



UKB – Olink Explore 3072 - Data Normalization Strategy

Samples were randomly assigned to 96-well plates, and fully randomized within plates. Each plate contained: i) 87 samples; ii) 1 empty well or overlapping sample; iii) 2 Olink control samples used for quality control; iv) 3 Olink negative control samples used to compute the baseline assay level of each plate; and v) 3 Olink plate control samples used for normalization of protein expression. All Olink samples were placed in column 12 of plates while the remaining 87 samples + 1 empty well were randomized across columns 1-11 and rows A to H (Figure 1).

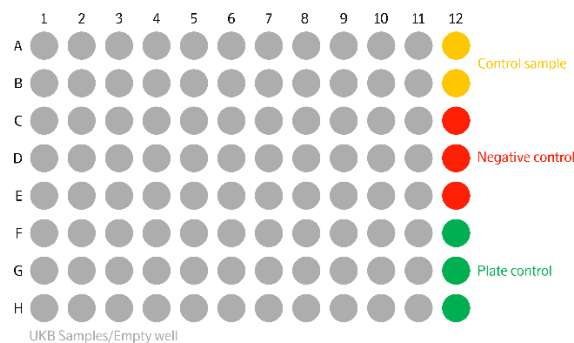


Figure 1: Layout of Olink plates with UKB samples. Each plate had one empty well between A1 and H11.

Plates were shipped in 7 batches (0-6) that consisted of different numbers of samples (Figure 2).

		45,640 samples						
		Batch 0	Batch 1	Batch 2	Batch 3	Batch 4	Batch 5	Batch 6
#Plates		1	20	72	160	100	100	100
#Samples (wBSDs)		87	1,740	6,264	13,920	8,700	8,699	8,700
#Additional Samples								

Figure 2: Schematic representation of the study design for UKB B0-B6.

Calculation of Normalized Protein eXpression (NPX) values was performed stepwise as described below. Initially, we calculated the \log_2 ratio of counts of each assay of each sample to the counts of the extension control

$$ExtNPX_{i,j} = \log_2 \left(\frac{Counts(Sample_j Assay_i)}{Counts(ExtensionControl_j)} \right)$$

, and the assay-specific median value of the plate controls was subtracted.

$$NPX_{i,j} = ExpNPX_{i,j} - median(ExtNPX(PlateControls_i))$$

This provided us with plate control (PC) normalized NPX values.

Next, we subtracted the assay-specific plate median NPX value, and we added the batch-specific median NPX value of each assay to account for effects within each batch.



$$NPX_{IntNorm_{i,j,k}} = NPX_{i,j,k} - plate_k \text{median}(NPX_i)$$

At this stage data was normalized within each batch. Next, we computed adjustment factors from the difference of the assay-specific median NPX value of each batch to the reference batch (batch 1). Batch 1 was the first to run chronologically, underwent extensive investigation on randomization and assay performance, hence, it was selected as the reference batch. However, selection of any batch as reference would result in identical results.

$$adj_factor_{batch_1-batch_{n_i}} = \text{median}(NPX_{batch_{1_i}}) - \text{median}(NPX_{batch_{n_i}})$$

Finally, adjustment factors from above were added to the NPX values of each batch.

$$NPX_{IntNorm_{i,j,n}}^{final} = NPX_{IntNorm_{i,j}} + adj_factor_{batch_1-batch_{n_i}}$$

In summary, this two-step normalization approach consisted of within-batch and across-batches intensity normalization.