Olink

Analysis Report

PROJECT NAME	UKB B0-B7 Analysis Report
CUSTOMER	Pharma Proteomics Project
BUSINESS DEVELOPMENT MANAGER	
STATISTICAL SERVICE REPRESENTATIVE	Olink Proteomics
DATE	March 3, 2023

Contents

1. Introduction

The UK biobank study contains 58718 EDTA plasma samples in 672 plates. Samples are divided into in 8 batches (0-7). Samples in batch 7 are divided into two sets, UKB samples and COVID-19 samples. UKB samples were selected at the same time for batches 0-6 and part of batch 7; COVID-19 samples were selected at a later time to be included in batch 7. For each set a different normalization strategy has been followed to best correct batch effects. Samples from the UKB set were subset normalized to batch 1; samples from the COVID-19 set were bridge-normalized to the subset-normalized set of batches 0-6 using 93 bridge samples (Figure 1 - marked in red).

A set of up to 32 samples from batch *n-1* was included in batch *n* as an approach to monitor batch-to-batch normalization (Figure 1 - marked in yellow). These samples have been marked with a suffix *"b"* in batches 2-6 and with a *"c"* in batch 7.

Please note that the batch 0 and PILOT refer to the same batch of samples.

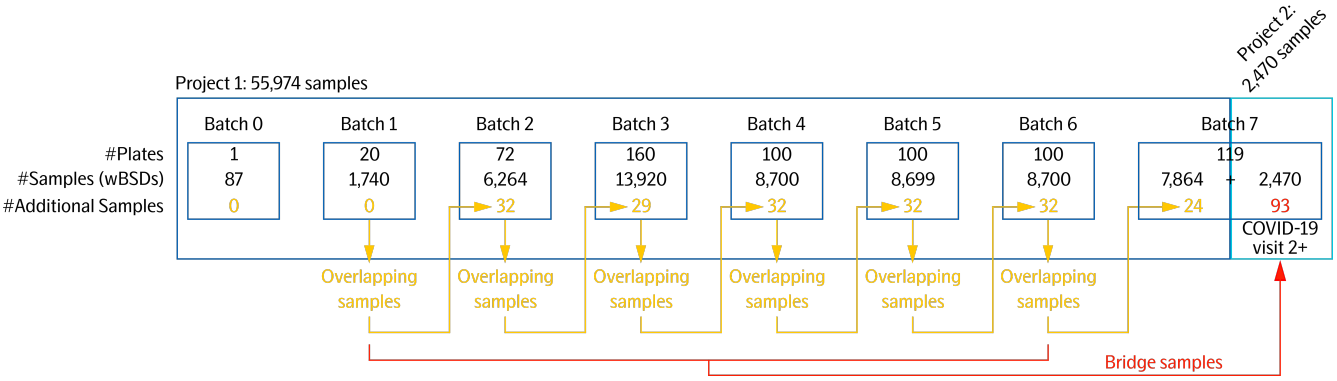

Figure 1: UKB study design and normalization strategy.

NB 1 Please consult with the accompanying document "UKB_Olink-Explore-1536_B0-B7_FAQ" for details on the samples belonging to the UKB set and the COVID set.

NB 2 Please consult with the accompanying document "UKB_Olink-Explore-1536_B0-B7_Data-Normalization-Strategy" for details on data normalization.

2 Quality Control (QC)

Three internal controls are added to each sample in the plate, the Incubation control, the Extension Control and the Amplification control. The Extension Control is used for the generation of the NPX values. The Incubation Control and the Amplification Control are used to monitor the quality of assay performance, as well as the quality of individual samples.

Three external controls are included in each run, the Plate Control (healthy pooled plasma), Sample Control (healthy pooled plasma) and Negative Control. The Plate Control is used for data normalization, the Sample Control is used to assess potential variation between runs and plates, and the Negative Control is used to calculate Limit of Detection for each assay and to assess potential contamination of assays. The following parameters are evaluated in the Quality Control (QC):

- Sample Quality Control (QC):
  - The average matched counts for each sample. To pass QC, there should be at least 500 counts, otherwise the sample receives a QC warning status.
  - The deviation from the median value of the Incubation- and Amplification Controls for each individual sample. To pass QC, the deviation should not exceed +/-0.3 NPX for either of the internal controls, otherwise the sample will receive a QC warning status.
- Assay QC:
  - The deviation of the median value of the Negative Controls from a predefined value set for each assay. To pass QC, the deviation of the median of the Negative Controls must be ≤5 standard deviations from the set predefined value, otherwise the assay will receive a warning status.

All samples included in the project are presented in the data output file. Samples that do not pass QC are indicated with "WARN" in the column named "QC_warning". Data points from samples that do not pass QC should be treated with caution. Failed samples are indicated with FAIL in the column named "QC_warning". Data points from failed samples are not included in the results file, the cells in the NPX column are empty. Section 2.1 reports the fraction of samples that pass QC for all assays per panel and the fraction of data points passing QC per panel. Assays that do not pass the QC are indicated with WARN in the column named "Assay_warning". Data points from assays that do not pass QC should be treated with caution.

2.1 QC summary

Olink Panel	Samples passed QC	Samples passed QC (%)	Datapoints passed QC	Datapoints passed QC (%)
Cardiometabolic	54523 / 58369	93.41	21152616 / 21538161	98.21
Inflammation	54281 / 58363	93.01	21011339 / 21477584	97.83
Oncology	54153 / 58366	92.78	20961212 / 21478688	97.59
Neurology	54133 / 58366	92.75	20913874 / 21420322	97.64

2.2 Intra- and Inter-assay Coefficient of Variance (%CV)

Intra- and inter-CVs are based on the Sample Controls (pooled plasma samples) included on each sample plate. Calculations are made for each assay using NPX-values. Average % CV for all assays on a panel is

presented in section 2.2.1. The number of assays with CVs within defined intervals are presented in sections 2.2.2 and 2.2.3. Distribution of % CV computed from blind-spiked duplicates, inter- and intra plate % CV computed from control samples and inter-instrument % CV computed from control samples are shown in section 2.2.4.

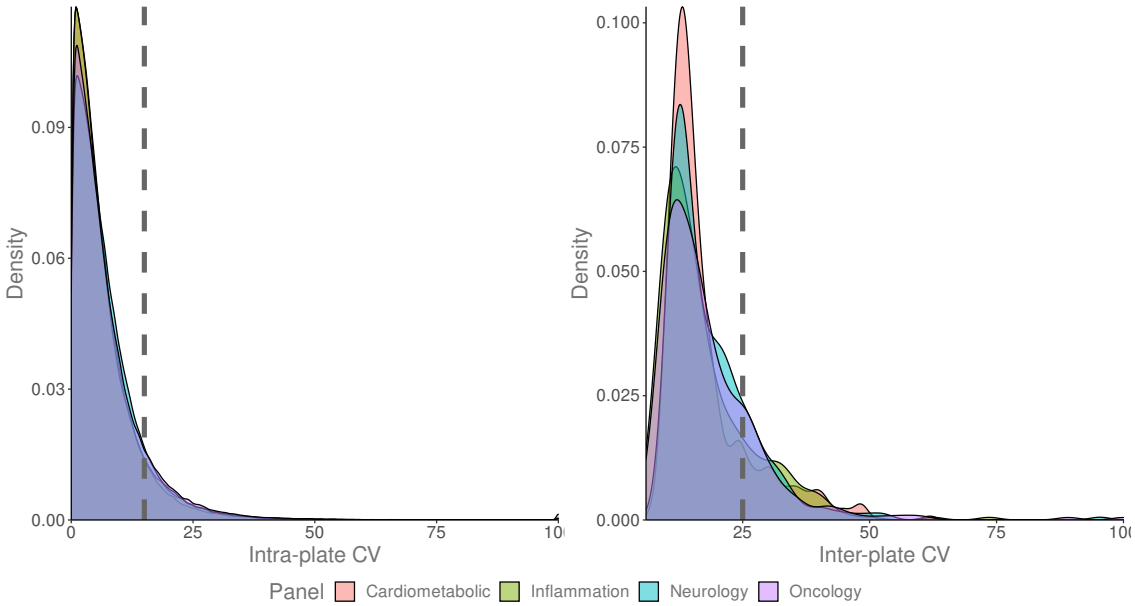

Figure 2: Distribution of intra- and inter-plate CVs.

2.2.1 Average %CV

Olink Panel	Intra-plate %CV	Inter-plate %CV
Explore 384 Cardiometabolic	7.59	17.83
Explore 384 Inflammation	7.61	17.59
Explore 384 Neurology	7.81	18.52
Explore 384 Oncology	8.04	18.37

2.2.2 Intra-assay %CV distribution

Olink Panel	0-5%	5-10%	10-15%	>15%	NA
Explore 384 Cardiometabolic	116	177	53	19	4
Explore 384 Inflammation	132	146	68	18	4
Explore 384 Neurology	59	220	62	19	7
Explore 384 Oncology	92	172	77	19	8

2.2.3 Inter-assay %CV distribution

Olink Panel	0-10%	10-20%	20-30%	>30%	NA
Explore 384 Cardiometabolic	17	262	47	41	2
Explore 384 Inflammation	57	210	59	42	0
Explore 384 Neurology	4	242	87	33	1
Explore 384 Oncology	47	206	81	33	1

2.2.4 Comparison of % CV

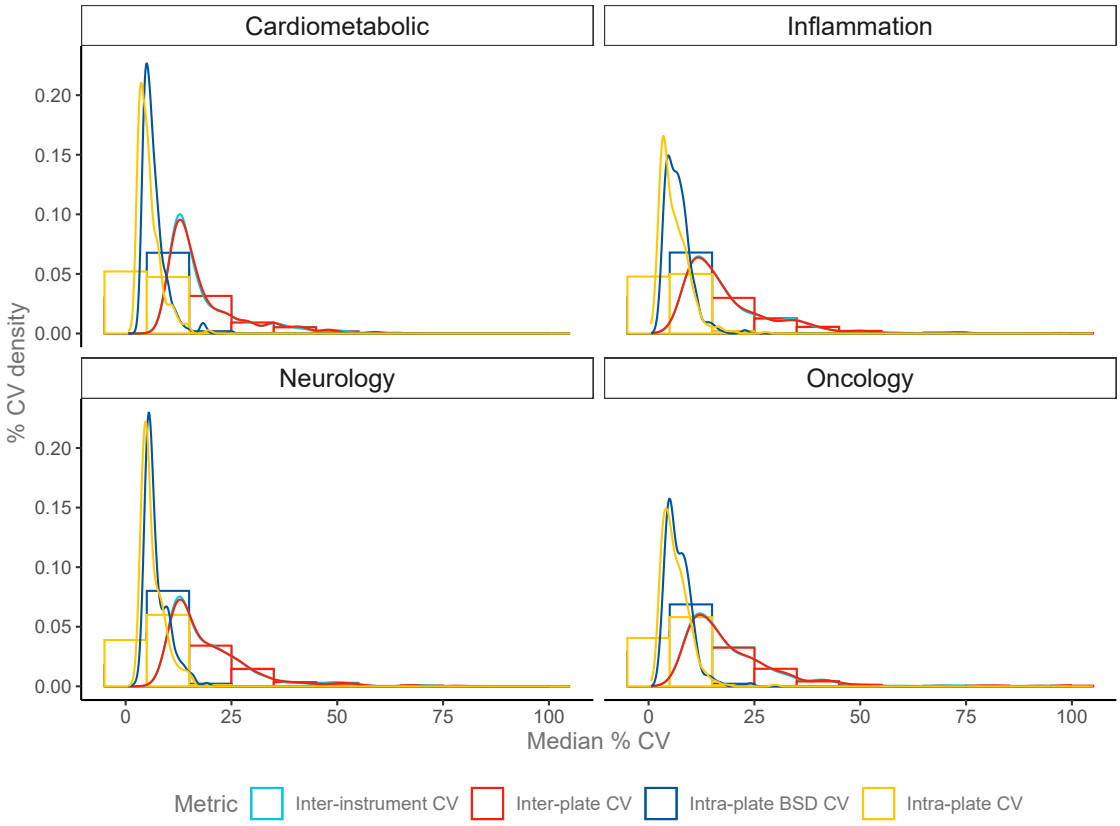

Figure 3: Barplot and density plot of the distribution of % CV computed from blind-spiked duplicates, inter- and intra plate % CV computed from control samples and inter-instrument % CV computed from control samples are shown in section 2.2.4.

3. Protein detection results

3.1 Number of proteins detected in >50% of the samples

Panel	Detected proteins	Detected proteins (%)
Cardiometabolic	350 / 369	94.85
Inflammation	334 / 368	90.76
Neurology	329 / 367	89.65
Oncology	349 / 368	94.84

4. Outlier detection

4.1 Principal component analysis (PCA)

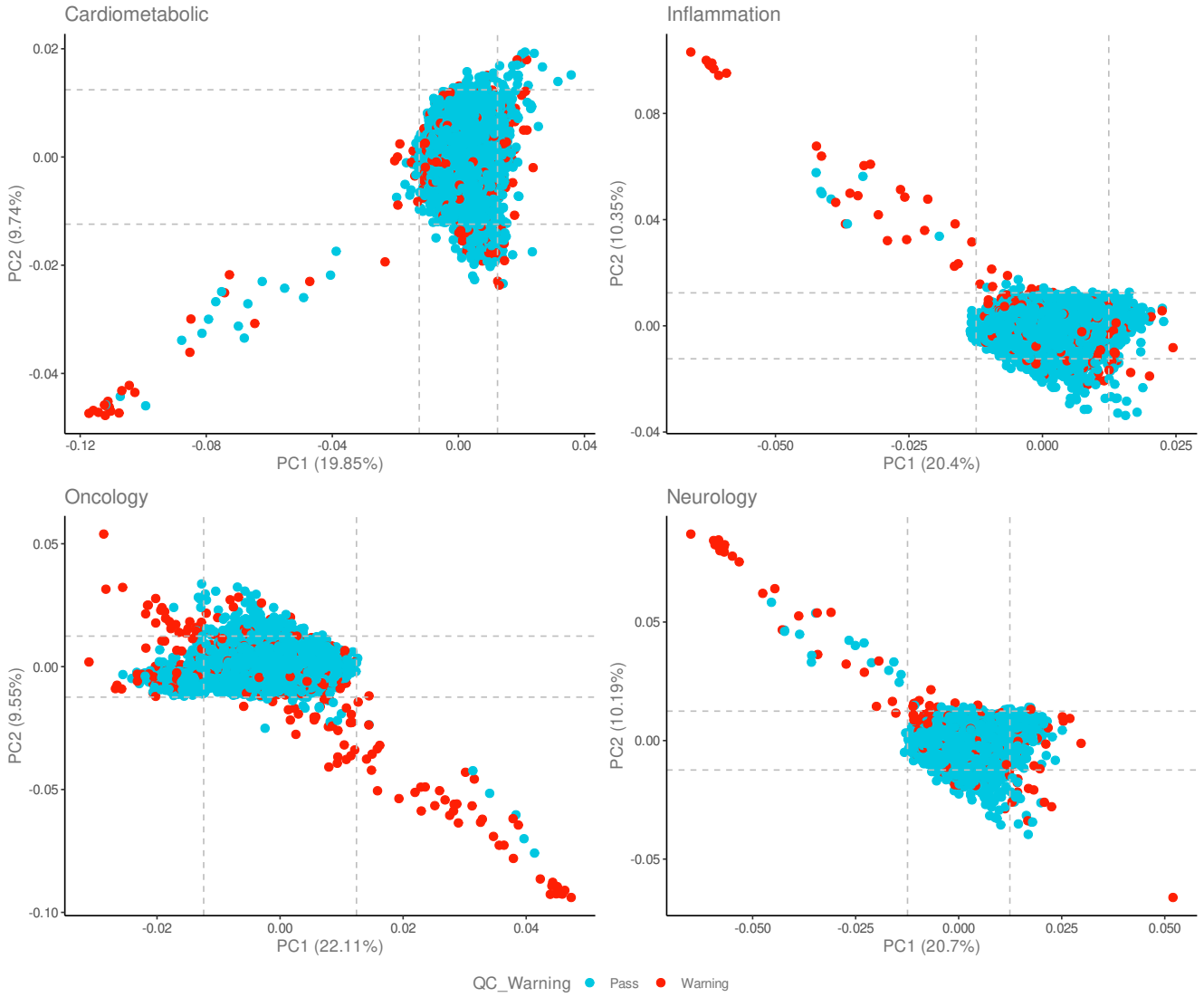

Figure 4: Projection along the first two principal components from a PCA performed on the NPX values. Dashed lines indicate ±3 SDs from the median of each axis. Axes titles indicate the percentage of the variance the principal component explains. Samples were colored by QC warning status.

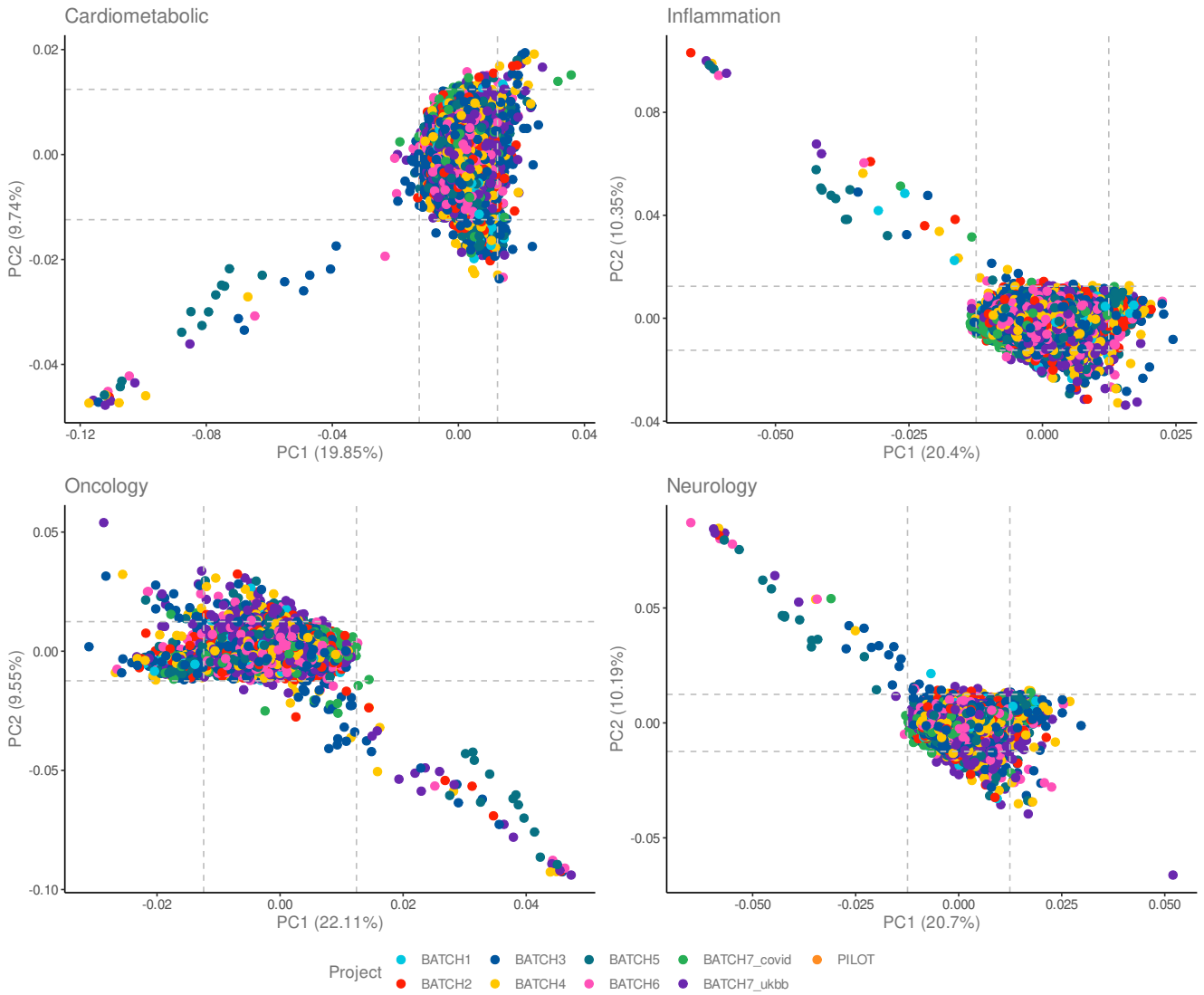

Figure 5: Projection along the first two principal components from a PCA performed on the NPX values. Dashed lines indicate ±3 SDs from the median of each axis. Axes titles indicate the percentage of the variance the principal component explains. Samples were colored by batch and normalization method.

4.2 Deviation from median and interquartile range (IQR)

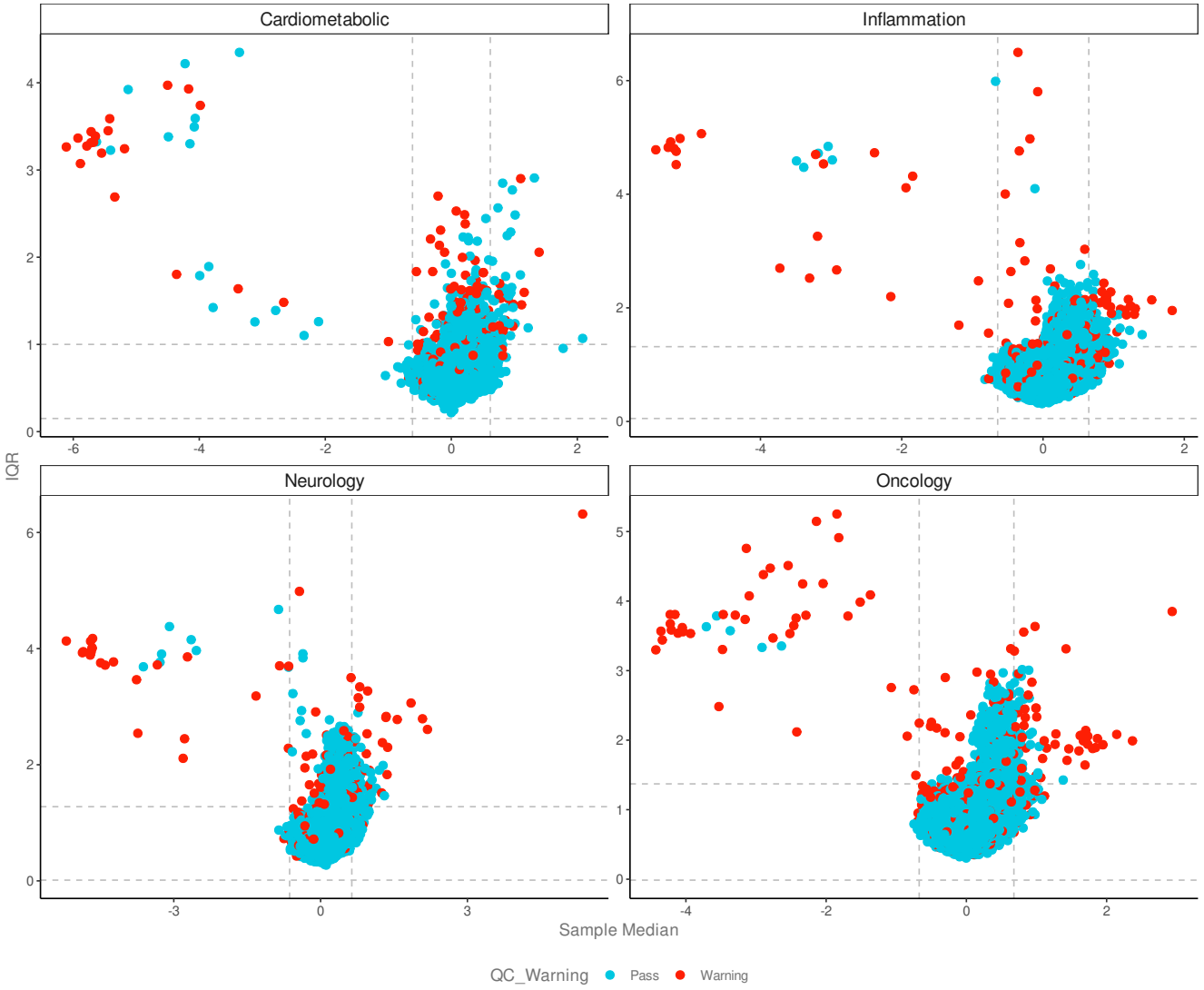

Figure 6: IQR vs. sample medians of each sample. Plot facets represent Olink panels. Horizontal dashed lines indicate ±3 standard deviations (SDs) from the average IQR NPX. Vertical dashed lines indicate ±3 SDs from the average sample median NPX. IDs of a subset of potential outliers are shown to avoid overcrowding the plot. Samples were colored by QC warnings status.

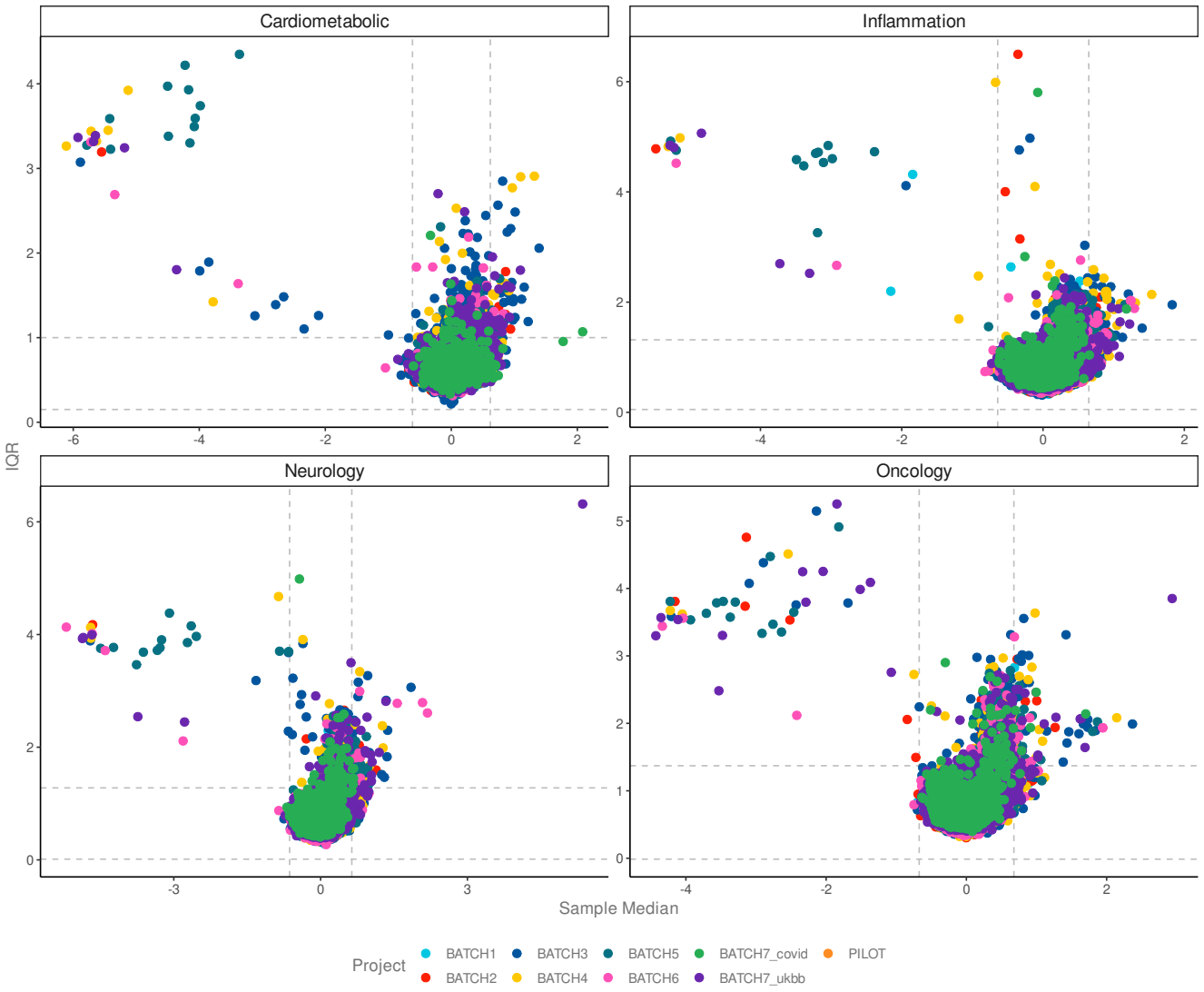

Figure 7: IQR vs. sample medians of each sample. Plot facets represent Olink panels. Horizontal dashed lines indicate ±3 standard deviations (SDs) from the average IQR NPX. Vertical dashed lines indicate ±3 SDs from the average sample median NPX. IDs of a subset of potential outliers are shown to avoid overcrowding the plot. Samples were colored by batch and normalization method.

5. Summary

Overall data is of very high quality. The average intra- and inter-plate %CVs are very low, and assay detectability is over 89% across panels. Over 92% of the samples have passed QC for all assays. PCA and median vs IQR plots indicate the presence of a small number of outliers (Figures 4 and 6).

The UKB samples from batch 7 were intensity normalized to samples from batch 1. Due to the different cohort design, samples from batches 2-6 and the pilot were also intensity normalized to samples from batch 1. The set of COVID-19 samples from batch 7 was bridge normalized to the UKB samples (pilot plate and batches 1-7) using the 93 bridging samples (Figure 1). The result of the normalization strategies is illustrated in (Figures 5 & 7) and suggests a successful approach. An additional investigation in individual batches showed that outliers are distributed across batches (Figures 4, 5, 6 and 7).