

La couverture est le nombre de lectures (*reads*) qui s'alignent sur, ou "couvrent", la séquence de référence connue. Le niveau de couverture du séquençage de nouvelle génération (NGS) détermine souvent si la découverte de variants peut être effectuée avec un certain degré de confiance à des positions de base particulières.

La profondeur de couverture de séquençage varie selon l'application, comme indiquée ci-dessous. À des niveaux de couverture élevés, chaque base est couverte par un grand nombre de lectures de séquences alignées, de sorte que les appels de base (base calls) peuvent être effectués avec un degré de confiance plus élevé.

La plupart des utilisateurs déterminent le niveau de couverture NGS nécessaire en fonction de l'application, ainsi que d'autres facteurs tels que la taille du génome de référence, le niveau d'expression génique, la littérature publiée et les meilleures pratiques définies par la communauté scientifique.

Voici des exemples de recommandations de couverture de séquençage pour certaines applications courantes:

- **Pour détecter les mutations du génome humain, les SNP et les réarrangements**
les publications recommandent souvent une profondeur de couverture de 10x à 30x, en fonction de l'application et du modèle statistique. Nous recommandons habituellement 15x pour les SNV homozygotes, 30x pour les SNV hétérozygotes et 60x pour les INDEL.
- **Pour le séquençage de l'ARN**
les chercheurs pensent généralement en termes de nombres de millions de *reads* à échantillonner. La détection de gènes rarement exprimés nécessite souvent une augmentation de la profondeur de couverture. Nous recommandons généralement des lectures à une seule extrémité (*single-end reads*) de 10 à 25 M pour les gènes exprimés à mi-hauteur et des lectures à une seule extrémité de 50 à 100 M pour une faible expression génique ou une détection de mutation. Utilisez des lectures appariées (*paired-end reads*) pour l'analyse transcriptomique de novo ou la détection d'événements d'épissage. En outre, utilisez au moins 3 réplicats biologiques pour chaque condition.
- **Pour les CHIP-Seq (séquençage d'immunoprécipitation de la chromatine)**
les publications recommandent souvent une couverture d'environ 100x. Habituellement, une lecture à une seule extrémité de 10 à 15 M pour l'analyse d'un facteur de transcription et de 20 à 40 M pour les histones.
- **Pour les petits ARN ou miR-seq**
les lectures simples de 1 à 5M sont généralement suffisantes. Une petite longueur de lecture est standard <50nt.
- **Pour la validation de bibliothèques CRISPR**
une bonne règle est d'avoir > 100 lectures par sgRNA. Contactez-nous avant de commencer votre préparation de bibliothèques comme certains facteurs doivent être pris en considération.

Pour estimer et atteindre votre couverture NGS souhaitée, veuillez nous contacter à l'avance afin d'avoir une meilleure idée de votre conception expérimentale et une estimation des coûts. Vous pouvez toujours utiliser ces outils pour vous guider:

Illumina Coverage Calculator:

https://support.illumina.com/downloads/sequencing_coverage_calculator.html

RNAseq longueur et lectures:

<https://support.illumina.com/bulletins/2017/04/considerations-for-rna-seq-read-length-and-coverage-.html>

Estimateur de couverture par application:

<https://genohub.com/recommended-sequencing-coverage-by-application/>