

# Sanofi - Olink Scout

An integrative platform for Olink proteomics data analysis and visualization

# Index

Application model

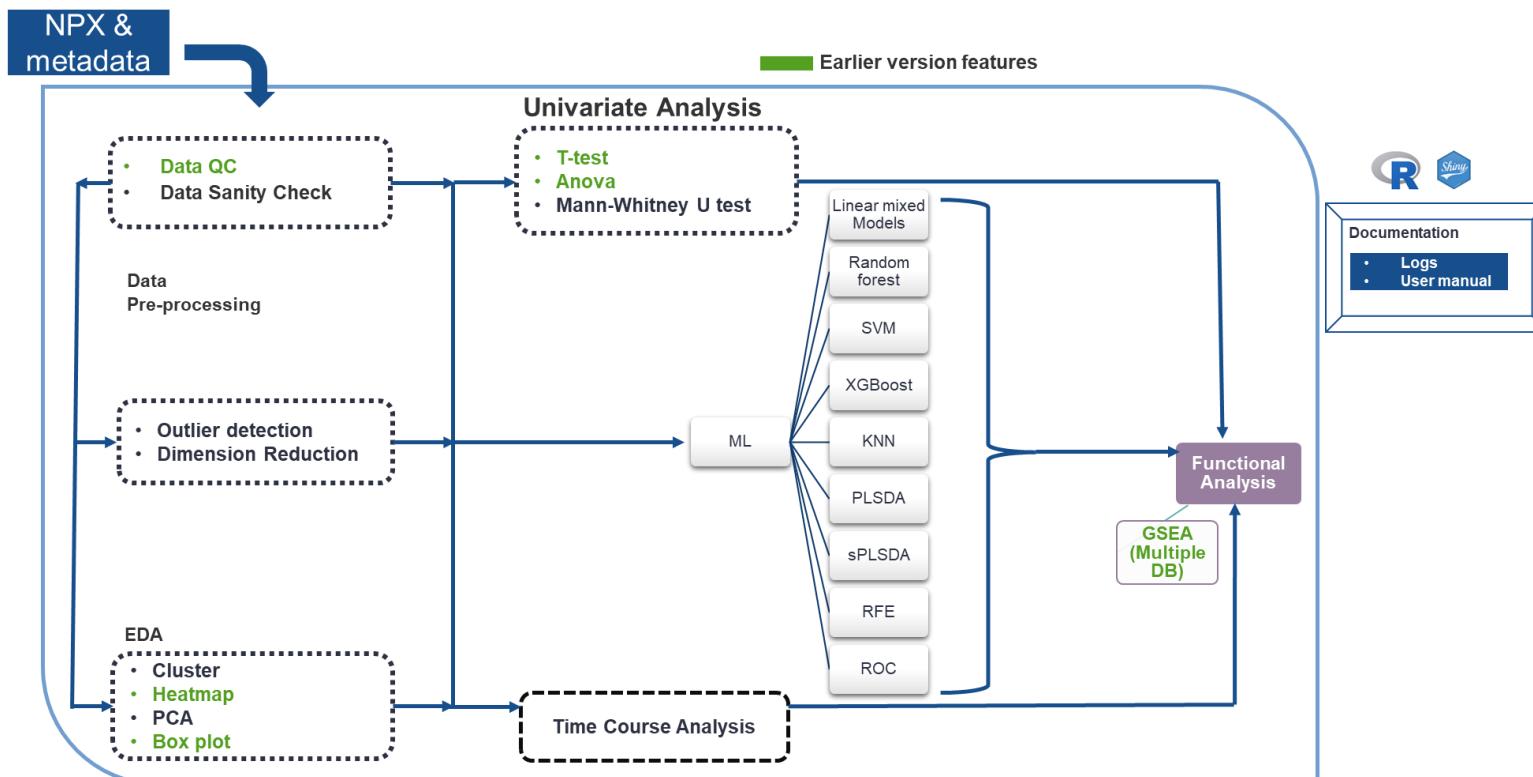
Feature comparison with Olink app

Application Manual (screenshots)

Internal Olink resources

# Sanofi- Olink Application (Olink Scout)

## Model



Features - Data Sanity Check, Data Pre-processing, EDA - Clustering & PCA, Mann-Whitney U test, Time Course Analysis and ML methods.

# Sanofi- Olink Application (Olink Scout)

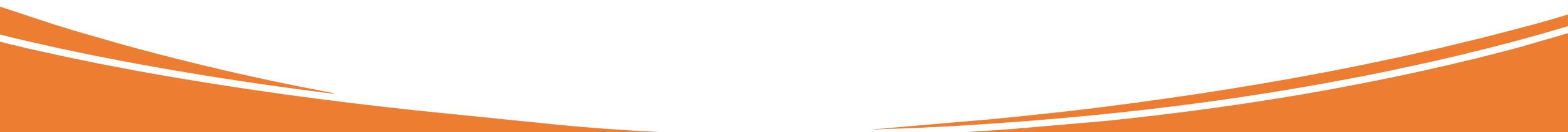
## Feature comparison with Olink app

	Olink app	Olink Scout
Data QC	✓	✓
Data Processing	✗	✓
EDA : Cluster, Box plot and Correlation	✗	✓
Univariate analysis : Mann-Whitney U Test	✗	✓
Time Course Analysis	✗	✓
Multivariate analysis : Multiple ML methods	✗	✓
Functional Analysis	✗	✓
EDA : PCA and Heatmap	✓	✓
Univariate analysis : T-Test & Anova	✓	✓
Analysis Report Generation	✗	✓

 Extra features in Scout

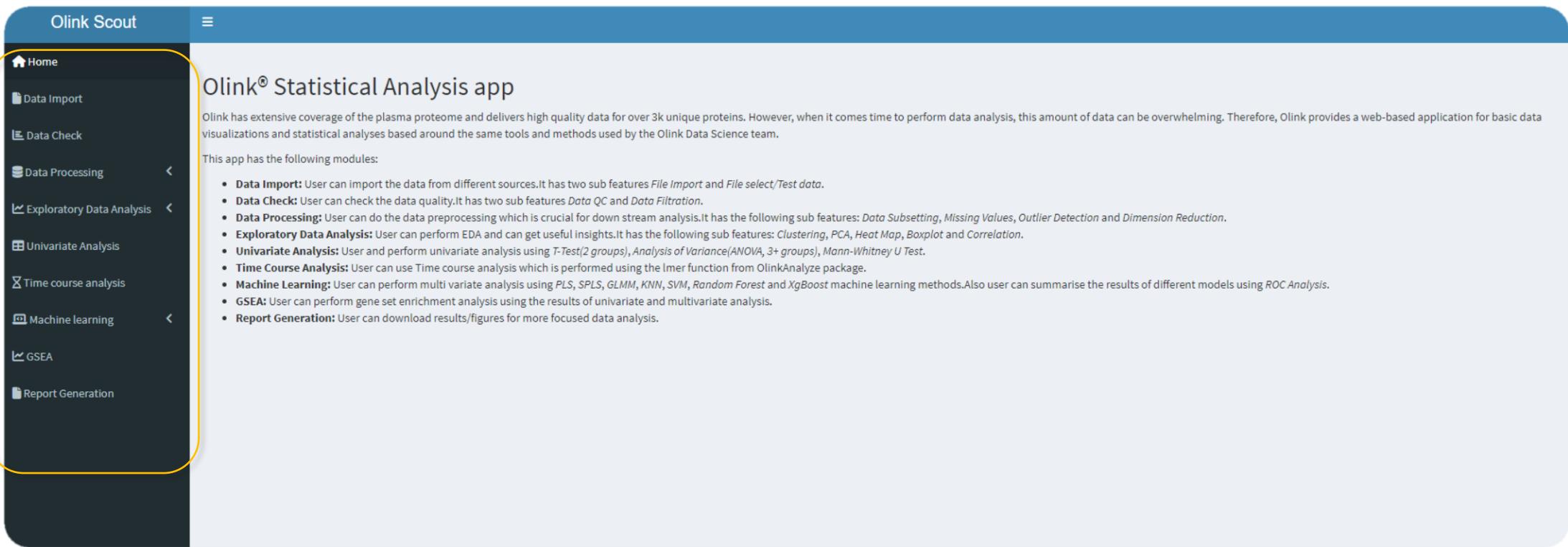
Seven new features added

# Application Manual



# Sanofi- Olink Application (Olink Scout)

 [Home Page](#)



The screenshot shows the Olink Scout web application interface. At the top, there's a blue header bar with the title "Olink Scout". Below it is a navigation bar with icons for Home, Data Import, Data Check, Data Processing, Exploratory Data Analysis, Univariate Analysis, Time course analysis, Machine learning, GSEA, and Report Generation. The "Data Import" item is highlighted with a yellow box. The main content area has a light gray background. It features a section titled "Olink® Statistical Analysis app" with a brief description of the tool's purpose. Below this, there's a list of modules with their descriptions:

- Data Import:** User can import the data from different sources. It has two sub features *File Import* and *File select/Test data*.
- Data Check:** User can check the data quality. It has two sub features *Data QC* and *Data Filtration*.
- Data Processing:** User can do the data preprocessing which is crucial for down stream analysis. It has the following sub features: *Data Subsetting*, *Missing Values*, *Outlier Detection* and *Dimension Reduction*.
- Exploratory Data Analysis:** User can perform EDA and can get useful insights. It has the following sub features: *Clustering*, *PCA*, *Heat Map*, *Boxplot* and *Correlation*.
- Univariate Analysis:** User can perform univariate analysis using *T-Test(2 groups)*, *Analysis of Variance(ANOVA, 3+ groups)*, *Mann-Whitney U Test*.
- Time Course Analysis:** User can use Time course analysis which is performed using the *lmer* function from *OlinkAnalyze* package.
- Machine Learning:** User can perform multi variate analysis using *PLS*, *SPLS*, *GLMM*, *KNN*, *SVM*, *Random Forest* and *XgBoost* machine learning methods. Also user can summarise the results of different models using *ROC Analysis*.
- GSEA:** User can perform gene set enrichment analysis using the results of univariate and multivariate analysis.
- Report Generation:** User can download results/figures for more focused data analysis.

*New feature panel*

# Sanofi- Olink Application (Olink Scout)

 Data Import Page

Olink Scout

☰

Home

Data Import

Data Check

Data Processing

Exploratory Data Analysis

Univariate Analysis

Time course analysis

Machine learning

GSEA

Report Generation

Data Import

Data Summary

NPX Data

Metadata

Data Import is to import the data from different sources:

- File Import - Import the data(.csv, .xlsx, .parquet) from local pc.
- File select/Test data - Select the data from snowflake database.
- Column names - Please make sure that SampleID column name must be same in the Olink & Meta data file. **Mandatory option**

Select a import method

File import  File select/Test data

**File Import**

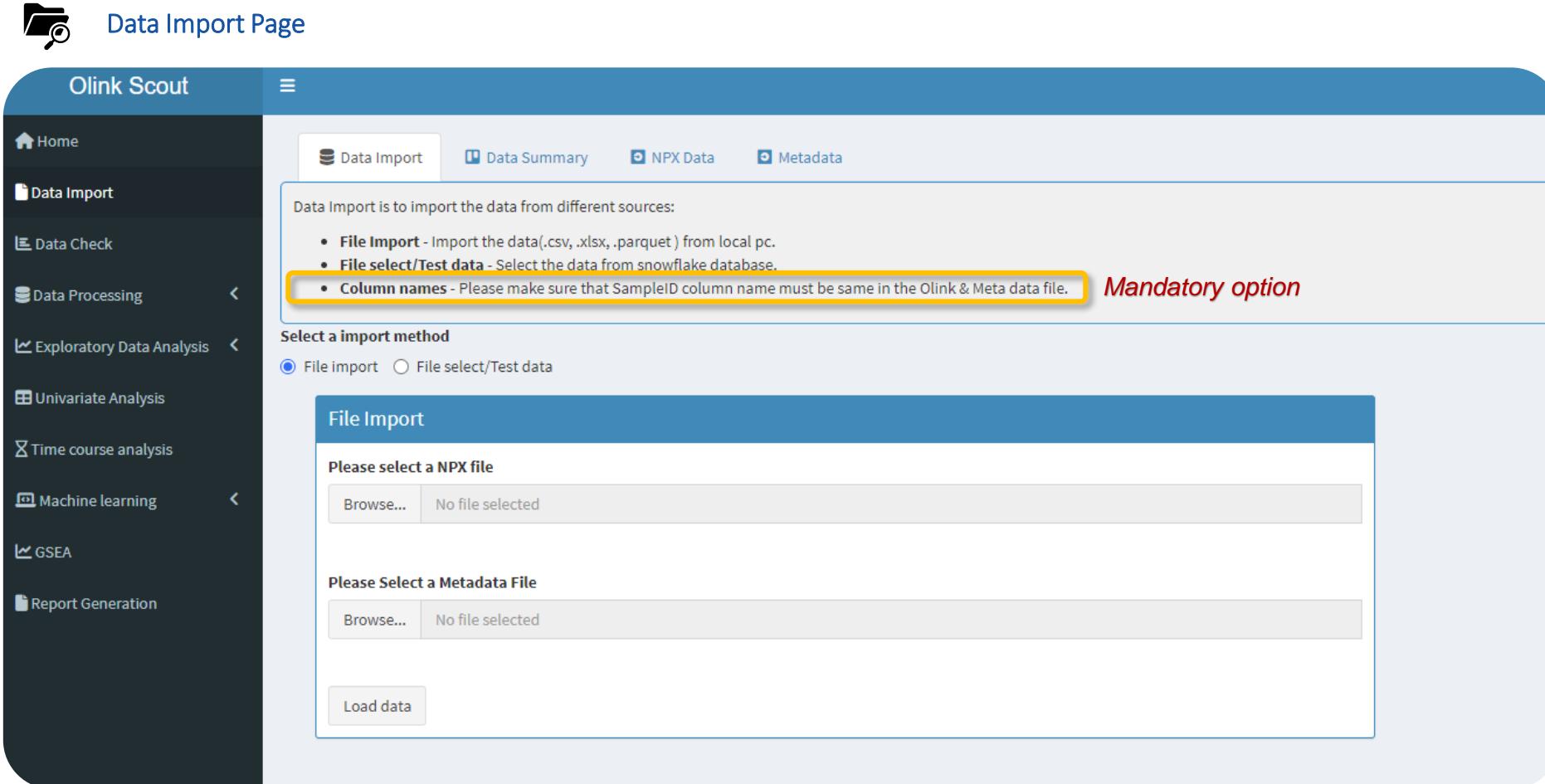
Please select a NPX file

Browse... No file selected

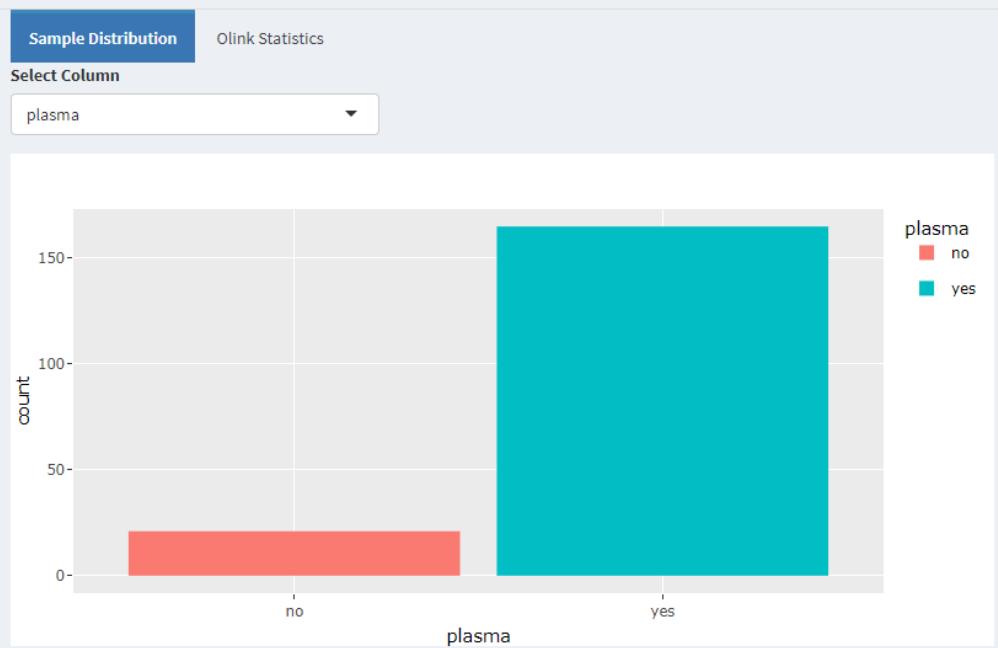
Please Select a Metadata File

Browse... No file selected

Load data



# Sanofi- Olink Application (Olink Scout)

A bar chart titled "Sample Distribution" showing the count of samples for each plasma status. The x-axis is labeled "plasma" with categories "no" and "yes". The y-axis is labeled "count" with ticks at 0, 50, 100, and 150. The "no" category is represented by a red bar reaching approximately 25 on the y-axis. The "yes" category is represented by a teal bar reaching approximately 160 on the y-axis.

plasma	count
no	~25
yes	~160

**Data Import - Data Summary Page**

Olink Scout

Home Data Import Data Check Data Processing Exploratory Data Analysis Univariate Analysis Time course analysis Machine learning GSEA Report Generation

Data Import Data Summary NPX Data Metadata

Data Processing Information

**Data Summary page**

**NPX Data:**

1. Total number of rows: 273792
2. Total number of columns: 8
3. Uploaded file format: .csv
4. Total proteins: 1472
5. Total samples: 186

**Meta Data:**

1. Total number of rows: 186
2. Total number of features: 9
3. Uploaded file format: .csv
4. Features Names: SampleID, CSF, plasma, sex, age, diagnosis\_at\_baseline, developed\_RRMS\_or\_not, disease\_duration\_time\_from\_first\_symptom\_to\_sampling, CSF\_findings\_Albumine\_ratio

Sample Distribution Olink Statistics

Select Column: plasma

plasma

- no
- yes

# Sanofi- Olink Application (Olink Scout)

 Data Import - NPX Data Page

Olink Scout

Data Import - NPX Data Page

NPX Data page

	SampleID	OlinkID	UniProt	Assay	MissingFreq	Panel	LOD	NPX
1	Dis_MS_46	OID20466	Q13574	DGKZ	0.9848	Inflammation	0.7489	-0.2893
2	Dis_MS_23	OID20466	Q13574	DGKZ	0.9848	Inflammation	0.7489	0.2046
3	Dis_MS_43	OID20466	Q13574	DGKZ	0.9848	Inflammation	0.7489	-0.1054
4	Rep_MS_41	OID20466	Q13574	DGKZ	0.9848	Inflammation	0.7489	-0.5311
5	Dis_HC_2	OID20466	Q13574	DGKZ	0.9848	Inflammation	0.7489	0.1005
6	Rep_HC_20	OID20466	Q13574	DGKZ	0.9848	Inflammation	0.7489	-0.3592
7	Dis_HC_6	OID20466	Q13574	DGKZ	0.9848	Inflammation	0.7489	-0.2171
8	Dis_MS_85	OID20466	Q13574	DGKZ	0.9848	Inflammation	0.7489	-0.0083
9	Dis_HC_19	OID20466	Q13574	DGKZ	0.9848	Inflammation	0.7489	0.3916
10	Dis_HC_16	OID20466	Q13574	DGKZ	0.9848	Inflammation	0.7489	-0.2355

Showing 1 to 10 of 242,880 entries

Previous 1 2 3 4 5 ... 24,288 Next

# Sanofi- Olink Application (Olink Scout)



## Data Import - Metadata Page

Olink Scout

☰

Home Data Import Data Check Data Processing Exploratory Data Analysis Univariate Analysis Time course analysis Machine learning GSEA Report Generation

Data Import Data Summary NPX Data Metadata

Metadata page

Search:

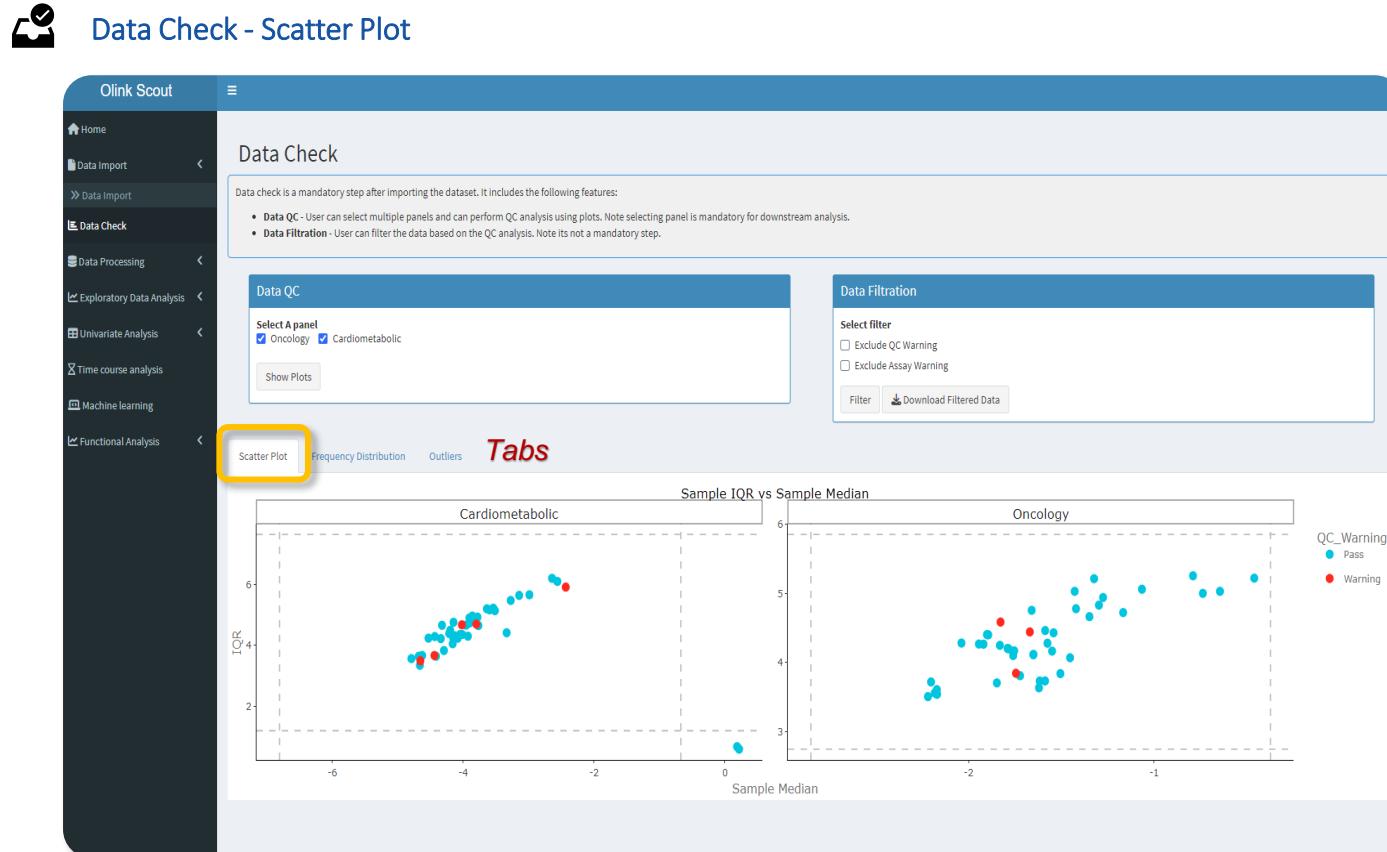
SampleID	CSF	plasma	sex	age	diagnosis_at_baseline	developed_RRMS_or_not	disease_duration_time_from_first_symptom_to_sampling	CSF_findings_Albumine_ratio
All	All	All	All	All	All	All	All	All
1 Dis_MS_1	yes	yes	m	30	RRMS	yes	11	2.9
2 Dis_MS_2	yes	yes	f	21	CIS	yes	1	4.4
3 Dis_MS_3	yes	yes	f	28	CIS	yes	2	3.3
4 Dis_MS_4	yes	yes	f	26	CIS	no	7	3.1
5 Dis_MS_5	yes	yes	f	28	CIS	yes	9	3.2

CSV Excel

Showing 1 to 5 of 186 entries

Previous 1 2 3 4 5 ... 38 Next

# Sanofi- Olink Application (Olink Scout)



Data Check page with tabs avoid scrolling down the page

# Sanofi- Olink Application (Olink Scout)

 Data Check - Frequency Distribution

**Olink Scout**

- Home
- Data Import
- Data Check**
- Data Processing
- Exploratory Data Analysis
  - Univariate Analysis
  - Time course analysis
- Machine learning
- GSEA
- Report Generation

**Data Check**

Data check is a mandatory step after importing the dataset. It includes the following features:

- **Data QC** - User can select multiple panels and can perform QC analysis using plots. Note selecting panel is mandatory for downstream analysis.
- **Data Filtration** - User can filter the data based on the QC analysis. Note its not a mandatory step.

**Data QC**

Select A panel  
 Oncology  Cardiometabolic

Show Plots

**Data Filtration**

Select filter  
 Exclude QC Warning  
 Exclude Assay Warning

Filter  Download Filtered Data

Scatter Plot **Frequency Distribution** Outliers **Frequency Distribution**

NPX Distribution per Sample

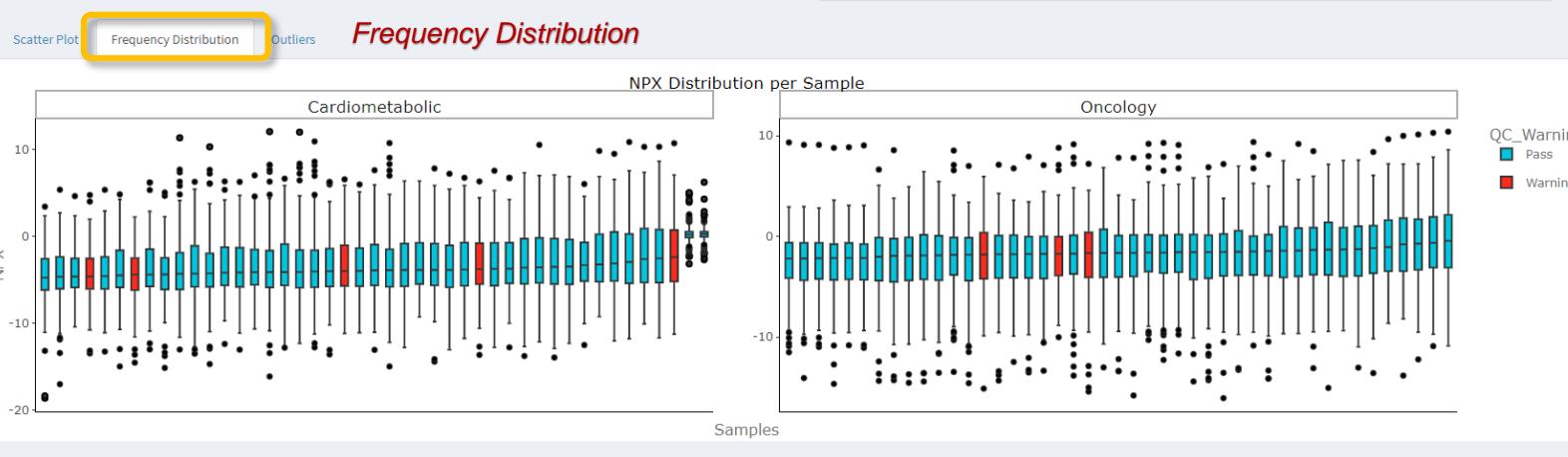
Cardiometabolic

NPX

Oncology

Samples

QC\_Warning  
Pass (Blue)  
Warning (Red)



# Sanofi- Olink Application (Olink Scout)

 Data Check - Outliers

**Olink Scout**

Home  
Data Import  
**Data Check**  
Data Processing  
Exploratory Data Analysis  
Univariate Analysis  
Time course analysis  
Machine learning  
GSEA  
Report Generation

**Data Check**

Data check is a mandatory step after importing the dataset. It includes the following features:

- **Data QC** - User can select multiple panels and can perform QC analysis using plots. Note selecting panel is mandatory for downstream analysis.
- **Data Filtration** - User can filter the data based on the QC analysis. Note its not a mandatory step.

**Data QC**

Select A panel  
 Oncology  Cardiometabolic  
Show Plots

**Data Filtration**

Select filter  
 Exclude QC Warning  
 Exclude Assay Warning  
Filter Download Filtered Data

**Scatter Plot** **Frequency Distribution** **Outliers** **Outliers**

Oncology

PC2 (13.78%)

PC1 (37.99%)

Cardiometabolic

FC2 (10.79%)

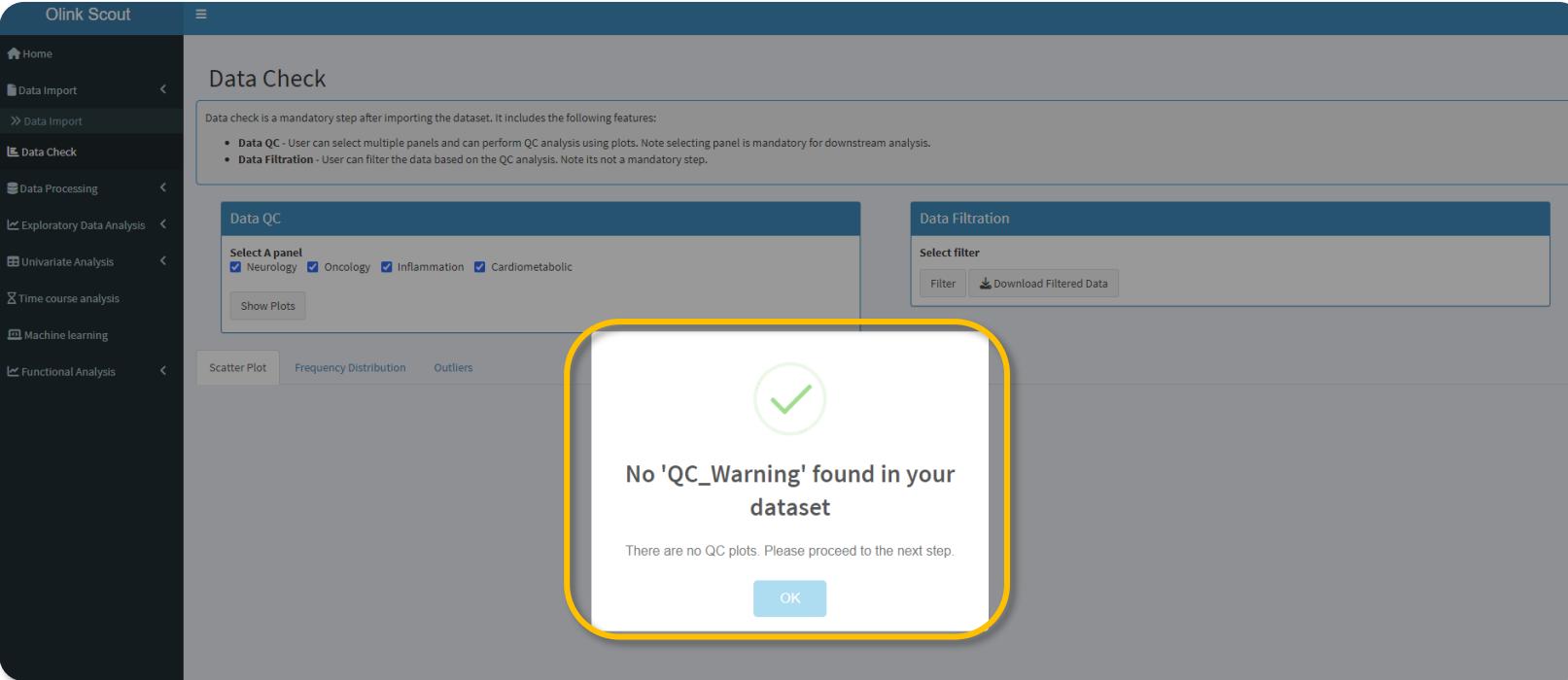
PC1 (33.51%)

QC\_Warning  
Pass (blue dots)  
Warning (red dots)



# Sanofi- Olink Application (Olink Scout)

 Data Check – Data QC

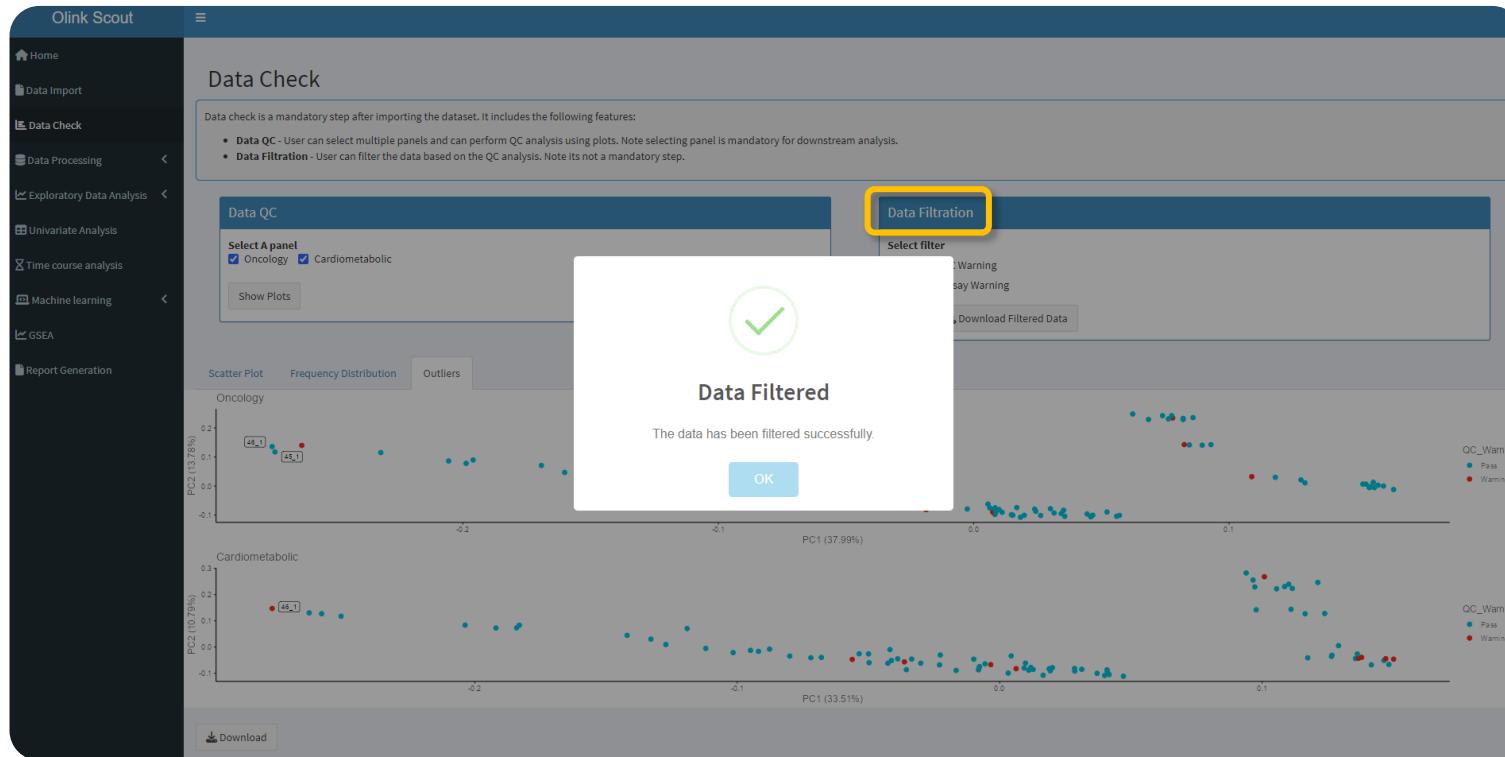


The screenshot shows the Olink Scout interface with the 'Data Check' section selected. The message pop-up is centered, stating: 'No 'QC\_Warning' found in your dataset'. It includes a green checkmark icon, a descriptive message, and an 'OK' button.

The Scout app has message pop-up feature.

# Sanofi- Olink Application (Olink Scout)

## ⌚ Data Check - Data Filtration



# Sanofi- Olink Application (Olink Scout)



## Data Processing – Data Subsetting

Data subsetting is used to select a subset of a data from the entire dataset. For data subsetting metadata has used. User can use the following features:

- User can filter the data in the view using search box given for different features(of the metadata). Note its not a mandatory step.
- User can adjust the view using show entries tab.

Subsetting Options

Select All Rows in View Clear All selected Rows

Show 10 entries Search:

SampleID	SubjectID	Dose., $\mu$ M.	Timepoint	Group
1	1_1	D50	0	8 Doner
2	5_1	D50	0	24 Doner
3	9_1	D50	0	48 Doner
4	13_1	D51	0	8 Doner
5	17_1	D51	0	24 Doner
6	21_1	D51	0	48 Doner
7	25_1	D52	0	8 Doner
8	29_1	D52	0	24 Doner
9	33_1	D52	0	48 Doner
10	37_1	D53	0	8 Doner

Showing 1 to 10 of 88 entries

Previous 1 2 3 4 5 ... 9 Next

Subset data

Select rows to subset

# Sanofi- Olink Application (Olink Scout)



## Data Processing – Check missingness

The screenshot shows the Olink Scout application interface with the following details:

- Left Sidebar:** A dark sidebar with a navigation menu:
  - Home
  - Data Import
  - Data Check
  - Data Processing** (selected)
    - >> Data Subsetting
    - >> Missing Values** (highlighted with a yellow box)
    - >> Outlier Detection
    - >> Dimension Reduction  - Exploratory Data Analysis
  - Univariate Analysis
  - Time course analysis
  - Machine learning
  - GSEA
  - Report Generation
- Main Content Area:**
  - Missing Values**: A section with instructions and three steps:
    - Step1.** Check missiness in your dataset.
    - Step2.** User can remove proteins by giving a threshold of missiness percentage. Use Sample missiness for details. Note its not a mandatory step.
    - Step3.** User can impute the remaining missings values by using **LOD**, **mean**, **median**, **min** methods. Note its not a mandatory step.
  - 1. Check missiness**: Contains a "Show Plots" button.
  - A pie chart titled "Sample Counts by Missingness Percentage Range" showing the distribution of sample counts across different missingness ranges:

Missingness Range	Percentage
(10,15]	43.8%
(5,10]	25%
(15,20]	18.8%
(20,25]	6.25%
(25,30]	9.5%
  - 2. Remove features(protein) with high missiness**: A form to exclude proteins based on missingness percentage.
    - Remove proteins** checkbox:  Exclude proteins having missiness greater than
    - Missiness value percentage**: A text input field.
    - Remove** button.
  - 3. Impute the remaining missings values**: A section to select an imputation method.
    - Select Imputation method**: A dropdown menu with "Replace by LOD" selected.
    - Impute** button.

Annotations with arrows:

- An arrow points from the "Select Imputation method" dropdown to the text "Select missing values imputation method".
- An arrow points from the "Missiness value percentage" input field to the text "Enter missiness percentage to remove proteins".

# Sanofi- Olink Application (Olink Scout)



Data Processing – IQR based Outlier Detection



New Data Processing options

Remove outliers

Plot Save/Zoom

Interactive plot

# Sanofi- Olink Application (Olink Scout)



## Data Processing – PCA based Outlier detection



# Sanofi- Olink Application (Olink Scout)



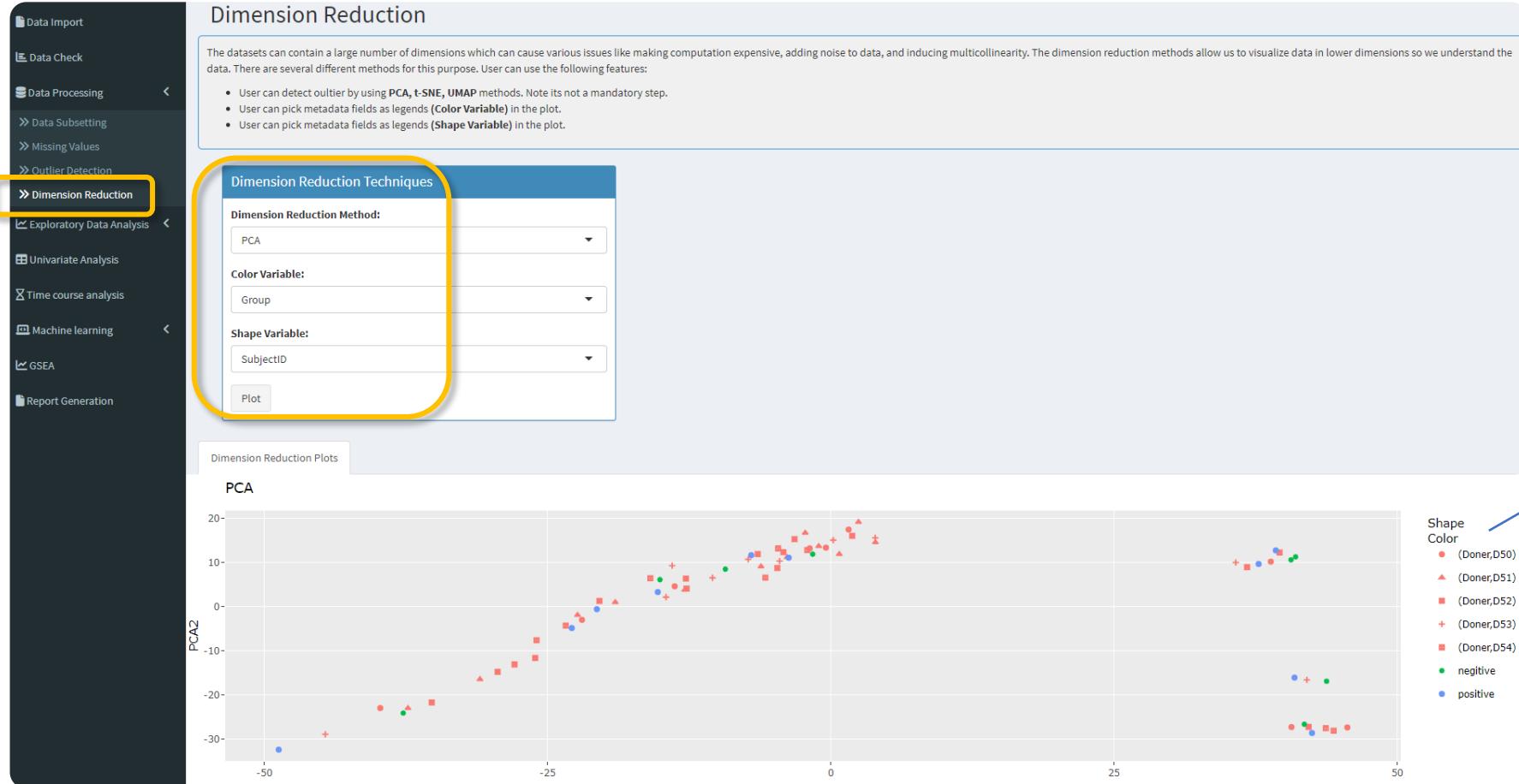
## Data Processing – Cook's distance based Outlier detection



# Sanofi- Olink Application (Olink Scout)



## Data Processing – Dimension Reduction

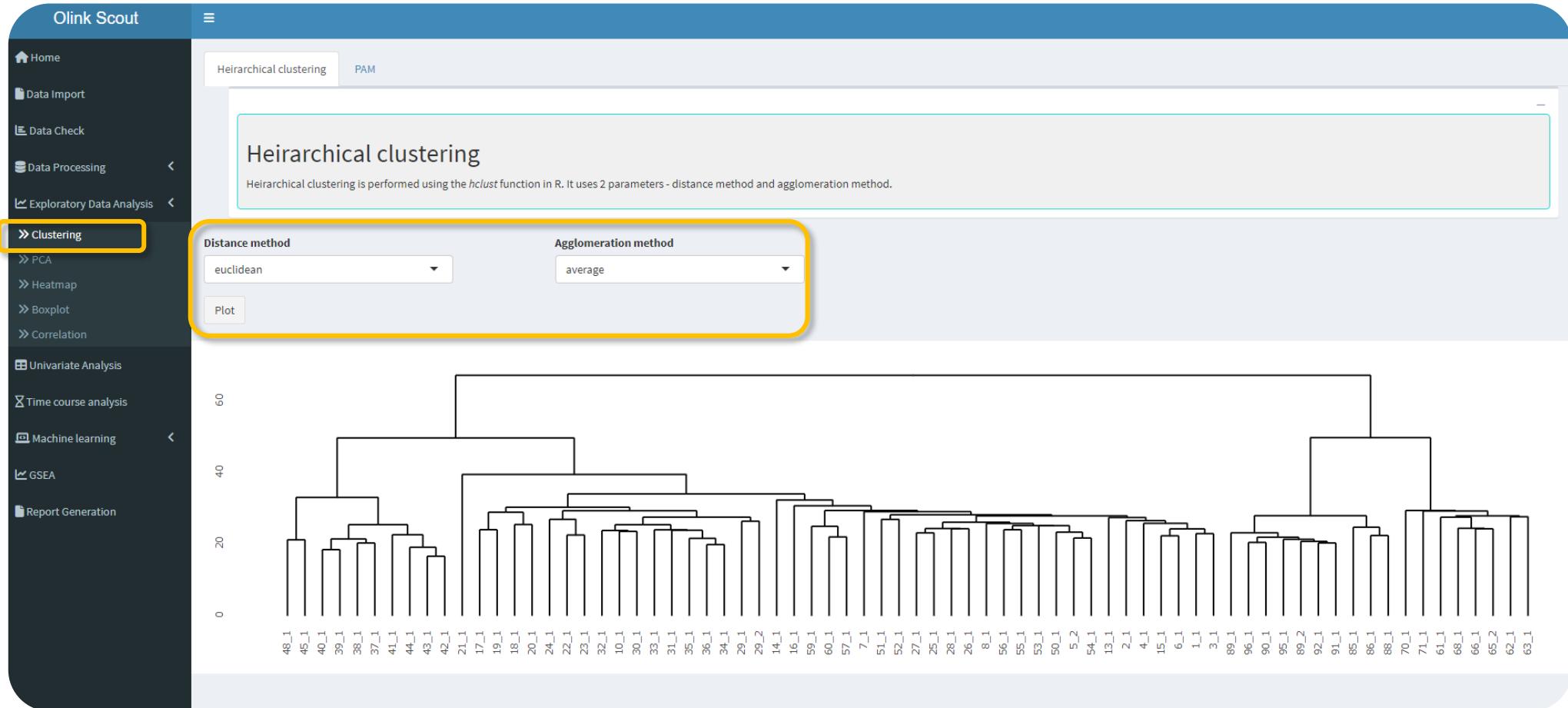


Shape, colour  
based on  
SubjectID & Group

# Sanofi- Olink Application (Olink Scout)



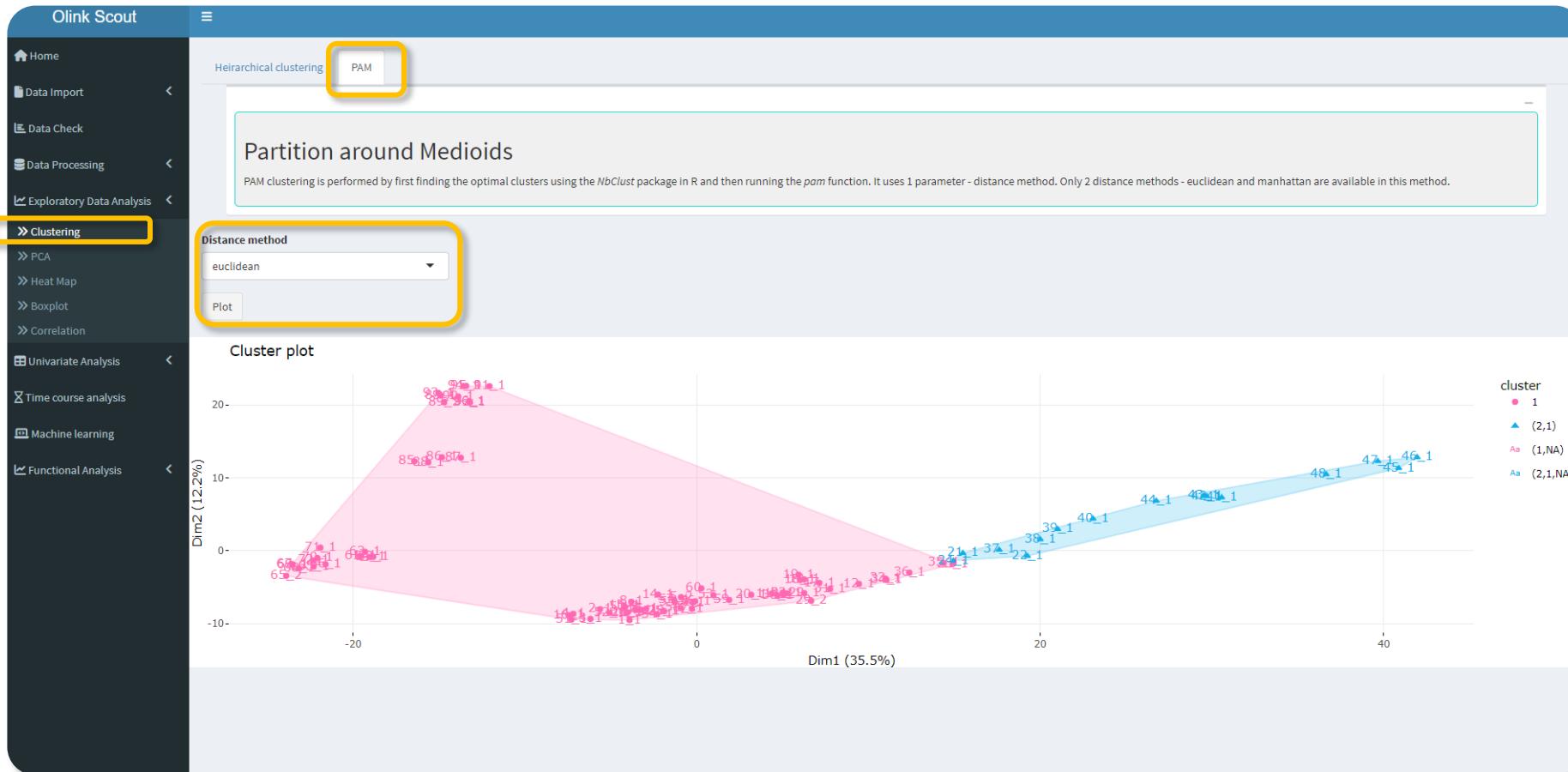
## Exploratory Data Analysis – Hierarchical clustering



# Sanofi- Olink Application (Olink Scout)

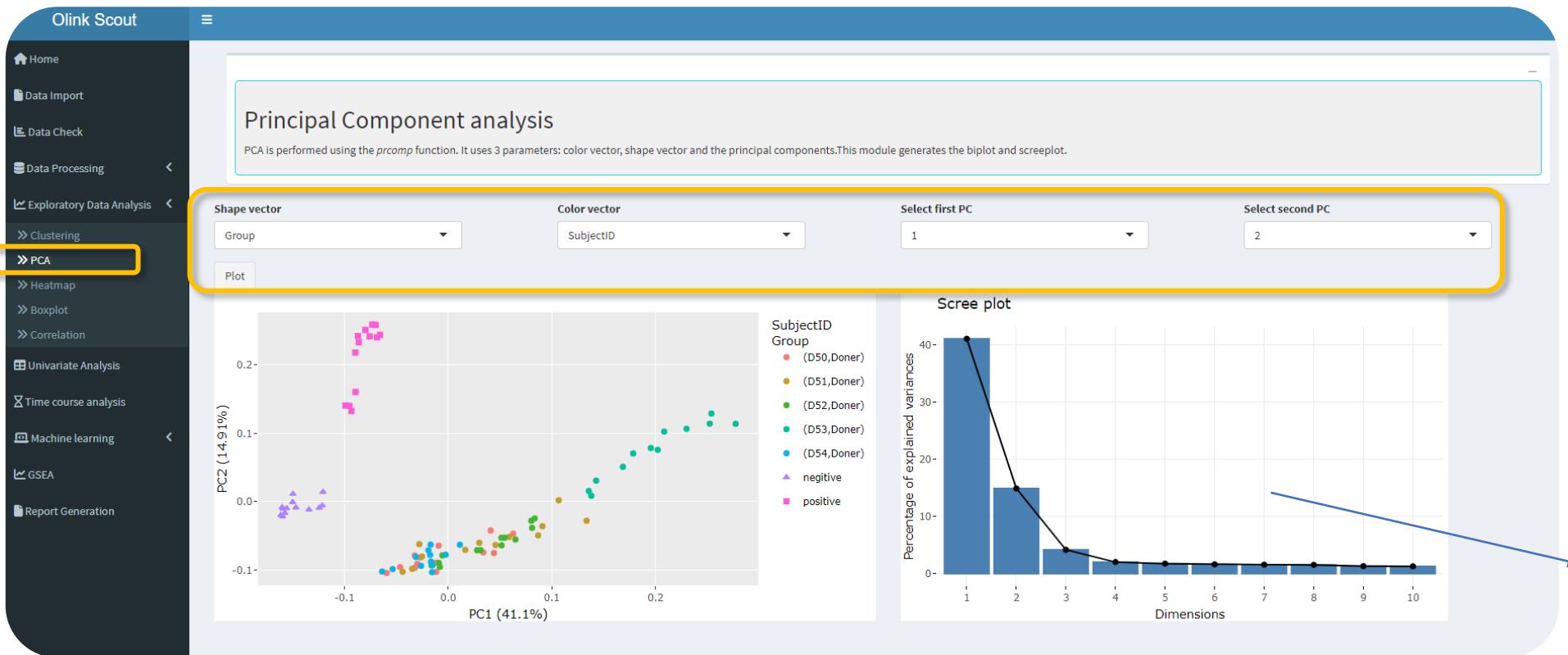


## Exploratory Data Analysis – Partition around Medoids Clustering



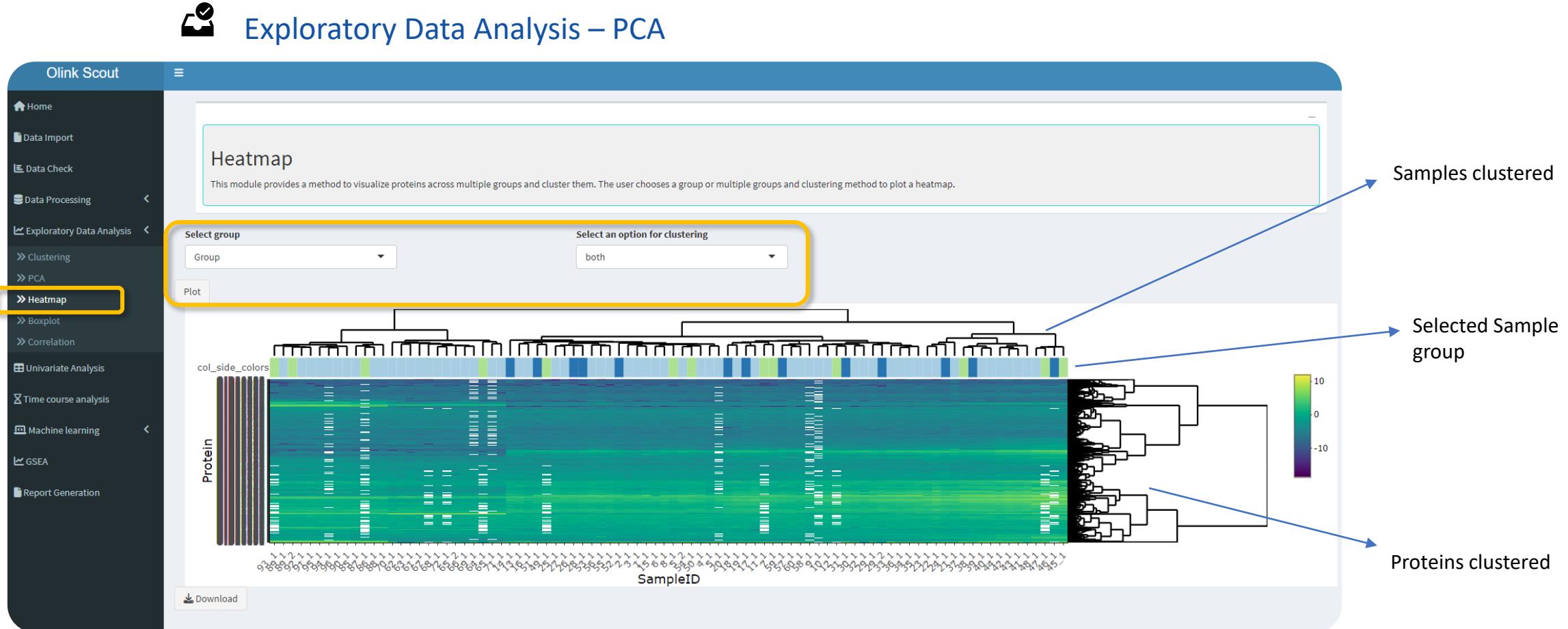
# Sanofi- Olink Application (Olink Scout)

## 📍 Exploratory Data Analysis – PCA



Variance explained  
by dimensions

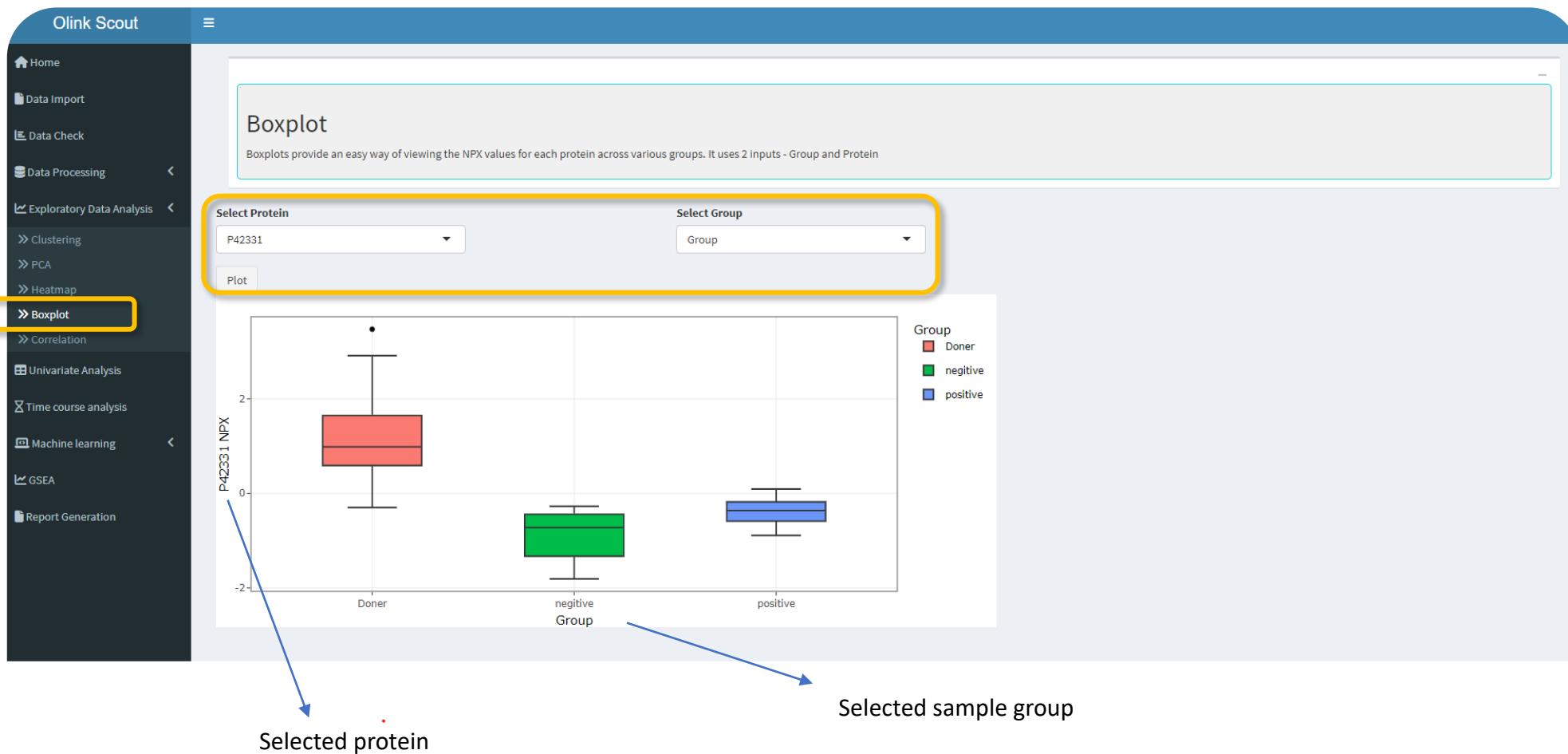
# Sanofi- Olink Application (Olink Scout)



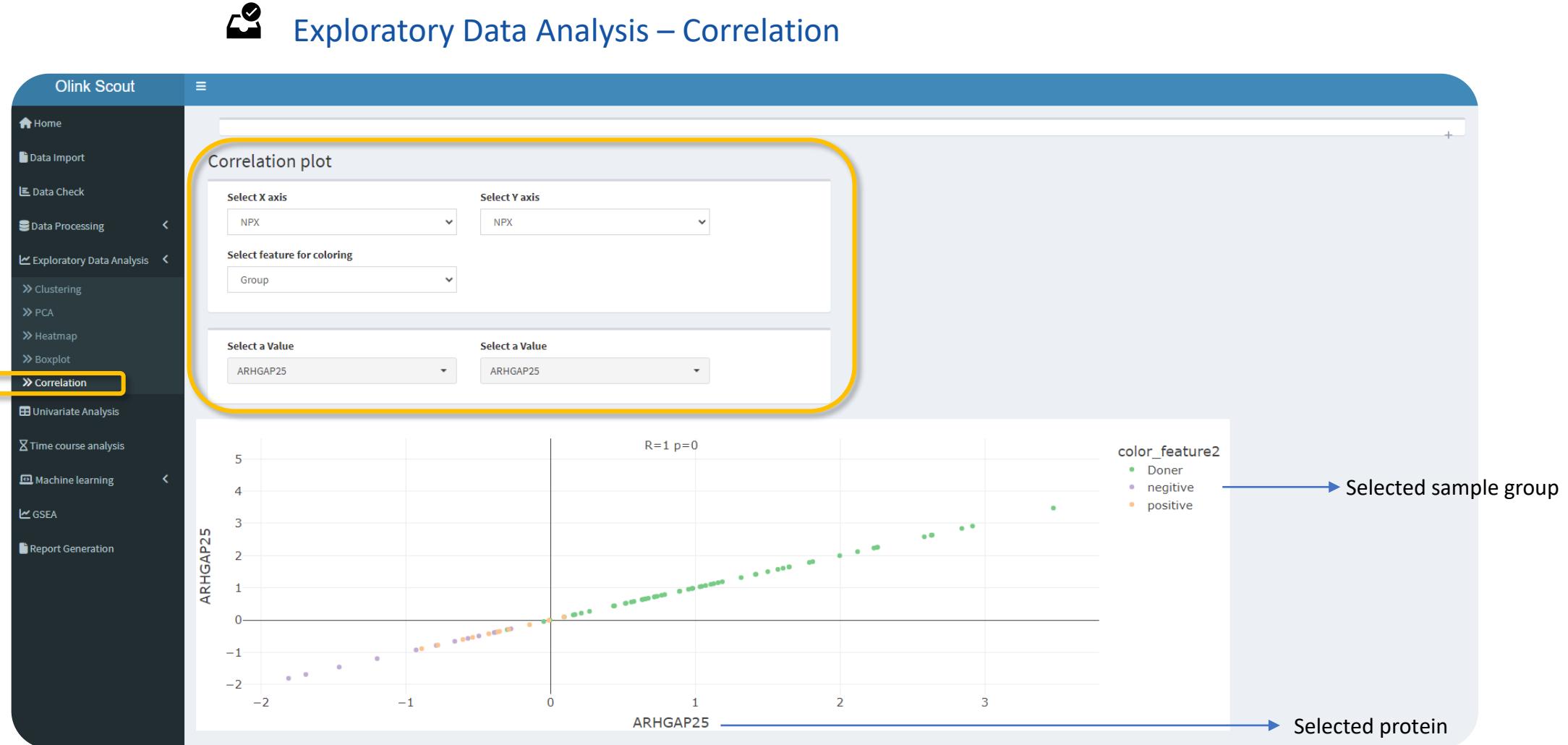
# Sanofi- Olink Application (Olink Scout)



## Exploratory Data Analysis – Boxplot



# Sanofi- Olink Application (Olink Scout)



# Sanofi- Olink Application (Olink Scout)



Univariate Analysis : T-Test

The screenshot displays the Olink Scout application interface for Univariate Analysis, specifically using a T-Test. The interface is divided into several sections:

- Top Left:** Diagnostic Markers section with a list of statistical tests: T-Test, ANOVA (Analysis of Variance), and Mann-Whitney U Test.
- Top Center:** A blue box labeled "T-Test" points to the "Select a test" dropdown which is set to "T-Test".
- Left Sidebar:** Navigation menu with options like Home, Data Import, Data Check, Data Processing, Exploratory Data Analysis, Univariate Analysis (which is selected and highlighted in orange), Diagnostic Marker, Time course analysis, Machine learning, and Functional Analysis.
- Parameters:** A yellow box labeled "Parameters" points to the "Comparison variable" dropdown set to "Group" and the "Case" dropdown set to "Doner" and "Control".
- Type of T-Test:** A blue box labeled "Type of T-Test" points to the "Type of T-Test" dropdown, which is set to "Unpaired".
- Run Button:** A small button labeled "Run" is located next to the "Type of T-Test" dropdown.
- Results tab:** A blue box labeled "Results tab" points to the main table area. The table has columns: Assay, OlinkID, UniProt, Panel, LogFC, p-value, method, alternative, and Adjusted\_pval. It shows 10 entries out of 737 total. An entry for IL6 is highlighted in blue.
- Heatmap:** A blue box labeled "Heatmap" points to a heatmap visualization where rows represent SampleID and columns represent Assay. A color scale on the right ranges from -4 (blue) to 4 (red). A yellow circle highlights the "p-value cutoff for heatmap" at 0.05.
- Volcano plot:** A blue box labeled "Volcano plot" points to a scatter plot showing -log10(p-value) on the y-axis versus LogFC on the x-axis. A horizontal line at approximately 1.30103 indicates the p-value cutoff.
- Proteins table:** A blue box labeled "Proteins table" points to a detailed table for the IL6 protein, showing NPK values for different conditions (Doner, negative; positive) across four dimensions (0, 0.3, 1, 3).
- Tool tip with information:** A blue box labeled "Tool tip with information" points to a tooltip providing details about the IL6 protein table.
- Box plot:** A blue box labeled "Selected Protein - box plot" points to a box plot for the IL6 protein, comparing NPK values between Doner and Control groups across different dimensions.

# Sanofi- Olink Application (Olink Scout)



Univariate Analysis : ANOVA

**Olink Scout**

Univariate Analysis:

- T-Test - Statistical test used to compare the means of two independent groups to determine if there is a significant difference between them. (Reference: OlinkAnalyze Package Vignette)
- ANOVA (Analysis of Variance) - Statistical method used to compare means across multiple groups to determine if there are significant differences between them. (Reference: ANOVA Vignette, ANOVA Post-Hoc Analysis Vignette)
- Mann-Whitney U Test - It is also known as the Wilcoxon rank-sum test, is a non-parametric statistical test used to compare the distributions of two independent groups. (Reference: Man-Whitney U Test Vignette)

Select a test:  
 T-Test  
 ANOVA  
 Mann-Whitney U Test

Comparison variable1: Group  
Comparison variable2: Dose. $\mu$ M.  
Covariate variable: Timepoint  
Perform Post-hoc:  
 Yes  
 No

Pair ID variable (This Column is used for removing singletons and replicates): SubjectID

Run

**Results tab**

Anova Results  
Show 10 entries

	Assay	OlinkID	UniProt	Panel	term	df	sumsq	meansq	statistic	p.value	Adjusted_pval
1	KRT18	OID21296	P05783	Oncology	Group	2	1,416.9684	708.4842	3,667.9031	0.0000	0.0000
2	FGFBP1	OID21507	Q14512	Oncology	Group	2	800.6298	400.3149	1,113.5957	0.0000	0.0000
3	EPS8L2	OID21397	Q9H653	Oncology	Group	2	178.2434	89.1217	1,013.0633	0.0000	0.0000
4	ADAMTS15	OID21275	Q8TE58	Oncology	Group	2	226.2207	113.1104	906.5397	0.0000	0.0000
5	PLAT	OID20247	P00750	Cardiometabolic	Group	2	398.9166	199.4583	934.2193	0.0000	0.0000
6	COL18A1	OID20332	P39060	Cardiometabolic	Group	2	343.3429	171.6714	905.6794	0.0000	0.0000
7	CA9	OID21417	Q16790	Oncology	Group	2	308.5412	154.4706	856.0778	0.0000	0.0000
8	S100A12	OID21374	P80511	Oncology	Group	2	816.7407	408.3704	642.7122	0.0000	0.0000
9	EDIL3	OID20081	O43854	Cardiometabolic	Group	2	914.9316	457.4658	717.1993	0.0000	0.0000
10	EDA2R	OID21451	Q9HAV5	Oncology	Group	2	62.3496	31.1748	577.8960	0.0000	0.0000

Showing 1 to 10 of 2,187 entries

Select cutoff for pval: 0.05

Select a group to sort: SampleID

**p-value heatmap**

Select variable1 for xaxis (Mandatory): Group  
Select variable2 for xaxis (optional): Timepoint

**KRT18**

Boxplot for KRT18 at timepoints 8, 24, 36, and 48. The y-axis is labeled "IPX". The x-axis shows "Donor" and "negative positive". The legend indicates "Group": red square for "positive", green square for "negative", and blue square for "Donor".

# Sanofi- Olink Application (Olink Scout)



Univariate Analysis : Mann-Whitney U Test

**Parameters**

**Heatmap**

**Mann-Whitney U Test**

**p-value cutoff for heatmap**

**Proteins table**

**Volcano plot**

**Selected Protein - box plot**

**Tool tip with information**

# Sanofi- Olink Application (Olink Scout)

## ⌚ Time course Analysis

The screenshot shows the 'Time course analysis' section of the Olink Scout application. On the left, a sidebar lists various analysis options. The 'Time course analysis' option is highlighted with a yellow box. The main panel has three input fields: 'Select Time variable' (disease\_duration\_weeks), 'Select Random variable' (sex), and 'Select Covariates' (CSF\_findings\_Albumine\_ratio). A 'Submit' button is below these. To the right is a 'Parameters' table with columns: Assay, OlinkID, UniProt, Panel, term, sumsq, meansq, NumDF, DenDF, and statistic. Below the table are two dot plots for CD27 and TNFRSF13B, each with multiple colored dots representing different protein levels across time points. At the bottom is a heatmap titled 'Selected Proteins plot'.

Run Posthoc

Advanced interface for Time course analysis.

## Time course Posthoc analysis

The screenshot shows the 'Time course Posthoc analysis' section of the Olink Scout application. The 'Time course analysis' parameters from the previous screen are repeated here. The 'Select effect' dropdown is set to 'disease\_duration\_weeks'. The 'Select proteins' dropdown is set to 'significant proteins'. A 'Submit' button is present. To the right is a 'Parameters' table with columns: Assay, OlinkID, UniProt, Panel, term, contrast, estimate, contlow, conthigh, and Adjusted\_pval. Below the table is a heatmap titled 'contrast for heatmap' with a dropdown menu set to '5\_weeks - No record'. The heatmap itself shows a grid of colored squares representing significant proteins across different time points and samples.

Significant proteins heatmap based on p-value

# Sanofi- Olink Application (Olink Scout)



# Machine Learning Analysis

## ML analysis to find proteins of interest.

# Sanofi- Olink Application (Olink Scout)



Functional Analysis

Sanofi- Olink Application (Olink Scout) - Functional Analysis

Gene Set Enrichment Analysis

Gene set enrichment can be performed on the results obtained from ttest, ANova, Man whitney test, Machine learning models to identify the pathways/genesets enriched in the samples. Current implementation supports all MSigDB genesets ( Molecular signature database). Alternately, custom genesets can be uploaded.

Select an option to perform GSEA

ttest  ANOVA  Man Whitney U-Test  Timecourse  Machine Learning

Select a model

PLSDA

Select a group

Overall

Select an option to perform GSEA

msigdb  Customized Gene Set

Please select species

Homo sapiens

Please Select a Category

H-hallmark gene sets  
 C1-positional gene sets  
 C2-curated gene sets  
 C3-motif gene sets  
 C4-computational gene sets  
 C5-GO gene sets  
 C6-oncogenic signatures  
 C7-immunologic signatures  
 C8-cell type signature

p adjust cutoff

0.05

Run GSEA

Parameters

Select Machine Learning/Univariate analysis/Timecourse analysis output proteins

GSEA results

Gene ontology table

Gene ratio dot plot

Enrichment score ridge plot

Functional analysis for proteins of interest.

Dot Plot

Ridge Plot

# Sanofi- Olink Application (Olink Scout)



## Report Generation

The screenshot shows the Olink Scout application interface. On the left, there is a dark sidebar with a list of analysis sections: Home, Data Import, Data Check, Data Processing, Exploratory Data Analysis, Univariate Analysis, Time course analysis, Machine learning, GSEA, and Report Generation. The 'Report Generation' button is highlighted with a yellow box. The main content area has a blue header bar with the title 'Generate report'. Below it, a text box states: 'Report Contains the plots which are generated by the user, if user doesn't performed or skipped any section, that will be showcased as NULL or Not selected'. A section titled 'Download Plots' follows, with a note: 'It will be downloaded as zipped file, which contains the plots generated by the user. if user doesn't performed or skipped any section during the analysis, those plots will be generated as Empty Plots.' At the bottom of this section are two buttons: 'Generate report' and 'Download Plots'. Blue arrows point from these buttons to the text 'Download html report' and 'Download zip folder with plots' respectively. At the very bottom of the page is a large blue button labeled 'Download analysis report and plots'.

Download analysis report and plots

# Internal Olink resources

