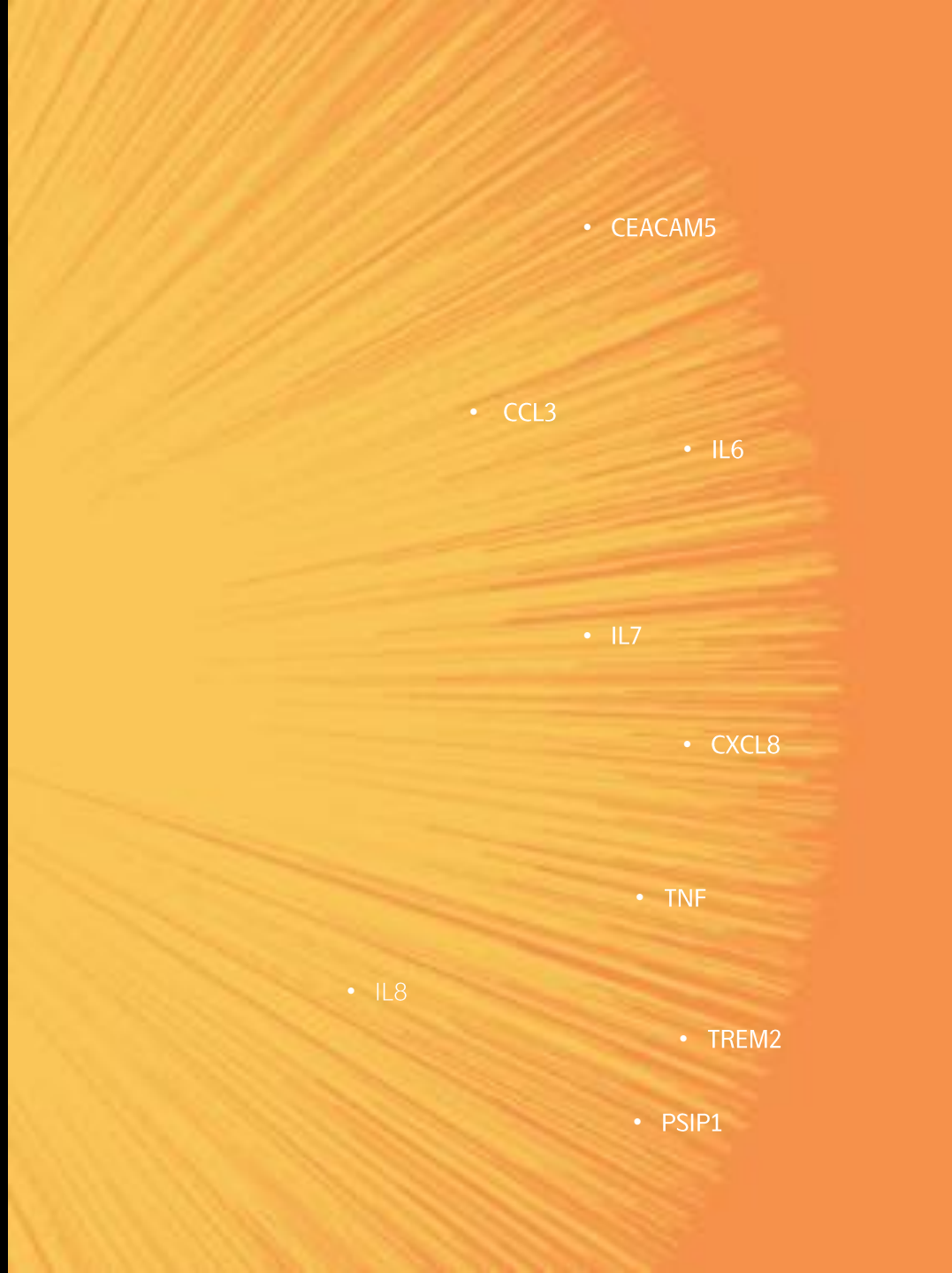
Welcome!

Olink® Explore HT Data QC

Olink®
Accelerating proteomics together

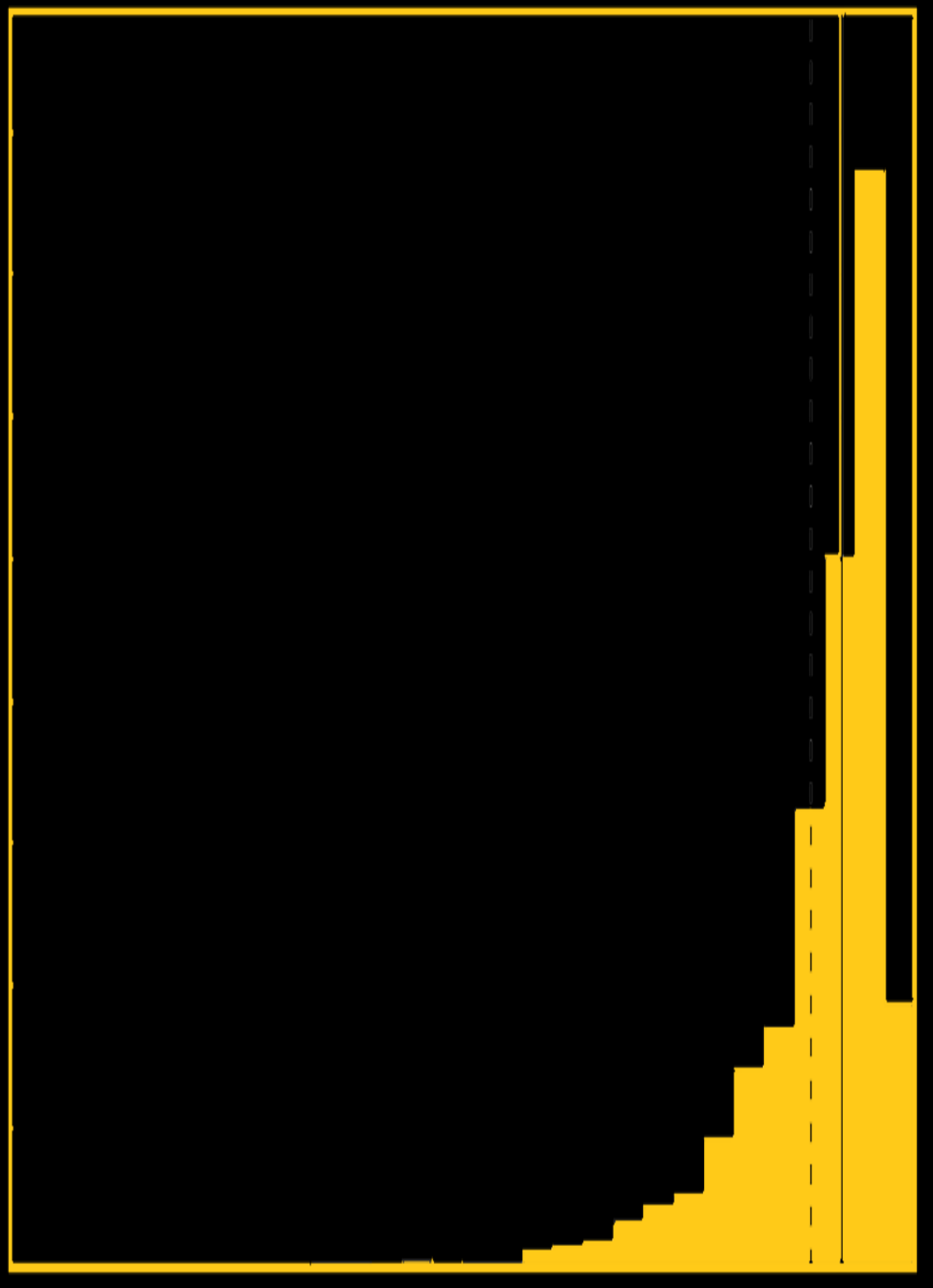
Agenda

- Explore HT Assay Validation
- Explore HT Normalization
- Explore HT QC workflow



Olink®
Accelerating proteomics together

Assay Validation



Olink®
Accelerating proteomics together

Unique biomarkers	New unique biomarkers	New customer “wish biomarkers”	Olink-developed antibodies
5,416	2,567	186	61%
>80% increase	Not in Explore 3072		

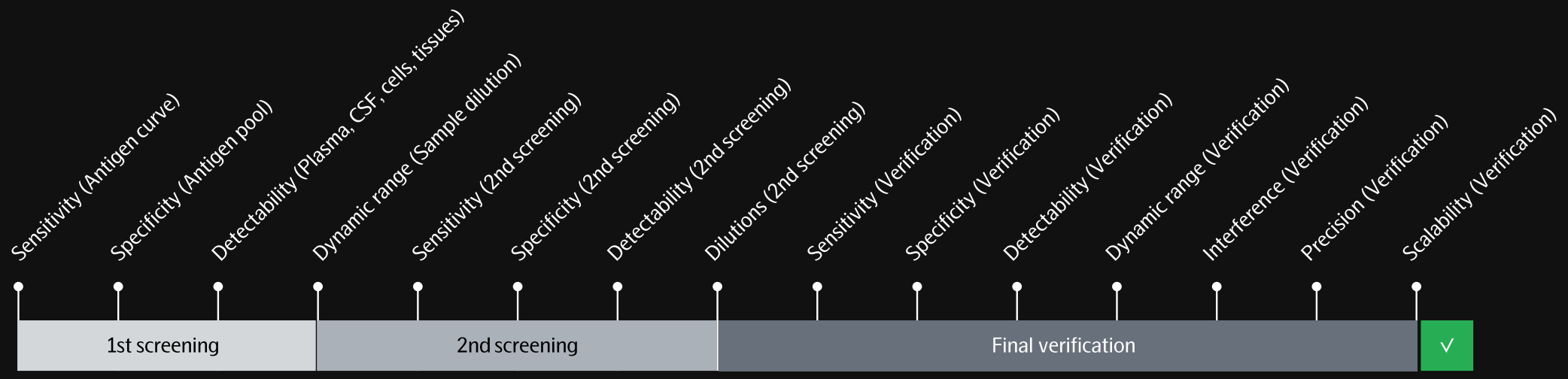
Excluding the 2 “overlapping assays” (GBP1, MAP2K1) that are included in 3 blocks, which are used to compare the performance of the different blocks. Total number of assays: 5420

Olink®
Accelerating proteomics together

All Olink assays undergo a rigorous three step,
15-factor analytical verification process

15,300 assays tested

5,400+ approved



Olink®
Accelerating proteomics together

Maintaining quality, ensuring reliability

qPCR screening

(96-plex)

- ✓ Plasma
- ✓ Serum
- ✓ CSF
- ✓ Tissue & cell lysates

NGS screening

(192-plex)

- ✓ Plasma
- ✓ Serum
- ✓ CSF

NGS data must be similar to qPCR data to pass

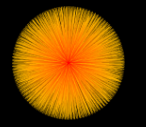
Verification

(5400+ plex)

Assays integrated into a product panel, and then undergo additional verification and validation studies

Passing criteria

1.Antigen curves	Range ~7 NPX
2. Specificity	No cross reactivity
3. Sample testing (Undiluted, 1:10, 1:00)	Uniform sample dilution + Antigen & samples hook at similar level, with similar dilution patterns



Olink Explore HT

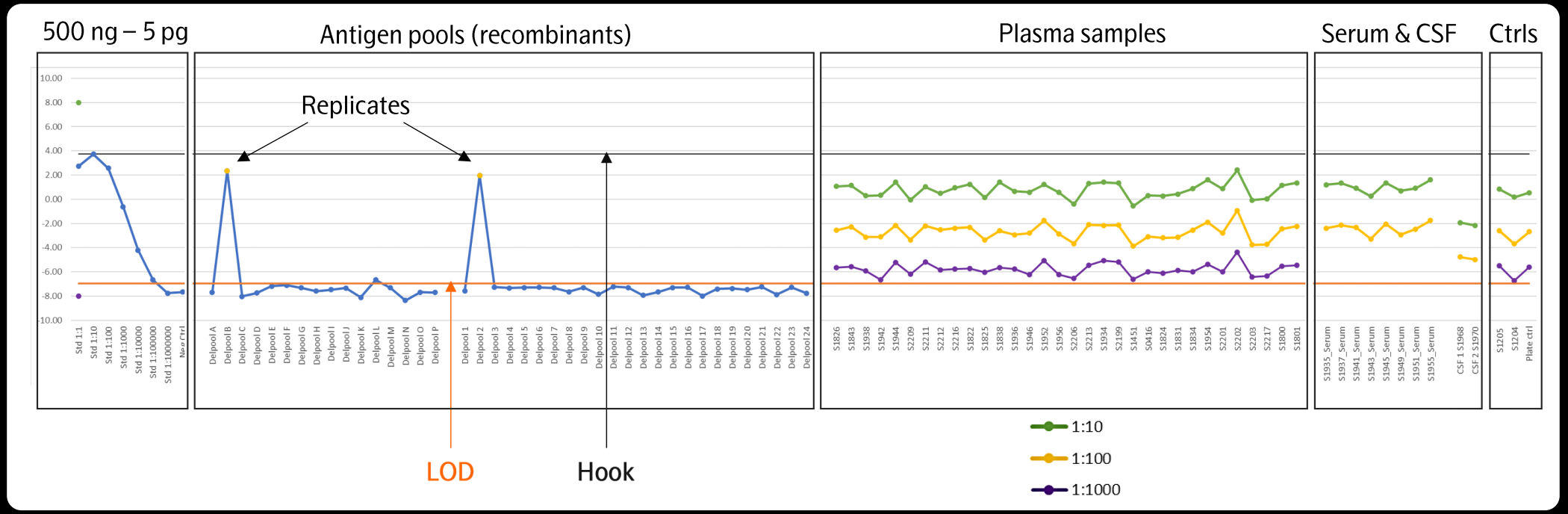
Olink®
Accelerating proteomics together

NGS screen: Passed biomarker assay

1 Antigen curve

2 Specificity

3 Dilution linearity



Olink®
Accelerating proteomics together

NGS screen: Failed biomarker assay

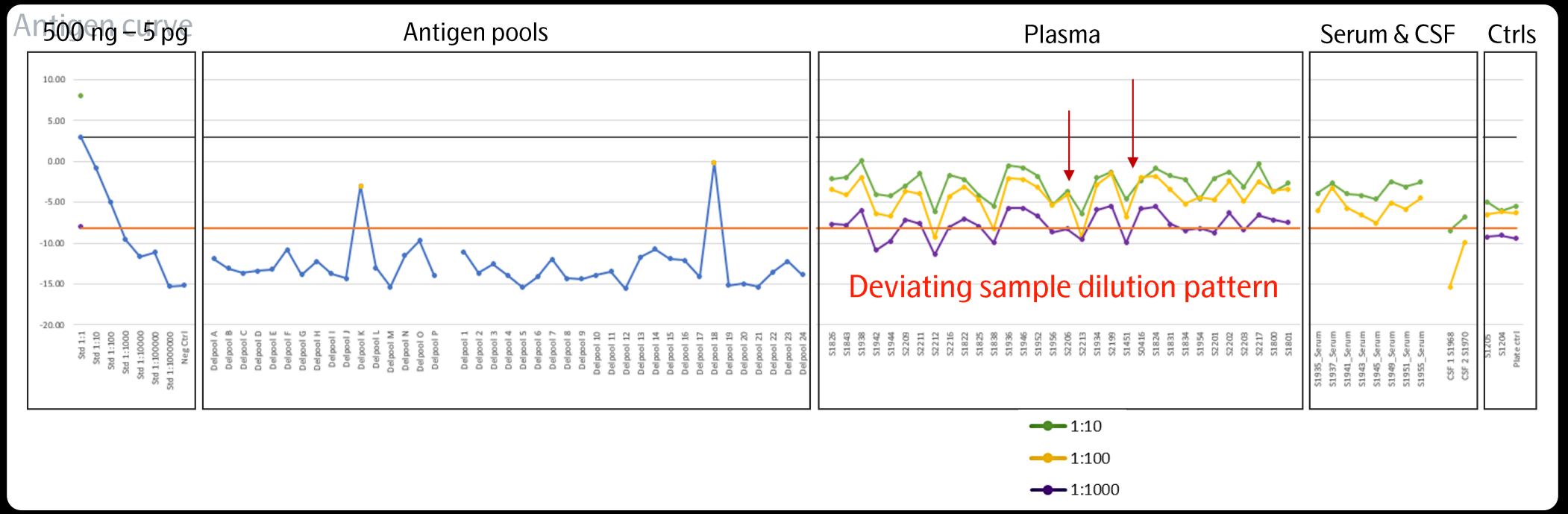
1

2

Specificity

3

Dilution linearity



Olink®
Accelerating proteomics together

Mean intra-CV
(within plate)

<11%

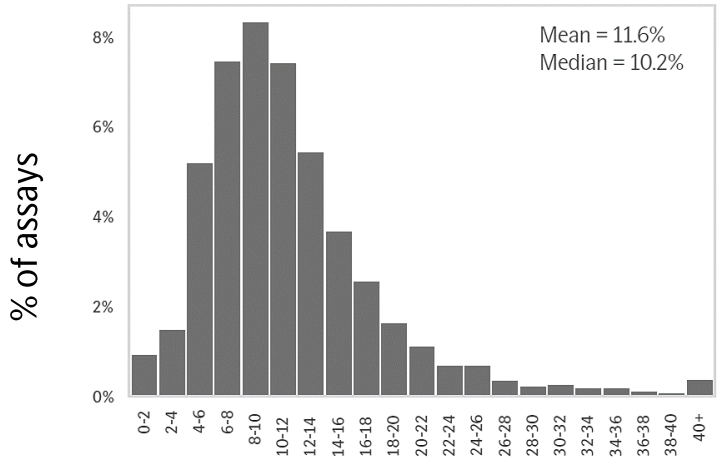
Explore 1536: <10%

Mean inter-CV
(between plates)

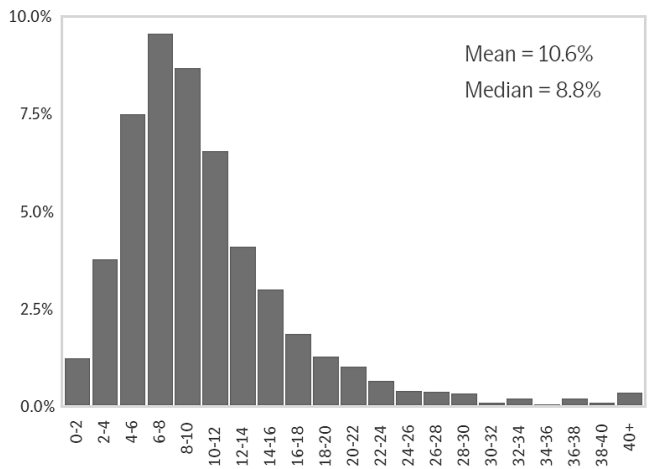
<9%

Explore 1536: <9%

Variation between Control Samples within a plate



Variation between plate means of Control Samples

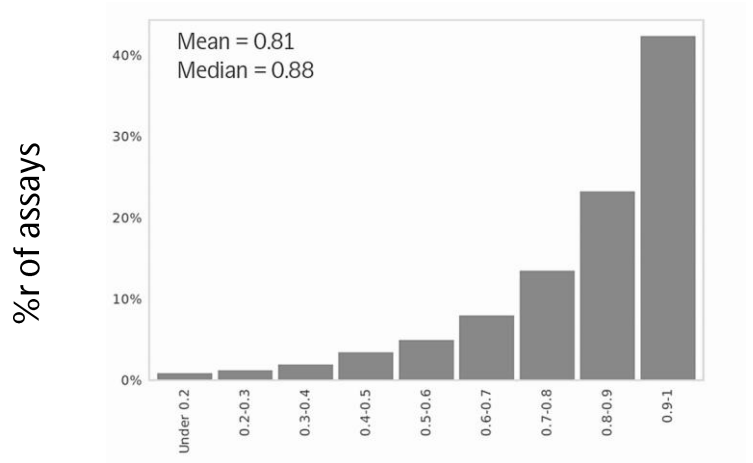


Coefficient of Variation (CV)

Olink®
Accelerating proteomics together

Per-assay correlation coefficient (R)
between Explore 3072 and Explore HT

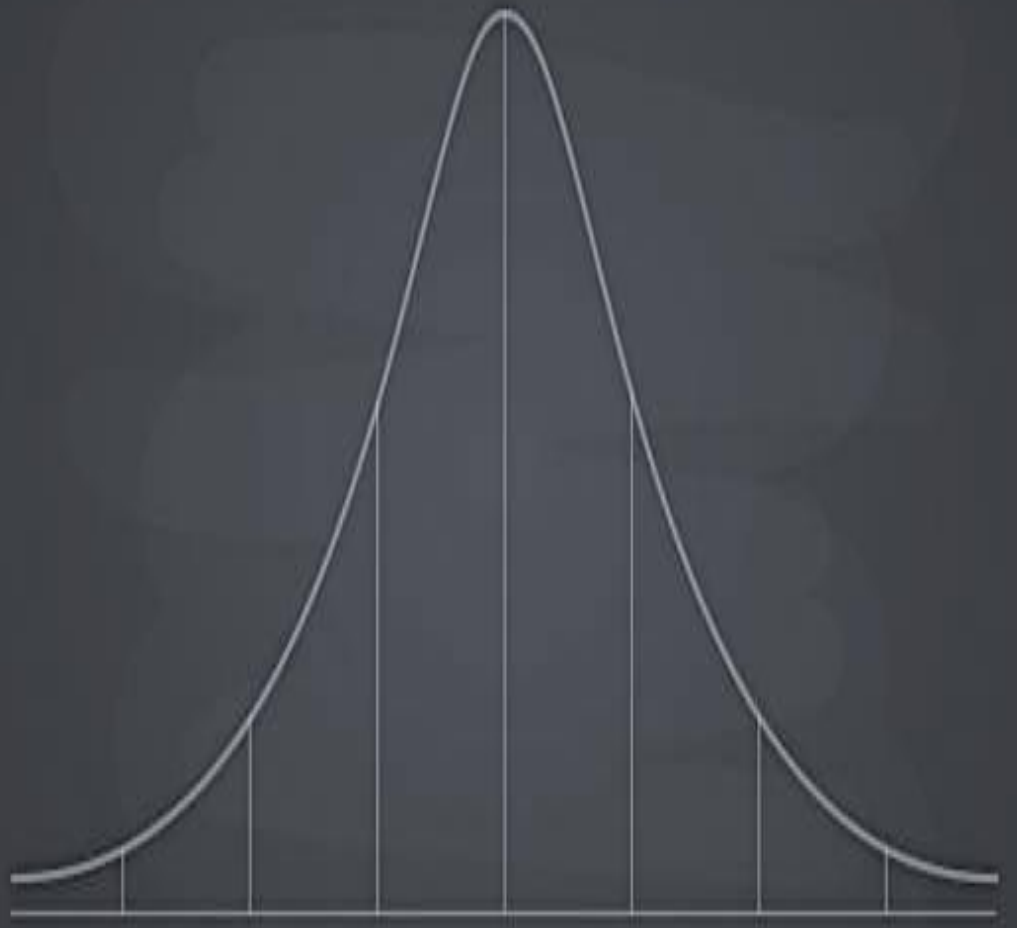
0.88 median



Correlation between Explore 3072 and Explore HT

Olink®
Accelerating proteomics together

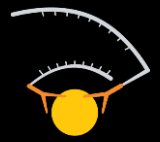
Data Normalization



Olink
Accelerating proteomics together

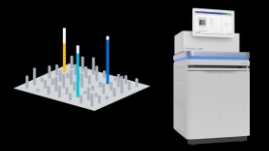
Explore HT quality control – Internal Controls

Incubation Reaction



Extension/ Amplification

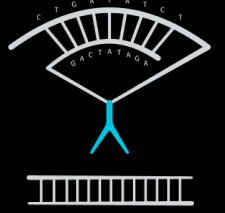
NGS Readout



Incubation control

Extension control

Amplification control



Incubation control

- 1 non-human antigen
- Monitors **all** steps of the assay
- Used for QC of samples and the entire run by comparing the experiment value in each sample with the plate

Extension control

- IgG conjugated with oligo pair
- Monitors **extension**, **pre-amplification** and **detection**
- Used for data normalization

Amplification control

- Synthetic double-stranded DNA
- Monitors **pre-amplification** and **detection**
- Used for QC of samples and the entire run by comparing the experiment value in each sample with the plate median

Olink®
Accelerating proteomics together

Explore HT quality control – External Controls

5 Plate Controls (PCs):

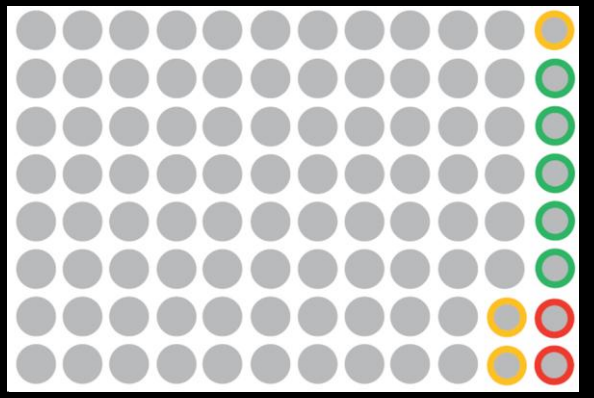
- Improve normalization performance, especially for low detecting assays
- Reduce need of bridging samples

3 Sample Controls:

- Ensure more robust CV calculation, even with one failed Sample Control

2 Negative Controls:

- Give more room to customer samples, while still sufficient for indicating contamination and plate swaps



Olink Plate Controls
Olink Sample Controls
Olink Negative Controls

Olink
Accelerating proteomics together

Calculating NPX

NPX represents the relative signal in log2 scale

1 $$ExtNPX_{i,j} = log2\left(\frac{counts\ (Sample_j\ Assay_i)}{counts\ (ExtCtr_j)}\right)$$

- For all assays and all samples (including negative controls, control samples, reference samples etc.)
- Relate counts to known standard (Extension control)
- Log2 transformation helps with more normalized distribution of data

2a $$NPX_{PC\ i,j} = ExtNPX_{i,j} - median\left(ExtNPX(plate\ controls_i)\right)$$

Plate control normalization:

- Normalized by the median of the plate controls on the same plate for the same assay
- Performs plate standardization

2b $$NPX_{IntNorm\ i,j} = ExtNPX_{i,j} - median\left(ExtNPX(samples_i)\right)$$

Intensity normalization:

- Normalized by the median of all samples on the same plate for the same assay (excluding the control strip)
- Performs plate standardization in randomized multi-plate projects

Olink®
Accelerating proteomics together

Bridging between platforms

- Bridging between Explore HT and 3K:
  - 40 – 64 bridging samples needed.
  - Bridging module available in Olink Analyze R
  - https://cran.r-project.org/web/packages/OlinkAnalyze/vignettes/bridging_E3072toEHT.html
  - Explore HT and Explore 3072 contain differences which need more bridging samples to address and perform optimal bridging correction than between the same products

- Bridging between Target and Explore:
  - Not recommended.
  - Target and Explore sample processing methods differs too much (qPCR vs Sequencing)

Olink®
Accelerating proteomics together

Bridging between platforms general rules

Recommended number of bridging samples Olink Platforms	
Platform	# Bridging samples
Target 96	8-16
Explore 384 Cardiometabolic, Inflammation, Neurology, and Oncology	8-16
Explore 384 Cardiometabolic II, Inflammation II, Neurology II, and Oncology II	16-24
Reveal	16-24
Explore HT	16-32
Explore 3072 to Explore HT	40-64
Explore 3072 to Reveal	32-48

https://cran.r-project.org/web/packages/OlinkAnalyze/vignettes/bridging_introduction.html

Olink®
Accelerating proteomics together

Olink® Explore HT software suite

Olink®
Accelerating proteomics together

Focus areas to develop a Software package that suites all Olink® products with NGS readout including Explore HT and addresses customer needs

Simplicity in Data QC

Automated handling of extra-large datasets

Flexibility in type of delivered data

Olink®
Accelerating proteomics together

An improved data QC

HIGHLY AUTOMATED

Reduces the number of manual steps resulting in a simpler and faster data QC

ROBUST

Independent of study design and sample quality

ACTIONABLE

Detects the most severe technical errors and helps identify failure root causes

Olink®
Accelerating proteomics together

Automated software solutions for a simplified & faster path to actionable insights

Olink® NPX MAP software

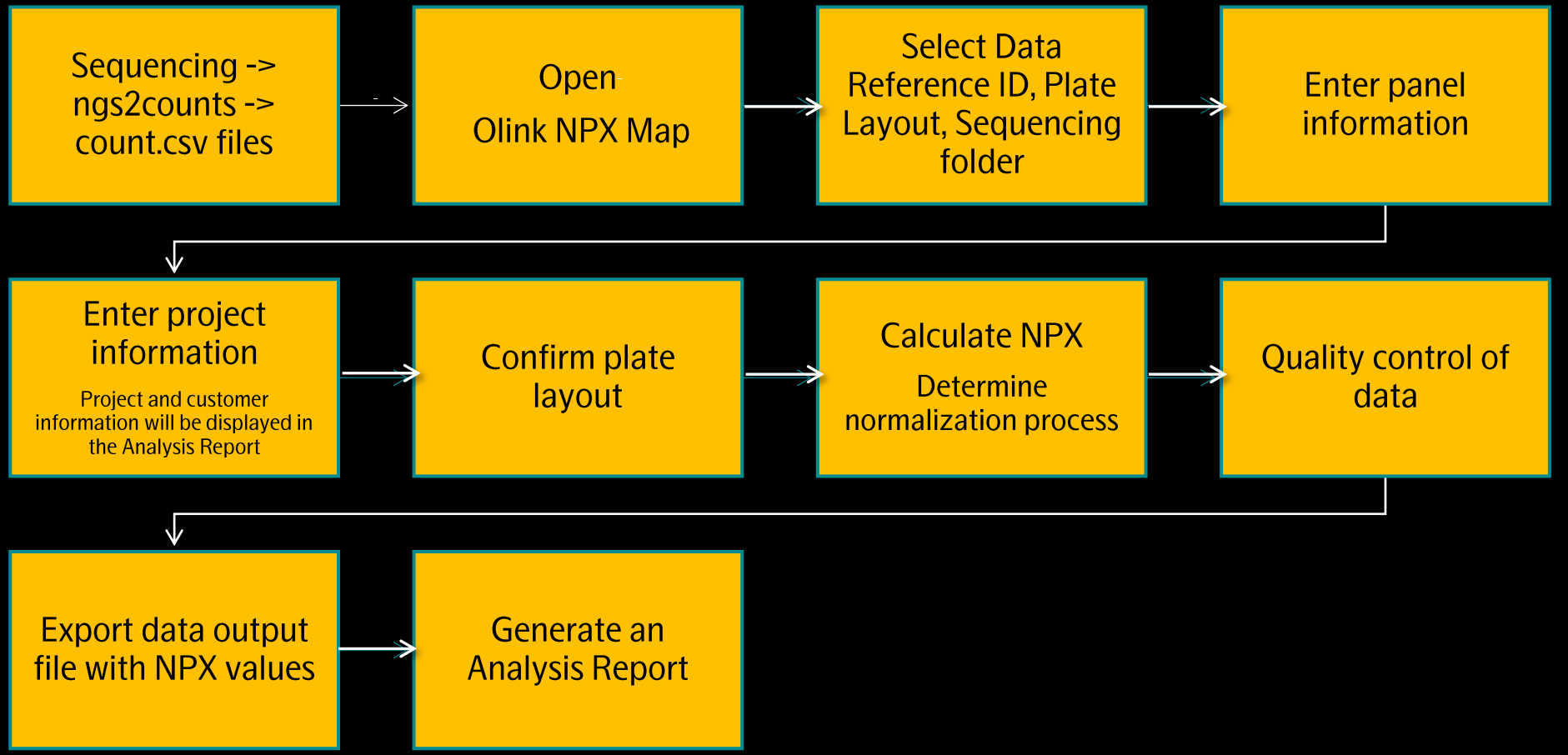
Desktop software

Olink® NPX MAP CLI software

Command Line Interface software

Olink
Accelerating proteomics together

Data Analysis and NPX MAP Workflow



Olink®
Accelerating proteomics together

Create Plate Layout

	A
1	well_id;sample_id;sample_type
2	A12;SC3;CONTROL
3	B12;PC1;PLATE_CONTROL
4	C12;PC2;PLATE_CONTROL
5	D12;PC3;PLATE_CONTROL
6	E12;PC4;PLATE_CONTROL
7	F12;PC5;PLATE_CONTROL
8	G12;NC1;NEGATIVE_CONTROL
9	H12;NC2;NEGATIVE_CONTROL
10	A1;S1;SAMPLE
11	B1;S2;SAMPLE
12	C1;S3;SAMPLE
13	D1;S4;SAMPLE
14	E1;S5;SAMPLE
15	F1;S6;SAMPLE
16	G1;S7;SAMPLE
17	H1;S8;SAMPLE

.csv format using semicolon as delimiter, template available

Header:
- well_id
- sample_id
- sample_type

well_id:
- A1, B1, C1, ...

sample_id:
- No character other than ”_”

sample_type:
- SAMPLE
- EMPTY
- CONTROL
- PLATE_CONTROL
- NEGATIVE_CONTROL
- NOT USED

Olink®
Accelerating proteomics together

NPX File content

- NPX File
  - Now contains counts files by default
- Extended NPX also contains
  - IntraCV
  - InterCV
  - SampleBlockQCWarn
  - SampleBlockQCFail
  - BlockQCFail
  - AssayQCWarn

Column	Description
SampleID	The annotated sample ID
Sample Type	Type of sample
WellID	Id for well
PlateID	Name of the plate the sample was run on
DataAnalysisRefID	Reference ID for data analysis
OlinkID	OlinkID for assay
UniProt	UniProt ID for assay
Assay	Gene name for assay
AssayType	Type of assay
Panel	Panel name
Block	Name of the block the sample was run on
Count	The total number of counts
ExtNPX	Intermediate value between count and NPX: log2 of the ratio between datapoint Count value and the count for the Extension Control assay for the same sample.
NPX	NPX value
Normalization	Type of normalization used in project
PCNormalizedNPX	NPX value displayed if plate control normalization has been chosen.
AssayQC	Overall QC status for an assay
SampleQC	Overall QC status for a sample in a block
MapVersion	Software version of the module in Olink NPX Map used for panel calculations and normalization

Olink®
Accelerating proteomics together

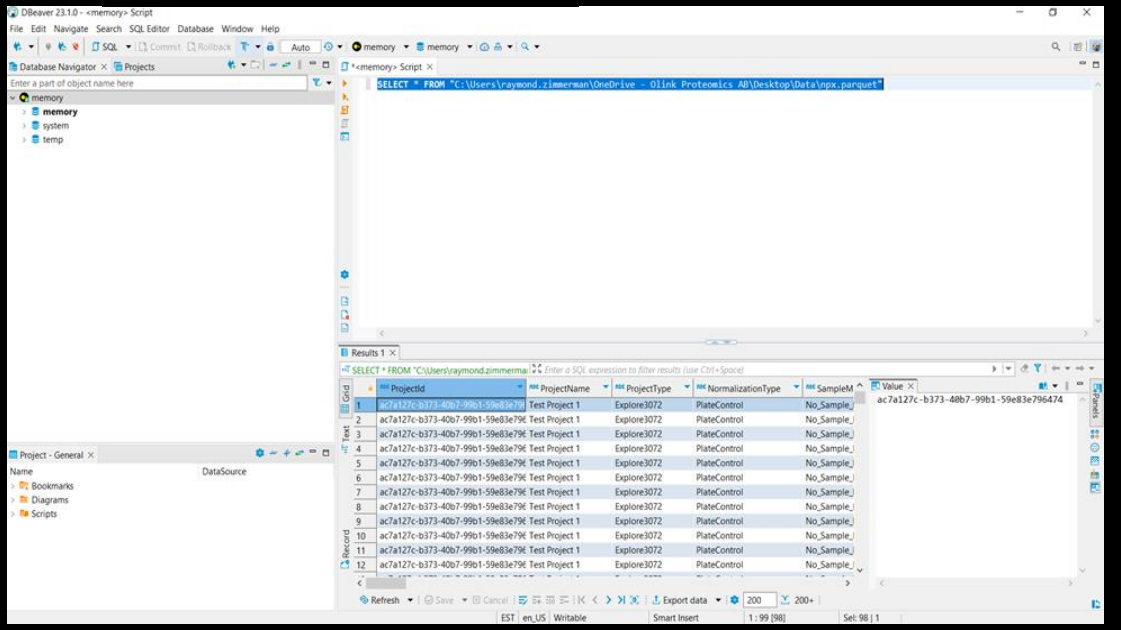
A new NPX file format

- Parquet
  - Data compression
    - 7-8Gb csv -> 250Mb parquet file
  - Easier to share with customers
  - Can be converted to CSV
  - Standard big data format sponsored by IBM, Apple, Microsoft, AWS and more
  - Increased data integrity
    - Data stored in binary form

Parquet

- Open with dBeaver or parquet-viewer software

DBeaver



Olink®
Accelerating proteomics together

Olink® NPX MAP CLI software

Software solution for Advanced kit users

A powerful Command Line Interface software

Facilitates efficient analysis of
large-scale data
through integration with LIMS

Generates both NPX and Counts to drive
innovation and Increase **transparency**

Olink®
Accelerating proteomics together

CLI Output file

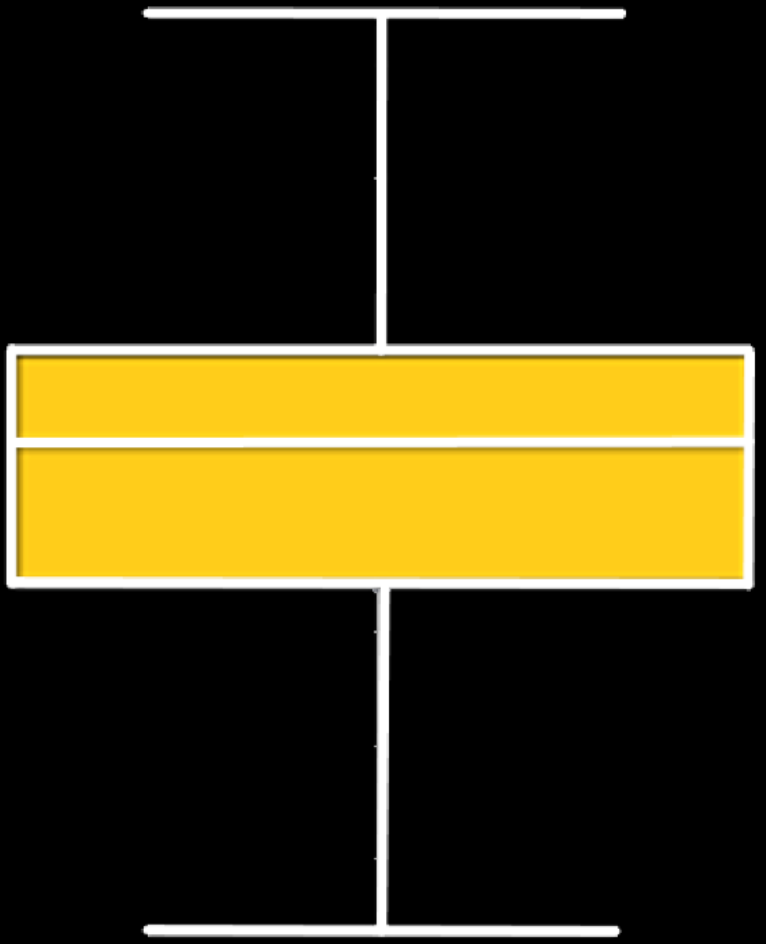
- NPX file
- Extended NPX file
- Analysis Report
- CLI Data Explore File
  - Contains all the columns in Extended NPX file , plus a set of additional columns.
  - Additional columns contain the information of run meta data, if the run unit and assay is included or excluded.

➢ Output data file in Parquet format.

Column
RunID
RunUnitID
RunUnitID
ExperimentName
FlowcellID
FlowcellType
FlowcellSide
InstrumentID
InstrumentType
InstrumentRunNumber
SequencingStartTimestamp
SequencingRecipeName
LibraryNumber
IndexPlate
SampleIndexVersion
MatchedCounts
Reads
Included
PreProcessingRunTimestamp
PreProcessingVersion
AssayCategory
ReadsPf
PercentReadsPf

Olink®
Accelerating proteomics together

QC Criteria



Olink®
Accelerating proteomics together

Sample QC

Criteria	FAIL *NPX is not calculated, exclude from statistical analysis*	WARN *NPX is calculated, assess further, use data with caution*
Sample QC		
Total counts per sample	< 10 000	N/A
Incubation control counts per sample	< 150	< 500
Extension control counts per sample	< 150	< 1000
Amplification control counts per sample	< 150	< 500
Internal control count fractions per sample	N/A	Log2 of incubation-to-amplification control count ration < -3.5 and,
	N/A	Log2 of incubation-to-extension control count ratio < -3.5 and,
	N/A	Absolute value of log2 of extension-to-amplification cotrol count ratio >3.5

Olink®
Accelerating proteomics together

External Control QC

Criteria	FAIL *NPX is not calculated, exclude from statistical analysis*	WARN *NPX is calculated, assess further, use data with caution*
External Control QC		
Total counts per sample	< 10 000	N/A
Incubation control counts per sample	< 500	N/A
Extension control counts per sample	< 1000	N/A
Amplification control counts per sample	< 500	N/A
Plate control internal control counts relative to assay counts	< or > internal control reference range to assay counts	N/A
Negative Control internal control counts relative to assay counts	Negative control fails	N/A
Internal control count fractions per sample	Log2 of incubation-to-amplification control count ratio < -3.5 and,	N/A
	Log2 of incubation-to-extension control count ratio < -3.5 and,	N/A
	Absolute value of log2 of extension-to-amplification control count ratio >3.5	N/A

Olink®
Accelerating proteomics together

Block and Assay QC

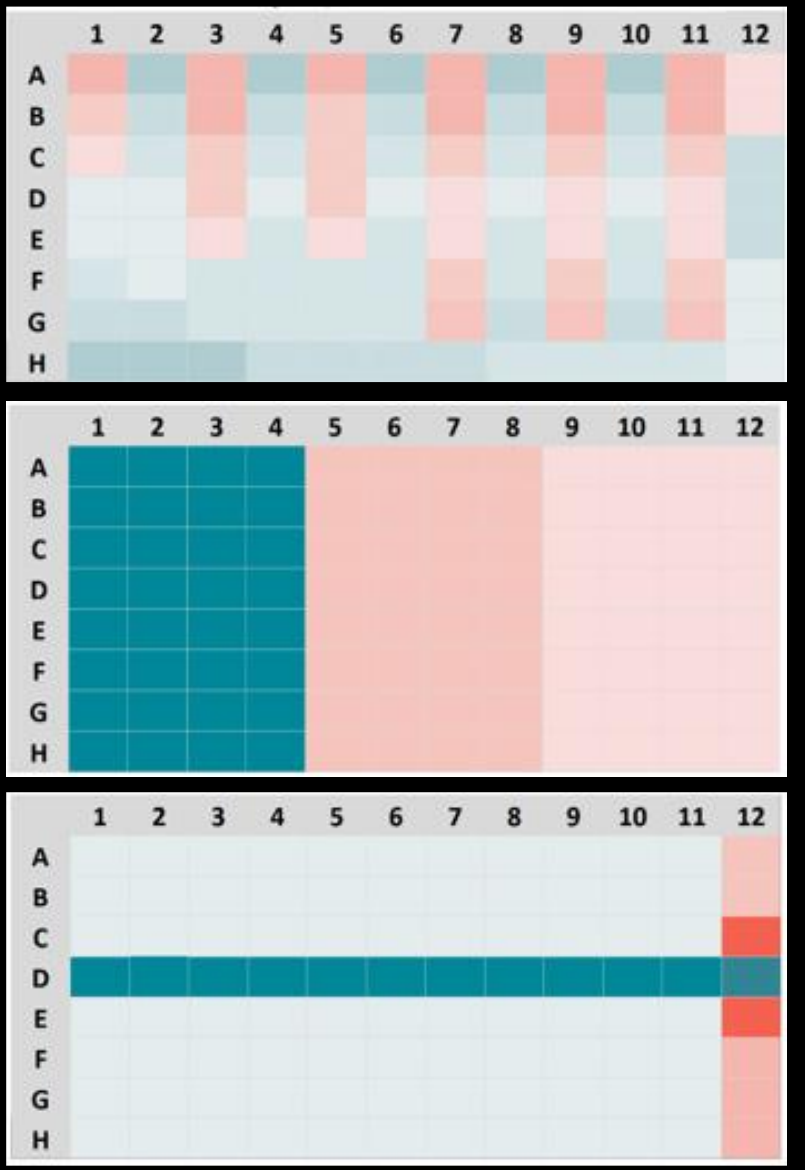
Criteria	FAIL *NPX is not calculated, exclude from statistical analysis*	WARN *NPX is calculated, assess further, use data with caution*
Block QC		
Plate Controls passing external sample QC	< 50% (minimum 3 Plate Controls must pass QC)	N/A
Negative Controls passing external sample QC	< 1	N/A
Systematic effect (NPX)	N/A	>10% of assays Systematic effect identified
Assay QC		
Assay count relative to internal control count in negative control	N/A	Assay count ≥ median of all internal control count in all negative controls

Olink®
Accelerating proteomics together

Systematic effect detection

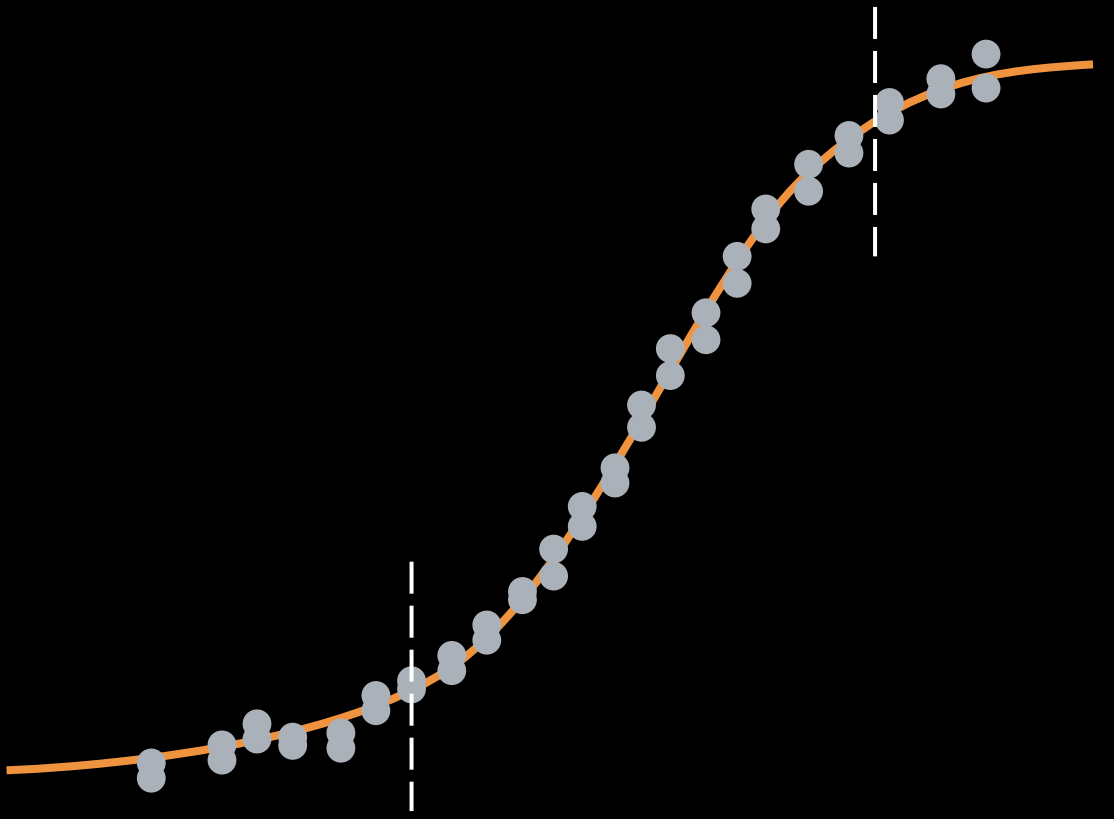
- Three main criteria
  - The frequency of assays representing the effect
  - The intensity of the effect on NPX deviation
  - Number of samples showing the effect

- A pattern is detected per assay/block/plate. If enough assays are affected, the block will get a systematic effects warning
- May be due to either non-randomized plate design or technical errors and should be investigated further



Olink®
Accelerating proteomics together

Limit of Detection



Olink®
Accelerating proteomics together

Limit of Detection (LOD)

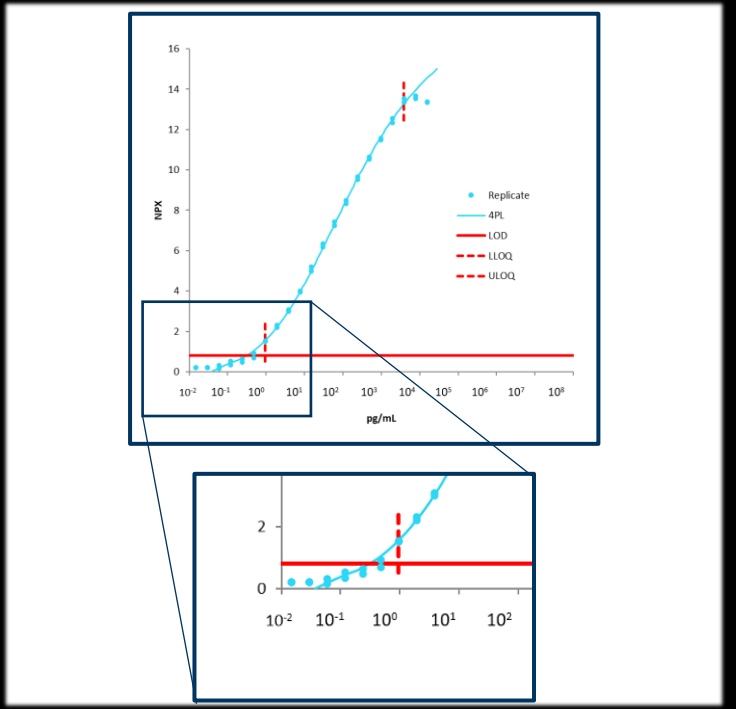
LOD is not reported by the NPX 3072 & HT Software

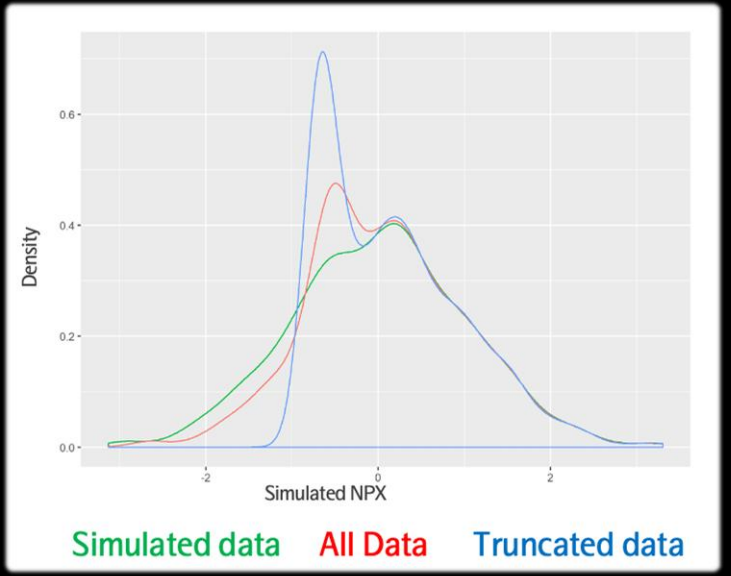
Data below LOD can be informative

- Expect low detectability for rare assays
- Can have perfect correlation for assays with all data below LOD
- Including all data gives better distribution from a statistical point

LOD not relevant for biological interpretation

- No risk of false positives including data below LOD
- Very small impact on false negatives (power)
- Some good biological findings have a lot of data <LOD
  Eg. non-expressed protein vs expressed protein



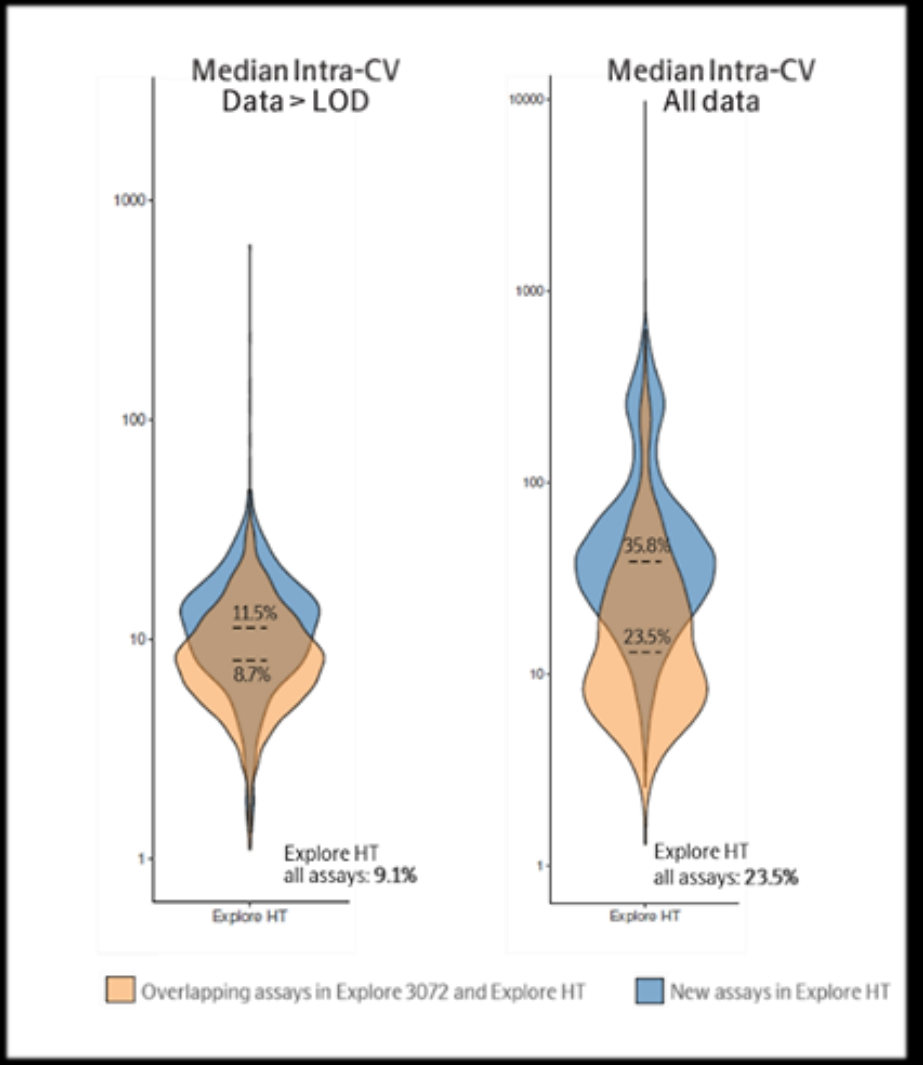


Olink®
Accelerating proteomics together

LOD is useful for technical evaluation

- LOD can be useful when performing technical evaluation of a dataset:
  - Calculate detectability
  - Calculate CVs using data > LOD

- LOD can be calculated via Olink Analyze:
  - Function olink lod() to calculate LOD adjusted to the normalization method used.



Based on 10 replicates of a pool of healthy plasma samples from 1 Concordance Test runs performed at AS Boston. Fixed LOD was used

Olink®
Accelerating proteomics together

Two types of LOD

The LOD function can calculate LOD from a dataset's NCs or a list of predetermined fixed LOD values:

- **Negative Control LOD**
  - Calculated on **NCs from customer runs**
  - Requires at least 10 NCs
- **Fixed LOD**
  - Calculated on **NCs from reference runs** (including 24-36 NCs)
  - Dependent on the Data Analysis Reference ID used
  - For both small sized studies (<10 NCs) and big studies
  - Fixed LOD .csv files for Explore 3072 and HT are available in the Download Center on the website
- **When to use Fixed LOD vs NC LOD**
  - For **smaller sized studies (<10 NCs)**, it is recommend using **fixed LOD** to integrate LOD values into a NPX dataset, as LOD calculations on fewer NCs may provide non-accurate values.
  - For **larger projects** we recommend calculating **LOD from NC** to obtain LOD values that are specific to the customer's project.

Olink®
Accelerating proteomics together

LOD calculations

1. **LOD for each assay is determined either at the counts level or at the NPX level**, depending on the assay's maximum counts in the NCs (of a reference dataset or customer run).
   - If at least one NC >150 counts, LOD is set at NPX level using the method of standard deviation of blank
   - If all NC ≤150 counts, LOD is set at count level using a read count threshold

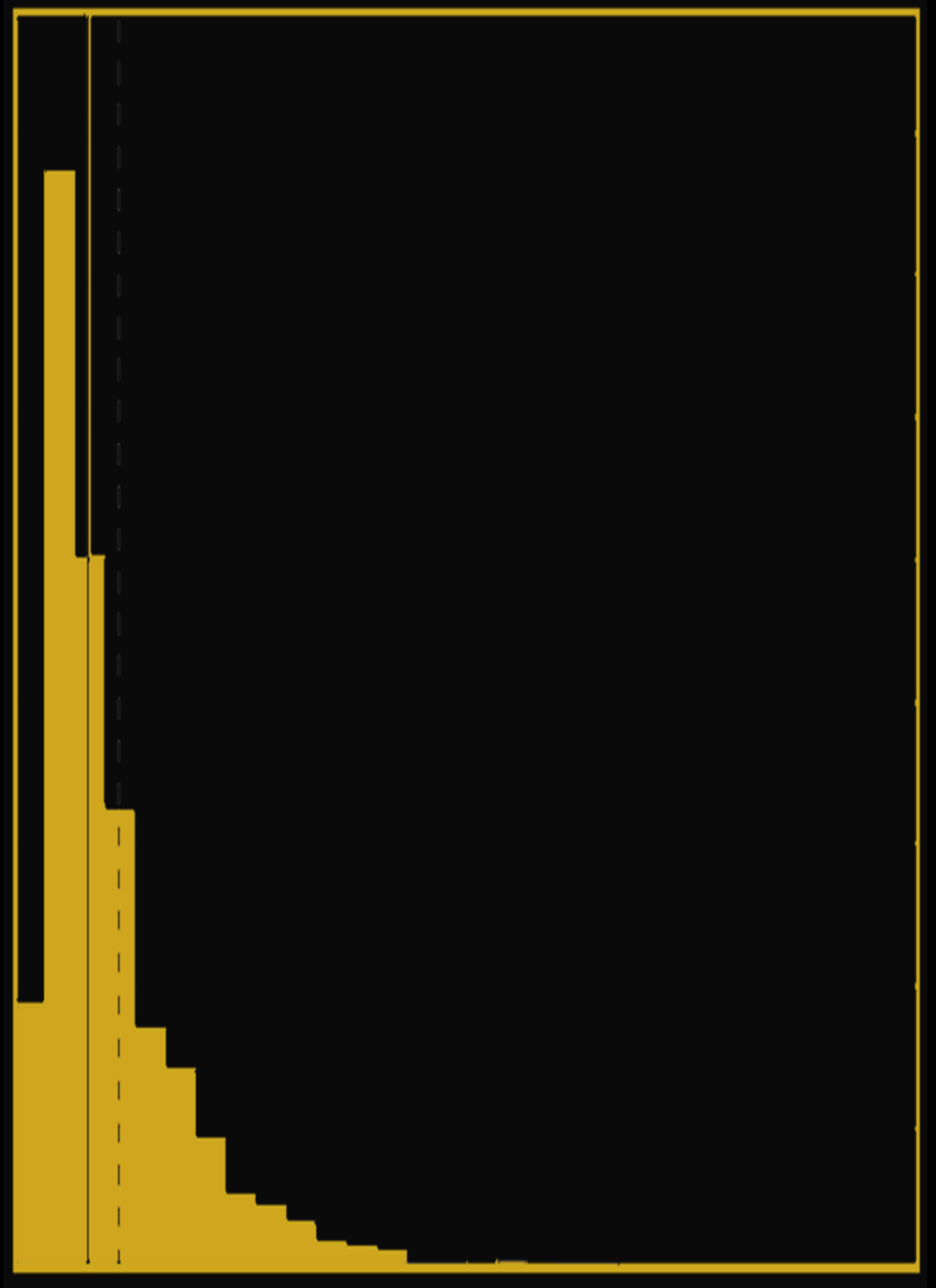
   Detailed formula:
   - If at least one NC > 150 counts: PC normalized LOD = median (PC normalized NPX) + (3SD|0.2NPX; whichever is the highest)
   - If all NCs ≤ 150 counts: PC normalized LOD = counts --> PC normalized NPX

2. **LOD needs to be adjusted If intensity normalized data is used:**
   - Intensity Normalized LOD = PC normalized LOD – adjustment factor
   - adjustment factor = median(NPX_excluding_external_control)

3. **LOD is reported in NPX in the parquet file**

Olink®
Accelerating proteomics together

CVs



Olink
Accelerating proteomics together

How are CVs determined?

CVs in Explore HT validation data

- Intra- and inter-CVs **for all assays*** were calculated on the Sample Controls in each of the 3 replicate runs performed
- This to evaluate the product performance (i.e., how precise each assay is detecting its target protein)

*Only data >LOD was used in CV calculations

CVs in customer runs

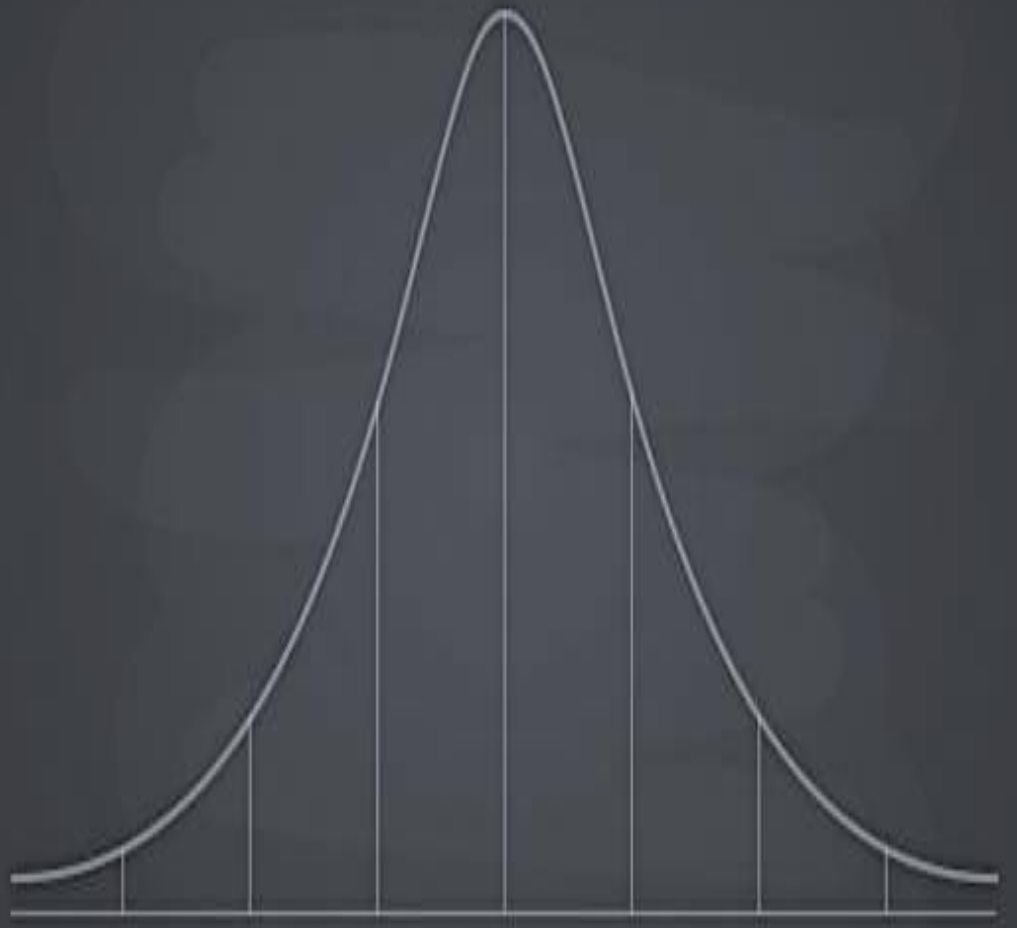
- Intra- and inter-CVs for **~300 selected assays**** calculated on the Sample Controls for each sample plate
- This to evaluate run performance and used for troubleshooting by looking at the reproducibility (=precision) between runs and plates

** Proteins typically well-expressed in healthy plasma, enabling calculation of reliable CVs

CVs are calculated differently in the validation data and in customer runs because they are **used for different purposes**

Olink®
Accelerating proteomics together

Data Normalization



Olink®
Accelerating proteomics together

Choosing normalization method – Simplified overview

INTENSITY NORMALIZATION

Used when **samples are randomized within a study** or **across multiple studies**

⊕ Preferred normalization when samples have a very different expression compared to Plate Controls (healthy plasma). E.g. certain diseases, alternative matrices.

⚠ **Requires sample randomization** across plates/runs/projects, etc

BRIDGING NORMALIZATION

Used when **samples are NOT randomized across multiple studies**

⊕ Preferred method to combine different studies

PLATE CONTROL NORMALIZATION

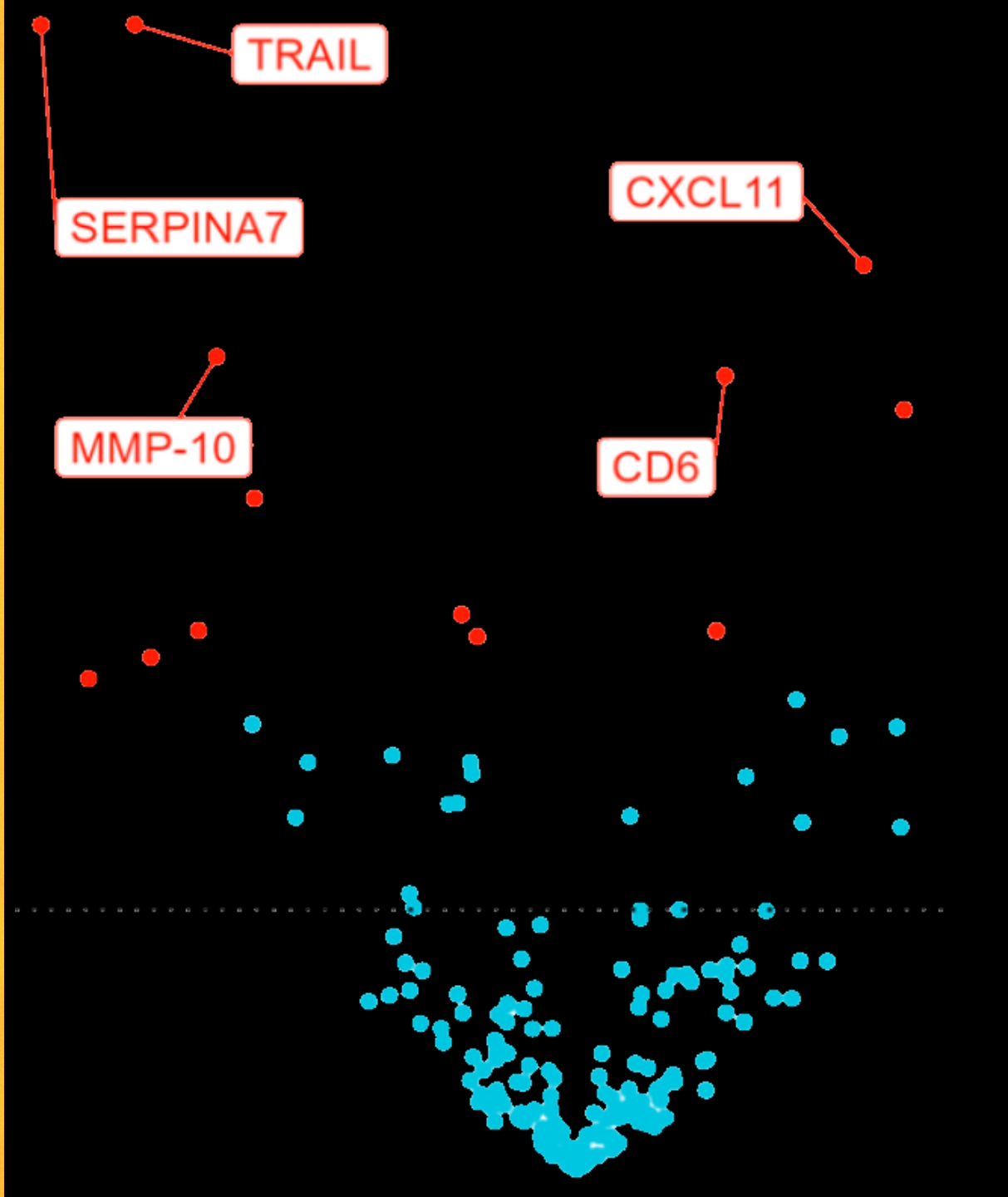
Used when **samples are NOT randomized within a study**

⚠ Does not work as well when the samples studied have a very different expression compared to Plate Control (healthy plasma). E.g. certain diseases, alternative matrices.

❗ **Randomization is preferred**

Olink®
Accelerating proteomics together

Statistical Analysis



Background on Olink data : high level review

- It is relative quantification
  - Look at differences, not absolute values
    - "Concentration of NEMO goes up across treatment, while IL-6 concentration goes down"
    - "Responders have higher MMP-7 expression than Non-Responders"
  - Only compares within assays, not between
- NPX is relative, but on a concrete quantitative scale
  - For a given assay, 1 NPX increase ~ 2 x increase in concentration
- Log2 scale
  - Much expression data appears more normally distributed on log scale
- Uses standard statistics

Olink®
Accelerating proteomics together

User-friendly and powerful tools available to support study design and data analysis

Olink® NPX MAP

A purpose-built software designed for data import, validate data quality, and normalization for subsequent statistical analysis

Olink® Insight

A platform for proteinbiomarker data discovery, collaborate with peers, and access a wealth of information.

Built-in Tools: Study Size Calculator, Stat Analyzer

Olink Analyze®

A versatile toolbox for handling of Olink data including QC, various statistical tests and visualization (R Package on Cran)

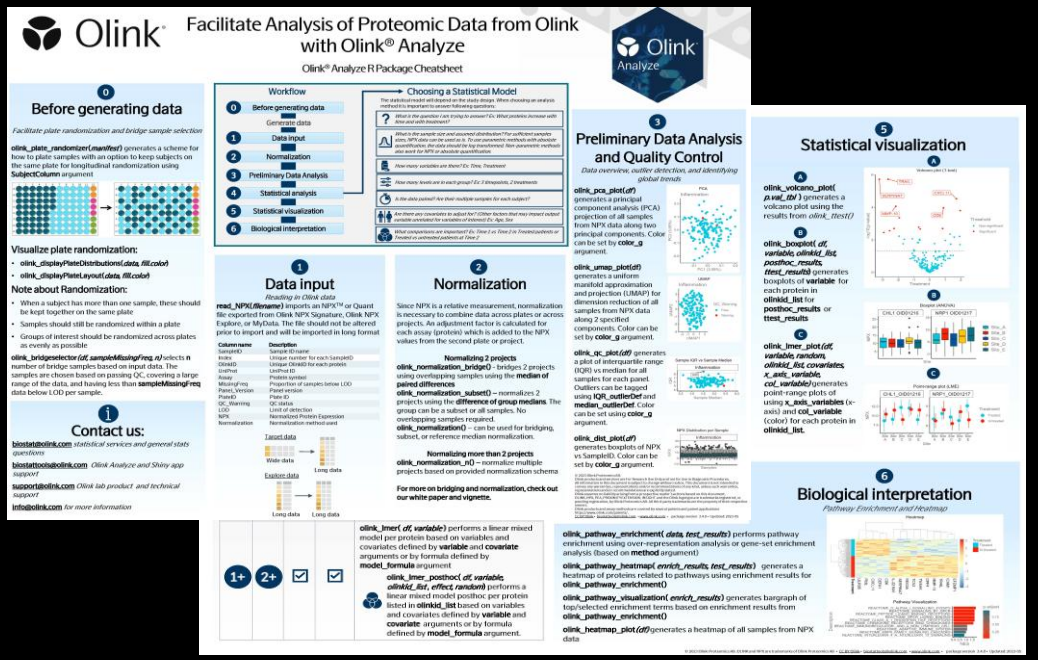
Olink® Statistical Services

Customized statistical analysis performed by experts experienced in handling this type of data

Olink
Accelerating proteomics together

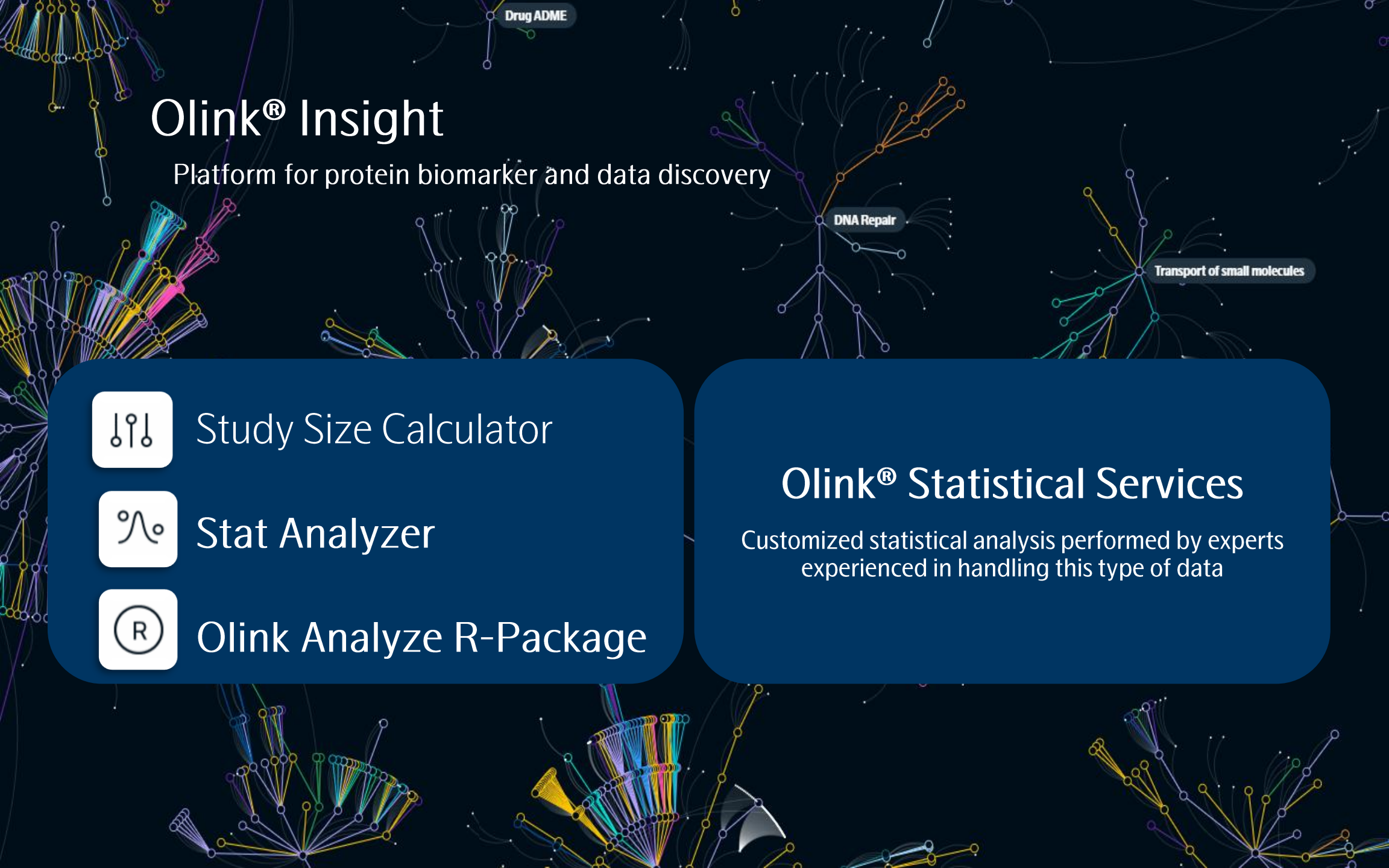
Olink® R Package: Olink® Analyze

Providing support at every step of the customer journey



Educational Resources:

- Cheat sheet
- Overview Vignette
- Tutorials on bridging, outlier detection and plate randomization
- Available on CRAN



Olink® Insight

Platform for protein biomarker and data discovery

- Study Size Calculator
- Stat Analyzer
- Olink Analyze R-Package

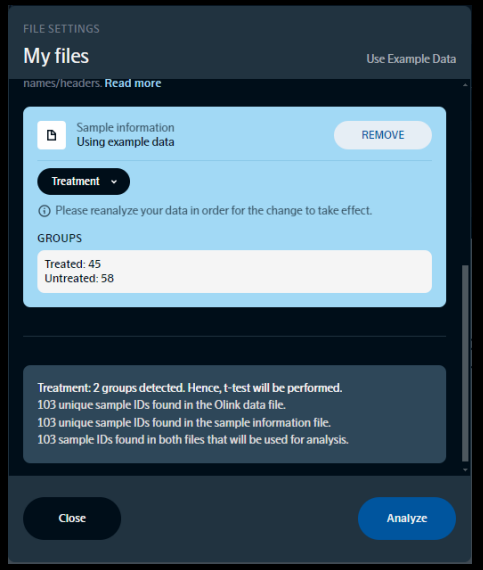
Olink® Statistical Services

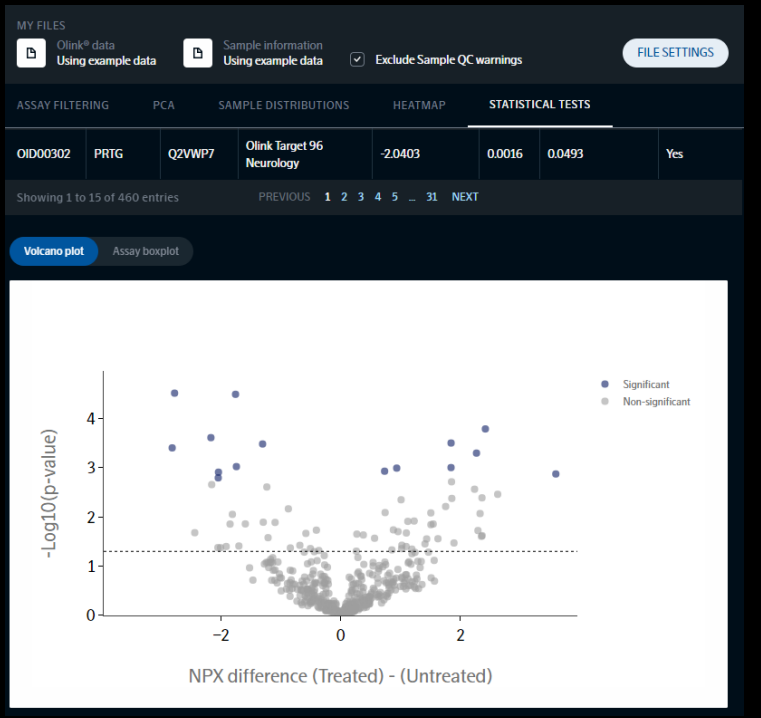
Customized statistical analysis performed by experts experienced in handling this type of data

Olink®
Accelerating proteomics together

Olink® Statistical Analysis App: Web Tool for Statistical Analysis

- In Olink Insight
- Training video and User Manual available
- Basic statistical analyses and visualizations of Olink data
  - PCA *– even without the sample information file*
  - T-Test
  - Anova

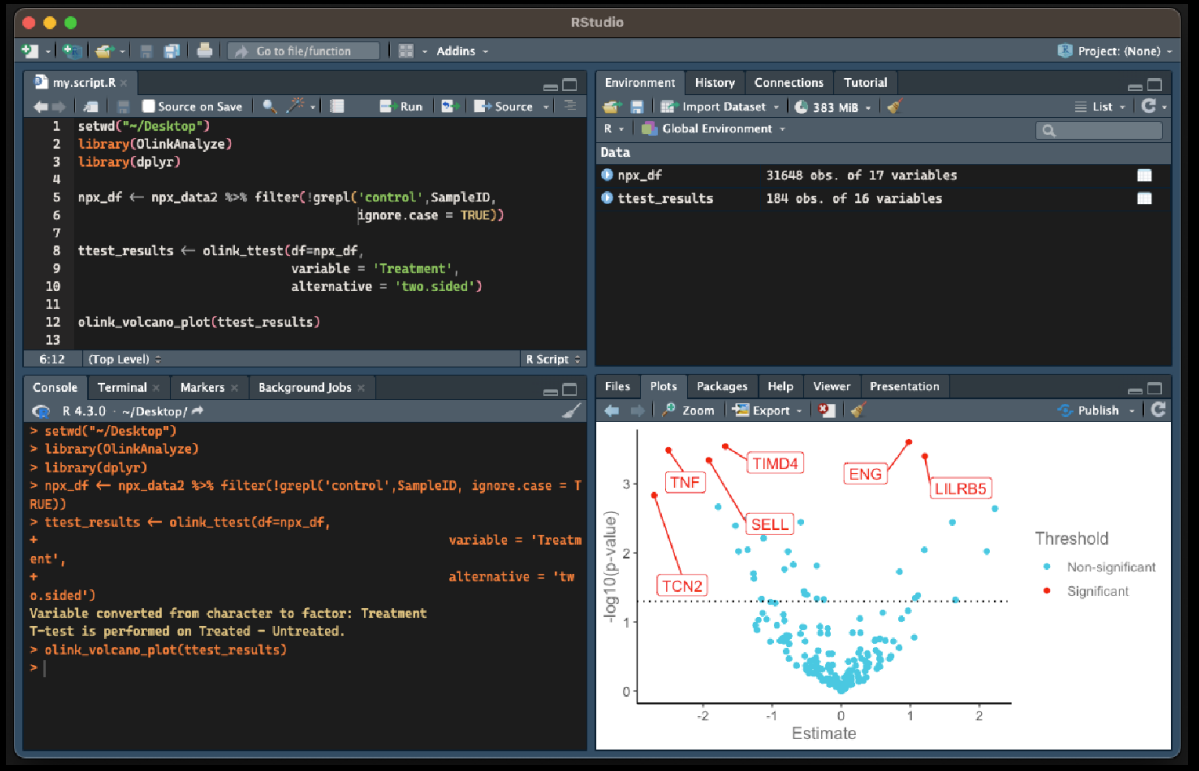




Olink®
Accelerating proteomics together

Olink® R Package : Olink Analyze®

- Data pre-processing
  - *Read in NPX data*
  - *Normalization*
- QC and exploratory data analysis
  - Outlier detection and data distribution:
    - QC plot – IQR vs Median
    - Distribution plots
    - PCA plots
- Stats Analysis
  - t-test
  - ANOVA
  - *Linear Mixed-Effect Models*



Olink®
Accelerating proteomics together

Olink® Explore HT QC Training

Questions ?