# Analysis Report

| | |
|---|---|
| PROJECT NAME | UKBB 3072 B0-B6 Analysis Report |
| CUSTOMER | Pharma Proteomics Project |
| BUSINESS DEVELOPMENT MANAGER | |
| STATISTICAL SERVICE REPRESENTATIVE | Olink Proteomics |
| DATE | September 22, 2023 |

# Contents

# 1. Introduction

The UK biobank study contains 48267 EDTA plasma samples in 553 plates. Samples are divided into 7 batches (0-6). Samples of batches 0 and 2-6 were subset normalized to batch 1 (Figure 1).

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). These samples have been marked with a suffix "b" in batches 2-6.

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

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 | 0 | 0 | 32 | 29 | 32 | 32 | 32 |

Overlapping samples
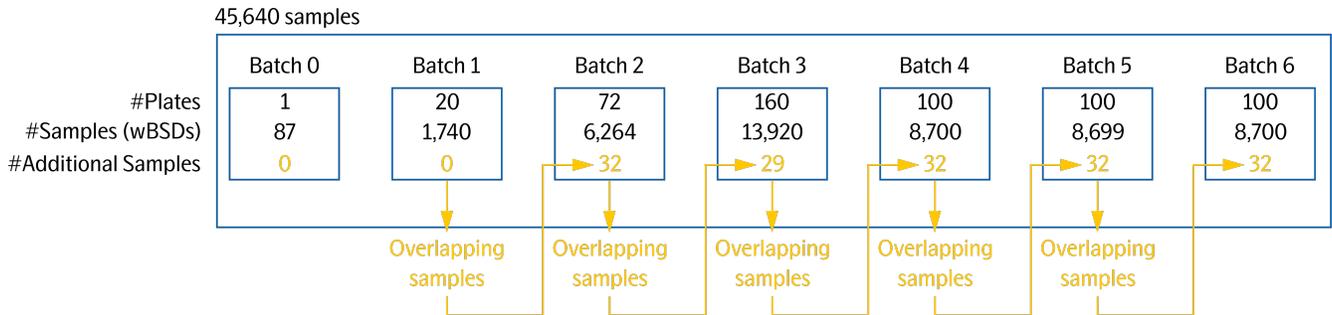
**Figure 1:** UKB study design and normalization strategy.

**NB:** Please consult with the accompanying document "UKB_Olink-Explore-3072_B0-B6_Data-Normalization-Strategy" for details on data normalization.

# 2 Quality Control (QC)

Three internal controls are added to each sample, 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_II | 43626 / 47831 | 91.21 | 17007870 / 17553977 | 96.89 |
| Neurology_II | 43900 / 47831 | 91.78 | 17073231 / 17506146 | 97.53 |
| Oncology_II | 44253 / 47831 | 92.52 | 17184539 / 17601808 | 97.63 |
| Inflammation_II | 43844 / 47831 | 91.66 | 17077469 / 17601808 | 97.02 |

## 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.
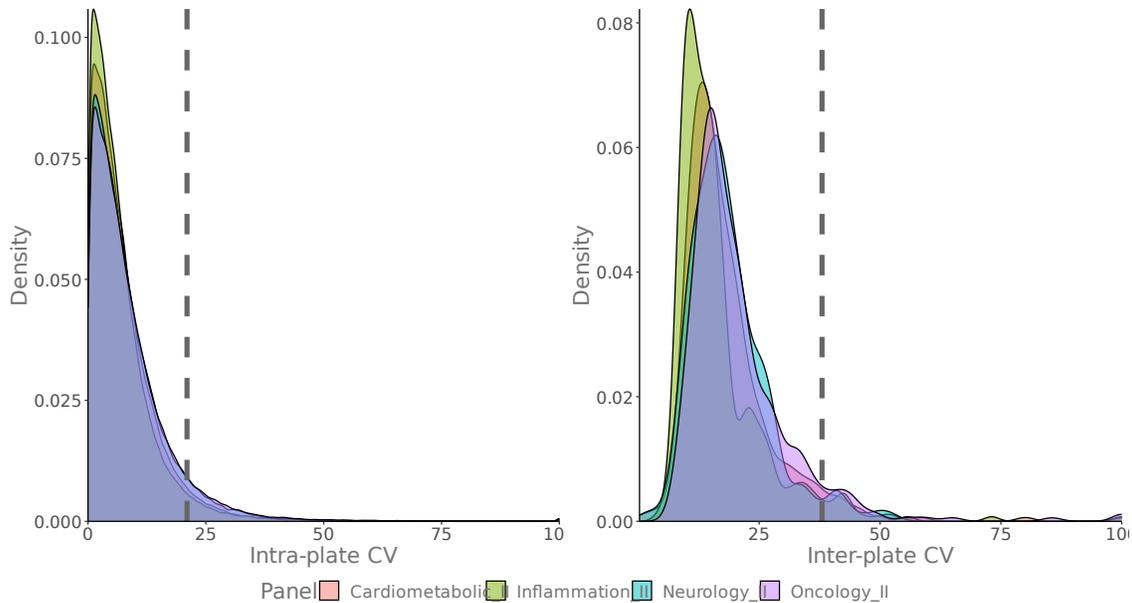


**Figure 2:** Distribution of intra- and inter-plate CVs.

### 2.2.1 Average %CV

| Olink Panel | Intra-plate %CV | Inter-plate %CV |
|---|---|---|
| Explore 384 Cardiometabolic_II | 7.92 | 18.62 |
| Explore 384 Inflammation_II | 7.42 | 16.75 |
| Explore 384 Neurology_II | 8.16 | 19.07 |
| Explore 384 Oncology_II | 8.60 | 20.92 |

### 2.2.2 Intra-assay %CV distribution

| Olink Panel | 0-10% | 10-20% | 20-30% | >30% | NA |
|---|---|---|---|---|---|
| Explore 384 Cardiometabolic_II | 273 | 58 | 5 | 2 | 29 |
| Explore 384 Inflammation_II | 284 | 61 | 2 | 1 | 20 |
| Explore 384 Neurology_II | 256 | 79 | 3 | 0 | 28 |
| Explore 384 Oncology_II | 242 | 95 | 5 | 0 | 26 |

## 2.2.3 Inter-assay %CV distribution

| Olink Panel | 0-15% | 15-30% | 30-40% | >40% | NA |
|---|---|---|---|---|---|
| Explore 384 Cardiometabolic_II | 158 | 160 | 25 | 14 | 10 |
| Explore 384 Inflammation_II | 207 | 122 | 18 | 12 | 9 |
| Explore 384 Neurology_II | 123 | 201 | 15 | 14 | 13 |
| Explore 384 Oncology_II | 118 | 194 | 31 | 19 | 6 |

## 2.2.4 Comparison of % CV
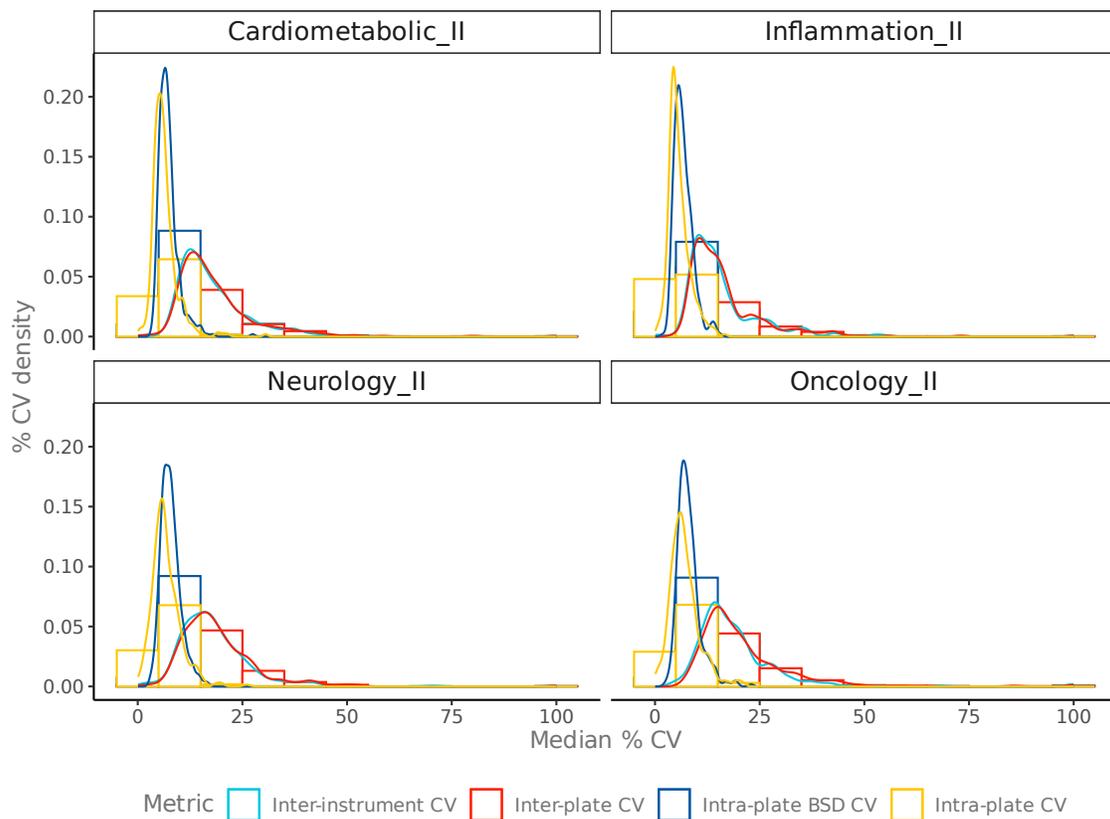


**Figure 3:** Barplot and density plot of the distribution of % CV computed from blind-spiked duplicates, inter- and intra plate % CV computed from control samples and inter-instrument % CV computed from control samples.
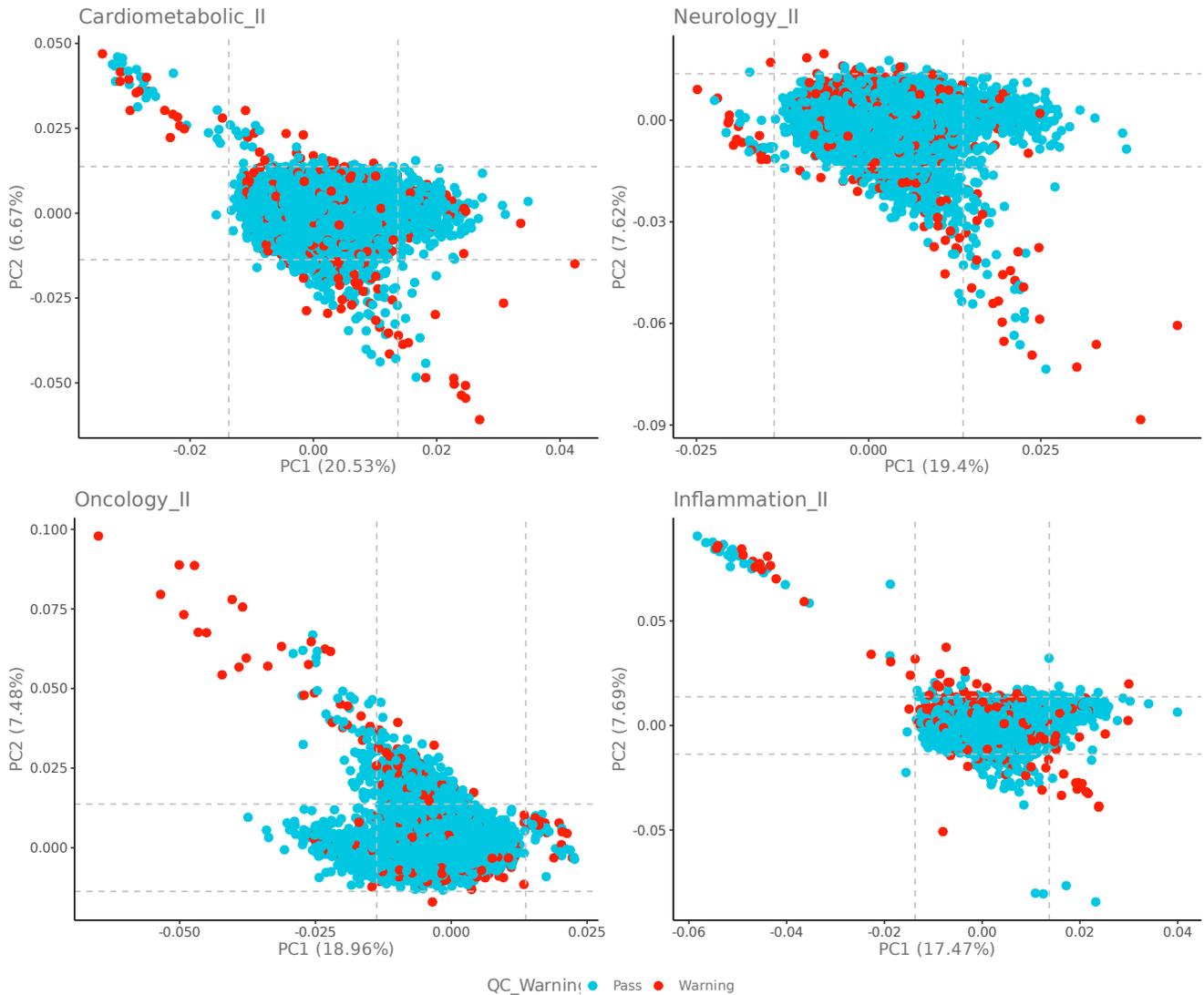
# 3. Protein detection results

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

| Panel | Detected proteins | Detected proteins (%) |
|---|---|---|
| Cardiometabolic_II | 248 / 367 | 67.57 |
| Inflammation_II | 281 / 368 | 76.36 |
| Neurology_II | 201 / 366 | 54.92 |
| Oncology_II | 190 / 368 | 51.63 |

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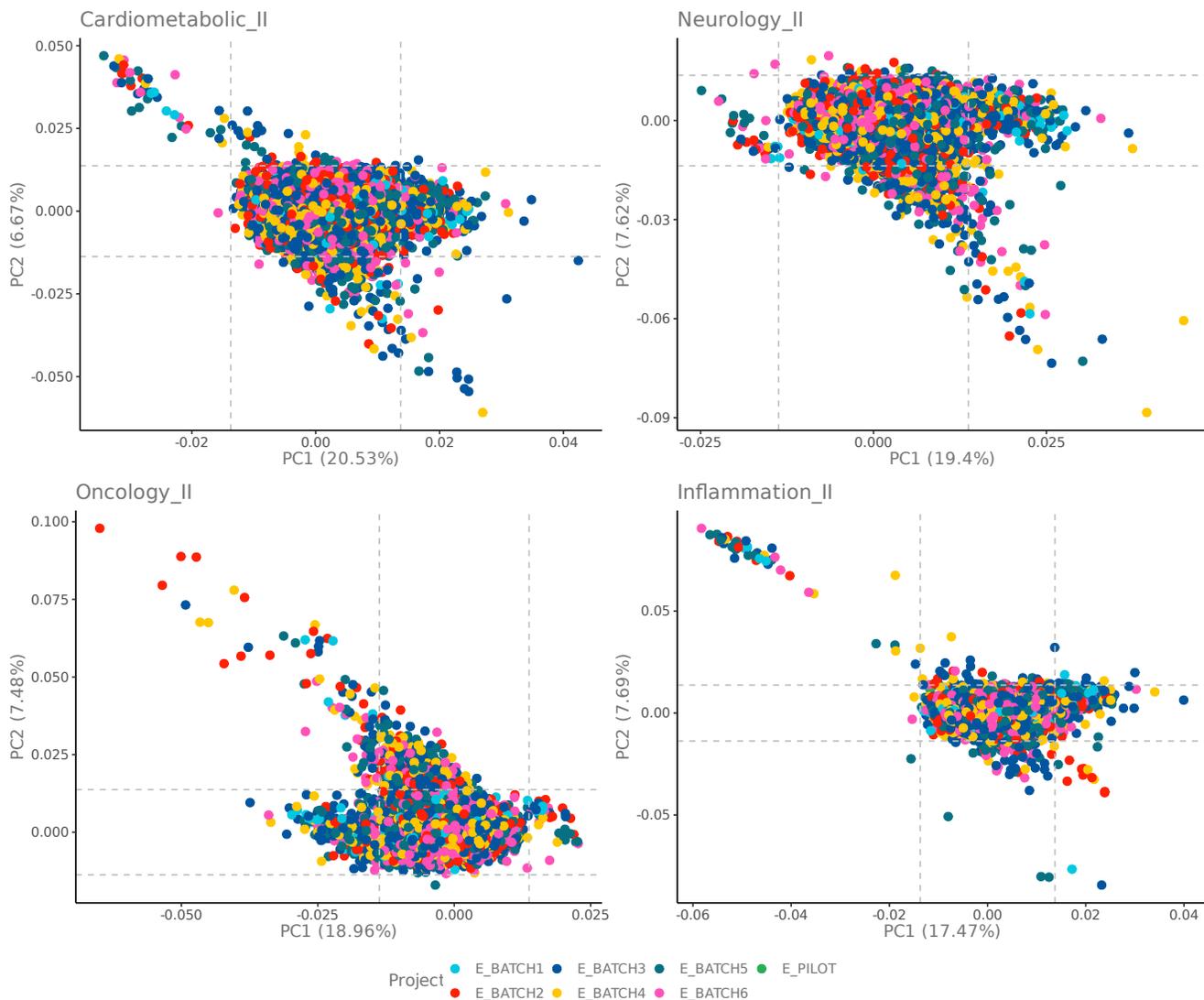
**Figure 5:** Projection along the first two principal components from a PCA performed on the NPX values. Dashed lines indicate ±3 standard deviations (SDs) from the median of each axis. Axes titles indicate the percentage of the variance the principal component explains. Samples were colored by batch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

# 4.2 Deviation from median and interquartile range (IQR)
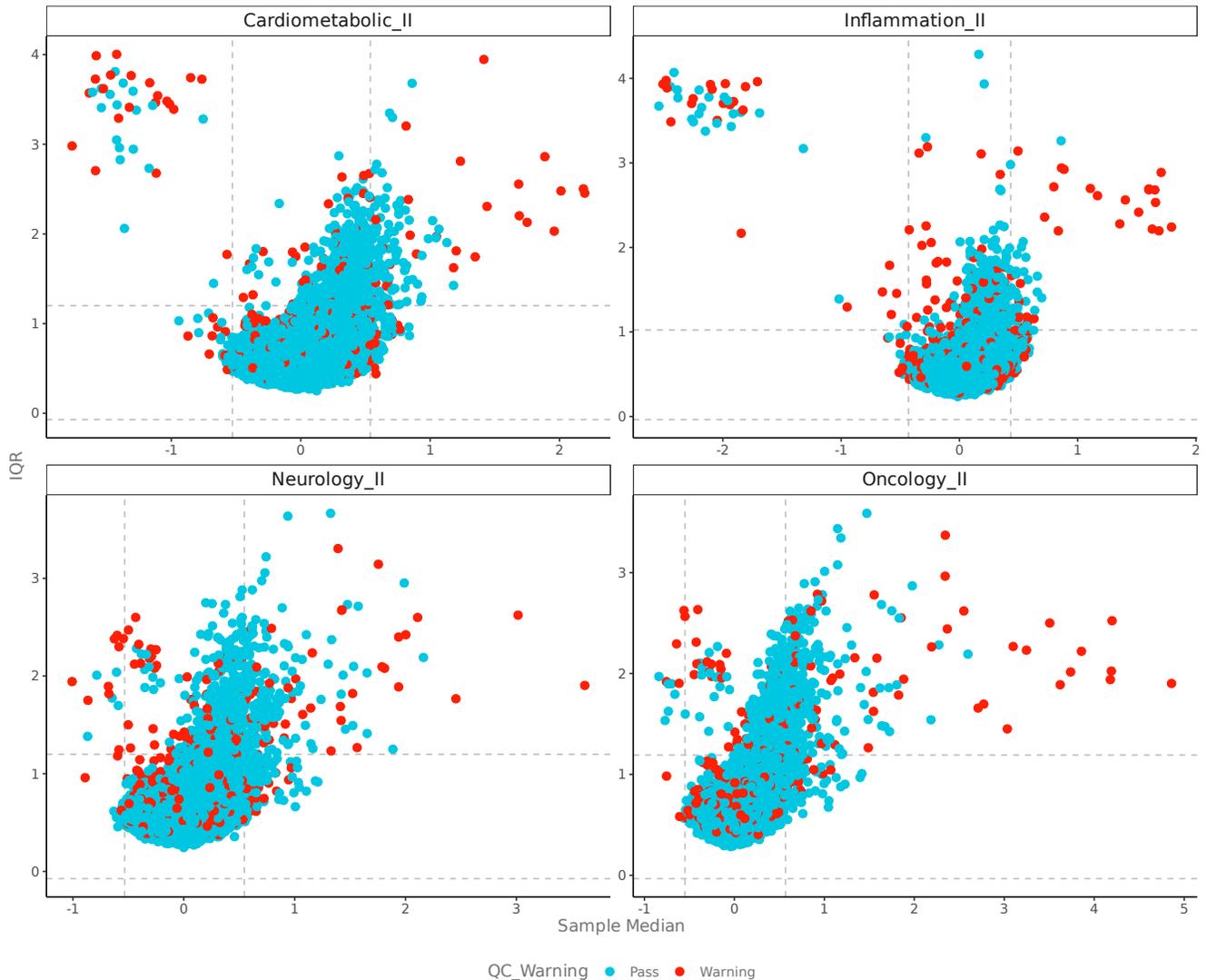


**Figure 6:** IQR vs. sample medians of each sample. Plot facets represent Olink panels. Horizontal dashed lines indicate ±3 standard deviations (SDs) from the average IQR NPX. Vertical dashed lines indicate ±3 SDs from the average sample median NPX. IDs of a subset of potential outliers are shown to avoid overcrowding the plot. Samples were colored by QC warnings status.
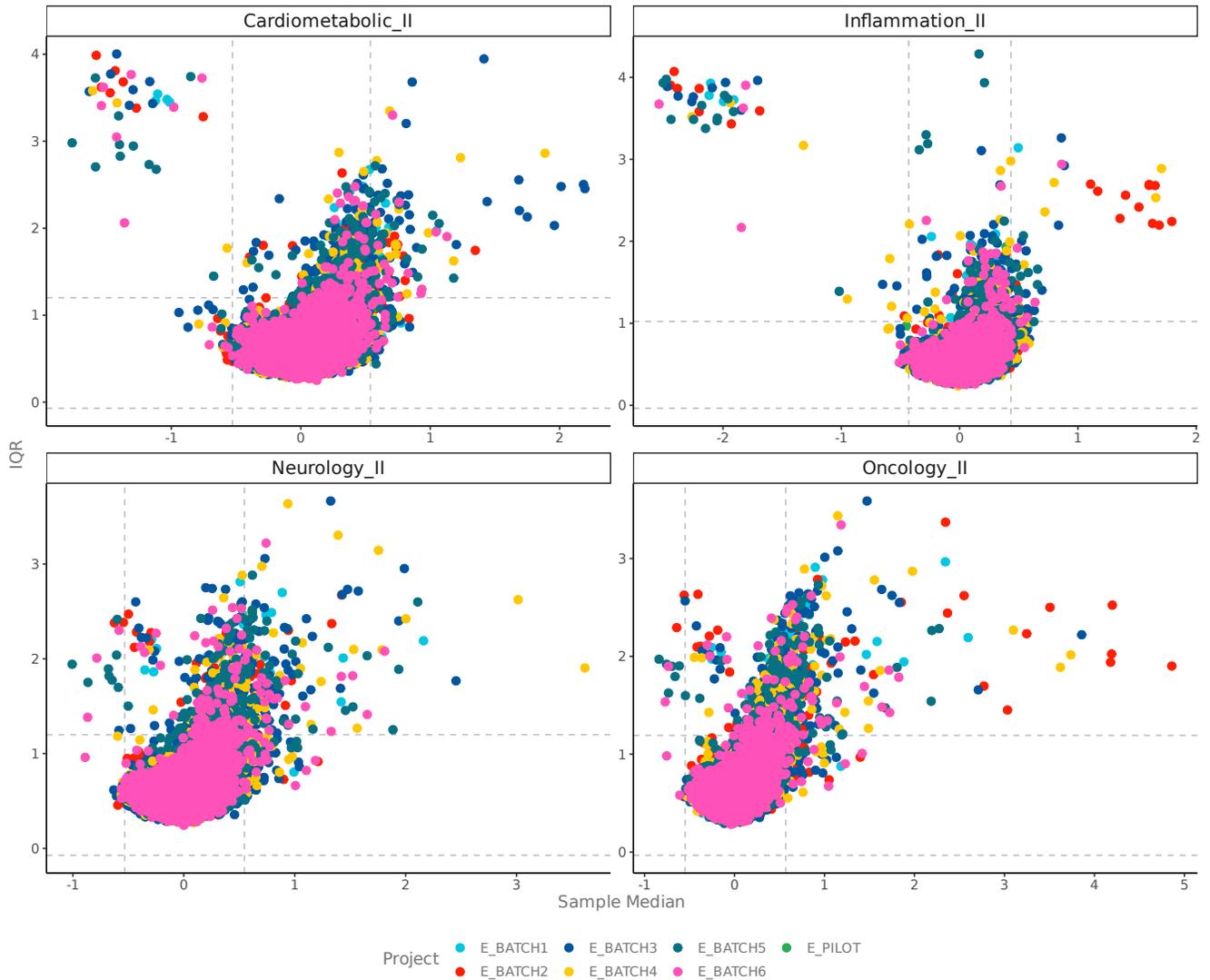
**Figure 7:** IQR vs. sample medians of each sample. Plot facets represent Olink panels. Horizontal dashed lines indicate ±3 standard deviations (SDs) from the average IQR NPX. Vertical dashed lines indicate ±3 SDs from the average sample median NPX. IDs of a subset of potential outliers are shown to 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 (Figures 2 and 3), and assay detectability is over 51% across panels. Over 91% 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, 5, 6 and 7). Samples from batches 2-6 and the pilot were intensity normalized to samples from batch 1. The result of the normalization strategies is illustrated in (Figures 5 and 7) and suggests a successful approach.