

Bilanciamento della copertura del campione per il sequenziamento dell'intero genoma

Strategie di correzione dell'indice per Illumina DNA PCR-Free Prep

- Aumento della copertura minima su librerie in multiplex
- Riduzione della variabilità nelle prestazioni dell'indice regolando i volumi nel punto di raggruppamento in pool della libreria
- Riduzione degli sprechi e dei costi di sequenziamento per gli studi a elevata processività



Aumento della copertura minima su campioni in multiplex

Illumina DNA PCR-Free Prep, Tagmentation (Illumina DNA PCR-Free) offre una soluzione ottimizzata per la preparazione delle librerie per una varietà di applicazioni di sequenziamento dell'intero genoma (WGS, whole-genome sequencing). In base alla chimica del trasposoma legato alle microsfere, Illumina DNA PCR-Free fornisce una copertura uniforme su tutto il genoma e supporta un facile raggruppamento in pool delle librerie basato sul volume.^{1,2} Grazie a Illumina DNA PCR-Free e NovaSeq™ 6000 System o NovaSeq X Series, i laboratori a elevata processività possono eseguire il multiplex delle librerie WGS per massimizzare l'efficienza. Il bilanciamento della resa di sequenziamento su campioni WGS in multiplex consente agli utenti di analizzare più campioni per cella a flusso, ottenendo nel contempo la copertura minima desiderata per l'identificazione sicura di varianti.

Sono molti i fattori che influiscono sulla profondità di copertura per campione, tra cui la resa totale della sequenza, il numero di campioni per cella a flusso e la variabilità della resa del campione. Per aumentare il numero di campioni per corsa che raggiungono la copertura di sequenza desiderata, gli utenti possono aumentare la copertura media su tutti i campioni o ridurre la variazione da campione a campione (Figura 1). A contribuire alla variabilità della resa delle sequenze per campione ci sono fattori come la qualità del DNA di input, la variabilità del pipettaggio e le prestazioni dell'indice. Questa nota tecnica si concentra su una strategia per mitigare le variazioni dovute alle prestazioni dell'indice.

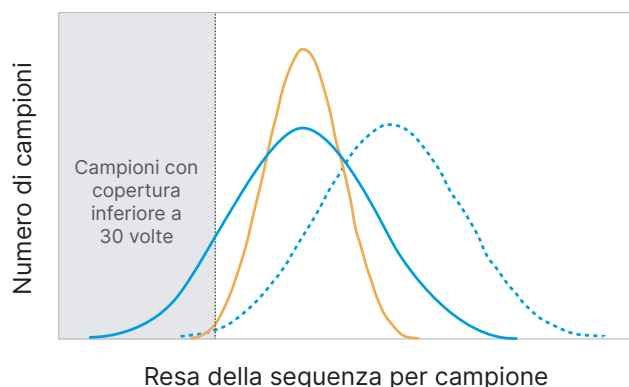


Figura 1: miglioamento della resa minima di sequenziamento per campione. Più campioni otterranno una copertura di almeno 30 volte con un aumento della resa di sequenziamento (linea tratteggiata blu) o una riduzione della variazione (linea arancione).

Influenza delle prestazioni dell'indice sulla copertura del campione

Illumina DNA/RNA Unique Dual (UD) Indexes, Tagmentation sono disponibili in quattro set di 96 coppie di indici (set A, B, C e D) per 384 coppie di indici totali. Quando vengono utilizzate con Illumina DNA PCR-Free* System e NovaSeq 6000 System o NovaSeq X Series,[†] alcune coppie di indici forniscono costantemente una resa di sequenza superiore o inferiore. Le coppie di indici con prestazioni limitate generano una copertura inferiore a quella desiderata per quei campioni, mentre le coppie di indici con prestazioni superiori possono ridurre la quota di letture per altri campioni.

Per dimostrare la variabilità della copertura del campione dovuta alla sequenza d'indice, abbiamo utilizzato una versione modificata del protocollo di preparazione delle librerie Illumina DNA PCR-Free. Il DNA genomico della linea cellulare umana (Coriell Institute, campione NA12878) è stato utilizzato come input e tutti i campioni sono stati preparati in massa per ridurre il rumore associato rispettivamente alla qualità del DNA di input e agli errori di pipettaggio. Nella fase di indicizzazione, i campioni sono stati distribuiti in aliquote su una piastra a 96 pozzetti per l'indicizzazione, quindi raggruppati nuovamente in pool per la pulizia della selezione delle dimensioni.[‡] Più operatori hanno condotto un totale di tre-cinque replicati per ogni set di UD Index utilizzando pipette manuali. Ogni set da 96 campioni è stato sequenziato su una singola corsia di una cella a flusso NovaSeq 6000 S4 con il flusso di lavoro Xp per ottenere la rappresentazione dell'indice.

Le prestazioni normalizzate dell'indice** sono mostrate per i set A e B (Tabella 1) e i set C e D (Tabella 2) degli UD Index. Per questo set di dati:

- In media, 32 coppie di indici per set di 96 elementi divergevano dalle prestazioni mediane di oltre il 15% (pozzetti blu).
- Alcune coppie di indici per piastra (pozzetti blu scuro) hanno generalmente avuto prestazioni limitate di almeno il 30%, esponendo questi indici a un rischio elevato di mancare gli obiettivi di copertura.

* Illumina DNA PCR-Free dispone di una chimica di indicizzazione univoca; i dati sulle prestazioni dell'indice non si applicano ad altri kit di preparazione delle librerie.

† Una certa variabilità dell'indice è dovuta alla generazione di cluster; i dati delle prestazioni dell'indice sono specifici per ciascun sistema di sequenziamento.

‡ La selezione delle dimensioni mediante immobilizzazione reversibile in fase solida (SPRI, solid-phase reversible immobilization)-purificazione delle microsfere impiega molte fasi di pipettaggio con un reagente viscoso che può introdurre ulteriori variazioni. Il raggruppamento dei campioni in pool subito dopo l'indicizzazione, anziché dopo la selezione delle dimensioni, ha ridotto la variazione interna (dati non mostrati).

Tabella 1: prestazioni dell'indice normalizzate nel layout di una piastra a 96 pozzetti per i set A e B di UD Index e Illumina DNA PCR-Free

Set 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

Set 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 preparazione delle librerie è stata eseguita manualmente con replicati di $n = 3$ (set A) o $n = 5$ (set B) utilizzando Illumina DNA PCR-Free Prep e sequenziati su NovaSeq 6000 System. I pozzetti evidenziati in blu mostrano una rappresentazione dell'indice che si è discostata dalla mediana di oltre il 15%. Le coppie di indici che hanno costantemente avuto prestazioni limitate di almeno il 30% (2 del set A; 3 del set B) sono evidenziate in blu scuro. Scaricare i dati aggiornati sulla rappresentazione degli indici e i fattori di correzione calcolati sotto forma di file Excel denominato "Illumina DNA PCR-Free Prep index correction" nella scheda "Documentation" (Documentazione) all'indirizzo illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html.

- Alcune singole colonne della piastra non mostravano coppie di indici, o ne mostravano solo una, con una deviazione maggiore del 15% dalla mediana, producendo prestazioni più elevate rispetto alla piastra complessiva.

Le prestazioni dell'indice erano coerenti tra i replicati ($R^2 = 0,54-0,80$). Ad esempio, se ci si focalizza su quattro replicati della seconda colonna della piastra a 96 pozzetti per il set B di UD Index (Figura 2), si osserva un pattern coerente di sovrarappresentazione e sottorappresentazione dell'indice.

Ciò suggerisce che le prestazioni dei valori erratici non sono semplicemente rumore, ma piuttosto una proprietà fondamentale delle coppie di indici utilizzate. La variabilità coerente nelle prestazioni dell'indice indica una strategia chiara per la "correzione dell'indice" in relazione alle prestazioni tipiche dell'indice. In linea di principio, la regolazione del volume di alcune librerie al raggruppamento in pool prima del sequenziamento potrebbe compensare la rappresentazione degli indici variabili. Si prevede che le coppie di indici con buone prestazioni forniscano intrinsecamente una bassa variabilità, mentre le coppie di indici che deviano costantemente dal valore mediano sono candidate per la correzione dell'indice.

Tabella 2: prestazioni dell'indice normalizzate nel layout di una piastra a 96 pozzetti per i set C e D di UD Index e Illumina DNA PCR-Free

Set 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

Set 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 preparazione delle librerie è stata eseguita manualmente con replicati di n = 4 (set C) o n = 3 (set D) utilizzando Illumina DNA PCR-Free Prep e sequenziati su NovaSeq 6000 System. I pozzetti evidenziati in blu mostrano una rappresentazione dell'indice che si è discostata dalla mediana di oltre il 15%. Le coppie di indici che hanno costantemente avuto prestazioni limitate di almeno il 30% (4 del set C; 1 del set D) sono evidenziate in blu scuro. Scaricare i dati aggiornati sulla rappresentazione degli indici e i fattori di correzione calcolati sotto forma di file Excel denominato "Illumina DNA PCR-Free Prep index correction" nella scheda "Documentation" (Documentazione) all'indirizzo illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html.

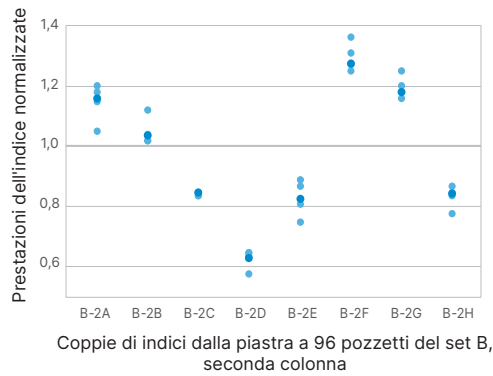


Figura 2: coppie di indici con prestazioni limitate o superiori mostrano prestazioni coerenti. Quattro replicati (azzurro) e prestazioni normalizzate medie (blu scuro) di otto coppie di indici dalla seconda colonna della piastra del set B di UD Index. Una coppia (B-2D) ha avuto prestazioni coerenti superiori al 30% sotto il valore mediano e una coppia (B-2F) ha avuto prestazioni coerenti superiori al 20% sopra il valore mediano.

Riduzione della variabilità nelle prestazioni dell'indice

I fattori di correzione dell'indice sono stati generati prendendo l'inverso del valore mediano di rappresentazione dell'indice per ciascuna coppia di indici mostrato nella [Tabella 1](#) e nella [Tabella 2](#). La moltiplicazione per questi fattori al fine di regolare il volume di ciascun campione durante il raggruppamento in pool, inserendo in modo efficace meno indici sovrarappresentati e più indici sottorappresentati, può riequilibrare il numero di letture di sequenziamento per campione.

Utilizzo del 10% della mediana come soglia per la correzione dell'indice

Per testare questa strategia, i fattori di correzione dell'indice sono stati applicati su tutti e quattro i set di UD Index per le coppie di indici che hanno deviato di almeno il 10% dalla mediana ([Figura 3](#)). L'ordine delle 384 coppie di indici mediante rappresentazione media non corretta degli indici ha rivelato un pattern lineare con una buona correlazione su quattro replicati ($R^2 = 0,75$) e un ampio intervallo di variabilità (coefficiente di variazione (CV, coefficient of variation) = 17%). Dopo la correzione dell'indice, la variazione delle prestazioni dell'indice è stata ridotta (CV = 11%) e i valori corretti non sono più correlati alle prestazioni precorrezione ($R^2 = 0,006$).

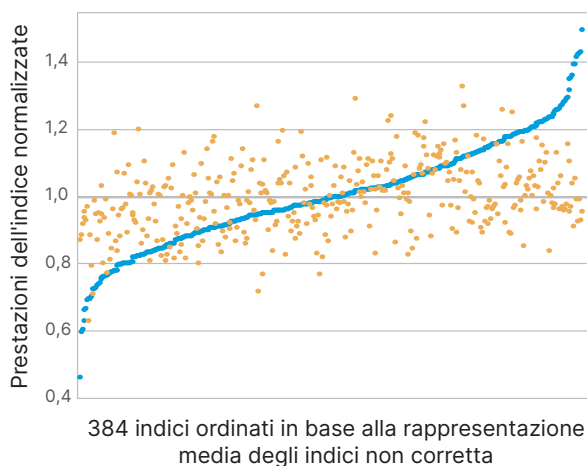


Figura 3: la correzione dell'indice basata sul volume rimuove la variazione specifica dell'indice nelle prestazioni. 384 coppie di indici ordinate dalla resa di sequenza più bassa (sinistra) a quella più alta (destra) per gli esperimenti non corretti. Gli indici non corretti (blu) hanno mostrato un CV del 17%. Gli indici corretti (arancione) hanno mostrato un CV migliorato pari all'11%.

Utilizzo del 15% della mediana come soglia per la correzione dell'indice

Come follow-up, un operatore ha preparato le librerie con tutte le 96 coppie di indici dal set B di UD Index ([Figura 4A](#)). Per ridurre il rumore dovuto alla fatica del pipettaggio manuale, la soglia per la correzione dell'indice è stata aumentata dal 10% al 15%. Dopo l'indicizzazione, le librerie sono state raggruppate in pool per la selezione delle dimensioni in volumi uguali (blu) o implementando fattori di correzione per tutti gli indici che hanno deviato almeno del 15% dalla mediana (arancione). La variazione complessiva è stata ridotta dopo la correzione dell'indice da un CV del 19% (non corretto) a uno del 10% (corretto). Concentrandosi nuovamente sulla seconda colonna della piastra per il set B di UD Index ([Figura 4B](#)), le coppie di indici che hanno costantemente avuto prestazioni limitate (B-2D) o superiori (B-2F) sono diventate performer più tipici.

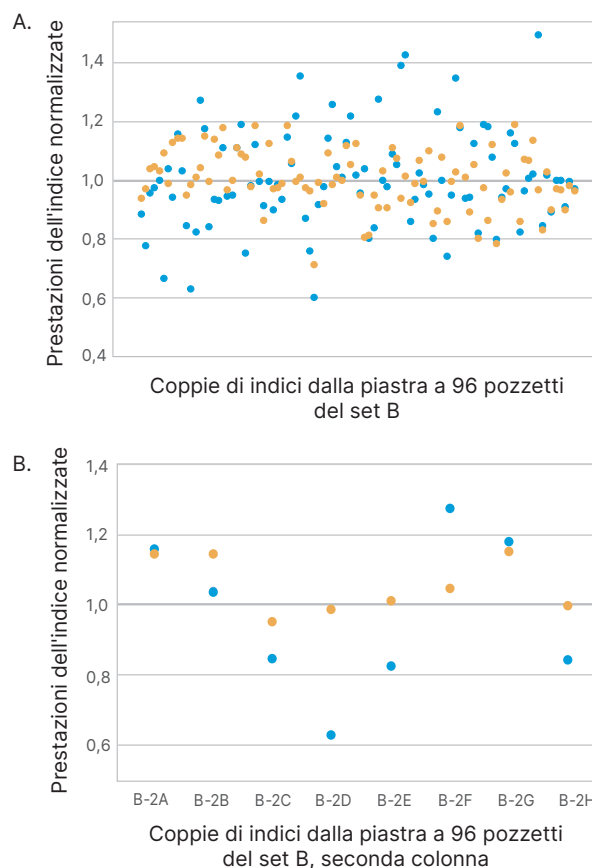


Figura 4: la correzione dell'indice riduce la variabilità delle prestazioni dell'indice. (A) 96 coppie di indici del set B di UD Index. (B) Otto coppie di indici dal set B di UD Index, colonna 2. Gli indici non corretti (blu) hanno mostrato un CV del 19%. Gli indici corretti (arancione) hanno mostrato un CV migliorato pari al 10%. Le coppie di indici che hanno deviato dalla mediana di oltre il 20% (B-2D e B-2F) sono tornate alle prestazioni mediane dell'indice dopo la correzione.

Un ultimo esperimento si è concentrato su tre colonne delle piastre (7, 8, 9) del set B di UD Index e ha corretto 11 coppie di indici che in genere differivano di almeno il 15% dalla mediana. Tutti i 24 campioni per ciascuna condizione (corretta o non corretta) sono stati eseguiti su una cella a flusso NovaSeq 6000 S4 (Figura 5). Come previsto, il CV della rappresentazione dell'indice è diminuito dal 18% (non corretto) al 7% (corretto). Sebbene la copertura media sia rimasta coerente su tutti i campioni, la copertura minima era maggiore per i campioni corretti (dati non mostrati).

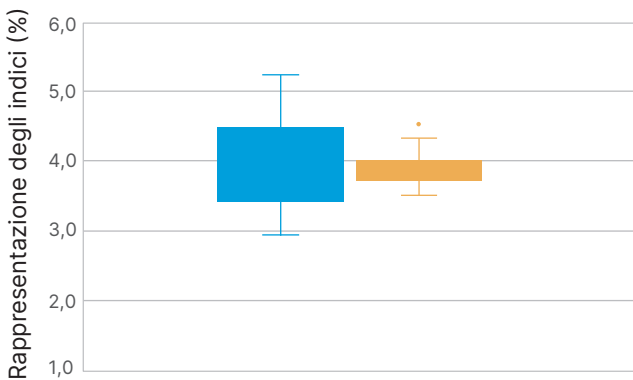


Figura 5: la correzione dell'indice ha ridotto la variabilità tra i campioni sequenziati sulla stessa cella a flusso. Prestazioni dell'indice per le librerie utilizzando 24 coppie di indici del set B di UD Index, colonne della piastra 7, 8, 9. Gli indici non corretti (blu) hanno mostrato un CV del 18%. Gli indici corretti (arancione) hanno mostrato un CV migliorato pari al 7%.

Riduzione degli sprechi e dei costi di sequenziamento per gli studi a elevata processività

Scelta delle coppie di indici con le migliori prestazioni

Questi numeri delle prestazioni dell'indice DNA PCR-Free di Illumina indicano quali indici si possono prevedere che forniscano una migliore copertura e quali colonne mostrano una bassa variazione nelle prestazioni dell'indice (Tabella 3, evidenziata in arancione). Su una piastra di 96 campioni, il CV tipico della rappresentazione dell'indice è pari al 15-20%. I clienti che utilizzano le sottosezioni di una piastra possono scegliere le colonne con il CV minimo per migliorare le prestazioni senza correzione.

Implementazione dei fattori di correzione dell'indice per ridurre la variabilità

Questi risultati dimostrano anche il principio secondo il quale la regolazione dei volumi delle librerie nel raggruppamento in pool può correggere le scarse prestazioni dell'indice. Il raggruppamento dei campioni in pool basato sul volume guidato dai fattori di correzione dell'indice da questo set di dati è stato sufficiente per ridurre la variazione in modo affidabile. Uno degli inconvenienti riscontrati dagli operatori manuali è la correzione di un numero davvero alto di coppie di indici. La correzione manuale degli indici è un processo noioso che può aumentare il rumore a causa di un errore di pipettaggio. Abbiamo osservato risultati migliori grazie all'uso di una soglia più elevata per la correzione degli indici (ovvero, correggendo gli indici con prestazioni superiori al 15% dalla mediana, rispetto a prestazioni superiori al 10% dalla mediana).

Tabella 3: variabilità delle prestazioni dell'indice (CV) per colonna della piastra e per piastra per Illumina DNA/RNA UD Indexes

Colonna ^a	1	2	3	4	5	6	7	8	9	10	11	12	Piastra ^b
Set 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%
Set 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%
Set 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%
Set 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 calcolato in base ai dati medi delle prestazioni dell'indice riportati nella Tabella 1 e nella Tabella 2. La preparazione delle librerie è stata eseguita manualmente con 3-5 replicati utilizzando Illumina DNA PCR-Free Prep e sequenziata su NovaSeq 6000 System. I valori evidenziati mostrano le colonne con CV di rappresentazione dell'indice inferiore al 10% (arancione scuro) o inferiore al 15% (arancione chiaro).
b. Mentre il CV della colonna varia dal 6,4% al 25,7%, il CV della piastra a 96 pozzetti varia solo dal 15,1% al 16,6%.

Anche la correzione di un piccolo numero di coppie di indici con le prestazioni peggiori dovrebbe "salvare" alcuni campioni dalla caduta a causa di una copertura inferiore rispetto al previsto. È possibile visualizzare la correzione dell'indice come un processo dinamico con regolazioni iterative man mano che si raccolgono più dati. Questi dati rappresentano un punto di partenza per ulteriori studi.[§] Per ridurre il rumore di fondo e ottenere tutti i vantaggi potrebbe essere necessaria un'automazione.

Riepilogo

Le prestazioni dell'indice sono un fattore chiave nella variabilità della copertura dei campioni per gli esperimenti che utilizzano Illumina DNA PCR-Free e NovaSeq 6000 System o NovaSeq X Series. Alcune sequenze di indici forniscono costantemente una resa superiore o inferiore. Regolando i volumi durante il raggruppamento in pool, è possibile ridurre la variabilità nella resa della sequenza tra i campioni sequenziati sulla stessa cella a flusso. Sebbene ciò non aumenti la resa complessiva del sequenziamento, aumenta la copertura minima, riducendo il numero di campioni che ricadono al di sotto di una soglia di copertura minima. I laboratori che conducono la ricerca genomica della popolazione e altre applicazioni WGS umane a elevata processività possono utilizzare i dati forniti come punto di partenza per ottimizzare la processività dei campioni e ridurre gli sprechi.

[§] Accedere ai dati aggiornati della rappresentazione degli indici e ai fattori di correzione calcolati per i set di indici v3 con DNA PCR-Free Illumina sia per NovaSeq 6000 System sia per NovaSeq X Series. Scaricare un file Excel denominato "Illumina DNA PCR-Free Prep index correction" nella scheda "Documentation" (Documentazione) all'indirizzo illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html.



Numero verde 1.800.809.4566 (U.S.A.) | Tel. +1.858.202.4566
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Tutti i diritti riservati. Tutti i marchi di fabbrica sono di proprietà di Illumina, Inc. o dei rispettivi proprietari. Per informazioni specifiche sui marchi di fabbrica, visitare la pagina web www.illumina.com/company/legal.html.
M-GL-00005 ITA v2.0.

Maggiori informazioni

[Illumina DNA PCR-Free Prep](#)

Bibliografia

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. Pubblicato nel 2020. Aggiornato nel 2023. Consultato il 6 settembre 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.