

# Equilibrado de la cobertura de muestras para la secuenciación del genoma completo

## Estrategias de corrección de índices para Illumina DNA PCR-Free Prep

- Aumente la cobertura mínima en librerías multiplexadas.
- Atenúe la variabilidad en el rendimiento de los índices ajustando los volúmenes en el momento de agrupar las librerías.
- Reduzca los costes de residuos y secuenciación en estudios de alta productividad.



## Aumente la cobertura mínima en muestras multiplexadas

Illumina DNA PCR-Free Prep, Tagmentation (Illumina DNA PCR-Free) ofrece una solución de preparación de librerías optimizada para diferentes aplicaciones de secuenciación del genoma completo (WGS, whole-genome sequencing). Mediante la química del transposoma ligado por bolas, Illumina DNA PCR-Free proporciona una cobertura uniforme en todo el genoma y admite una agrupación sencilla de librerías por volumen.<sup>1,2</sup> Mediante Illumina DNA PCR-Free y NovaSeq™ 6000 System o NovaSeq X Series, los laboratorios con una elevada productividad pueden multiplexar librerías de WGS para lograr una máxima eficiencia. El equilibrio del rendimiento de la secuenciación en muestras de WGS multiplexadas permite a los usuarios procesar más muestras por celda de flujo al tiempo que logran la cobertura mínima deseada para realizar la llamada de variantes con confianza.

Hay diversos factores que afectan a la profundidad de cobertura por muestra, como el rendimiento total de la secuencia, el número de muestras por celda de flujo y la variabilidad en el rendimiento de la muestra. Para aumentar el número de muestras por experimento que logran la cobertura de secuencia deseada, los usuarios pueden aumentar la cobertura media en todas las muestras o reducir la variación entre muestras (figura 1). La calidad del aporte de ADN, la variabilidad del pipeteo y el rendimiento de los índices contribuyen a la variabilidad en el rendimiento de la secuenciación por muestra. Esta nota técnica se centra en una estrategia para reducir la variación debida al rendimiento de los índices.

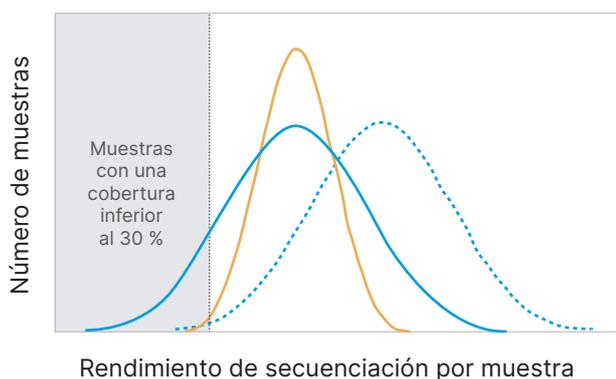


Figura 1: Mejora del rendimiento mínimo de secuenciación por muestra. Se logra una cobertura de al menos 30x en un mayor número de muestras, con un aumento del rendimiento de las secuenciación (línea de puntos azul) o una disminución de la variación (naranja).

## El rendimiento de los índices afecta a la cobertura de las muestras

Illumina DNA/RNA Unique Dual Indexes (UD), Tagmentation están disponibles en cuatro juegos de 96 pares de índices (juegos A, B, C y D) para un total de 384 pares de índices. Cuando se utiliza con Illumina DNA PCR-Free\* y NovaSeq 6000 System o NovaSeq X Series,<sup>†</sup> algunos pares de índices proporcionan de forma sistemática un rendimiento de secuenciación mayor o menor. Los pares de índices con un rendimiento que no alcanza las expectativas no dan lugar a la cobertura deseada para esas muestras, mientras que los pares de índices con un rendimiento por encima de lo esperado pueden reducir la proporción de lecturas para otras muestras.

Para demostrar la variabilidad de la cobertura de las muestras debido a la secuencia de índice, empleamos una versión modificada del protocolo de preparación de librerías Illumina DNA PCR-Free. Se utilizó ADN genómico de línea celular humana (Coriell Institute, muestra NA12878) como aporte y todas las muestras se prepararon a granel para reducir el ruido asociado a la calidad del ADN de aporte y a los errores de pipeteo, respectivamente. En el paso de indexado, las muestras se dividieron en alícuotas en una placa de 96 pocillos para indexado y, a continuación, se agruparon de nuevo para la limpieza de selección por tamaño.<sup>‡</sup> Diversos operadores llevaron a cabo un total de tres a cinco réplicas para cada juego de índices UD con pipetas manuales. Para obtener la representación de índices, cada conjunto de 96 muestras se secuenció en un único carril de una celda de flujo S4 de NovaSeq 6000 con el flujo de trabajo de Xp.

Se muestra el rendimiento normalizado de los índices\*\* para los juegos de índices UD A y B (tabla 1) y los juegos C y D (tabla 2). Para este conjunto de datos:

- De media, 32 pares de índices por juego de 96 mostraron divergencias de más del 15 % respecto a la mediana del rendimiento (pocillos en azul)
- Unos pocos pares de índices por placa (pocillos en azul oscuro) normalmente registraron un rendimiento inferior al 30 %, lo que pone a estos índices en alto riesgo de incumplimiento de los objetivos de cobertura

\* Illumina DNA PCR-Free tiene una química de indexado única; los datos de rendimiento de los índices no se aplican a otros kits de preparación de librerías.

† Parte de la variabilidad de los índices se debe a la agrupación; los datos de rendimiento de los índices son específicos para cada sistema de secuenciación.

‡ Selección por tamaño mediante inmovilización reversible en fase sólida (SPRI, Solid-Phase reversible Immobilization). La purificación de bolas emplea muchos pasos de pipeteo con un reactivo viscoso que puede introducir variaciones adicionales. La agrupación de muestras inmediatamente después de la indexación, en lugar de después de la selección por tamaño, redujo la variación interna (no se muestran los datos).

Tabla 1: Rendimiento de índices normalizado en la disposición de placas de 96 pocillos para Illumina DNA PCR-Free y los juegos de índices UD A y B

Juego 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

Juego 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

La preparación de librerías se realizó manualmente con réplicas de n = 3 (juego A) o n = 5 (juego B) con Illumina DNA PCR-Free Prep y se secuenciaron en NovaSeq 6000 System. Los pocillos resaltados en azul muestran una representación de índices con una divergencia respecto a la mediana de más del 15 %. Los pares de índices que registraron un rendimiento inferior de forma sistemática en al menos un 30 % (2 del juego A; 3 del juego B) se resaltan en azul más oscuro. Descargue un archivo de Excel, llamado «Illumina DNA PCR-Free Prep index correction» (Corrección de índices de Illumina DNA PCR-Free Prep) en la pestaña «Documentation» (Documentación) en [illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html](https://illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html).

- Algunas columnas individuales de placas no mostraron pares de índices, o solo un par de índices, con una desviación superior al 15 % de la mediana, lo que produjo un mayor rendimiento que la placa en su conjunto.

El rendimiento de los índices fue uniforme entre las réplicas ( $R^2 = 0,54-0,80$ ). Por ejemplo, si nos fijamos en cuatro réplicas de la segunda columna de la placa de 96 pocillos para el juego de índices UD B (figura 2), se observa un patrón uniforme de sobre e infrarrepresentación de los índices.

Esto sugiere que el rendimiento de los valores atípicos no es simplemente ruido, sino una propiedad fundamental de los pares de índices utilizados. La variabilidad constante en el rendimiento de los índices apunta a una estrategia clara para la «corrección de índices» en relación con el rendimiento típico de los índices. En principio, el ajuste del volumen de ciertas librerías en la agrupación antes de la secuenciación podría compensar la representación variable de los índices. Se espera que los pares de índices con un buen rendimiento proporcionen una baja variabilidad inherente, mientras que los pares de índices que se desvían sistemáticamente del valor medio son candidatos para la corrección del índice.

Tabla 2: Rendimiento de índices normalizado en la disposición de placas de 96 pocillos para Illumina DNA PCR-Free y los juegos de índices UD C y D

Juego 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

Juego 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

La preparación de librerías se realizó manualmente con réplicas de n = 4 (juego C) o n = 3 (juego D) con Illumina DNA PCR-Free Prep y se secuenciaron en NovaSeq 6000 System. Los pocillos resaltados en azul muestran una representación de índices con una divergencia respecto a la mediana de más del 15 %. Los pares de índices que registraron un rendimiento inferior de forma sistemática en al menos un 30 % (4 del juego C; 1 del juego D) se resaltan en azul más oscuro. Descargue un archivo de Excel, llamado «Illumina DNA PCR-Free Prep index correction» (Corrección de índices de Illumina DNA PCR-Free Prep) en la pestaña «Documentation» (Documentación) en [illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html](https://illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html).

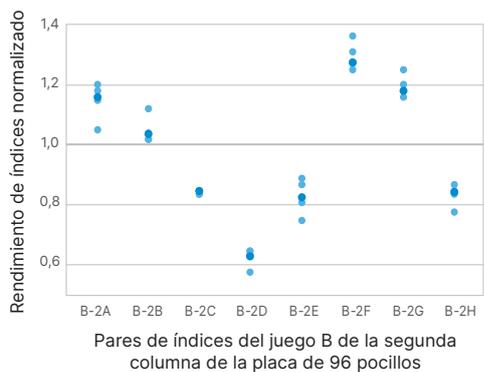


Figura 2: Los pares de índices que tienen un rendimiento inferior o superior a lo esperado lo hacen de forma sistemática. Cuatro réplicas (azul claro) y rendimiento normalizado medio (azul oscuro) de ocho pares de índices de la segunda columna de la placa del juego de índices UD B. Un par (B-2D) tuvo un rendimiento constante más del 30 % por debajo del valor de la mediana y un par (B-2F) tuvo un rendimiento constante más del 20 % por encima del valor de la mediana.

## Reducción de la variabilidad en el rendimiento de los índices

Los factores de corrección de los índices se generaron tomando el valor inverso de la mediana de la representación de cada par de índices que se muestran en la [tabla 1](#) y la [tabla 2](#). La multiplicación por estos factores para ajustar el volumen de cada muestra en la agrupación, con lo que se introducen de forma eficaz menos índices sobrerrepresentados y más índices infrarrepresentados, reequilibra el número de lecturas de secuenciación por muestra.

### Uso del 10 % de la mediana como umbral para la corrección de índices

Para probar esta estrategia, se aplicaron factores de corrección de índices en los cuatro juegos de índices UD para los pares de índices que se desviaron al menos un 10 % de la mediana ([figura 3](#)). La disposición de los 384 pares de índices por representación del índice medio no corregido reveló un patrón lineal con una buena correlación entre las cuatro réplicas ( $R^2 = 0,75$ ) y un amplio intervalo de variabilidad (coeficiente de variación; CV = 17 %). Después de la corrección de índices, se redujo la variación en el rendimiento de los índices (CV = 11 %) y los valores corregidos ya no mostraron correlación con el rendimiento previo a la corrección ( $R^2 = 0,006$ ).

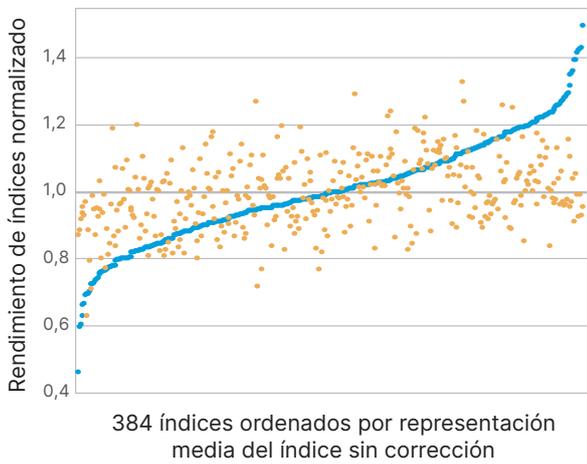


Figura 3: La corrección de índices basada en el volumen elimina la variación específica de los índice en el rendimiento. Se muestran 384 pares de índices ordenados del rendimiento de secuencia más bajo (izquierda) al más alto (derecha) para experimentos no corregidos. Los índices no corregidos (azul) mostraron un CV del 17 %. Los índices corregidos (naranja) mostraron un CV mejorado del 11 %.

### Uso del 15 % de la mediana como umbral para la corrección de índices

Como seguimiento, un operador preparó librerías con los 96 pares de índices del juego de índices UD B ([figura 4A](#)). Para reducir el ruido debido a la fatiga del pipeteo manual, el umbral para la corrección de índices se aumentó del 10 % al 15 %. Después de la indexación, las librerías se agruparon para la selección por tamaño en volúmenes iguales (azul) o mediante la implementación de factores de corrección para todos los índices que se desviaron al menos un 15 % de la mediana (naranja). Tras la corrección de los índices, la variación global se redujo de un CV del 19 % (sin corregir) al 10 % (corregido). Si nos fijamos de nuevo en la segunda columna de la placa del juego de índices UD B ([figura 4B](#)), se observan que los pares de índices que registraron un rendimiento inferior (B-2D) o superior (B-2F) de forma sistemática se convirtieron en valores más típicos.

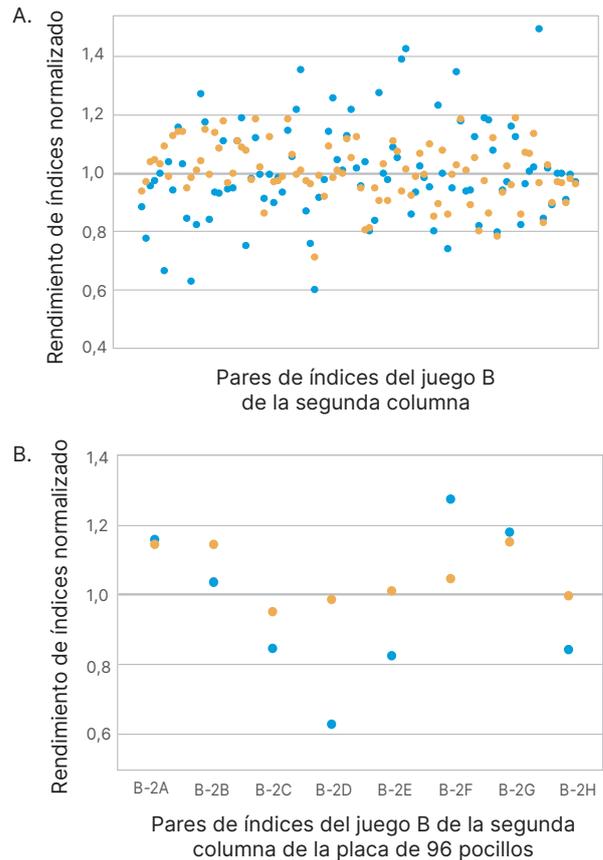


Figura 4: La corrección de índices reduce la variabilidad del rendimiento de los índices. (A) 96 pares de índices del juego de índices UD B. (B) Ocho pares de índices del juego de índices UD B, columna 2. Los índices no corregidos (azul) mostraron un CV del 19 %. Los índices corregidos (naranja) mostraron un CV mejorado del 10 %. Los pares de índices que se desviaron de la mediana en más del 20 % (B-2D y B-2F) volvieron a la mediana del rendimiento del índice después de la corrección.

Un último experimento se centró en tres columnas (7, 8, 9) de la placa del juego de índices UD B, y en él se corrigieron 11 pares de índices que normalmente divergían al menos un 15 % de la mediana. Las 24 muestras para cada condición (corregidas o no corregidas) se procesaron en una celda de flujo S4 de NovaSeq 6000 (figura 5). Como cabía esperar, el CV de la representación de los índices disminuyó del 18 % (sin corregir) al 7 % (corregido). Aunque la cobertura media se mantuvo constante en todas las muestras, la cobertura mínima fue mayor para las muestras corregidas (no se muestran los datos).

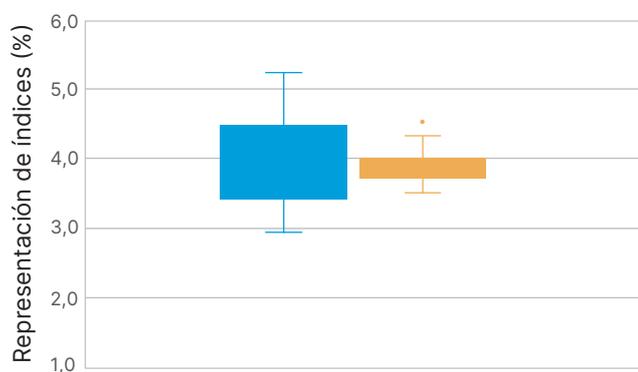


Figura 5: La corrección de índices redujo la variabilidad entre las muestras secuenciadas en la misma celda de flujo. Rendimiento de los índices para las librerías, utilizando 24 pares de índices del juego de índices UD B, columnas 7, 8 y 9 de la placa. Los índices no corregidos (azul) mostraron un CV del 18 %. Los índices corregidos (naranja) mostraron un CV mejorado del 7 %.

## Reduzca los costes de residuos y secuenciación en estudios de alta productividad.

### Elija los pares de índices de mayor rendimiento

Estas cifras de rendimiento de los índices de Illumina DNA PCR-Free indican qué índices se espera que proporcionen una mejor cobertura y qué columnas muestran una baja variación en el rendimiento de los índices (tabla 3, resultado en naranja). En una placa de 96 muestras, el CV típico de la representación de los índices es del 15 al 20 %. Los clientes que utilizan subsecciones de una placa pueden elegir columnas con el CV más bajo para mejorar el rendimiento sin corrección.

### Implemente factores de corrección de índices para reducir la variabilidad

Estos resultados también demuestran el principio de que el ajuste de los volúmenes de las librerías en la agrupación puede corregir el bajo rendimiento de los índices. La agrupación de muestras basada en el volumen guiada por los factores de corrección de los índices de este conjunto de datos fue suficiente para reducir de forma fiable la variación. La necesidad de corregir demasiados pares de índices puede ser un inconveniente para los operadores manuales. La corrección manual de índices es un proceso tedioso que puede aumentar el ruido debido a un error de pipeteo. Observamos mejores resultados utilizando un umbral más alto para la corrección de índices (es decir, corrección de índices con un rendimiento superior al 15 % con respecto a la mediana, frente a superior al 10 % con respecto a la mediana). Se espera que incluso la corrección de un pequeño número de los pares de índices de peor rendimiento evite la retirada de algunas muestras debido a una cobertura inferior a la esperada.

Tabla 3: Variabilidad del rendimiento de los índices (CV) por columna de placa y por placa para los índices UD de ADN/ARN de Illumina

Columna <sup>a</sup>	1	2	3	4	5	6	7	8	9	10	11	12	Placa <sup>b</sup>
Juego 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 %
Juego 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 %
Juego 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 %
Juego 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 calculado a partir de los datos de rendimiento medio de los índices en la tabla 1 y la tabla 2. La preparación de librerías se realizó manualmente con de 3 a 5 réplicas con Illumina DNA PCR-Free Prep y se secuenciaron en NovaSeq 6000 System. Las partes resaltadas muestran columnas con un CV de representación de los índices inferior al 10 % (naranja oscuro) o inferior al 15 % (naranja claro).  
 b. Aunque el CV de la columna oscila entre el 6,4 y el 25,7 %, el CV de la placa de 96 pocillos solo oscila entre el 15,1 y el 16,6 %.

Considere la corrección de índices como un proceso dinámico con ajustes reiterados a medida que recopila más datos. Estos datos representan un punto de partida para estudios posteriores.<sup>§</sup> La automatización puede ser necesaria para reducir el ruido de fondo y obtener todas las ventajas.

## Resumen

El rendimiento de los índices es un factor clave en la variabilidad de la cobertura de muestras para experimentos en los que se utiliza Illumina DNA PCR-Free y NovaSeq 6000 System o NovaSeq X Series. Algunas secuencias de índices proporcionan de forma sistemática un rendimiento mayor o menor. Al ajustar los volúmenes durante la agrupación, puede reducir la variabilidad del rendimiento de la secuenciación entre las muestras secuenciadas en la misma celda de flujo. Aunque esto no aumenta el rendimiento general de la secuenciación, sí aumenta la cobertura mínima, lo que reduce el número de muestras que quedan por debajo de un umbral de cobertura mínima. Los laboratorios que llevan a cabo investigaciones genómicas poblacionales y otras aplicaciones de WGS humano de alta productividad pueden utilizar los datos proporcionados como punto de partida para optimizar el rendimiento de las muestras y reducir los residuos.

§ Acceda a los datos de representación de índices actualizados y a los factores de corrección calculados para la v3 de los juegos de índices con Illumina DNA PCR-Free, tanto para NovaSeq 6000 System como para NovaSeq X Series. Descargue un archivo de Excel, llamado «Illumina DNA PCR-Free Prep index correction» (Corrección de índices de Illumina DNA PCR-Free Prep) en la pestaña «Documentation» (Documentación) en [illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html](https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html).

## Más información

[Illumina DNA PCR-Free Prep](#)

## Bibliografía

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](https://www.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). Año de publicación: 2020. Año de actualización: 2023. Fecha de consulta: 6 de septiembre de 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.



1 800 809 4566 (llamada gratuita, EE. UU.) | tel.: +1 858 202 4566  
techsupport@illumina.com | [www.illumina.com](https://www.illumina.com)

© 2024 Illumina, Inc. Todos los derechos reservados. Todas las marcas comerciales pertenecen a Illumina, Inc. o a sus respectivos propietarios. Si desea consultar información específica sobre las marcas comerciales, consulte [www.illumina.com/company/legal.html](https://www.illumina.com/company/legal.html).  
M-GL-00005 ESP v2.0.