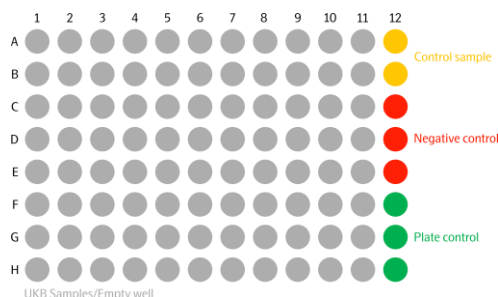




## UKB – Olink Explore 1536 - FAQ

1. **What is NPX?** Normalized Protein eXpression (NPX), is Olink's arbitrary unit which is in  $\log_2$  scale. Further details [here](#).
2. **How does Olink's plate layout look like?**



Olink's plate format consists of a 96-sample plate, which is composed of up to 88 UKB samples in columns 1-11, and 8 external controls that are provided by Olink in column 12. A negative control run in triplicate is used to calculate the limit of detection (LOD) of each assay in every plate, and a plate control sample, consisting of a pooled plasma sample is run in triplicate to adjust the levels between plates. A duplicate pooled sample control is included to estimate precision within and between runs.

3. **How were the samples shipped and how was the data delivered?** Due to the size of the project, samples were shipped pre-plated to Olink in batches and data was delivered to UKB as batches.
4. **How many batches and plates are UKB samples randomized on?** See table below with number of plates per batch:

Batch	Alternative BatchID	Number of Plates
B1	P101	20
B2	P102	72
B3	P103	160
B4	P104	100
B5	P105	100
B6	P106	100
B7	P107	119
Pilot	PILOT	1
<b>Total</b>	<b>8</b>	<b>672</b>



5. **What is the difference between overlapping and bridging samples?** All bridge and overlapping samples consist of samples within the UKB cohort. Bridge samples have been used to form the study design by normalizing between sub-parts of the project while overlapping samples are used only to monitor the quality of the normalization.
6. **What is the difference between overlapping and control samples?** Control samples are plasma pools provided and used by Olink for quality control (e.g. computation of coefficients of variation). Overlapping samples are UKB samples that were transferred from one batch to the next to monitor batch-to-batch variation and normalization.
7. **How was the data normalized within and across batches?** UKB samples between batches 1 and 6 and the pilot plate were totally randomized. Batch 7 consists of UKB samples and COVID samples. Sample selection in batch 7 was biased as COVID samples were either heavily overrepresented or heavily underrepresented in sample plates. Details regarding data normalization are available in the accompanying document “Data Normalization Strategy”.
8. **What format is the Olink data delivered in?** UK Biobank reformats the data for delivery through the RAP and Showcase systems, and information on the structure is provided in additional resources. For researchers who wish to reconstruct the original Olink data format, Olink delivers data in an NPX file format, which contains the following columns:

Column name	Description	Format
SampleID	Sample ID	Alpharithmetic
Index	Plate index	Integer 1-96
OlinkID	Olink ID	Alpharithmetic
UniProt	UniProt identifier	Alpharithmetic
Assay	Gene name	Alpharithmetic
MissingFreq	Missing frequency	Numeric with 4 decimals
Panel	Olink Panel	Character
Panel_Lot_Nr	Panel lot number	Alpharithmetic
PlateID	Plate ID	Alpharithmetic
WellID	Plate well ID	A1, A2, ..., H12
QC_Warning	Quality control warning	PASS, WARN or MANUAL_WARN
LOD	Limit of detection	Numeric with 4 decimals
NPX	NPX value	Numeric with 4 decimals
Normalization	Normalization method	Type of normalization of data
Assay_Warning	Assay warning	PASS, WARN or MANUAL_WARN
Assay_Version	Assay version	Integer



9. **Why are there plates with all NPX data null?** Some plates were contaminated, and they were not run on Olink Explore 1536. Datapoints for these plates have been reported with null NPX values. Plate identifiers of contaminated plates are shown in the table below.

Batch	PlateID
B5	0890000000432
B6	0890000000548
B6	0890000000562
B6	0890000000494

10. **How many assays and panels does UKB data contain?** The table below shows the total number of assays per batch. The table also shows the Panel\_Lot\_Nr used in each case.

Panel	Panel_Lot	B1	B2	B3	B4	B5	B6	B7	Pilot
Cardiometabolic	B04413	369	369	369	369	369	369	369	369
Inflammation	B04411	368	368	368	368	368	368	368	368
Neurology	B04414	367	367	367	367	367	367	367	367
Oncology	B04412	368	368	368	368	368	368	368	368
<b>Total assays</b>		<b>1472</b>	<b>1472</b>	<b>1472</b>	<b>1472</b>	<b>1472</b>	<b>1472</b>	<b>1472</b>	<b>1472</b>
<b>Unique assays*</b>		<b>1463</b>	<b>1463</b>	<b>1463</b>	<b>1463</b>	<b>1463</b>	<b>1463</b>	<b>1463</b>	<b>1463</b>

\*Note, the 4 panels from Olink Explore 1536 share 3 common overlapping assays, IL8, CXCL8 and TNF, resulting in 1463 unique assays. All overlapping assays have unique Olink identifiers.

11. **How was the sample selection performed?** Each of the 13 partners of the Pharma Proteomics Project selected a set of 500 samples of their preference. Additional samples originating from the COVID study were included in batch 7. All other samples were randomly selected from the UKB population.



12. **How many samples are there in the UKB data?** Please see the table below for details. Also note that one subject whose sample was included in batch 5 withdrew from the study.

Batch	Total Samples Sent to Olink	Withdrawn	Contaminated	Total Usable Samples
Pilot	87	0	0	87
B1	1740	0	0	1740
B2	6264	0	0	6264
B3	13920	0	0	13920
B4	8700	0	0	8700
B5	8700	1	87*	8612
B6	8700	0	261*	8439
B7	10334	0	0	10334
<b>Subtotal</b>	<b>58445</b>	<b>1</b>	<b>348</b>	<b>58096</b>

\*See point 9 above for plate identifiers of the contaminated plates.

13. **Why are there more than 2 different sample identifiers for control samples of the same plate?** Due to reruns, identifiers of control samples of the same plate might be different among panels. As a result, plates might have 2, 4, 6 or 8 different identifiers for the control samples. If unique identifiers are required for control samples, we recommend combining the columns PlateID and WellID.
14. **Why are there plates that have the same sample identifier for control samples?** Due to its size the study was processed into subprojects which led to a subset of plates with shared control sample identifiers. We recommend creating unique sample identifiers by combining the columns SampleID and PlateID.
15. **Why is there fewer data than the ones reported in Olink's analysis report?** The Pharma Proteomics Project identified a set of samples there were marked as outliers. This data has been delivered by Olink but removed from the final dataset as outlined in the accompanying Pharma Proteomics Project resource document UKB-UKB\_PPP\_Phase\_1\_QC\_dataset\_companion\_document.pdf. Olink's analysis report is available in the file UKB\_Olink\_Explore\_1536\_B0\_to\_B7\_Analysis\_Report.pdf.



16. **Why is there a different number of reported samples between sections 1 and 2.1 in Olink's analysis report?** Data for contaminated plates was reported with null NPX values (see point 9 above). In addition, a small fraction of overlapping/bridging samples was not included in some panels due to low volume and the sample of a withdrawn subject was removed. The table below shows the breakdown of samples in batches 0-7 for each panel.

Panel	UKB	Overlapping/ Bridging	Contaminated	Withdrawn	Usable	Total
Cardiometabolic	58,095	273	349	1	58,369	58,718
Inflammation	58,095	267	349	1	58,363	58,712
Neurology	58,095	270	349	1	58,366	58,715
Oncology	58,095	270	349	1	58,366	58,715