

Équilibrage de la couverture des échantillons pour le séquençage du génome entier

Stratégies de correction d'index pour Illumina DNA PCR-Free Prep

- Augmentez la couverture minimale des librairies multiplexées
- Atténuez la variabilité des performances des index en ajustant les volumes au point de regroupement des librairies
- Réduisez les déchets et les coûts de séquençage pour les études à haut débit

Augmenter la couverture minimale des échantillons multiplexés

ILLUMINA DNA PCR-Free Prep, Tagmentation (ILLUMINA DNA PCR-Free) offre une solution de préparation de bibliothèques optimisée pour une variété d'applications de séquençage du génome entier (WGS, Whole-Genome Sequencing). Basé sur la chimie transposomique liée aux billes, ILLUMINA DNA PCR-Free fournit une couverture uniforme sur l'ensemble du génome et permet un regroupement facile des bibliothèques en fonction du volume^{1,2}. À l'aide d'ILLUMINA DNA PCR-Free et du NovaSeq[™] 6000 System ou de la série NovaSeq X, les laboratoires à débit élevé peuvent multiplexer les bibliothèques de WGS pour maximiser leur efficacité. Équilibrer le rendement du séquençage des échantillons de WGS multiplexés permet aux utilisateurs d'analyser plus d'échantillons par Flow Cell tout en atteignant la couverture minimale souhaitée pour un appel des variants fiable.

Plusieurs facteurs ont un impact sur la profondeur de couverture par échantillon, notamment le rendement de l'intégralité de la séquence, le nombre d'échantillons par Flow Cell et la variabilité du rendement de l'échantillon. Pour augmenter le nombre d'échantillons par analyse qui obtiennent la couverture de séquence souhaitée, les utilisateurs peuvent soit augmenter la couverture moyenne de tous les échantillons, soit réduire la variation d'un échantillon à l'autre (figure 1). La qualité de l'ADN d'entrée, la variabilité du pipetage et la performance des index contribuent toutes à la variabilité du rendement de la séquence par échantillon. Cette note technique se concentre sur une stratégie visant à atténuer les variations dues à la performance des index.

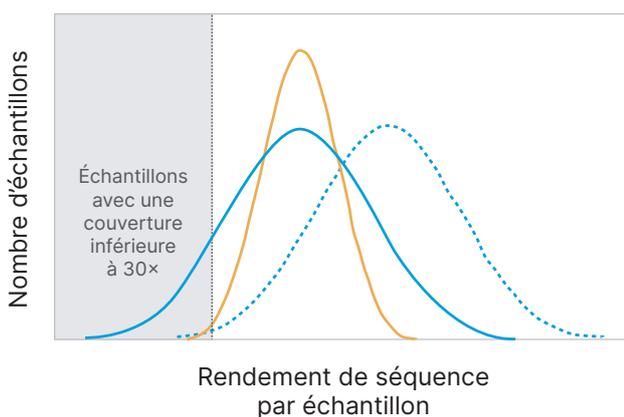


Figure 1 : Amélioration du rendement de séquençage minimal par échantillon : davantage d'échantillons obtiendront une couverture d'au moins 30x avec une augmentation du rendement de séquençage (pointillés bleus) ou une diminution de la variation (orange).

La performance des index affecte la couverture de l'échantillon

Les index ILLUMINA DNA/RNA Unique Dual (UD) Indexes Tagmentation sont disponibles en quatre ensembles de 96 paires d'index (ensembles A, B, C et D) pour un total de 384 paires d'index. Lorsqu'elles sont utilisées avec ILLUMINA DNA PCR-Free^{*} et NovaSeq 6000 System ou la série NovaSeq X[†], certaines paires d'index fournissent systématiquement un rendement de séquence plus élevé ou plus faible. Les paires d'index dont la performance est insuffisante entraînent une couverture inférieure à celle souhaitée pour ces échantillons, tandis que les paires d'index dont la performance dépasse les attentes peuvent réduire la part des lectures pour d'autres échantillons.

Pour démontrer la variabilité de la couverture des échantillons en raison de la séquence d'indexage, nous avons utilisé une version modifiée du protocole de préparation de bibliothèques ILLUMINA DNA PCR-Free. De l'ADN génomique de lignées cellulaires humaines (Coriell Institute, échantillon NA12878) a été utilisé comme entrée et tous les échantillons ont été préparés en lots pour réduire le bruit associé à la qualité de l'ADN d'entrée et aux erreurs de pipetage, respectivement. À l'étape d'indexage, les échantillons ont été aliquotés sur une plaque à 96 puits pour l'indexage, puis regroupés à nouveau pour le nettoyage et la sélection de la taille[‡]. Plusieurs opérateurs ont effectué un total de trois à cinq réplicats pour chaque ensemble d'index doubles uniques (UD, Unique Double) à l'aide de pipettes manuelles. Chaque ensemble de 96 échantillons a été séquençé sur une seule ligne de Flow Cell S4 NovaSeq 6000 avec le flux de travail Xp pour obtenir une représentation des index.

La performance normalisée des index^{††} est indiquée pour les ensembles d'index UD A et B (tableau 1) et les ensembles C et D (tableau 2). Pour cet ensemble de données :

- En moyenne, 32 paires d'index par ensemble de 96 échantillons divergeaient de la performance médiane de plus de 15 % (puits bleus).
- Quelques paires d'index par plaque (puits bleu foncé) ont généralement sous-performé d'au moins 30 %, exposant ainsi ces index à un risque élevé de manquer des cibles de couverture.

* ILLUMINA DNA PCR-Free dispose d'une chimie d'indexage unique; les données de performance des index ne s'appliquent pas aux autres trousseaux de préparation de bibliothèques.

† Une certaine variabilité des index est due à la génération d'amplifiats; les données de performance des index sont spécifiques à chaque système de séquençage.

‡ La sélection de la taille à l'aide de la purification des billes par immobilisation réversible en phase solide (SPRI, Solid-Phase Reversible Immobilization) nécessite de nombreuses étapes de pipetage avec un réactif visqueux qui peut introduire une variation supplémentaire. Regrouper les échantillons immédiatement après l'indexage, par rapport à après la sélection de la taille, a permis de réduire la variation interne (données non présentées).

Tableau 1 : Performances normalisées des index dans un schéma de disposition de plaque à 96 puits pour Illumina DNA PCR-Free et les ensembles A et B des index UD

Ensemble A	1	2	3	4	5	6	7	8	9	10	11	12
A	1,19	1,18	1,07	1,02	0,84	1,20	1,00	0,90	0,98	1,15	1,03	1,03
B	0,82	1,09	0,74	1,02	1,09	1,03	0,77	0,73	1,11	1,09	1,08	0,96
C	0,95	0,82	0,81	0,83	0,78	1,09	0,89	0,99	0,95	1,07	0,96	0,70
D	1,07	0,76	0,92	1,42	1,08	0,96	1,01	1,18	1,10	0,86	1,05	1,28
E	0,95	0,89	1,01	0,85	1,03	0,87	1,00	0,88	1,42	0,88	0,92	1,01
F	1,21	1,07	1,24	1,04	0,91	0,70	1,32	0,97	1,22	1,36	0,95	1,09
G	1,07	0,98	1,21	0,86	0,84	1,20	1,05	1,27	0,95	0,94	1,08	0,98
H	1,16	0,93	1,14	0,80	0,97	1,09	0,91	0,94	1,15	0,73	0,96	1,10

Ensemble B	1	2	3	4	5	6	7	8	9	10	11	12
A	0,89	1,16	0,94	0,99	1,15	0,98	0,96	1,06	0,81	0,95	0,98	0,85
B	0,78	1,04	0,94	1,12	1,06	1,15	1,04	1,39	1,24	1,13	1,17	1,02
C	0,96	0,85	1,11	1,00	1,22	1,26	0,80	1,43	1,00	0,82	1,13	0,90
D	0,98	0,63	0,95	0,92	1,36	1,05	0,84	0,86	0,75	1,19	0,83	1,00
E	1,00	0,83	0,95	1,00	0,87	1,01	1,28	0,94	0,95	1,19	0,97	1,00
F	0,67	1,27	1,12	0,90	0,76	1,13	1,00	1,03	1,35	1,08	1,01	0,91
G	1,04	1,18	1,19	0,99	0,60	1,22	0,98	0,99	1,18	0,80	1,02	1,00
H	0,94	0,85	0,75	0,94	0,92	1,02	1,09	0,95	0,94	0,94	1,50	0,97

Les librairies ont été préparées manuellement avec des réplicats de n = 3 (ensemble A) ou n = 5 (ensemble B) à l'aide d'Illumina DNA PCR-Free Prep et séquencées sur NovaSeq 6000 System. Les puits surlignés en bleu montrent une représentation des index qui divergeait de la médiane de plus de 15 %. Les paires d'index qui ont systématiquement sous-performé d'au moins 30 % (2 de l'ensemble A; 3 de l'ensemble B) sont surlignées en bleu plus foncé. Téléchargez les données de représentation des index à jour et les facteurs de correction calculés sous la forme d'un fichier Excel intitulé « Illumina DNA PCR-Free Prep index correction » (Correction des index Illumina DNA PCR-Free Prep) dans l'onglet « Documentation » à l'adresse illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html.

- Certaines colonnes de plaques individuelles ne présentaient aucune paire d'index, ou une seule paire d'index, avec un écart > 15 % par rapport à la médiane, ce qui a entraîné une performance plus élevée que la plaque globale.

La performance des index était uniforme entre les réplicats ($R^2 = 0,54$ à $0,80$). Par exemple, l'aperçu détaillé de quatre réplicats de la deuxième colonne de la plaque à 96 puits pour l'ensemble d'index UD B (figure 2) montre un modèle cohérent de surreprésentation et de sous-représentation des index.

Cela suggère que la performance des valeurs aberrantes n'est pas simplement du bruit, mais plutôt une propriété fondamentale des paires d'index utilisées. La variabilité constante de la performance des index indique une stratégie claire pour la « correction des index » par rapport à la performance typique des index. En principe, l'ajustement du volume de certaines librairies lors du regroupement avant le séquençage pourrait compenser la représentation variable des index. Les paires d'index performantes devraient fournir une faible variabilité par nature, tandis que les paires d'index qui dévient systématiquement de la valeur médiane sont candidates à la correction d'index.

Tableau 2 : Performances normalisées des index dans un schéma de disposition de plaque à 96 puits pour Illumina DNA PCR-Free et les ensembles C et D des index UD

Ensemble C	1	2	3	4	5	6	7	8	9	10	11	12
A	0,98	0,84	1,12	0,92	0,88	1,40	1,04	0,78	0,89	0,89	1,19	0,85
B	0,70	0,97	0,88	1,13	0,80	1,07	1,21	1,03	1,05	0,86	0,97	1,00
C	0,67	0,96	1,17	1,19	1,05	1,09	1,14	1,07	1,20	1,21	0,86	1,05
D	0,80	1,00	1,23	1,02	1,26	0,89	1,21	0,85	0,60	1,01	1,05	0,92
E	1,15	1,07	0,95	0,85	0,84	0,98	1,16	1,10	0,78	0,88	1,04	1,06
F	1,02	1,07	1,26	1,13	0,47	1,09	1,16	0,91	0,95	1,11	0,80	1,22
G	0,77	0,96	1,17	0,84	1,06	1,11	0,96	0,90	1,14	1,13	1,24	1,09
H	0,78	0,91	1,00	1,02	0,91	1,03	0,98	0,98	0,93	0,99	0,89	0,95

Ensemble D	1	2	3	4	5	6	7	8	9	10	11	12
A	0,91	0,74	1,01	0,98	0,96	0,73	1,03	1,04	0,97	0,77	0,82	0,84
B	1,22	0,95	1,03	1,19	0,84	0,83	1,02	0,80	0,79	1,05	0,84	0,76
C	1,19	1,14	0,99	0,90	0,93	0,93	1,01	0,80	1,26	1,11	1,02	0,96
D	0,86	0,91	1,16	1,25	1,02	1,12	0,95	1,06	0,99	1,09	1,23	0,91
E	0,69	0,88	1,13	1,03	1,43	0,90	1,23	1,12	1,05	1,30	0,88	1,18
F	0,88	1,13	1,29	0,91	0,82	1,15	0,93	0,93	1,18	1,02	1,24	1,20
G	1,18	0,99	1,07	1,02	1,16	0,93	0,98	0,91	1,03	0,77	1,03	1,03
H	0,96	1,12	1,02	1,23	0,99	0,90	0,95	0,88	1,20	0,80	0,88	0,83

Les librairies ont été préparées manuellement avec des réplicats de n = 4 (ensemble C) ou n = 3 (ensemble D) à l'aide d'Illumina DNA PCR-Free Prep et séquencées sur NovaSeq 6000 System. Les puits surlignés en bleu montrent une représentation des index qui divergeait de la médiane de plus de 15 %. Les paires d'index qui ont systématiquement sous-performé d'au moins 30 % (4 de l'ensemble C; 1 de l'ensemble D) sont surlignées en bleu plus foncé. Téléchargez les données de représentation des index à jour et les facteurs de correction calculés sous la forme d'un fichier Excel intitulé « Illumina DNA PCR-Free Prep index correction » (Correction des index Illumina DNA PCR-Free Prep) dans l'onglet « Documentation » à l'adresse illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html.

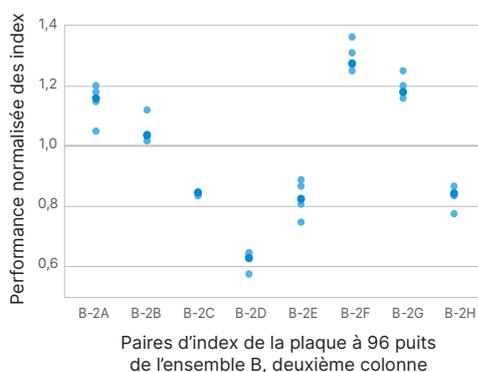


Figure 2 : Les paires d'index qui sous-performent ou surperforment le font de manière systématique : quatre réplicats (bleu clair) et la performance normalisée moyenne (bleu foncé) de huit paires d'index de la deuxième colonne de plaques de l'ensemble d'index UD B. Une paire (B-2D) a systématiquement obtenu une performance > 30 % en dessous de la valeur médiane, et une paire (B-2F) a systématiquement obtenu une performance > 20 % au-dessus de la valeur médiane.

Atténuer la variabilité des performances des index

Les facteurs de correction des index ont été générés en prenant l'inverse de la valeur de représentation médiane des index pour chaque paire d'index indiquée dans le [tableau 1](#) et le [tableau 2](#). La multiplication par ces facteurs pour ajuster le volume de chaque échantillon lors du regroupement, en ajoutant de manière efficace moins d'index surreprésentés et plus d'index sous-représentés, peut rééquilibrer le nombre de lectures de séquençage par échantillon.

Utilisation de 10 % de la médiane comme seuil pour la correction d'index

Pour tester cette stratégie, les facteurs de correction d'index ont été appliqués aux quatre ensembles d'index UD pour les paires d'index qui déviaient d'au moins 10 % de la médiane ([figure 3](#)). Le classement des 384 paires d'index par représentation des index moyens non corrigés a révélé un modèle linéaire avec une bonne corrélation entre quatre réplicats ($R^2 = 0,75$) et une large plage de variabilité (coefficient de variation; CV = 17 %). Après la correction des index, la variation de la performance des index a été réduite (CV = 11 %) et les valeurs corrigées n'étaient plus corrélées à la performance avant correction ($R^2 = 0,006$).

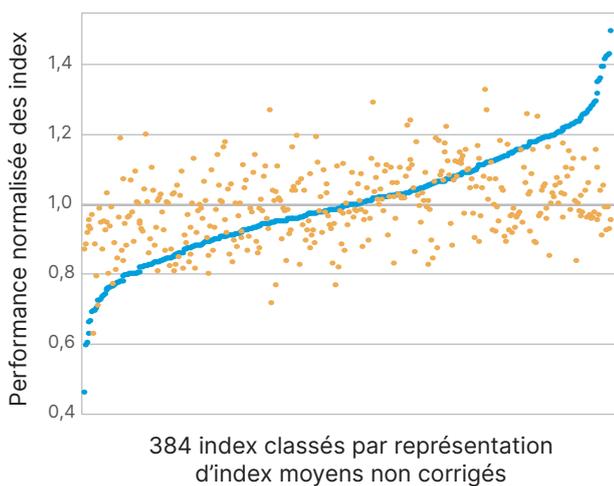


Figure 3 : La correction des index en fonction du volume élimine la variation de performance spécifique aux index : 384 paires d'index classées du rendement le plus faible (gauche) au rendement le plus élevé (droite) pour les expériences non corrigées. Les index non corrigés (bleus) ont présenté un CV de 17 %. Les index corrigés (orange) ont présenté une amélioration du CV de 11 %.

Utilisation de 15 % de la médiane comme seuil pour la correction d'index

À titre de suivi, un opérateur a préparé des bibliothèques avec les 96 paires d'index de l'ensemble d'index UD B ([figure 4A](#)). Pour réduire le bruit dû à la fatigue associée au pipetage manuel, le seuil de correction des index a été augmenté de 10 % à 15 %. Après l'indexage, les bibliothèques ont été regroupées pour sélection de la taille en volumes égaux (bleu) ou en appliquant des facteurs de correction pour tous les index qui déviaient d'au moins 15 % de la médiane (orange). La variation globale a été réduite après correction des index, passant d'un CV de 19 % (non corrigé) à 10 % (corrigé). La deuxième colonne de la plaque pour l'ensemble d'index UD B ([figure 4B](#)) montre également que les paires d'index qui ont constamment sous-performé (B-2D) ou surperformé (B-2F) ont obtenu une performance plus typique.

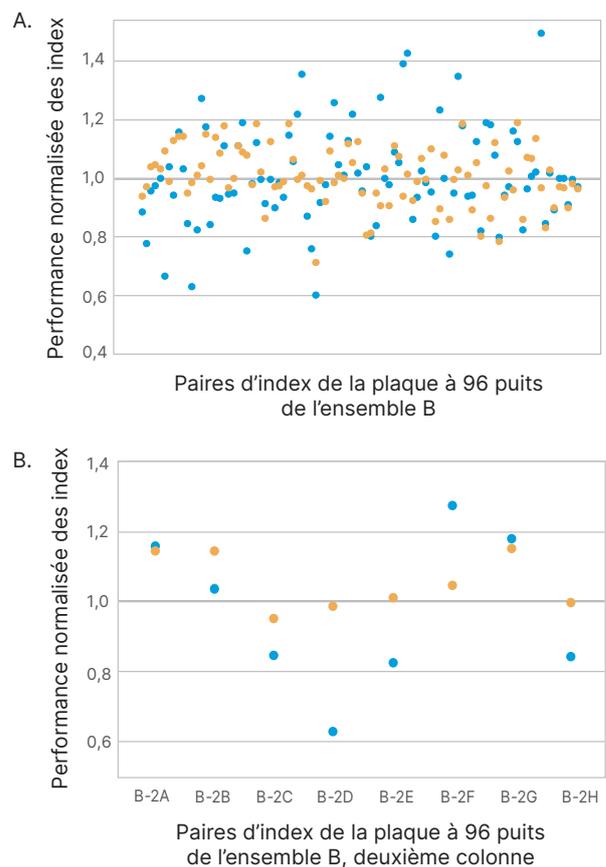


Figure 4 : La correction des index réduit la variabilité des performances des index : (A) 96 paires d'index de l'ensemble d'index UD B. (B) Huit paires d'index de l'ensemble d'index UD B, colonne 2. Les index non corrigés (bleus) ont présenté un CV de 19 %. Les index corrigés (orange) ont présenté une amélioration du CV de 10 %. Les paires d'index qui déviaient de la médiane de plus de 20 % (B-2D et B-2F) sont revenues à la performance médiane des index après correction.

La dernière expérience portait sur trois colonnes de plaques (7, 8, 9) de l'ensemble d'index UD B et sur 11 paires d'index corrigées qui divergeaient généralement d'au moins 15 % de la médiane. Les 24 échantillons pour chaque condition (corrigée ou non) ont été analysés sur une Flow Cell S4 NovaSeq 6000 (figure 5). Comme prévu, le CV de la représentation des index est passé de 18 % (non corrigé) à 7 % (corrigé). Bien que la couverture moyenne soit restée uniforme pour tous les échantillons, la couverture minimale était plus élevée pour les échantillons corrigés (données non présentées).

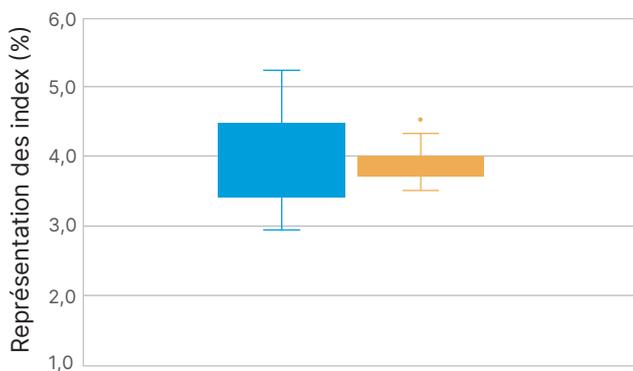


Figure 5 : La correction des index a réduit la variabilité entre les échantillons séquencés sur la même Flow Cell : performance des index pour les bibliothèques utilisant 24 paires d'index de l'ensemble d'index UD B, colonnes de plaques 7, 8, 9. Les index non corrigés (bleus) ont présenté un CV de 18 %. Les index corrigés (orange) ont présenté une amélioration du CV de 7 %.

Réduire les déchets et les coûts de séquençage pour les études à haut débit

Choisir les paires d'index les plus performantes

Ces chiffres de performance des index Illumina DNA PCR-Free indiquent quels index devraient offrir une meilleure couverture et quelles colonnes montrent une faible variation de la performance des index (tableau 3, surligné en orange). Sur une plaque de 96 échantillons, le CV typique de la représentation des index est de 15 à 20 %. Les clients qui utilisent des sous-sections d'une plaque peuvent choisir des colonnes avec le CV le plus bas pour améliorer les performances sans correction.

Mettre en œuvre des facteurs de correction des index pour réduire la variabilité

Ces résultats démontrent également le principe selon lequel l'ajustement des volumes de bibliothèques dans le regroupement peut corriger une mauvaise performance des index. Le regroupement d'échantillons basé sur le volume et guidé par les facteurs de correction d'index de cet ensemble de données était suffisant pour réduire de manière fiable les variations. L'un des pièges pour les opérateurs manuels est de corriger un trop grand nombre de paires d'index. La correction manuelle des index est un processus fastidieux qui peut augmenter le bruit en raison d'une erreur de pipetage. Nous avons observé de meilleurs résultats en utilisant un seuil plus élevé pour la correction des index (c.-à-d., la correction des index avec une performance > 15 % par rapport à la médiane, contre > 10 % par rapport à la médiane). Même la correction d'un petit nombre des paires d'index les moins performantes devrait « sauver » certains échantillons de la perte en raison d'une couverture inférieure à celle attendue.

Tableau 3 : Variabilité des performances des index (CV) par colonne de plaque et par plaque pour les index Illumina DNA/RNA UD Indexes

Colonne ^a	1	2	3	4	5	6	7	8	9	10	11	12	Plaque ^b
Ensemble A	12,9 %	14,7 %	18,1 %	20,7 %	12,5 %	16,7 %	16,1 %	17,5 %	14,4 %	19,8 %	6,4 %	16,3 %	15,5 %
Ensemble B	13,8 %	22,6 %	14,1 %	7,0 %	25,2 %	9,4 %	14,9 %	19,7 %	20,6 %	15,4 %	18,6 %	6,7 %	16,6 %
Ensemble C	19,9 %	7,9 %	12,6 %	13,3 %	25,7 %	13,4 %	9,1 %	11,8 %	20,5 %	12,8 %	15,5 %	11,2 %	15,7 %
Ensemble D	19,3 %	14,4 %	9,3 %	13,1 %	19,3 %	14,8 %	9,5 %	12,7 %	14,2 %	19,4 %	17,1 %	16,8 %	15,1 %

a. CV calculé à partir des données de performance moyenne des index dans le tableau 1 et le tableau 2. Les bibliothèques ont été préparées manuellement avec 3 à 5 réplicats à l'aide d'Illumina DNA PCR-Free Prep et séquencées sur NovaSeq 6000 System. Les surbrillances montrent les colonnes dont le CV de représentation d'index est inférieur à 10 % (orange foncé) ou inférieur à 15 % (orange clair).

b. Alors que le CV de la colonne varie de 6,4 % à 25,7 %, le CV de la plaque à 96 puits ne varie que de 15,1 % à 16,6 %.

Considérez la correction d'index comme un processus dynamique avec des ajustements itératifs à mesure que vous recueillez davantage de données. Ces données représentent un point de départ pour d'autres études[§]. L'automatisation peut être nécessaire pour réduire le bruit de fond et obtenir tous les avantages.

Résumé

La performance des index est un facteur clé de variabilité de la couverture des échantillons pour les expériences utilisant Illumina DNA PCR-Free et NovaSeq 6000 System ou la série NovaSeq X. Certaines séquences d'index fournissent systématiquement un rendement plus élevé ou plus faible. En ajustant les volumes pendant le regroupement, vous pouvez réduire la variabilité du rendement de séquence des échantillons séquencés sur la même Flow Cell. Bien que cela n'augmente pas le rendement global du séquençage, cela augmente la couverture minimale, réduisant ainsi le nombre d'échantillons qui tombent en dessous du seuil de couverture minimal défini. Les laboratoires menant des recherches génomiques sur la population et d'autres applications de séquençage du génome humain entier à débit élevé peuvent utiliser les données fournies comme point de départ pour optimiser le débit des échantillons et réduire les déchets.

§ Accédez aux données de représentation des index à jour et aux facteurs de correction calculés pour les ensembles d'index v3 avec Illumina DNA PCR-Free pour NovaSeq 6000 System et la série NovaSeq X. Téléchargez un fichier Excel intitulé « Illumina DNA PCR-Free Prep index correction » (Correction des index Illumina DNA PCR-Free Prep) dans l'onglet « Documentation » à l'adresse illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html.

En savoir plus

[Illumina DNA PCR-Free Prep](#)

Références

1. Illumina. Illumina DNA PCR-Free Prep, Tagmentation. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-dna-pcr-free-data-sheet-m-gl-00679/illumina-dna-pcr-free-data-sheet-m-gl-00679.pdf. Publié en 2020. Mis à jour en 2023. Consulté le 6 septembre 2024.
2. Bruinsma S, Burgess J, Schlingman D, et al. [Bead-linked transposomes enable a normalization-free workflow for NGS library preparation](#). *BMC Genomics*. 2018;19(1):722. doi:10.1186/s12864-018-5096-9.



Numéro sans frais aux États-Unis : + (1) 800 809-4566 | Téléphone : + (1) 858 202-4566
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Tous droits réservés. Toutes les marques de commerce sont la propriété d'Illumina, Inc. ou de leurs détenteurs respectifs. Pour obtenir des renseignements sur les marques de commerce, consultez la page www.illumina.com/company/legal.html.
M-GL-00005 FRA v2.0.