et 29 selon le modèle suivant :

"L'élément [12,15/18/19/28/29] est un fichier [SAM/VCF/pileup/BAM/fastq] qui stocke les informations de [reads /mapping/variants] dans un format [binaire/lisible par l'homme]".