TECHNICAL NOTE

Balancing sample coverage for whole-genome sequencing

Index correction strategies for Illumina DNA PCR-Free Prep

- Increase minimum coverage across multiplexed libraries
- Mitigate variability in index performance by adjusting volumes at the point of library pooling
- Reduce waste and sequencing costs for high-throughput studies

illumina®

For Research Use Only. Not for use in diagnostic procedures.

Increase minimum coverage across multiplexed samples

Illumina DNA PCR-Free Prep, Tagmentation (Illumina DNA PCR-Free) offers an optimized library prep solution for a variety of whole-genome sequencing (WGS) applications. Based on bead-linked transposome chemistry, Illumina DNA PCR-Free provides even coverage across the genome and supports easy volume-based library pooling.[1,2] Using Illumina DNA PCR-Free and the NovaSeq™ 6000 System or the NovaSeq X Series, high-throughput labs can multiplex WGS libraries to maximize efficiency. Balancing sequencing yield across multiplexed WGS samples enables users to run more samples per flow cell while achieving the minimum coverage desired for confident variant calling.

Several factors impact coverage depth per sample, including total sequence yield, the number of samples per flow cell, and variability in sample yield. To increase the number of samples per run that achieve the desired sequence coverage, users can either increase mean coverage across all samples or reduce sample-to-sample variation (Figure 1). Input DNA quality, pipetting variability, and index performance all contribute to variability in sequence yield per sample. This technical note focuses on a strategy to mitigate variation due to index performance.

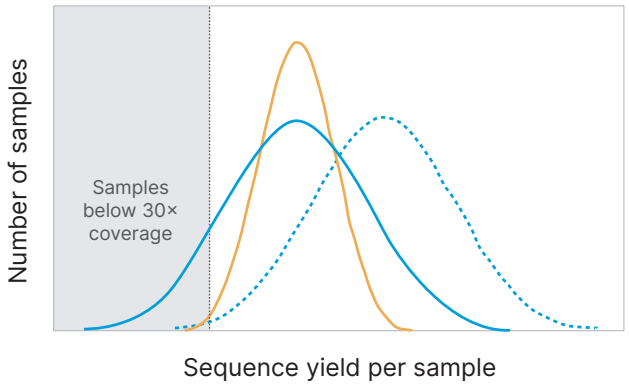

Figure 1: Improving minimum sequencing yield per sample—More samples will achieve at least 30× coverage with increasing sequencing yield (dotted blue) or decreasing variation (orange).

Index performance affects sample coverage

Illumina DNA/RNA Unique Dual (UD) Indexes, Tagmentation are available in four sets of 96 index pairs (Sets A, B, C, and D) for 384 total index pairs. When used with Illumina DNA PCR-Free* and the NovaSeq 6000 System or the NovaSeq X Series,† certain index pairs consistently give higher or lower sequence yield. Index pairs that underperform lead to lower than desired coverage for those samples, while index pairs that overperform can reduce the share of reads for other samples.

To demonstrate sample coverage variability due to the index sequence, we employed a modified version of the Illumina DNA PCR-Free library preparation protocol. Human cell line genomic DNA (Coriell Institute, sample NA12878) was used as input and all samples were prepared in bulk to reduce noise associated with input DNA quality and pipetting errors, respectively. At the indexing step, samples were aliquoted across a 96-well plate for indexing, then pooled again for size-selection cleanup.‡ Multiple operators conducted a total of three to five replicates for each set of UD Indexes using manual pipettes. Each 96-sample set was sequenced on a single lane of a NovaSeq 6000 S4 flow cell with the Xp workflow to obtain index representation.

Normalized index performance*† is shown for UD Index Sets A and B (Table 1) and Sets C and D (Table 2). For this data set:

- On average, 32 index pairs per set of 96 diverged from median performance by more than 15% (blue wells)
- A few index pairs per plate (darker blue wells) typically underperformed by at least 30%, putting those indexes at high risk of missing coverage targets

* Illumina DNA PCR-Free has unique indexing chemistry; index performance data does not apply to other library prep kits.

† Some index variability is due to clustering; index performance data are specific for each sequencing system.

‡ Size selection using solid-phase reversible immobilization (SPRI)–bead purification employs many pipetting steps with a viscous reagent that may introduce additional variation. Pooling samples immediately after indexing, instead of after size selection, reduced internal variation (data not shown).

For Research Use Only. Not for use in diagnostic procedures.

Table 1: Normalized index performance in 96-well plate layout for Illumina DNA PCR-Free and UD Index Sets A and B

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

Library preparation was performed manually with replicates of n = 3 (Set A) or n = 5 (Set B) using Illumina DNA PCR-Free Prep and sequenced on the NovaSeq 6000 System. Blue highlighted wells show index representation that diverged from the median by more than 15%. Index pairs that consistently underperformed by at least 30% (2 from Set A; 3 from Set B) are highlighted in darker blue. Download up-to-date index representation data and calculated correction factors as an Excel file, named "Illumina DNA PCR-Free Prep index correction" under the "Documentation" tab at illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html.

- Some individual plate columns exhibited no index pairs, or only one index pair, with a deviation > 15% from median, yielding higher performance than the overall plate

Index performance was consistent across replicates (R^2 = 0.54–0.80). For example, a focus on four replicates from the second column of the 96-well plate for UD Index Set B (Figure 2) shows a consistent pattern of index over- and under-representation.

This suggests that the performance of outliers is not simply noise, but rather a fundamental property of the index pairs used. The consistent variability in index performance points to a clear strategy for "index correction" in relation to typical index performance. In principle, adjusting the volume of certain libraries at pooling before sequencing could compensate for variable index representation. Well-performing index pairs are expected to give low variability inherently, while index pairs that consistently deviate from the median value are candidates for index correction.

Table 2: Normalized index performance in 96-well plate layout for Illumina DNA PCR-Free and UD Index Sets C and D

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

Library preparation was performed manually with replicates of n = 4 (Set C) or n = 3 (Set D) using Illumina DNA PCR-Free Prep and sequenced on the NovaSeq 6000 System. Blue highlighted wells show index representation that diverged from the median by more than 15%. Index pairs that consistently underperformed by at least 30% (4 from Set C; 1 from Set D) are highlighted in darker blue. Download up-to-date index representation data and calculated correction factors as an Excel file, named "Illumina DNA PCR-Free Prep index correction" under the "Documentation" tab at illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html.

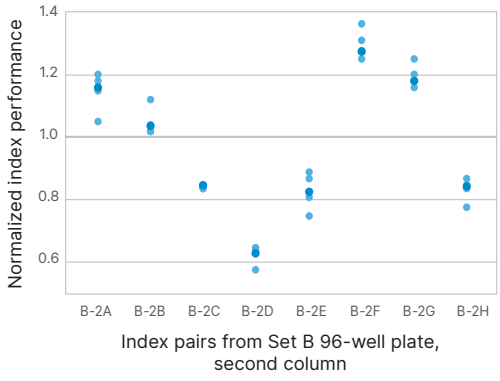

Figure 2: Index pairs that underperform or overperform do so consistently—Four replicates (light blue) and mean (dark blue) normalized performance of eight index pairs from the second plate column of UD Index Set B. One pair (B-2D) consistently performed > 30% below the median value, and one pair (B-2F) consistently performed > 20% above the median value.

For Research Use Only. Not for use in diagnostic procedures.

Mitigate variability in index performance

Index correction factors were generated by taking the inverse of the median index representation value for each index pair shown in Table 1 and Table 2. Multiplying by these factors to adjust the volume of each sample when pooling—effectively putting in less of over-represented indexes and more of under-represented indexes—can rebalance the number of sequencing reads per sample.

Using 10% from the median as the threshold for index correction

To test this strategy, index correction factors were applied across all four UD Index sets for index pairs that deviated at least 10% from the median (Figure 3). Ordering the 384 index pairs by uncorrected mean index representation revealed a linear pattern with good correlation across four replicates (R^2 = 0.75) and a wide range of variability (coefficient of variation; CV = 17%). After index correction, variation in index performance was reduced (CV = 11%) and corrected values no longer correlated with precorrection performance (R^2 = 0.006).

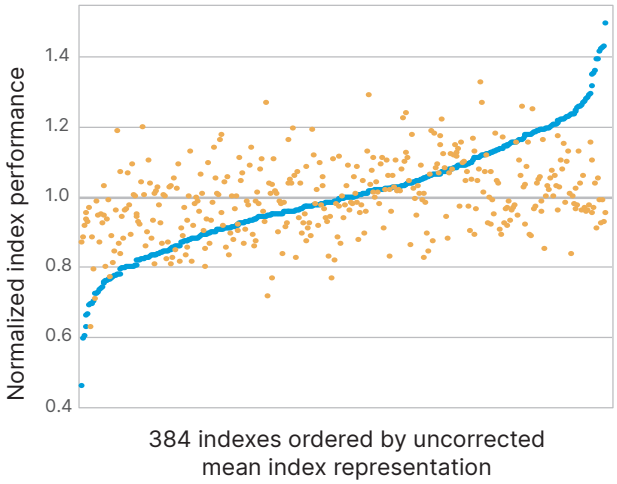

Figure 3: Volume-based index correction removes index-specific variation in performance—384 index pairs ordered from lowest (left) to highest (right) sequence yield for uncorrected experiments. Uncorrected indexes (blue) showed a CV of 17%. Corrected indexes (orange) showed an improved CV of 11%.

Using 15% from the median as the threshold for index correction

As a follow up, one operator prepared libraries with all 96 index pairs from UD Index Set B (Figure 4A). To reduce the noise due to manual pipetting fatigue, the threshold for index correction was increased from 10% to 15%. After indexing, libraries were pooled for size selection in equal volumes (blue) or by implementing correction factors for all indexes that deviated at least 15% from the median (orange). Overall variation was reduced after index correction from a CV of 19% (uncorrected) to 10% (corrected). Focusing again on the second plate column for UD Index Set B (Figure 4B) shows index pairs that consistently underperformed (B-2D) or overperformed (B-2F) became more typical performers.

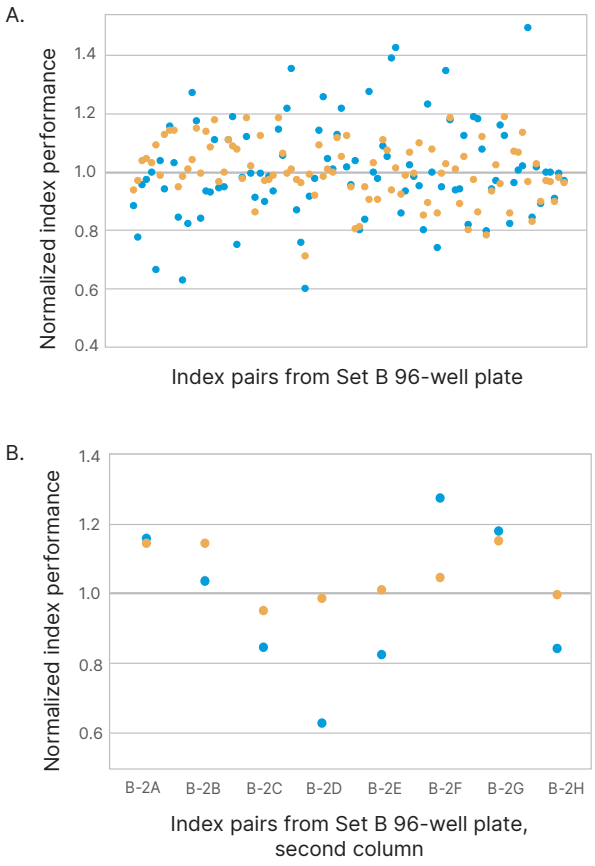

Figure 4: Index correction reduces index performance variability—(A) 96 index pairs from UD Index Set B. (B) Eight index pairs from UD Index Set B, column 2. Uncorrected indexes (blue) showed a CV of 19%. Corrected indexes (orange) showed an improved CV of 10%. Index pairs that deviated from the median by over 20% (B-2D and B-2F) returned to median index performance after correction.

A last experiment focused on three plate columns (7, 8, 9) from UD Index Set B and corrected 11 index pairs that typically diverged by at least 15% from median. All 24 samples for each condition (corrected or uncorrected) were run on a NovaSeq 6000 S4 flow cell (Figure 5). As expected, the CV of index representation decreased from 18% (uncorrected) to 7% (corrected). Although mean coverage stayed consistent across all samples, minimum coverage was higher for the corrected samples (data not shown).

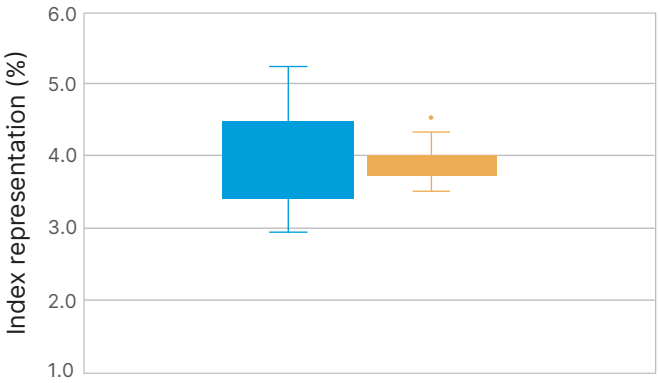

Figure 5: Index correction reduced variability across samples sequenced on the same flow cell—Index performance for libraries using 24 index pairs from UD Index Set B, plate columns 7, 8, 9. Uncorrected indexes (blue) showed a CV of 18%. Corrected indexes (orange) showed an improved CV of 7%.

Reduce waste and sequencing costs for high-throughput studies

Choose the best-performing index pairs

These Illumina DNA PCR-Free index performance numbers indicate which indexes may be expected to give better coverage and which columns show low variation in index performance (Table 3, highlighted in orange). Across a plate of 96 samples, the typical CV of index representation is 15–20%. Customers using subsections of a plate can choose columns with the lowest CV to improve performance without correction.

Implement index correction factors to reduce variability

These results also demonstrate the principle that adjusting library volumes in pooling can correct for poor index performance. Volume-based sample pooling guided by index correction factors from this data set were sufficient to reliably reduce variation. One pitfall for manual operators is correcting too many index pairs. Manually correcting indexes is a tedious process that may increase noise due to pipetting error. We observed better results using a higher threshold for index correction (ie, correcting indexes with performance > 15% from median, versus > 10% from median). Even correcting a small number of the worst-performing index pairs would be expected to "rescue" some samples from dropping out due to lower

Table 3: Index performance variability (CV) by plate column and by plate for Illumina DNA/RNA UD Indexes

Column[a]	1	2	3	4	5	6	7	8	9	10	11	12	Plate[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 calculated from mean index performance data in Table 1 and Table 2. Library preparation was performed manually with 3-5 replicates using Illumina DNA PCR-Free Prep and sequenced on the NovaSeq 6000 System. Highlights show columns with index representation CV lower than 10% (dark orange) or lower than 15% (light orange).
b. While column CV ranges from 6.4%–25.7%, 96-well plate CV only ranges from 15.1%–16.6%.

For Research Use Only. Not for use in diagnostic procedures.

than expected coverage. View index correction as a dynamic process with iterative adjustments as you gather more data. These data represent a starting point for further studies.§ Automation may be necessary to reduce background noise and obtain full benefits.

Summary

Index performance is a key driver in sample coverage variability for experiments using Illumina DNA PCR-Free and the NovaSeq 6000 System or NovaSeq X Series. Some index sequences consistently give a higher or lower yield. By adjusting volumes during pooling, you can reduce variability in sequence yield across samples sequenced on the same flow cell. Although this does not increase overall sequencing yield, it increases the minimum coverage, reducing the number of samples that fall below a minimum coverage threshold. Labs conducting population genomics research and other high-throughput human WGS applications can use the provided data as a starting point to optimize sample throughput and reduce waste.

§ Access up-to-date index representation data and calculated correction factors for v3 index sets with Illumina DNA PCR-Free for both the NovaSeq 6000 System and the NovaSeq X Series. Download an Excel file, named "Illumina DNA PCR-Free Prep index correction" under the "Documentation" tab at illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html.

Learn more

Illumina DNA PCR-Free Prep

References

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. Published 2020. Updated 2023. Accessed September 6, 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.

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners. For specific trademark information, see www.illumina.com/company/legal.html.
M-GL-00005 v2.0.