# Bridging Olink® Explore 3072 to Olink® Explore HT

**Compiled: August 08, 2024**



## Introduction

Individual Olink® NPX™ data sets are generally normalized using either plate control normalization or intensity normalization methods. Since NPX is a relative measurement, in the case when a study is separated into multiple batches, an additional normalization step is needed to allow the data to be comparable. The following tutorial is designed to give you an overview of the Olink bridging procedure for combining data sets from Olink® Explore 3072 and Olink® Explore HT products.

### Within- and between-product bridging

The joint analysis of two or more NPX datasets run on the same Olink product often requires an additional batch correction step to remove technical variation, which is referred to as overlapping sample reference normalization, bridge normalization, or just simply bridging. For more information on within-product bridging, see the Introduction to Bridging tutorial.

Olink Explore 3072 and Olink Explore HT are both products that use PEA technology combined with next generation sequencing to calculate NPX for thousands of proteins. Since many of the assays profiled in Olink Explore 3072 are also found on Olink Explore HT, bridging data across products enables increased power in studies consisting of both Explore 3072 and Explore HT data sets, rather than limiting these studies to meta-analysis. However, these products differ in the number of assays measured and in the **antibodies and reagents used**, which means that normalization cannot be performed in the same way that within-product bridging is performed.

In the case where a study consists of separate batches run on Olink Explore 3072 and Olink Explore HT, an additional batch correction step is required to allow data from these two products to be analyzed together, which is referred to as between-**product bridging, or Olink Explore 3072 to Olink Explore HT bridging**.

### Considerations for between-product bridging

Product bridging allows the NPX values of an Olink Explore 3072 project to be normalized and made comparable to the NPX values of an Olink Explore HT project. This process is one-directional, and normalizing Olink Explore HT NPX values to Olink Explore 3072 is not supported.

The product bridging normalization uses the **~3000** assays that are overlapping between Olink Explore 3072 and Olink Explore HT. Each overlapping assay undergoes a series of checks that evaluate the number of counts, correlation, and difference of NPX ranges between the two data sets. If an assay has enough counts and comparable metrics between the two data sets, it is determined to be suitable for bridging (referred to as a "bridgeable assay"). The set of "bridgeable assays" across platforms will vary from data set to datas et, based on the samples present within the studies. Depending on the NPX distribution of each bridgeable assay in the two data sets, the assay is normalized using either median normalization or **QQ normalization**.

Product bridging between an Explore 3072 and an Explore HT NPX dataset requires **40-60 bridging samples**. Bridging samples are shared samples among datasets - that is that samples that are analyzed in both datasets. Olink® NPX datasets without shared samples should not be combined using the bridging approach described below. Before starting bridging, it is important to check if the same sample IDs were assigned to the bridging samples. More information on bridge sample selection can be found in the selecting bridging samples section of the Introduction to Bridging tutorial.

## Bridge Sample Selection

Prior to running the second study, bridging samples must be selected from the Explore 3072 study and be added to the Explore HT study. These samples can be selected using the `olink_bridgeselector()` function in Olink Analyze. The bridge selection function will select a number of bridge samples based on the expression in Olink Explore 3072. This function selects samples which passes QC and does not include external control samples. External controls are not selected as bridge samples as they are not necessarily representative of the study and therefore may not cover the dynamic range of assays that would be expressed within the samples. Note that due to naming convention differences, it is necessary to exclude the control samples using either `SampleType == "SAMPLE"` if available or `stringr::str_detect()` as shown below.

To select samples across the range of the data, the samples are ordered by mean NPX value and selected across this range. In the case where LOD data is available, Olink recommends starting at `sampleMissingFreq = 0.10` which represents a maximum of 10% data below LOD per sample. If there are not enough samples output, increase to 20%. For alternative matrices and specific disease types, it may be needed to increase the sampleMissingFreq to higher levels. For cases where LOD data is not available, `sampleMissingFreq` can be set to 1.

In the example below we demonstrate how to select **45** bridging samples using `npx_data1` which will act as the Explore 3072 data. The selected bridge samples are displayed in Table 2.

```
bridge_Samples<- npx_data1 %>%
  # Excluding control samples. Naming convention may differ.
  dplyr::filter((stringr::str_detect(SampleID, "CONTROL", negate = TRUE))) %>%
  olink_bridgeselector(sampleMissingFreq = 0.1,
                       n = 45)
```

It's important to make sure that the select bridge samples are representative of the overall samples within the study. This can be done by generating a PCA plot with the following code and make sure that the bridge samples are evenly dispersed among the other samples, as shown in Figure 1.

```
npx_data1 %>%
  filter(!str_detect(SampleID, 'CONT')) %>%
```

```
mutate(Bridge = ifelse(SampleID %in% bridge_Samples$SampleID, "Bridge", "Sample")) %>%
olink_pca_plot(color_g = "Bridge")
```

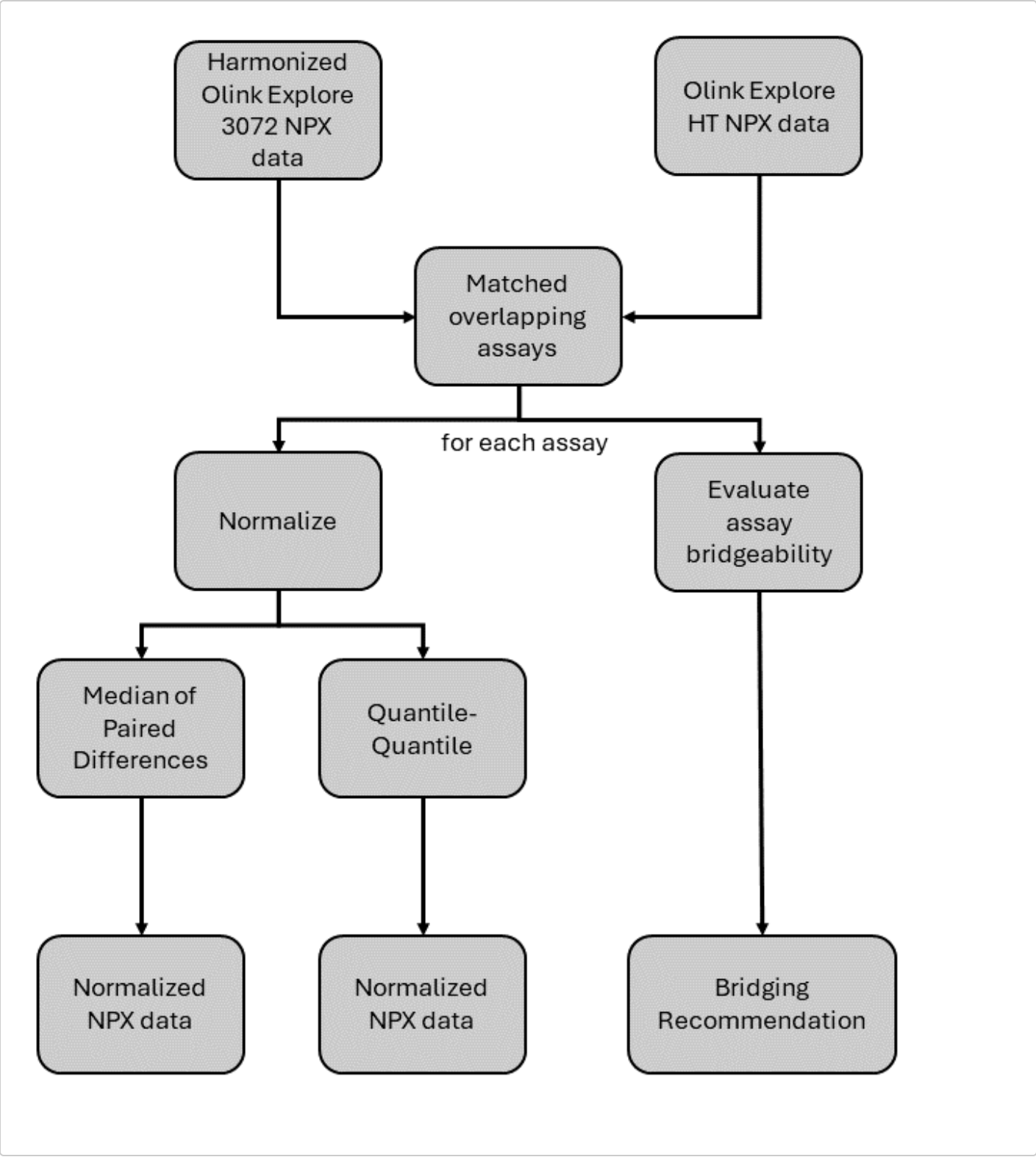
## Workflow Overview

Olink Explore 3072 to Olink Explore HT bridging required harmonized Explore 3072 data and Explore HT data which has at least **40 to 60** overlapping bridge samples. For studies containing multiple batches of Explore 3072 data, the Explore 3072 data sets should be bridged using within product bridging as detailed in the Introduction to bridging tutorial prior to performing between product bridging.

The assays from Explore 3072 are matched to the corresponding assays in Explore HT and evaluated to determine if the assay is bridgeable. In parallel, the assays are normalized using **quantile-quantile normalization** and normalization using the median of paired differences. The result is an adjusted Explore 3072 dataset with three additional columns:

- a flag which indicates if the assay is bridgeable and, if so, which normalization method is recommended
- NPX values for normalization using the median of paired differences
- NPX values using **quantile-quantile normalization**

A visual representation of the workflow is shown below.




Schematic of Explore 3072 to Explore HT Bridging Workflow

Note that regardless of the bridging recommendation, NPX values will be available for both normalization methods.

## Import NPX files

To normalize Explore 3072 data to Explore HT data, first the two data sets are read into R using `read_NPX()`. If more that two data sets are being normalized, all Explore 3072 studies should be normalized together prior to normalizing between products. In the case of multiple Explore HT studies, only one Explore HT study should be chosen as the reference data set. The data can be loaded using `read_NPX()` function with default Olink Software NPX file as input, as shown below.

```
data_E3072 <- read_NPX("~/NPX_Explore3072_location.parquet") # Could also be a CSV file
data_EHT <- read_NPX("~/NPX_ExploreHT_location.parquet")
```

## Checking input datasets and bridging samples

First, confirm that there are overlapping sample IDs within the study. The bridge samples should have the same Sample IDs in both platforms. Note that if external controls share the same SampleID in both data sets, these samples should not be included in the list of bridging samples. External control samples often share the same naming convention across data sets but may represent different samples due to reagent batch differences. Appending the project name to the end of the control samples can ensure unique Sample IDs.

```
#### Extract bridging samples

overlapping_samples <-  data.frame(SampleID = intersect(npx_1$SampleID, npx_2$SampleID)) %>%
  filter(!str_detect(SampleID, "CONTROL_SAMPLE")) %>% #Remove control samples
  pull(SampleID)

npx_before_br <- npx_data1 %>%
  dplyr::filter(!str_detect(SampleID, "CONTROL_SAMPLE")) %>% #Remove control samples
  dplyr::mutate(Type = if_else(SampleID %in% overlapping_samples,
                        paste0("data1 Bridge"),
                        paste0("data1 Sample"))) %>%
  rbind({
    npx_data2 %>%
      filter(!str_detect(SampleID, "CONTROL_SAMPLE")) %>% #Remove control samples %>%
      mutate(Type = if_else(SampleID %in% overlapping_samples,
                            paste0("data2 Bridge"),
                            paste0("data2 Sample"))) %>%
      mutate(SampleID = if_else(SampleID %in% overlapping_samples,
                                paste0(SampleID, "_new"),
                                SampleID))
  })

### PCA plot
OlinkAnalyze::olink_pca_plot(df          = npx_before_br,
                             color_g     = "Type",
                             byPanel     = TRUE)
```

## Normalization

The `olink_normalization_bridge_product()` function is used to determine which assays are bridgeable and to bridge across products as described in the Workflow Overview and in the sections below. Within this function the bridging recommendations for each assay are determined and the NPX values are normalized using the two methods described below.

```
# Find shared samples
npx_1 <- npx_data1 %>%
  mutate(Project = "data1")
npx_2 <- npx_data2 %>%
  mutate(Project = "data2")
```

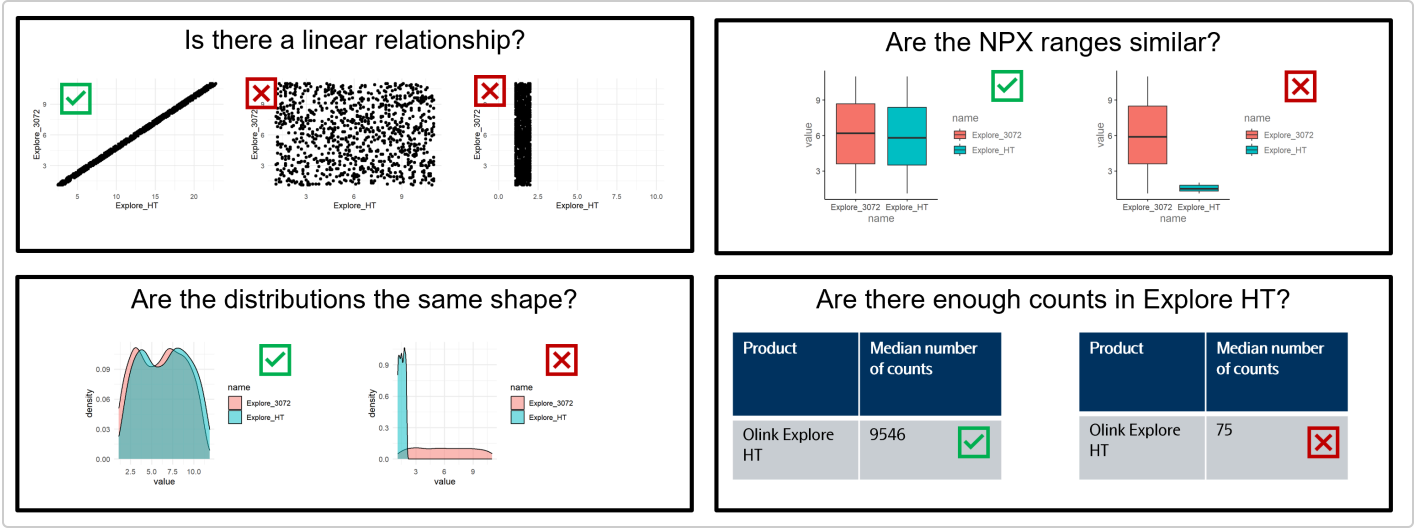
```
overlap_samples <-data.frame(SampleID = intersect(npx_1$SampleID, npx_2$SampleID)) %>%
  filter(!str_detect(SampleID, "CONTROL_SAMPLE")) %>% #Remove control samples
  pull(SampleID)

overlap_samples_list <- list("DF1" = overlap_samples,
                             "DF2" = overlap_samples)
olink_normalization_bridge_product(exploreht_df,
                                   explore3072_df,
                                   bridge_samples,
                                   exploreht_name = "reference",
                                   explore3072_name = "new")
```

## Determining bridging recommendations

There are four criteria that are used to determine if an assay is bridgeable and what normalization method should be used:

- Is there a linear relationship between products?
  - **Assessing linearity across products:** To determine if there is a linear relationship between products for an assay, the linear coefficient of determination is calculated using Pearson correlation. In this correlation counts below 10 are excluded due to lack of signal. The $R^2$ value is calculated and an assay is considered to have a linear relationship across products if the $R^2$ value is above the cutoff. The default cutoff is set to $R^2 > 0.8$.
- Are the NPX ranges in the two products similar?
  - **Assessing similarity of NPX ranges:** To determine if the NPX ranges are similar across products, the difference in NPX values from the 10% to 90% quantile is calculated for each platform, excluding data points with counts less than 10. If the difference in range between platforms is greater than the cutoff than the ranges are not considered similar across platforms. By default, the cutoff is set to a difference of less than 1 NPX between platforms.
- Are there sufficient counts in the Explore HT data?
  - **Assessing if there are sufficient counts:** To determine if there are sufficient counts in an assay for bridging, the median number of counts from Explore HT is calculated, excluding counts less than 10. If the median number of counts is less than the cutoff than the assay does not have sufficient counts to be used for bridging. The default cutoff is set to 150 counts.
- Are the distributions between products the same shape?
  - **Assessing similarity of NPX distribution across products:** If the three criteria outlined above are met than the assay is considered bridgeable. Otherwise, bridging is not recommended for that assay. If an assay is bridgeable, the similarity of the NPX distribution is used to determine which method is recommended for bridging. The Kolmogorov-Smirnov test, or KS test, is used to assess the similarity of two distributions by calculating the distances between the NPX distributions for each product. If the distance is above the cutoff than quantile-quantile normalization is recommended. If the distance is less than the cutoff than normalization using the median of paired differences is recommended. By default this cutoff difference is set to 0.2.

An overview of these criteria is visualized below.




Criteria to determine the bridging recommendation for an assay

Prior to assessment, outlier bridge samples are excluded. A sample is considered an outlier if the NPX value is more than 3 times the interquartile range above or below the median on either platform.

After assessment, an assay is considered bridgeable if it meets the first three criteria. The fourth criteria determines which normalization method is recommended for bridging. If all four criteria are met than the recommended method is normalization using the median of paired differences. If only the first three criteria are met then quantile-quantile normalization is recommended. If any of the first three criteria are not met than bridging is not recommended for that assay. Note that bridgeable assays will differ between projects based on the expression of bridge samples in the studies.

## Normalization using the median of paired differences

If it is expected that both the kind of distribution and the equal variance per test between runs are the same, then the normalization using the median of paired differences will be preferred. The normalization using the median of paired differences based on the bridge samples is performed in the following steps:

1. For each assay and the Explore 3072 project, calculate the pairwise difference for each of the overlapping samples with the Explore HT project.
2. Estimate the normalization factor of each assay by finding the median of the pairwise differences.
3. Use the assay-specific normalization factor for each assay to normalize each data point from Explore 3072 to Explore HT.

## Quantile-Quantile normalization

Since Explore HT and Explore 3072 are two distinct products and the assays differ considerably between products, some of the assays exist in corresponding but distinct NPX spaces. For those assays, a Quantile-Quantile normalization will be favored. The normalization using the Quantile-Quantile Normalization based on the bridge samples is performed in the following steps:

1. For each assay of Explore 3072, calculate the empirical distribution function.
2. Map the empirical distribution function from Explore 3072 to the Explore HT space using the quantiles.
3. Construct a spline regression model using the sorted Explore 3072 data and the mapped Explore 3072 data along with the anchor points of the spline function.
4. Use the spline regression model to predict all the data points from Explore 3072 to Explore HT.

## Function Output

The output from `olink_normalization_bridge_product()` function is a dataframe with concatenated data from the two products and additional columns including adjusted NPX values, bridging recommendations, mapping information, and project names. The adjusted NPX values are notated in the columns `MedianAdjustedNPX` and `QQNormalizedNPX`. For each assay a recommendation is listed in the `BridgingRecommendation` column and lists what method, if any should be used for that assay. Additional columns including `OlinkID_HT` and `OlinkID_3072` map the assays across products and the `Project` column lists the name of the project based on the `exploreht_name` and `explore3072_name` arguements. The resulting dataset will contain the Explore HT data set, which will be identical to the input reference data and the newly bridged Explore 3072 data set.

## Evaluating the quality of bridging

PCA is used to assess the quality of bridging by determining if the sample controls SCs) and bridging samples appear closer after bridging. Two PCAs can be generated, one containing the SCs and one containing the bridging samples. Prior to bridging there will be a noticeable separation between products which should decrease after bridging.

```
## After bridging

### Generate unique SampleIDs

npx_after_br <- npx_br_data %>%
  dplyr::mutate(Type = ifelse(SampleID %in% overlapping_samples,
                              paste(Project, "Bridge"),
                              paste(Project, "Sample"))) %>%
  dplyr:::mutate(SampleID = paste0(Project, PlateID, SampleID))

### PCA plot
OlinkAnalyze::olink_pca_plot(df          = npx_after_br,
                             color_g     = "Type",
                             byPanel     = TRUE)
```

## Exporting Normalized Data

The normalized Explore 3072 data can be exported using `arrow::write_parquet()` to create a long format Olink Explore file.

```
df <- normalized_data |>
    dplyr::filter(Project == "Explore_3072_NAME") |>
    arrow::as_arrow_table()

df$metadata$FileVersion <- "NA"
df$metadata$ExploreVersion <- "NA"
df$metadata$ProjectName <- "NA"
df$metadata$SampleMatrix <- "NA"
df$metadata$DataFileType <- "Olink Analyze Export File"
df$metadata$ProductType <- "Explore3072"
df$metadata$Product <- "Explore3072"
arrow::write_parquet(x = df, sink = "path_to_output.parquet")
```

## FAQs

### Correlation Assays

Both the Explore 3072 and Explore HT platforms contain correlation assays in their biomarker portfolios. In Explore 3072, these present as overlapping assays across panels. In Explore HT, these are overlapping assays across blocks. The correlation assays are included for QC purposes and allow users to evaluate data performance across panels in Explore 3072 and across blocks in Explore HT.

Within each platform, the correlation assays contain unique OlinkID values for each of their corresponding panels and blocks in Explore 3072 and Explore HT, respectively. It should be noted that the correlation assays themselves differ between each platform. For more information on overlapping assays, you can access the Olink FAQ page.

Explore 3072 to Explore HT bridging generates and returns adjusted values for all data points corresponding to each of the Explore 3072 correlation assays. In other words, each instance of an overlapping assay will contain an adjusted value in its respective Explore 3072 panel *For consistency in reference values, only the Block 5 data points from Explore HT's correlation assays are used in the Explore 3072 to Explore HT bridging process. For downstream analysis, it is recommended to include the data from all instances of the correlation assays, barring any clear outliers.*

### Downstream Analysis

## Contact Us

We are always happy to help. Email us with any questions:

- biostat@olink.com for statistical services and general stats questions
- support@olink.com for Olink lab product and technical support
- info@olink.com for more information

## Legal Disclaimer



Olink products and services are For Research Use Only and not for Use in Diagnostic Procedures.

All information in this document is subject to change without notice. This document is not intended to convey any warranties, representations and/or recommendations of any kind, unless such warranties, representations and/or recommendations are explicitly stated.

Olink assumes no liability arising from a prospective reader's actions based on this document.



Olink products and assay methods are covered by several patents and patent applications https://olink.com/legal/patents.