Comparison of Illumina and Ultima Sequencing Platforms for Olink Explore HT Libraries

INTRODUCTION

Widespread adoption of new sequencing technologies is often hampered in part due to high costs and uncertainty in data quality compared to gold standard assays. Olink Proteomics has pioneered Explore HT, a multiplex protein assay targeting over 5,000 proteins in a single panel [1]. Each protein assay uses a process called Proximity Extension Assay, where two antibodies bind nearby epitopes on the same target protein, complimentary single stranded DNA oligos ligated to the antibodies bind each other, are extended, amplified by PCR, and then sequenced, allowing accurate and specific detection of very low abundance proteins [2].

However, adoption of the Olink Explore HT assay has been limited given the high costs associated with next-generation sequencing. Recently, Ultima Genomics revealed the UG 100, which utilizes "flow-based" sequencing combined with data capture cycles where only one nucleotide at a time is "flowed" through [3]. This flow system reduces errors by not allowing base-calling errors as only one nucleotide is present at a time during the sequencing run. Next, high-speed cameras acquire data as sequencing is performed on a flat, rotating silicon wafer [3]. Using a moving wafer allows even reagent dispensing across the entire surface area, which yields higher quality and larger datasets per run. This directly relates to cost savings: the UG 100 10B read wafer provides end users with cost savings over traditional next-generation sequencing by synthesis (SBS) based flow cell chemistry, with savings of several thousand dollars per Olink Explore HT run.

Here, we compare four different Olink Explore HT projects representing three different biosamples for their data quality and consistency between the Illumina NovaSeq X Plus and the Ultima UG 100 and report comparable yields between both platforms.

MATERIALS AND METHODS

Sequencing

Four Olink Explore HT libraries from a variety of biological sources, previously prepared for Illumina sequencing, were chosen for sequencing on the UG 100: AN00022234 cell lysate (6 samples), AN00022234 plasma (76 samples), AN00022234 tissue lysate (10 samples), and AN00022466 plasma (79 samples). Libraries were converted to Ultima format by adaptor ligation, then efficiency was verified by length increase of ~35 bp by BioAnalyzer HS and TapeStation D1000. The four converted libraries were sequenced at two different concentrations, 600 pM (conc. recommended by Ultima and Olink; run ID 418126) and 700 pM (run ID 418271).

Data Format

CRAM files were initially converted to Olink counts format on the UG 100 but gave readability errors in the Olink NPX Map v1.0.2 software. Instead, on Olink's recommendation, future runs should export the CRAM files off the UG 100 and then be run through the Olink Map CLI v1.0.2 tool on the Psomagen server to generate counts files. Additionally, Ultima conversion tools only take Olink Explore HT new Index Plates A/C (released December 2024) as input. Should the Olink plates use the old Index A/B, please contact Ultima technical support for assistance converting the A/B barcodes to A/C format.

Data Analysis

Counts files (16 per run) and "run_metadata.json" files were downloaded from the Psomagen server into local folders then read into the Olink NPX Map v1.0.2 software, along with sample manifest lists. Raw counts data were converted to, then exported as, NPX values and data were analyzed in R v4.4.1 using the OlinkAnalyze v4.0.1, tidyverse v2.0.0, ggpubr v0.6.0, DESeq2 v1.46.0, and pheatmap v1.0.12 packages. Unless otherwise stated, counts were used for comparison purposes across platforms and sequencing runs, as NPX values are arbitrary normalizations for a specific Olink project and cannot be compared or extrapolated between projects.

RESULTS

Count distributions are comparable between Illumina and Ultima

To determine if the Ultima UG 100 gave equivalent sequencing depth compared to the Illumina NovaSeq X Plus, we converted four Olink Explore HT Illumina libraries to Ultima format and sequenced each library at two different concentrations (600 pM and 700 pM). First, comparison of raw counts revealed that total counts on all Illumina were in between those from the Ultima 600 and 700 pM libraries (Figures 1, 2).

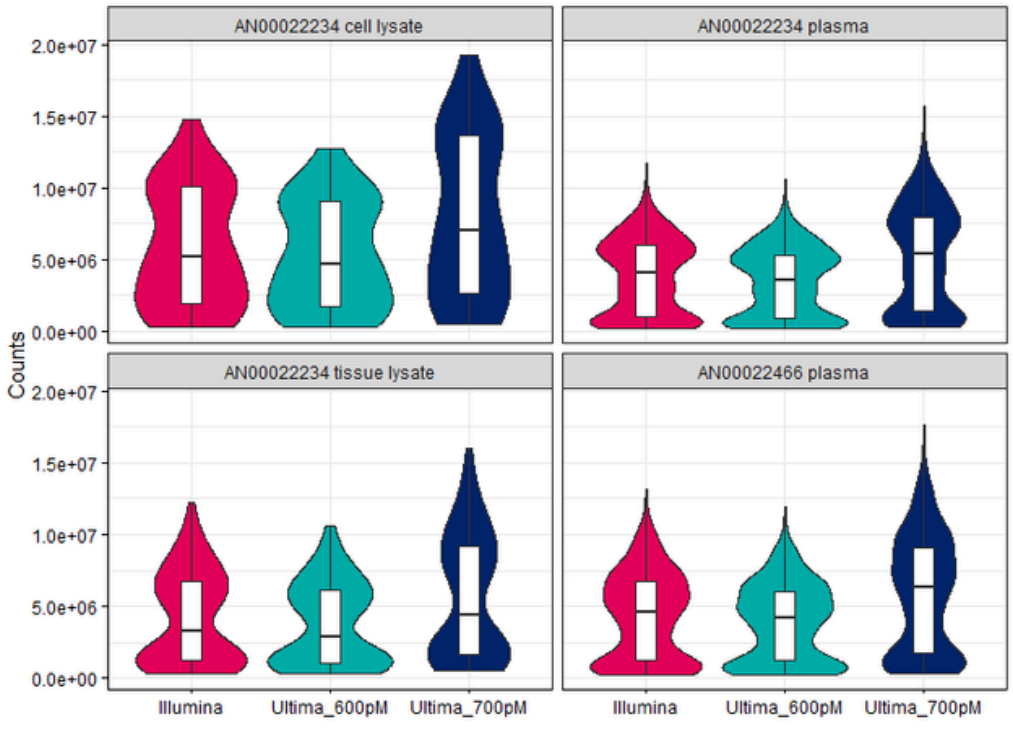

Figure 1: Distribution of total counts per platform. Sum of assay counts per sequencing platform from samples and controls; graphs are split by order ID and sample type.

Figure 2: Total counts per sample. Assay counts were summed per sample and control for each library separated by individual order and tissue type. Illumina counts are shown in pink, Ultima 600 pM in light blue, and Ultima 700 pM in dark blue.

Then, we compared the correlation between samples to determine if the observed differences in counts were consequential. Negative, Plate, and Sample Controls were strongly correlated between Illumina and both Ultima runs (R = 0.97 – 1.0, $p < 2.2x10^{-16}$) (data not shown). Next, we examined the correlation of all protein assays between samples. These counts were highly concordant (R = 0.99 – 1.0, $p < 2.2x10^{-16}$) (Figure 3), demonstrating high reproducibility between platforms.

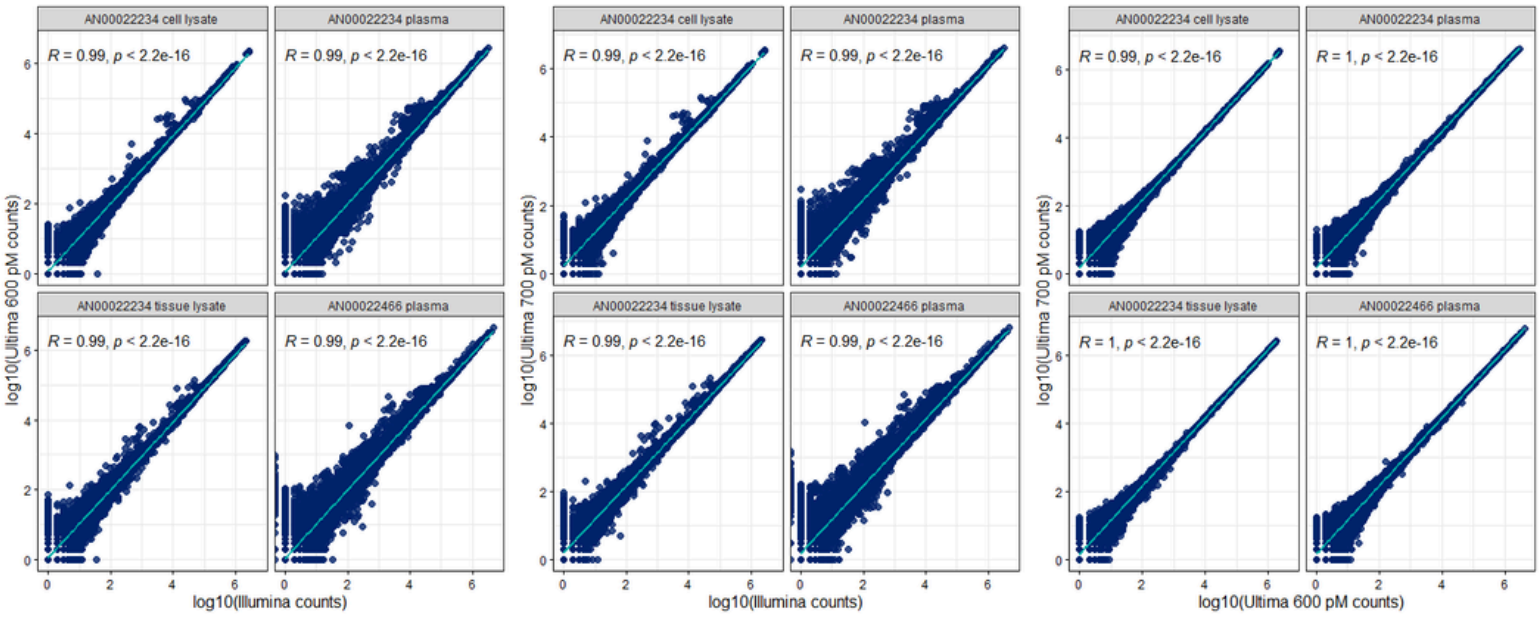

Figure 3: Counts are strongly correlated between sequencing runs. Pearson correlation coefficient for all protein assays per sample, separated by individual order/tissue type. Comparison groups include Illumina versus Ultima 600 pM (left two columns), Illumina versus Ultima 700 pM (middle two columns), and Ultima 600 pM versus 700 pM (right two columns).

Sample protein abundances are highly consistent across runs

Having shown that the counts for each sample were comparable, we asked if samples' protein abundance were more similar to themselves across platforms or to other samples sequenced together on the same platform. We performed principal component analysis (PCA) and examined distribution of controls and samples. As shown in Figure 4, controls consisted clustered together by type (negative, plate, and sample) independent of the sequencing platform. Similarly for each dataset, the pattern of sample distribution was highly similar, although separation between each run was found, indicative of sequencing batch effects (Figure 4, next page page). Taken together, these data indicate that the protein abundances in each sample are highly consistent across sequencing runs.

Only 3 out of 5,416 protein assays were differentially expressed (padj < 0.1) between the two Ultima libraries, however, this increased to 54 proteins between the Illumina and Ultima 600 pM library and 46 proteins between the Illumina and Ultima 700 pM library (Figure 5, next page). Overall, 46 (82%) differentially expressed proteins (DEPs) were common in these comparisons, suggesting minimal differences were caused by sequencing at different library concentrations.

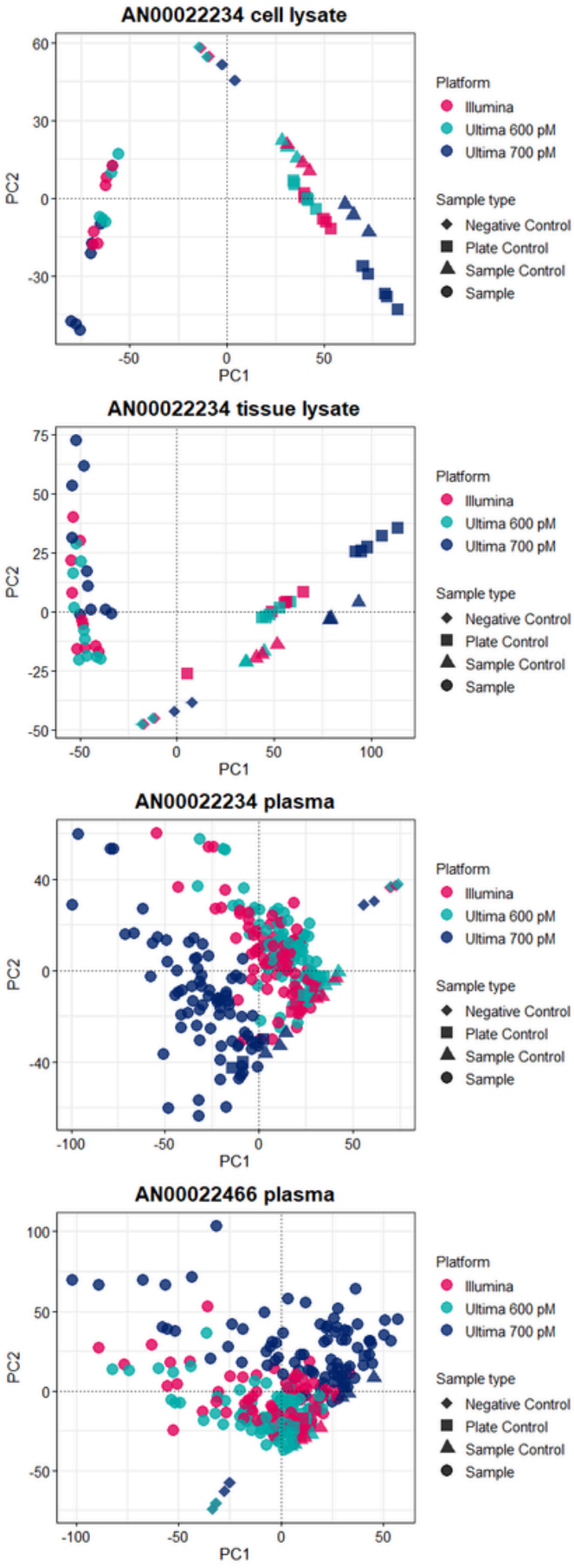

Figure 4 (left): PCA reveals similar clustering patterns across sequencing runs. Illumina sequencing runs are shown in pink while Ultima 600 pM and 700 pM sequencing runs are shown in light and dark blue, respectively. Diamonds indicate Negative Controls, squares indicate Plate Controls, triangles indicate Sample Controls, and circles indicate Samples.

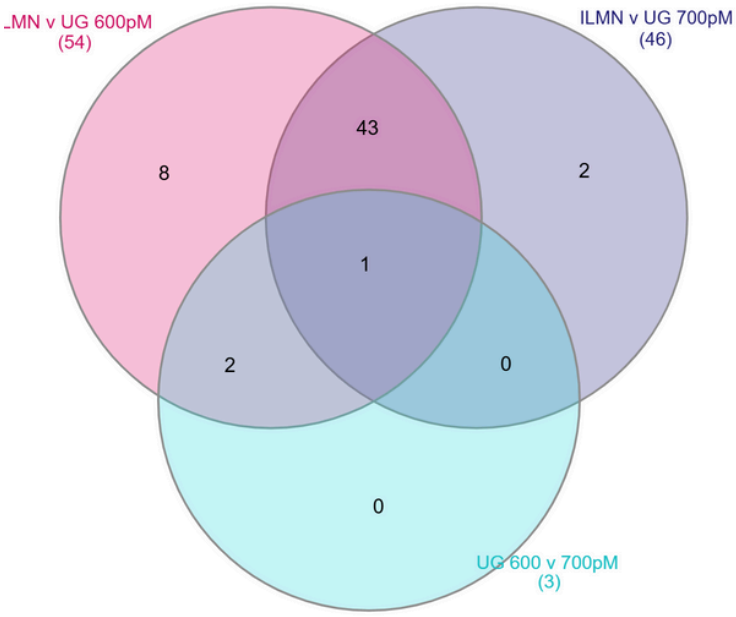

Figure 5: (above) Comparison of number of differentially expressed proteins. AN00022466 comparison groups from pairwise T-test: Illumina vs Ultima 600 pM (pink), Illumina vs Ultima 700 pM (navy), and Ultima 600 pM vs 700 pM (light blue). Total number of differentially expressed proteins (DEPs) (up and down regulated) for each comparison are indicated in each circle fragment or overlap region.

Bridging samples removes batch effects observed between sequencing runs

As the protein abundances were highly consistent between sequencing and PCA plots showed comparable dispersion patterns, we hypothesized these dispersions were likely caused by sequencing run batch effects. To test this, we performed bridging between Illumina and Ultima 600 pM libraries only, as bridging is limited to two datasets and the strong concordance observed between the two Ultima libraries (Figure 3).

This procedure uses the NPX values from 16-32 samples overlapping between Olink Explore HT runs to normalize these values across all samples and proteins [4].

For AN00022234 cell (n = 6) and tissue (n = 10) lysate projects, all samples were used, while 24 overlapping samples were randomly selected for each of AN00022234 and AN00022466 plasma projects. Following bridging and NPX adjustment, we found that all samples showed strong overlap (Figure 6), indicating that the observed differences shown in Figure 4 can be solely attributed to batch effects and that we find no difference in read quality between libraries sequenced on the Illumina NovaSeq X Plus and the Ultima UG 100. We then compared counts for each protein assay between the three sequencing platforms to determine if and how measured abundance may fluctuate.

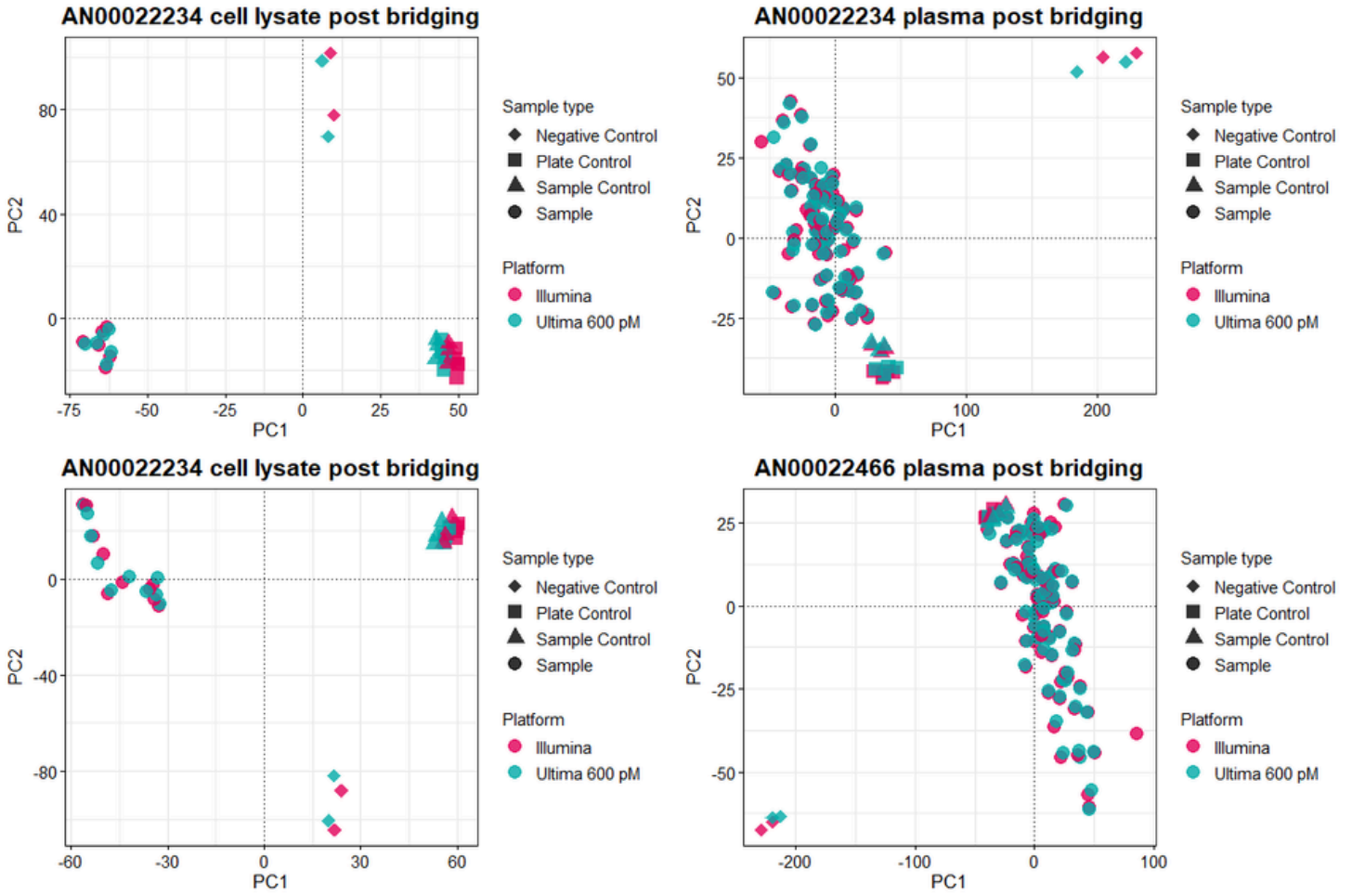

Figure 6: Bridging samples corrects for sequencing-caused batch effects. NPX values for 6, 24, 10, and 24 samples present in AN00022234 cell lysate, plasma, tissue lysate, and AN00022466 plasma projects, respectively, for both Illumina and Ultima 600 pM libraries were randomly chosen for bridging other samples. Illumina sequencing runs are shown in pink while Ultima 600 pM sequencing runs are shown in light blue. Diamonds indicate Negative Controls, squares indicate Plate Controls, triangles indicate Sample Controls, and circles indicate Samples.

CONCLUSIONS

Cost-effective solutions are critical for emerging next-generation sequencing technologies. Here, we compared the use of the "flow-based" sequencing of the Ultima UG 100 against the gold standard Illumina NovaSeq X Plus for efficacy using the Olink Explore HT proteomics panel as a test subject. We found that Olink libraries, originally prepared for Illumina sequencing, could be adapted for the UG 100 and returned comparable numbers of raw counts compared to the Illumina across multiple datasets. This was consistent across controls and samples.

Furthermore, PCA showed highly similar, but separated, clustering patterns which likely result solely from batch effects. Bridging would be a viable strategy to reduce these batch effects, by including 16-32 samples between Olink Explore HT panels, particularly for longitudinal studies. Collectively, these data show that the Ultima UG 100 performs comparably to the Illumina NovaSeq X Plus for the Olink Explore HT assay, but at much lower cost.

LITERATURE CITED

1. Olink Explore HT. https://olink.com/products/olink-explore-ht. Accessed March 13, 2025.
2. Wik L, N Nordberg, J Broberg, J Björkesten, E Assarsson, S Henriksson, I Grundberg, E Pettersson, C Westerberg, E Liljeroth, A Falck, M Lundberg. 2021. Mol Cell Proteomics 20:100168.
3. Almogy G, M Pratt, F Oberstrass, L Lee, D Mazur, N Beckett, O Barad, I Soifer, E Perelman, Y Etzioni, M Sosa, A Jung, T Clark, E Trepagnier, G Lithwick-Yanai, S Pollock, G Hornung, M Levy, M Coole, T Howd, M Shand, Y Farjoun, J Emery, G Hall, S Lee, T Sato, R Magner, S Low, A Bernier, B Gandi, J Stohlman, N Lennon, S Gabriel, D Lipson. 2022. Cost-efficient whole genome-sequencing using novel mostly natural sequencing-by-synthesis chemistry and open fluidics platform. bioRxiv 2022.05.29.493900.
4. Bridging across NGS-based Olink products. https://cran.r-project.org/web/packages/OlinkAnalyze/vignettes/bridging_crossproduct.html. Accessed April 25, 2025.