

Equilíbrio da cobertura da amostra para o sequenciamento do genoma completo

Estratégias de correção de índice para o Illumina DNA PCR-Free Prep

- Aumentar a cobertura mínima em bibliotecas multiplexadas.
- Reduzir a variabilidade no desempenho do índice através do ajuste dos volumes no ponto do pool de bibliotecas.
- Reduzir os custos de resíduos e do sequenciamento em estudos de alto rendimento.



Aumento da cobertura mínima em amostras multiplexadas

O Illumina DNA PCR-Free Prep, Tagmentation (Illumina DNA PCR-Free) oferece uma solução otimizada de preparação de bibliotecas para uma variedade de aplicações de sequenciamento do genoma completo (WGS). Com base no processo químico do transposossoma ligado a beads, o Illumina DNA PCR-Free fornece cobertura uniforme em todo o genoma e oferece suporte ao fácil pool de bibliotecas baseado em volume.^{1,2}

Através do Illumina DNA PCR-Free, do NovaSeq™ 6000 System ou do NovaSeq X Series, os laboratórios de alto rendimento podem multiplexar bibliotecas de WGS para maximizar a eficiência. Ao equilibrar o rendimento do sequenciamento em amostras de WGS multiplexadas, os usuários poderão fazer o sequenciamento de mais amostras por lâmina de fluxo, alcançando a cobertura mínima desejada para identificação confiável de variantes.

Vários fatores afetam a profundidade de cobertura por amostra, incluindo o rendimento total da sequência, o número de amostras por lâmina de fluxo e a variabilidade no rendimento da amostra. Para aumentar o número de amostras por sequenciamento que atingem a cobertura de sequência desejada, os usuários podem aumentar a cobertura média em todas as amostras ou reduzir a variação de cada amostra (Figura 1). A qualidade do DNA de entrada, a variabilidade da pipetagem e o desempenho do índice contribuem para a variabilidade no rendimento da sequência por amostra. Esta nota técnica se concentra em uma estratégia para mitigar a variação relacionada ao desempenho do índice.

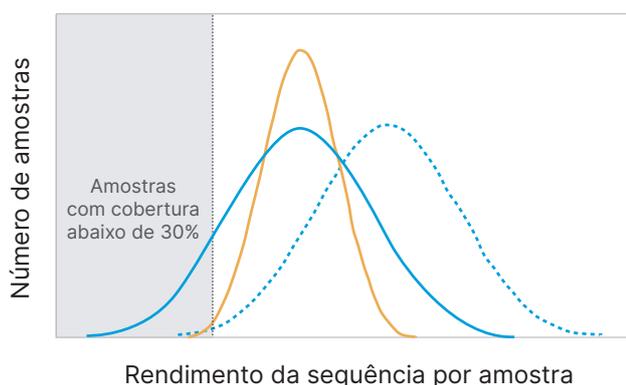


Figura 1: Melhoria do rendimento mínimo de sequenciamento por amostra. Mais amostras atingirão pelo menos 30 vezes a cobertura com aumento do rendimento de sequenciamento (azul pontilhado) ou diminuição da variação (laranja).

O desempenho do índice afeta a cobertura da amostra

O Illumina DNA/RNA Unique Dual (UD) Indexes, Tagmentation está disponível em quatro conjuntos de 96 pares de índices (conjuntos A, B, C e D) para um total de 384 pares de índices. Quando usado com o Illumina DNA PCR-Free* e o NovaSeq 6000 System ou o NovaSeq X Series,† certos pares de índices fornecem consistentemente maior ou menor rendimento de sequência. Os pares de índices que apresentam desempenho inferior resultam em uma cobertura inferior à desejada para essas amostras, enquanto os pares de índices que apresentam desempenho superior podem reduzir a proporção de leituras para outras amostras.

Para demonstrar a variabilidade da cobertura da amostra devido à sequência de índice, empregamos uma versão modificada do protocolo de preparação da biblioteca Illumina DNA PCR-Free. O DNA genômico da linhagem celular humana (Coriell Institute, amostra NA12878) foi usado como entrada e todas as amostras foram preparadas em massa para reduzir o ruído associado à qualidade do DNA de entrada e erros de pipetagem, respectivamente. Na etapa de indexação, as amostras foram distribuídas em uma placa de 96 poços para indexação e, em seguida, agrupadas novamente para limpeza de seleção do tamanho.‡ Vários operadores conduziram um total de três a cinco réplicas para cada conjunto do UD Indexes mediante o uso de pipetas manuais. Cada conjunto de 96 amostras foi sequenciado em uma cavidade única da lâmina de fluxo NovaSeq 6000 S4 com o fluxo de trabalho Xp para obtenção da representação do índice.

O desempenho do índice normalizado** é mostrado para os conjuntos A e B do UD Indexes (Tabela 1) e os conjuntos C e D (Tabela 2). Para este conjunto de dados:

- Em média, 32 pares de índices por conjunto de 96 divergiram do desempenho mediano em mais de 15% (poços azuis).
- Determinados pares de índices por placa (poços azuis mais escuros) normalmente apresentam desempenho inferior em pelo menos 30%, colocando esses índices em alto risco de ocorrência de alvos sem cobertura.

* O Illumina DNA PCR-Free conta com um processo químico de indexação exclusiva; os dados de desempenho do índice não se aplicam a outros kits de preparação de bibliotecas.

† Alguma variabilidade do índice se deve à clusterização; os dados de desempenho do índice são específicos para cada sistema de sequenciamento.

‡ Seleção de tamanho usando imobilização reversível de fase sólida (SPRI): a purificação de beads emprega muitas etapas de pipetagem com um reagente viscoso que pode acrescentar uma variação adicional. O pool de amostras é realizado imediatamente após a indexação, em vez de após a seleção do tamanho, com variação interna reduzida (dados não mostrados).

Tabela 1: Desempenho do índice normalizado no layout da placa de 96 poços para o Illumina DNA PCR-Free e os conjuntos A e B do UD Index.

Conjunto 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

Conjunto 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

A preparação da biblioteca foi realizada manualmente com réplicas de n = 3 (conjunto A) ou n = 5 (conjunto B) através do Illumina DNA PCR-Free Prep e sequenciada no NovaSeq 6000 System. Os poços destacados em azul mostram uma representação de índice que divergiu da mediana em mais de 15%. Os pares de índices que tiveram um desempenho consistentemente inferior em pelo menos 30% (2 do Conjunto A; 3 do Conjunto B) foram destacados em azul mais escuro. Baixe os dados atualizados sobre a representação de índices e fatores de correção calculados como um arquivo do Excel, nomeado como "Illumina DNA PCR-Free Prep index correction" (Correção de índice do Illumina DNA PCR-Free Prep) na guia "Documentation" (Documentação) em illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html.

- Algumas colunas de placas individuais não exibiram pares de índices, ou apenas um par de índices, com um desvio acima de 15% da mediana, produzindo um desempenho maior do que a placa geral.

O desempenho do índice foi consistente entre réplicas ($R^2 = 0,54$ a $0,80$). Por exemplo, um foco em 4 réplicas da segunda coluna da placa de 96 poços para o conjunto B do UD Index (Figura 2) mostra um padrão consistente de super e sub-representação do índice.

Isso sugere que o desempenho de valores atípicos não é simplesmente um ruído, mas sim uma propriedade fundamental dos pares de índices usados. A variabilidade consistente no desempenho do índice aponta para uma estratégia clara de "correção do índice" em relação ao seu desempenho típico. A princípio, ajustar o volume de certas bibliotecas durante a realização do pool antes do sequenciamento pode compensar a representação de índices variáveis. Espera-se que os pares de índices de bom desempenho proporcionem baixa variabilidade inerentemente, enquanto os pares de índices que se desviam consistentemente do valor mediano são candidatos à correção do índice.

Tabela 2: Desempenho de índice normalizado no layout da placa de 96 poços para o Illumina DNA PCR-Free e os conjuntos C e D do UD Index.

Conjunto 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

Conjunto 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

A preparação da biblioteca foi realizada manualmente com réplicas de n = 4 (conjunto C) ou n = 3 (conjunto D) através do Illumina DNA PCR-Free Prep e sequenciada na NovaSeq 6000 System. Os poços destacados em azul mostram uma representação de índice que divergiu da mediana em mais de 15%. Os pares de índices que tiveram desempenho consistentemente inferior em pelo menos 30% (4 do conjunto C; 1 do conjunto D) foram destacados em azul mais escuro. Baixe os dados atualizados sobre a representação de índices e fatores de correção calculados como um arquivo do Excel, nomeado como "Illumina DNA PCR-Free Prep index correction" (Correção de índice do Illumina DNA PCR-Free Prep) na guia "Documentation" (Documentação) em illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html.

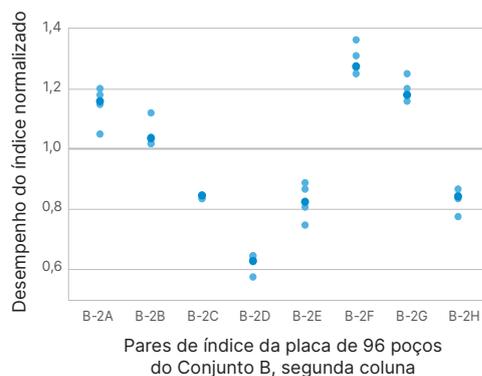


Figura 2: Os pares de índices com desempenho inferior ou superior fazem isso de forma consistente: quatro réplicas (azul claro) e desempenho normalizado médio (azul escuro) de oito pares de índice da segunda coluna da placa do conjunto B do UD Index. Um par (B-2D) teve desempenho consistente com mais de 30% abaixo do valor mediano e um par (B-2F) teve desempenho consistente com mais de 20% acima do valor mediano.

Mitigação da variabilidade no desempenho do índice

Os fatores de correção do índice foram gerados ao adotar o inverso do valor mediano de representação do índice para cada par de índices mostrado na [Tabela 1](#) e na [Tabela 2](#). A multiplicação por esses fatores para ajustar o volume de cada amostra durante a realização do pool, colocando efetivamente menos índices super-representados e mais índices sub-representados, pode reequilibrar o número de leituras de sequenciamento por amostra.

Uso de 10% da mediana como limite para correção do índice

Para testar essa estratégia, fatores de correção de índice foram aplicados em todos os quatro conjuntos do UD Index para pares de índices que desviaram pelo menos 10% da mediana ([Figura 3](#)). A ordenação dos 384 pares de índices por representação média de índices não corrigida revelou um padrão linear com boa correlação em quatro réplicas ($R^2 = 0,75$) e uma ampla faixa de variabilidade (coeficiente de variação; $CV = 17\%$). Após a correção do índice, a variação no desempenho do índice foi reduzida ($CV = 11\%$) e os valores corrigidos não se correlacionaram mais com o desempenho antes da correção ($R^2 = 0,006$).

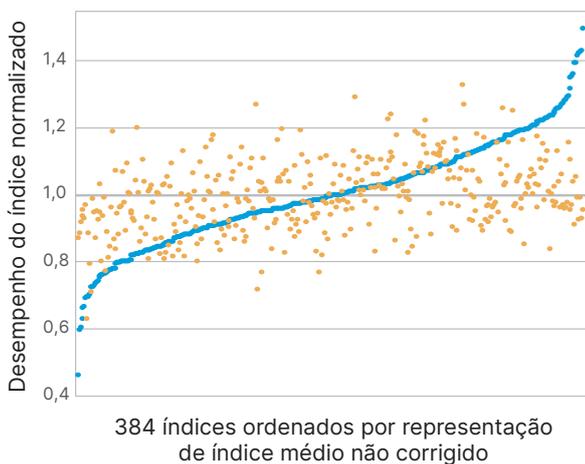


Figura 3: A correção do índice baseada em volume remove a variação específica do índice no desempenho: 384 pares de índices ordenados do menor (esquerdo) ao maior (direito) rendimento de sequência para experimentos não corrigidos. Índices não corrigidos (azul) mostraram um CV de 17%. Índices corrigidos (laranja) mostraram um CV melhorado de 11%.

Uso de 15% da mediana como limite para correção do índice

Como acompanhamento, um operador preparou bibliotecas com todos os 96 pares de índices do conjunto B do UD Index ([Figura 4A](#)). Para reduzir o ruído devido à fadiga de pipetagem manual, o limite para correção do índice foi aumentado de 10% para 15%. Após a indexação, as bibliotecas foram agrupadas para seleção de tamanho em volumes iguais (azul) ou mediante a implementação de fatores de correção para todos os índices que desviaram pelo menos 15% da mediana (laranja). A variação geral foi reduzida após a correção do índice de um CV de 19% (não corrigido) para 10% (corrigido). Focando novamente na segunda coluna da placa para o conjunto B do UD Index ([Figura 4B](#)), é possível ver que pares de índices que tiveram desempenho consistentemente inferior (B-2D) ou superior (B-2F) se tornaram desempenhos mais típicos.

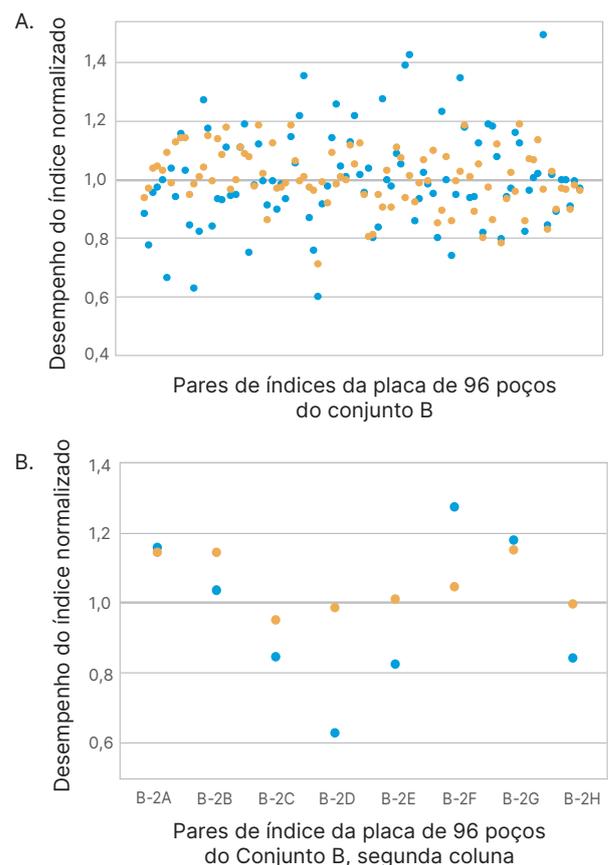


Figura 4: A correção do índice reduz a variabilidade do desempenho do índice. (A) 96 pares de índices do conjunto B do UD Index. (B) Oito pares de índices do conjunto B do UD Index, coluna 2. Índices não corrigidos (azul) mostraram um CV de 19%. Índices corrigidos (laranja) mostraram um CV melhorado de 10%. Os pares de índices que se desviaram da mediana em mais de 20% (B-2D e B-2F) retornaram ao desempenho mediano do índice após a correção.

Um último experimento se concentrou em três colunas de placa (7, 8 e 9) do conjunto B do UD Index e corrigiu 11 pares de índices que normalmente divergiram em pelo menos 15% da mediana. Todas as 24 amostras para cada condição (corrigidas ou não corrigidas) foram sequenciadas em uma lâmina de fluxo NovaSeq 6000 S4 (Figura 5). Conforme esperado, o CV da representação do índice diminuiu de 18% (não corrigido) para 7% (corrigido). Embora a cobertura média tenha permanecido consistente em todas as amostras, a cobertura mínima foi maior para as amostras corrigidas (dados não mostrados).

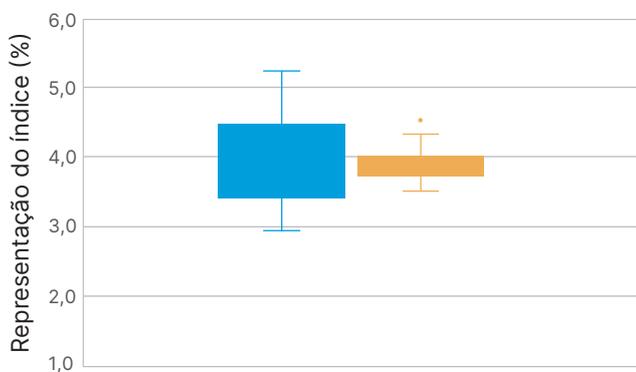


Figura 5: A correção do índice reduziu a variabilidade entre as amostras sequenciadas na mesma lâmina de fluxo: desempenho do índice para bibliotecas que usaram 24 pares de índices do conjunto B do UD Index, colunas de placa 7, 8, 9. Índices não corrigidos (azul) mostraram um CV de 18%. Índices corrigidos (laranja) mostraram um CV melhorado de 7%.

Redução dos custos de resíduos e de sequenciamento para estudos de alto rendimento

Seleção dos pares de índices de melhor desempenho

Esses números de desempenho do índice Illumina DNA PCR-Free indicam quais índices podem oferecer melhor cobertura e quais colunas mostram baixa variação no desempenho do índice (Tabela 3, destacada em laranja). Em uma placa de 96 amostras, o CV típico da representação do índice é de 15 a 20%. Os clientes que usam subseções de uma placa podem escolher colunas com o CV mais baixo para melhorar o desempenho sem correção.

Implementação de fatores de correção de índice para reduzir a variabilidade

Esses resultados também demonstram o princípio de que o ajuste dos volumes da biblioteca no pooling pode corrigir o desempenho ruim do índice. O pool de amostras baseado em volume guiado por fatores de correção de índice desse conjunto de dados foi suficiente para reduzir de forma confiável a variação. A correção de muitos pares de índices pode se tornar uma armadilha para operadores manuais. O processo de correção manual de índices é árduo, e pode aumentar o ruído devido a um erro de pipetagem. Observamos melhores resultados ao usarmos um limite mais alto para correção de índices (ou seja, correção de índices com desempenho superior a 15% em relação à mediana, contra o desempenho de mais de 10% em relação à mediana).

Tabela 3: Variabilidade de desempenho (CV) do índice por coluna de placa e por placa para o UD Indexes de DNA/RNA da Illumina.

Coluna ^a	1	2	3	4	5	6	7	8	9	10	11	12	Placa ^b
Conjunto 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%
Conjunto 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%
Conjunto 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%
Conjunto 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 dos dados médios de desempenho do índice na Tabela 1 e na Tabela 2. A preparação da biblioteca foi realizada manualmente por meio de 3 a 5 réplicas através do Illumina DNA PCR-Free Prep e foi sequenciada no NovaSeq 6000 System. Os destaques mostram colunas com CV de representação de índice inferior a 10% (laranja escuro) ou inferior a 15% (laranja claro).
 b. Enquanto o CV da coluna varia de 6,4% a 25,7%, o CV da placa de 96 poços varia apenas de 15,1% a 16,6%.

Mesmo a correção de um pequeno número dos pares de índices de pior desempenho deve "resgatar" algumas amostras da falha causada por uma cobertura inferior à esperada. A correção do índice pode ser visualizada como um processo dinâmico com ajustes iterativos à medida que mais dados são coletados. Esses dados representam um ponto de partida para estudos adicionais.[§] A automação pode ser necessária para reduzir o ruído de fundo e obter todos os benefícios.

Resumo

O desempenho do índice é um fator essencial na variabilidade da cobertura de amostras para experimentos que utilizam o Illumina DNA PCR-Free e o NovaSeq 6000 System ou o NovaSeq X Series. Algumas sequências de índice proporcionam consistentemente um rendimento maior ou menor. Ao ajustar os volumes durante a realização do pool, é possível reduzir a variabilidade no rendimento da sequência entre as amostras sequenciadas na mesma lâmina de fluxo. Embora isso não aumente o rendimento geral do sequenciamento, a cobertura mínima é aumentada, reduzindo o número de amostras que estão abaixo de um limite mínimo de cobertura. Os laboratórios que realizam pesquisa genômica populacional e outras aplicações de WGS humanos de alto rendimento podem usar os dados fornecidos como ponto de partida para otimizar o rendimento da amostra e reduzir resíduos.

[§] Acesso a dados atualizados de representação de índices e fatores de correção calculados para conjuntos de índices v3 com o Illumina DNA PCR-Free para o NovaSeq 6000 System e o NovaSeq X Series. Baixe o arquivo do Excel nomeado como "Illumina DNA PCR-Free Prep index correction" (Correção de índice do Illumina DNA PCR-Free Prep) na guia "Documentation" (Documentação) em [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).

Saiba mais

[Illumina DNA PCR-Free Prep](#)

Referências

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). Publicado em 2020. Atualizado em 2023. Acessado em 6 de setembro 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, ligação gratuita (EUA) | tel. +1 (858) 202-4566
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Todos os direitos reservados. Todas as marcas comerciais pertencem à Illumina, Inc. ou aos respectivos proprietários. Para obter informações específicas sobre marcas comerciais, consulte www.illumina.com/company/legal.html.
M-GL-00005 PTB v2.0.