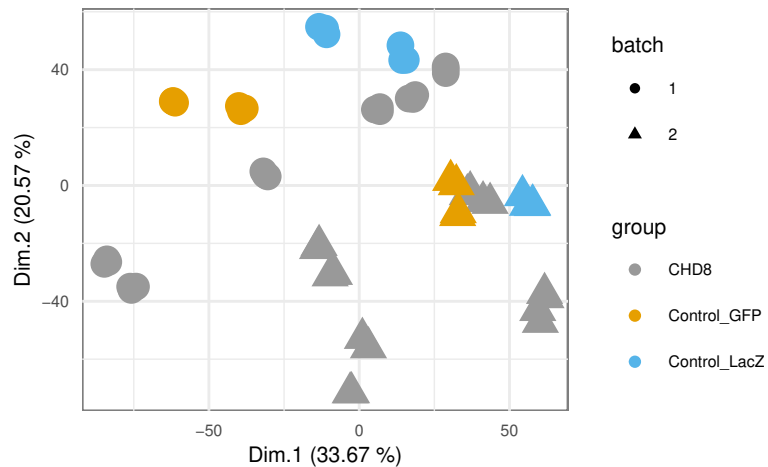


Supplementary material S1

From top left to bottom right: PCA Abundance, shows the uncorrected PCA of the rlog normalized counts quantified by salmon and imported to a `deseq2` object, next to it in the top right panel a bar plot shows the Low-Quality probability P_{low} for each sample. The PCA corrected with batch uses the `acPCA` package to return principal components that were computed with the true batch as a confounding factor. The PCA corrected with P_{low} uses the `acPCA` package likewise, but P_{low} as a confounding factor. To the right of either corrected PCA we see a corrected PCA on the basis of the data without outliers. For the correction with the real batch, we removed outliers that showed up either in the base PCA or in the corrected version. In P_{low} we also removed outliers of base or corrected PCA, but additionally added a threshold for P_{low} depending on the barplot on the top right. The last two panels are the PCA corrected by both batch and P_{low} and the Boxplot of Batch against P_{low} .

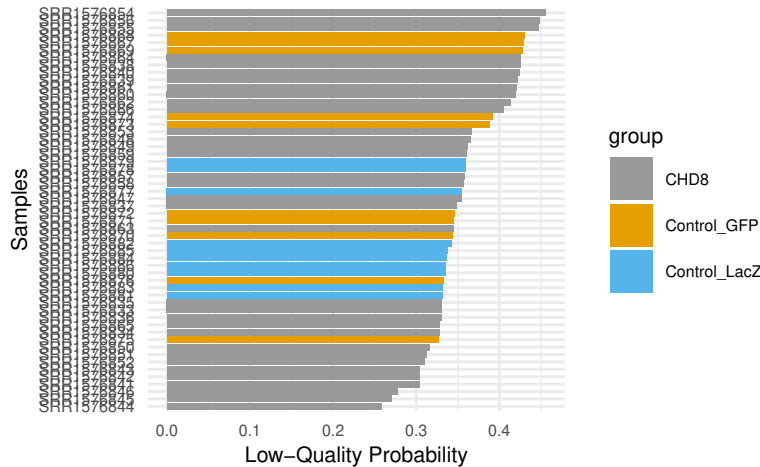
PCA Abundance: GSE61491

DesignBias=0.45, Gamma=0.06, Dunn1=0.03, WbRatio=0.88, DEGs=322



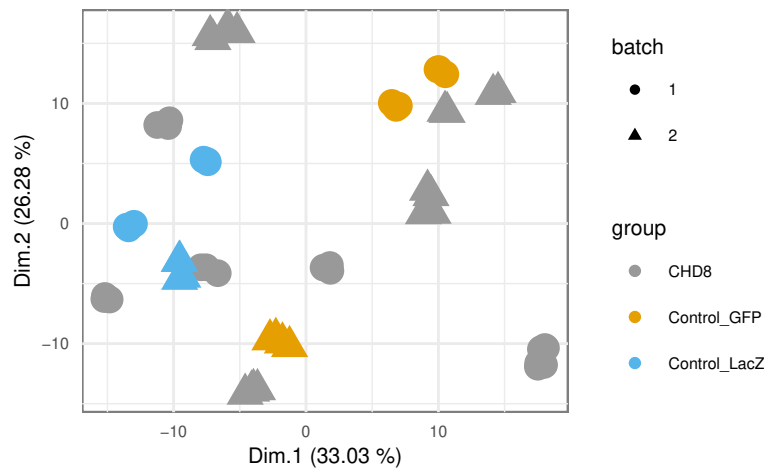
Low Quality Probabilities in GSE61491

Correlation coefficient of P_low vs Group: 0.4474



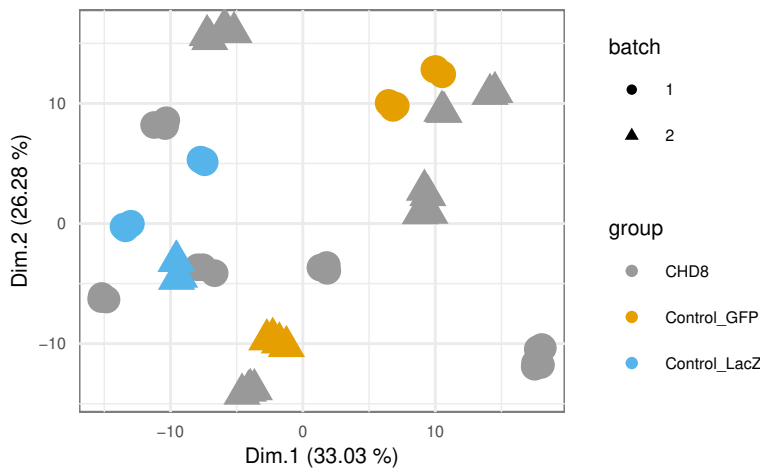
PCA corrected Batch: GSE61491

DesignBias=0.45, Gamma=-0.01, Dunn1=0.04, WbRatio=0.88



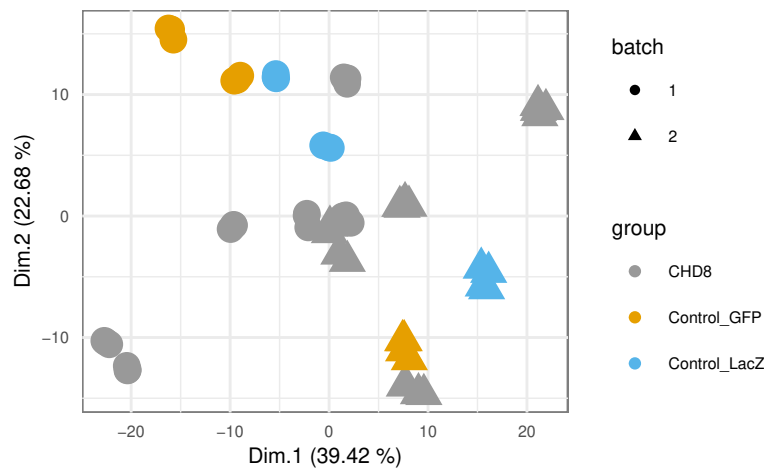
PCA corrected Batch and no Outliers: GSE61491

DesignBias=0.45, Gamma=-0.01, Dunn1=0.04, WbRatio=0.88



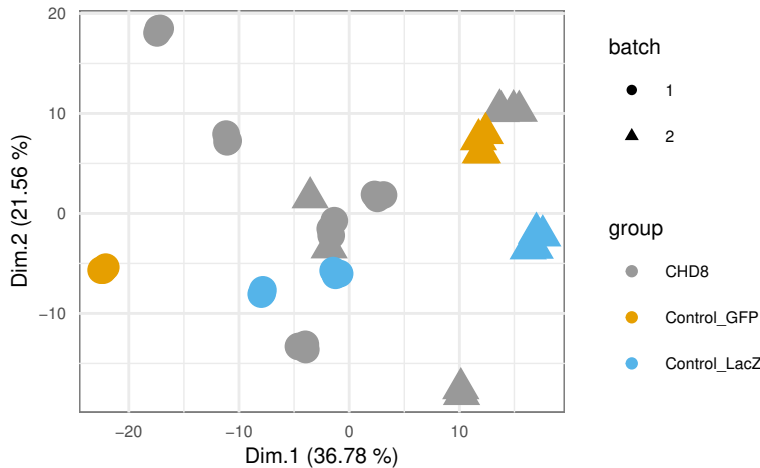
PCA corrected Plow: GSE61491

DesignBias=0.45, Gamma=0.07, Dunn1=0.04, WbRatio=0.9



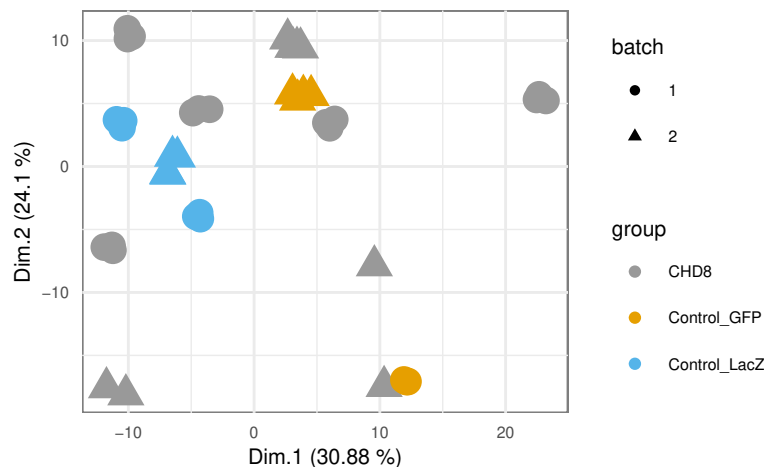
PCA corrected Plow and no outliers: GSE61491

DesignBias=0.45, Gamma=0.08, Dunn1=0.05, WbRatio=0.9

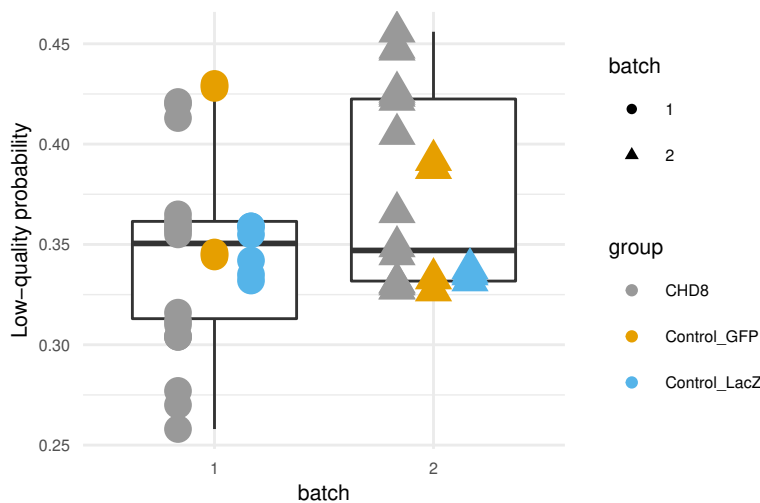


PCA corrected Batch and Plow and no outliers: GSE61491

DesignBias=0.45, Gamma=0.03, Dunn1=0.04, WbRatio=0.85

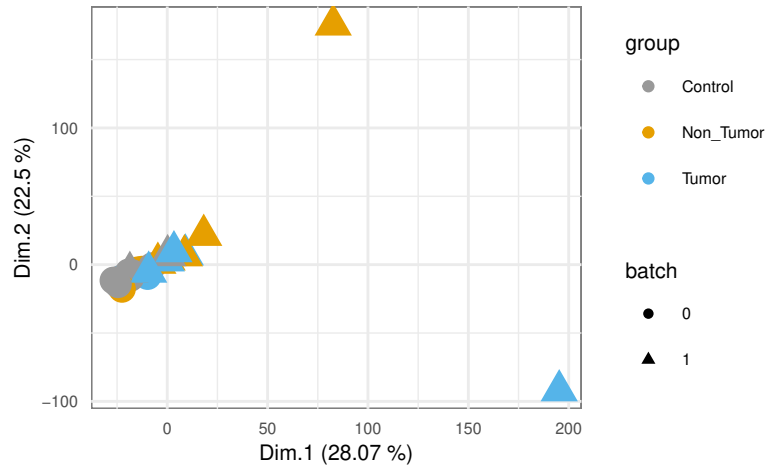


GSE61491: Kruskal-Wallis' p-value: 2.33e-01



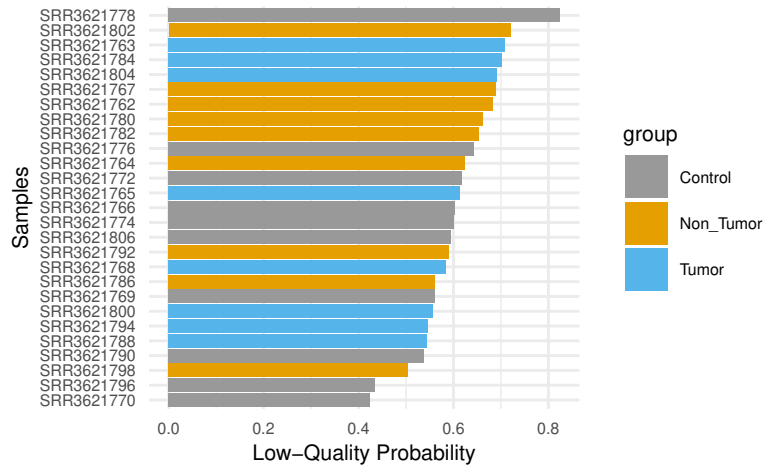
PCA Abundance: GSE82177

DesignBias=0.49, Gamma=0.03, Dunn1=0, WbRatio=0.95, DEGs=5



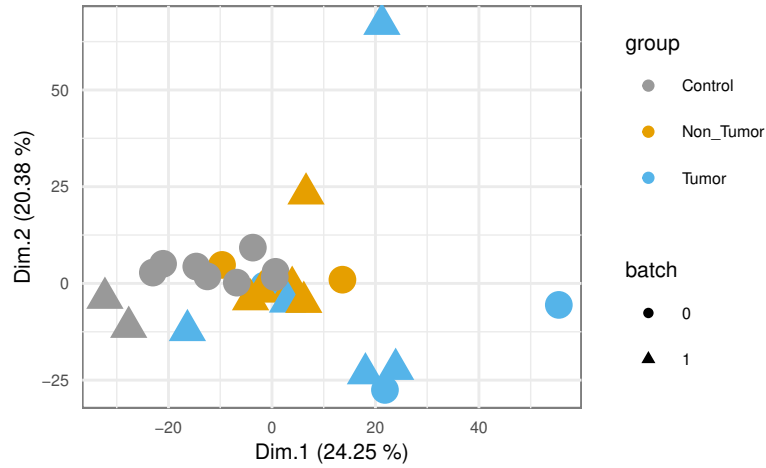
Low Quality Probabilities in GSE82177

Correlation coefficient of P_low vs Group: 0.4945



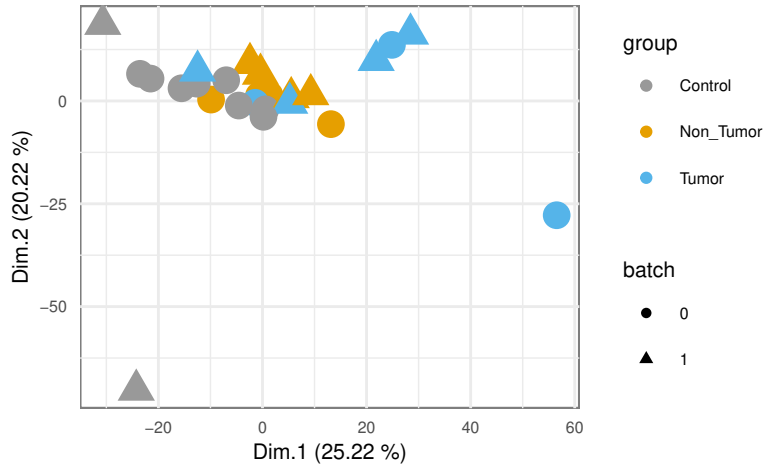
PCA corrected Batch: GSE82177

DesignBias=0.49, Gamma=0.18, Dunn1=0.01, WbRatio=0.76



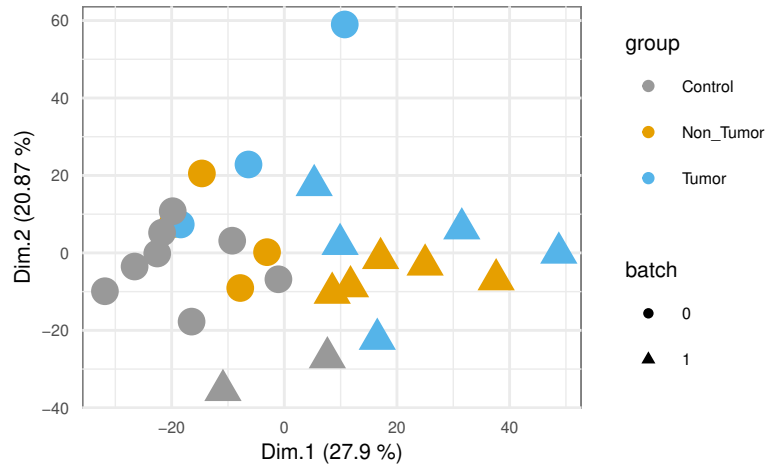
PCA corrected Batch and no Outliers: GSE82177

DesignBias=0.49, Gamma=0.11, Dunn1=0.02, WbRatio=0.81



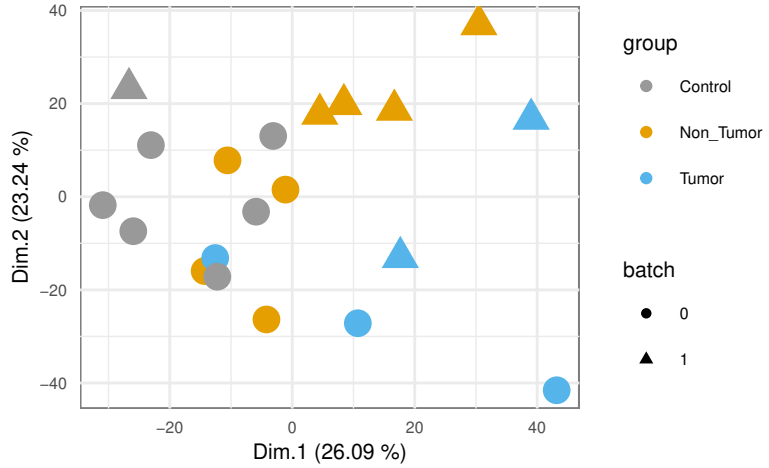
PCA corrected Plow: GSE82177

DesignBias=0.49, Gamma=0.18, Dunn1=0.02, WbRatio=0.83



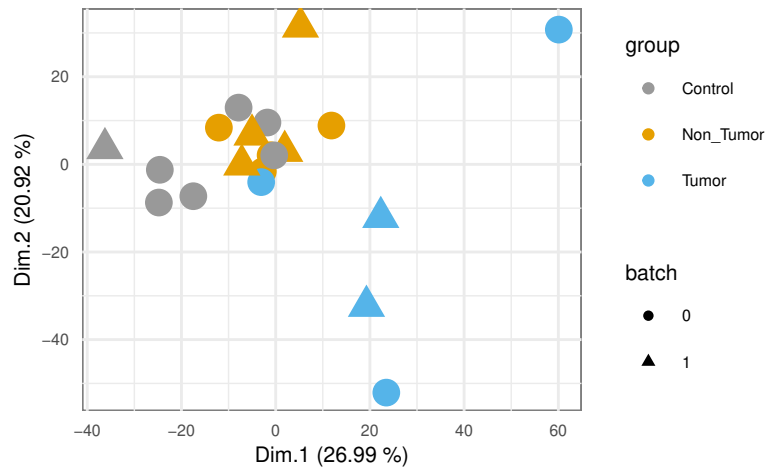
PCA corrected Plow and no outliers: GSE82177

DesignBias=0.49, Gamma=0.23, Dunn1=0.03, WbRatio=0.79

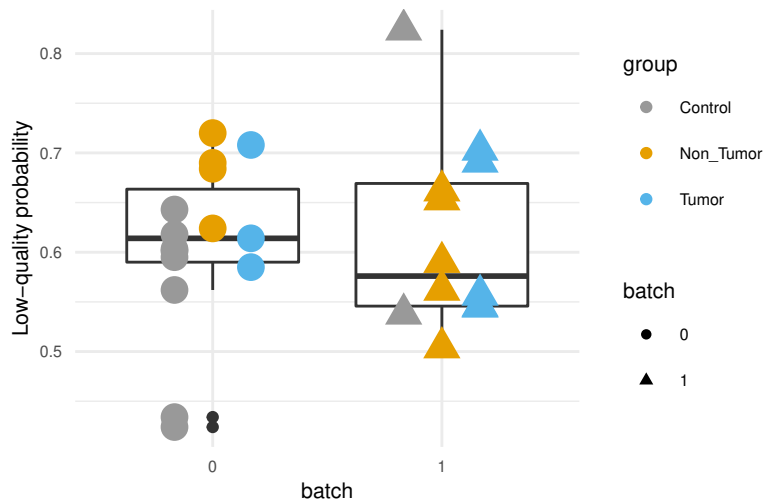


PCA corrected Batch and Plow and no outliers: GSE82177

DesignBias=0.49, Gamma=0.28, Dunn1=0.01, WbRatio=0.7

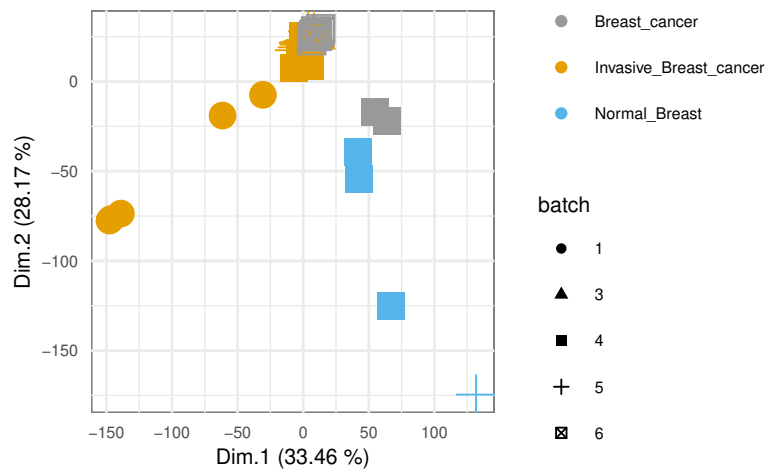


GSE82177: Kruskal-Wallis' p-value: 6.08e-01



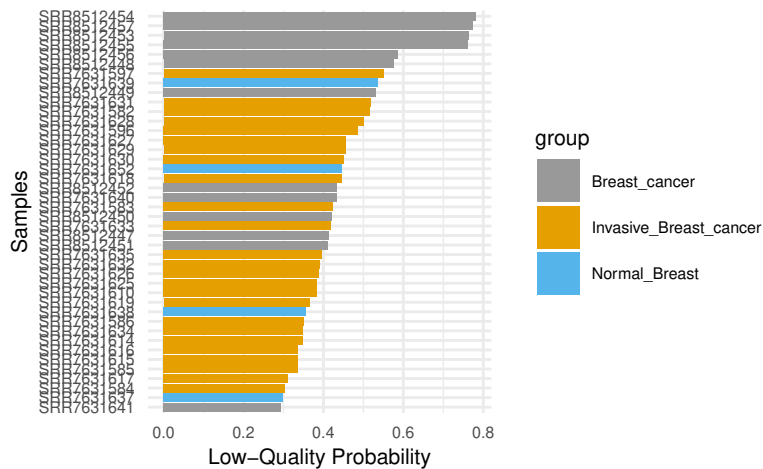
PCA Abundance: GSE117970

DesignBias=0.62, Gamma=0.21, Dunn1=0, WbRatio=0.6, DEGs=1844



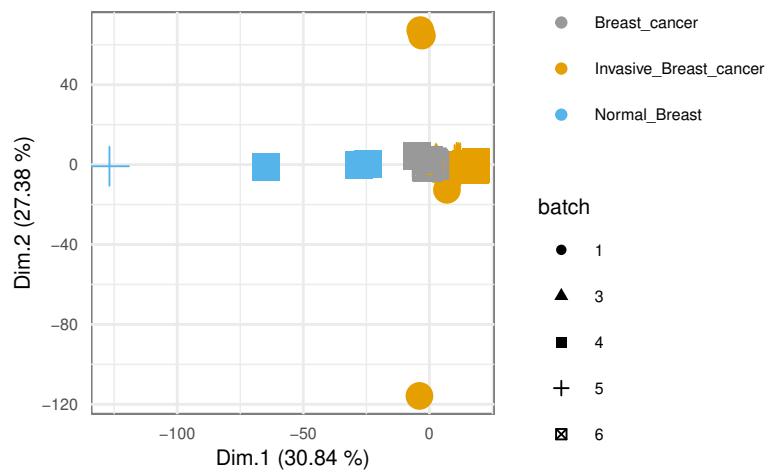
Low Quality Probabilities in GSE117970

Correlation coefficient of P_low vs Group: 0.6196



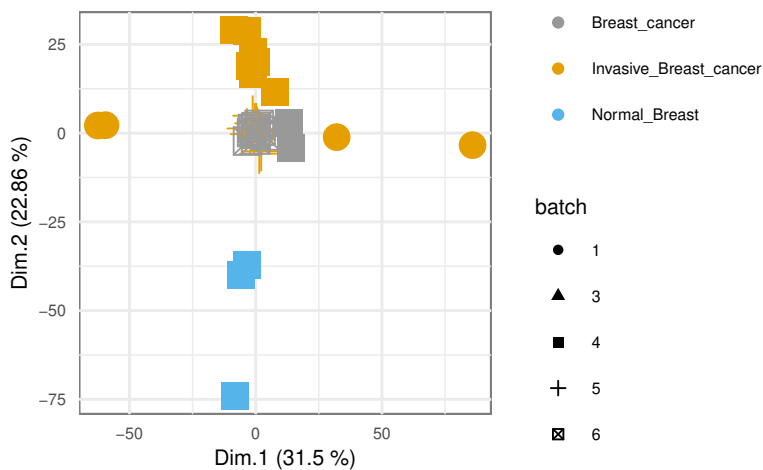
PCA corrected Batch: GSE117970

DesignBias=0.62, Gamma=0.21, Dunn1=0.01, WbRatio=0.57



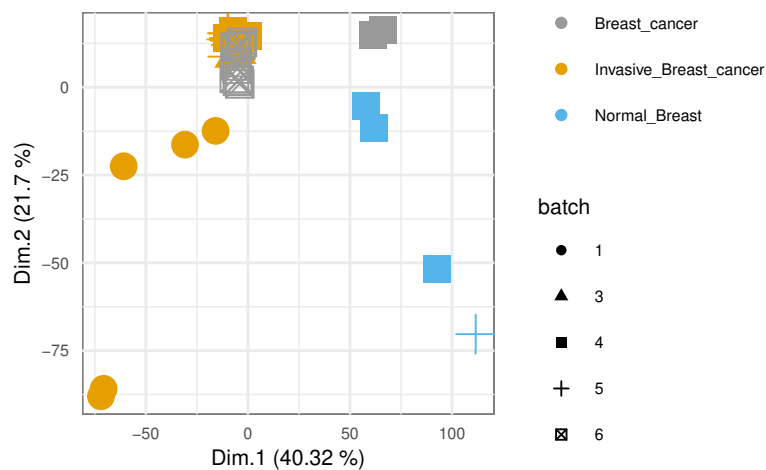
PCA corrected Batch and no Outliers: GSE117970

DesignBias=0.62, Gamma=0.11, Dunn1=0, WbRatio=0.72



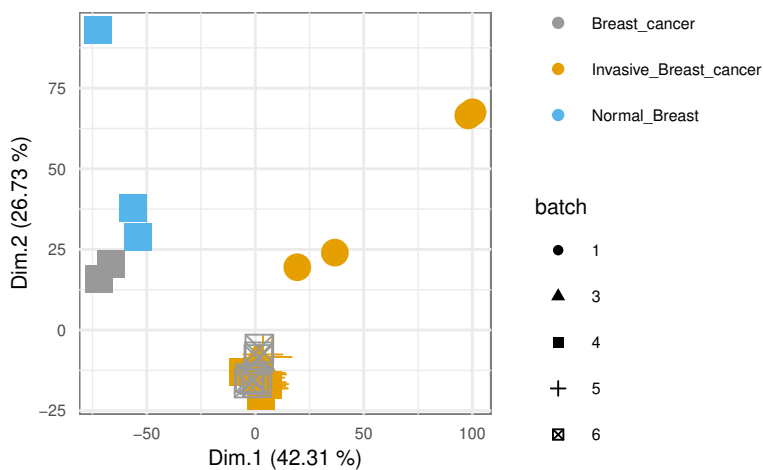
PCA corrected Plow: GSE117970

DesignBias=0.62, Gamma=0.26, Dunn1=0.01, WbRatio=0.55



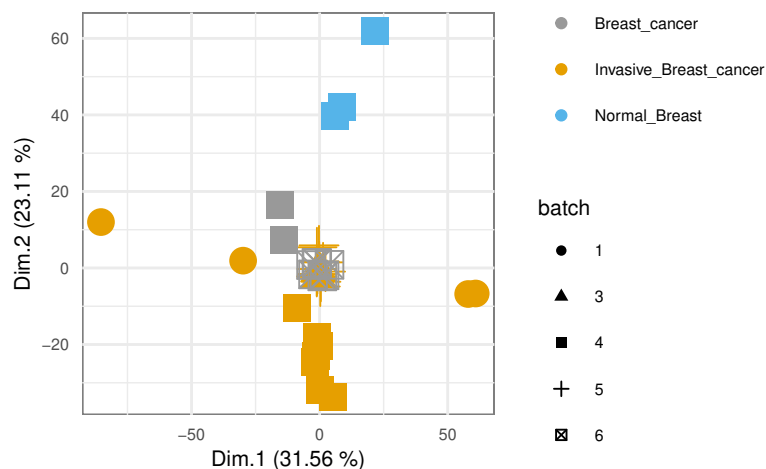
PCA corrected Plow and no outliers: GSE117970

DesignBias=0.62, Gamma=0.25, Dunn1=0.01, WbRatio=0.59

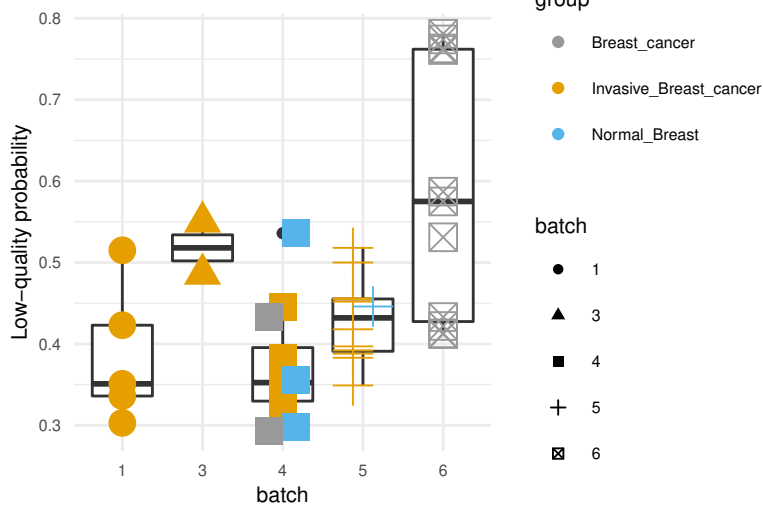


PCA corrected Batch and Plow and no outliers: GSE117970

DesignBias=0.62, Gamma=0.11, Dunn1=0.01, WbRatio=0.71

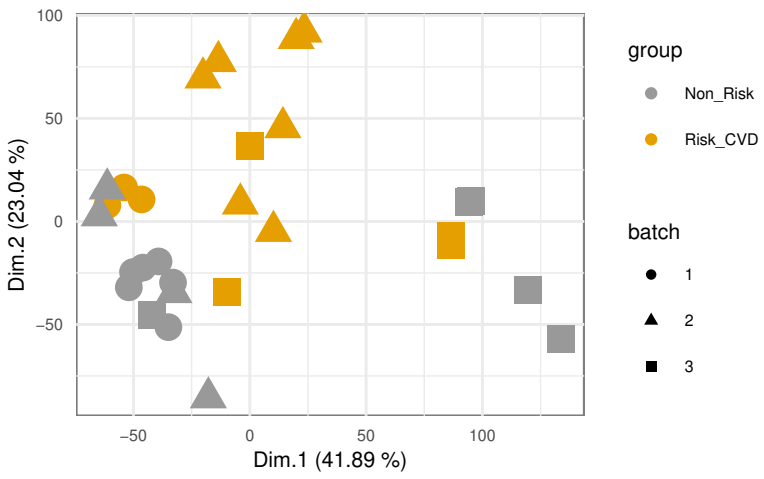


GSE117970: Kruskal-Wallis' p-value: 1.11e-03



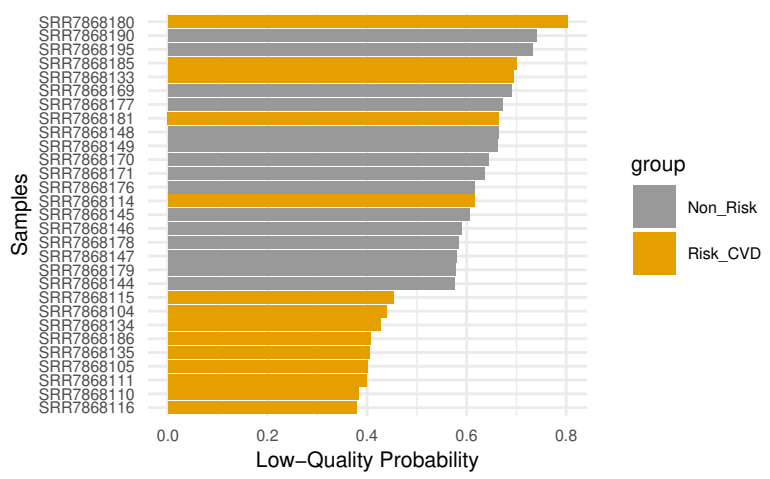
PCA Abundance: GSE120099

DesignBias=0.66, Gamma=0.16, Dunn1=0.03, WbRatio=0.83, DEGs=721



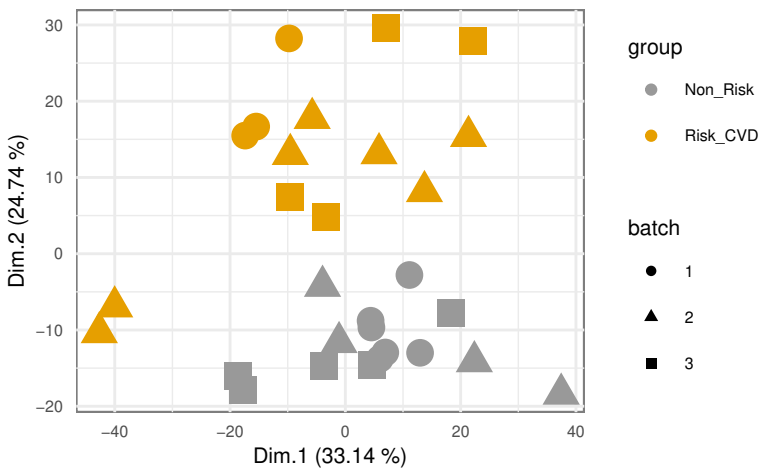
Low Quality Probabilities in GSE120099

Correlation coefficient of P_low vs Group: 0.6562



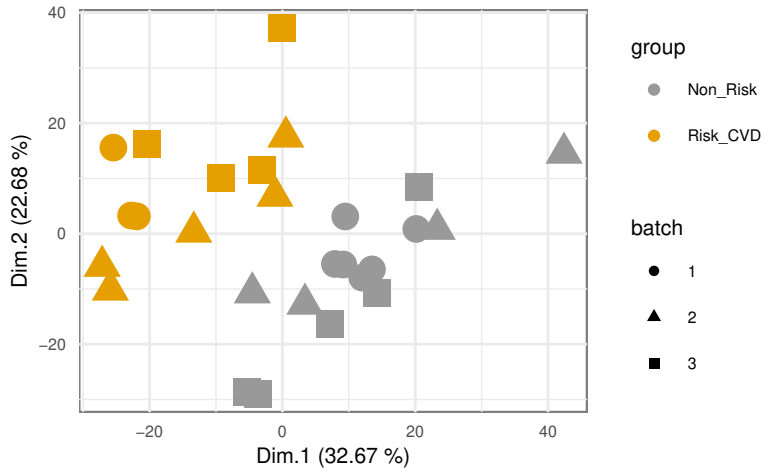
PCA corrected Batch: GSE120099

DesignBias=0.66, Gamma=0.45, Dunn1=0.12, WbRatio=0.63



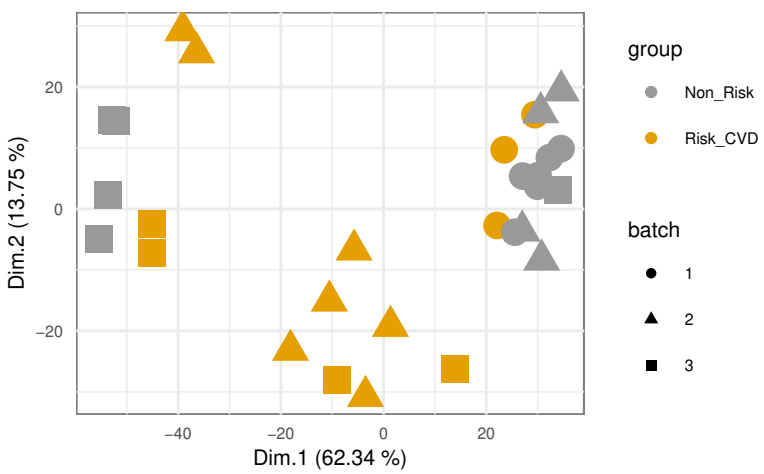
PCA corrected Batch and no Outliers: GSE120099

DesignBias=0.66, Gamma=0.54, Dunn1=0.17, WbRatio=0.57



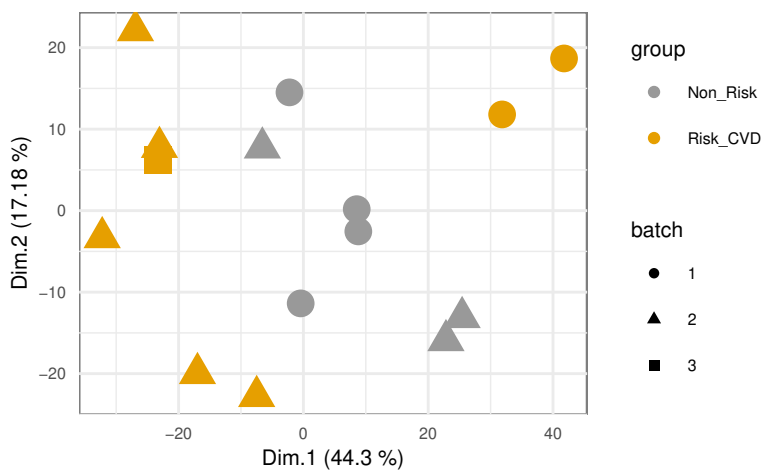
PCA corrected Plow: GSE120099

DesignBias=0.66, Gamma=0.12, Dunn1=0.01, WbRatio=0.87



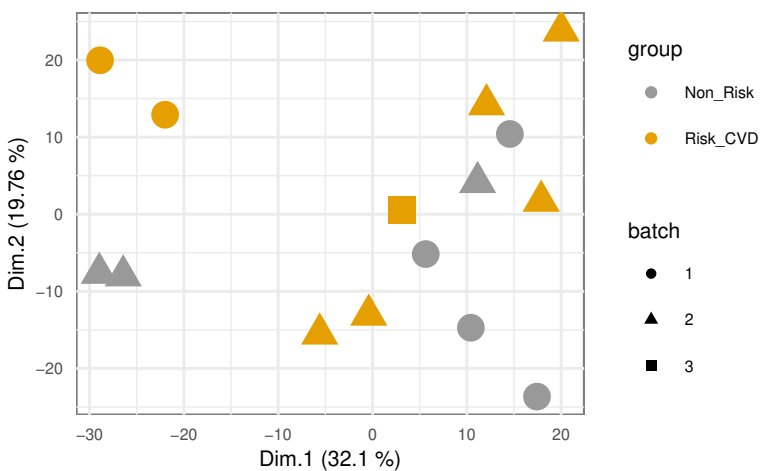
PCA corrected Plow and no outliers: GSE120099

DesignBias=0.66, Gamma=0.12, Dunn1=0.17, WbRatio=0.87

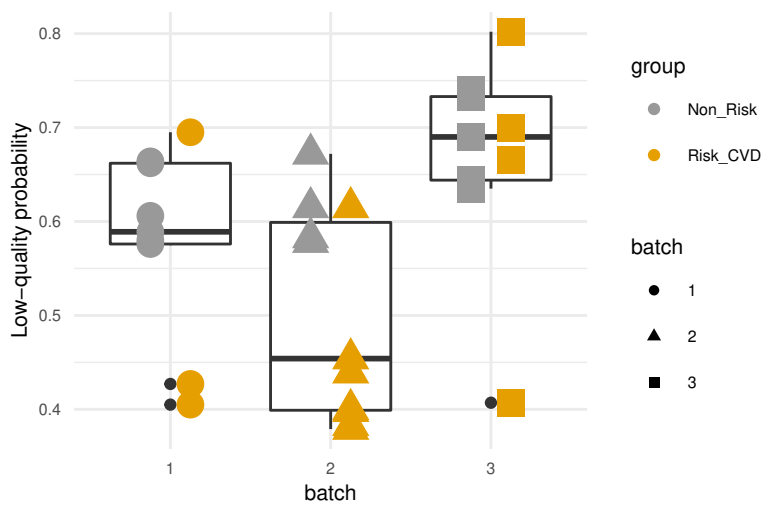


PCA corrected Batch and Plow and no outliers: GSE120099

DesignBias=0.66, Gamma=-0.02, Dunn1=0.09, WbRatio=1.02

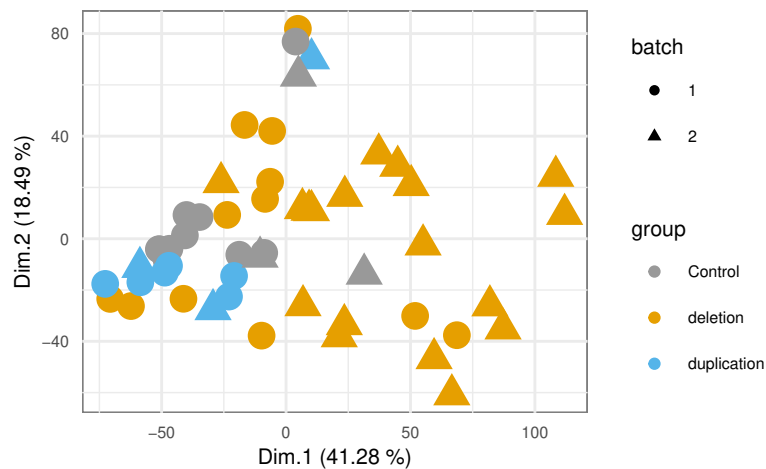


GSE120099: Kruskal-Wallis' p-value: 5.11e-03



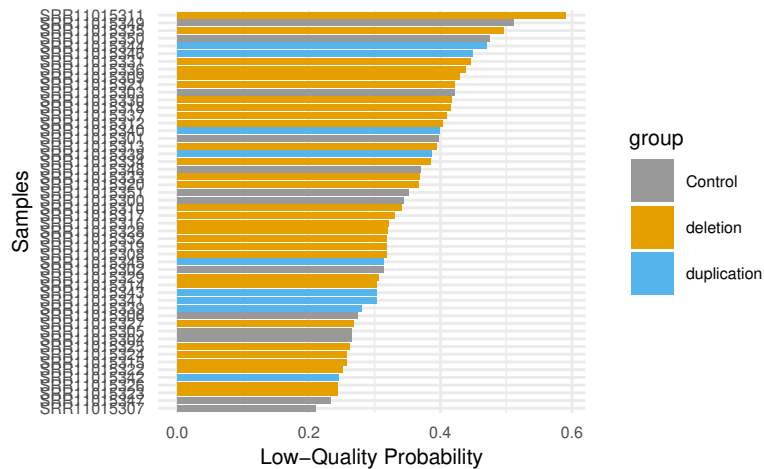
PCA Abundance: GSE144736

DesignBias=0.49, Gamma=0.07, Dunn1=0.01, WbRatio=0.84, DEGs=10



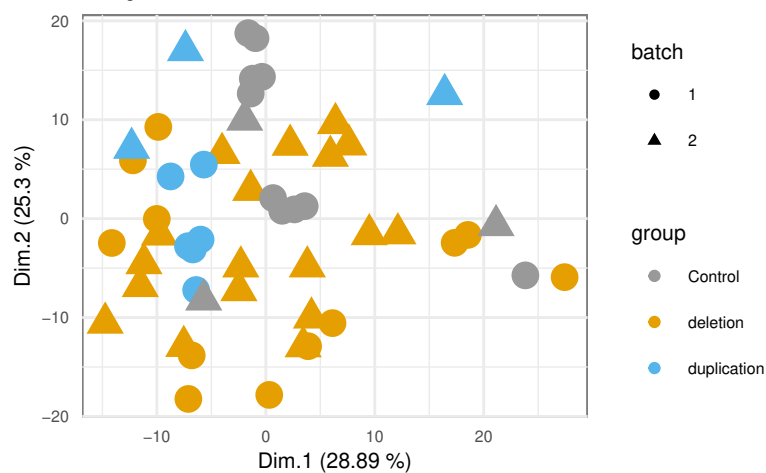
Low Quality Probabilities in GSE144736

Correlation coefficient of P_low vs Group: 0.4945



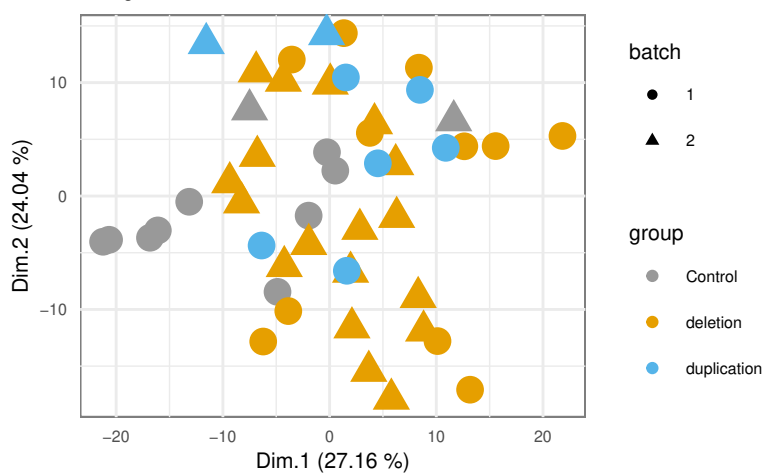
PCA corrected Batch: GSE144736

DesignBias=0.49, Gamma=0.07, Dunn1=0.03, WbRatio=0.9



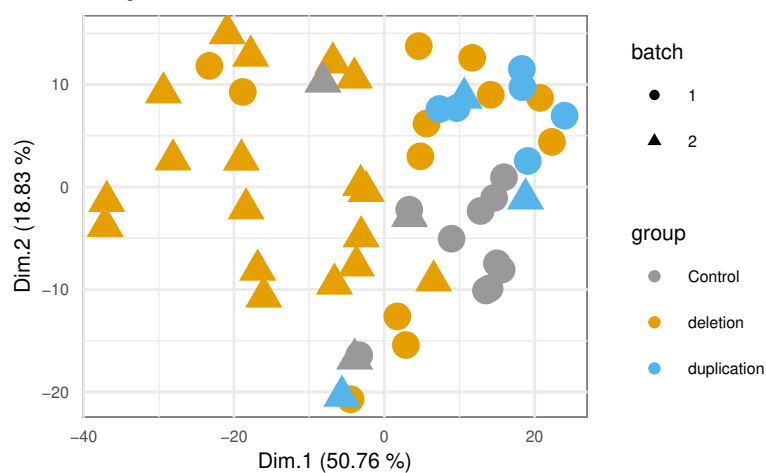
PCA corrected Batch and no Outliers: GSE144736

DesignBias=0.49, Gamma=0.08, Dunn1=0.01, WbRatio=0.9



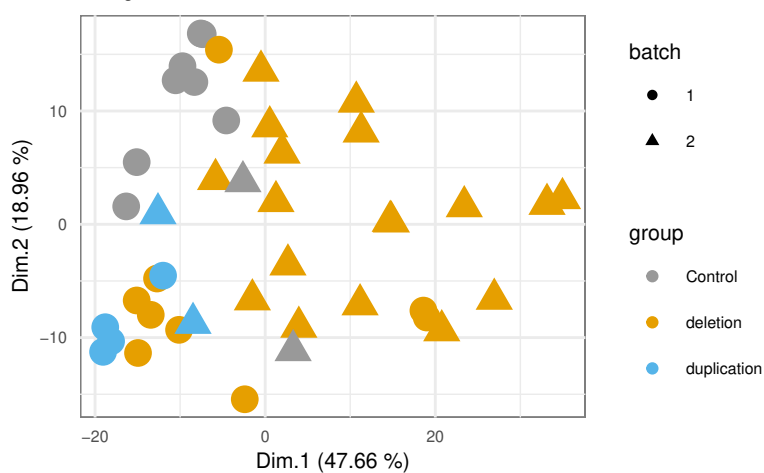
PCA corrected Plow: GSE144736

DesignBias=0.49, Gamma=0.11, Dunn1=0.02, WbRatio=0.8



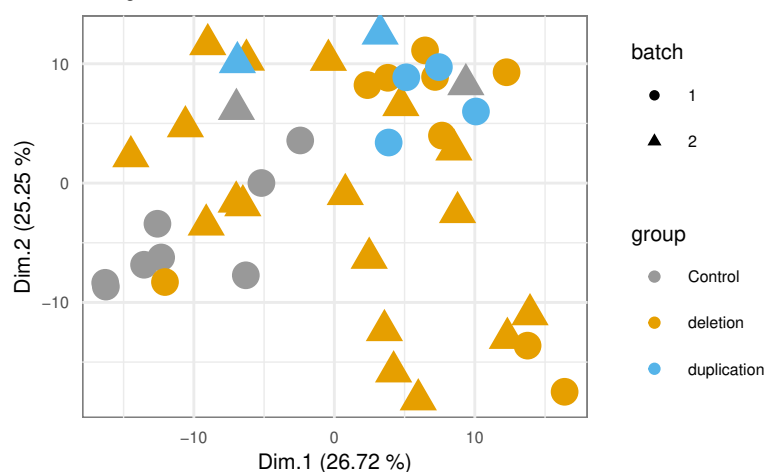
PCA corrected Plow and no outliers: GSE144736

DesignBias=0.49, Gamma=0.16, Dunn1=0.01, WbRatio=0.74

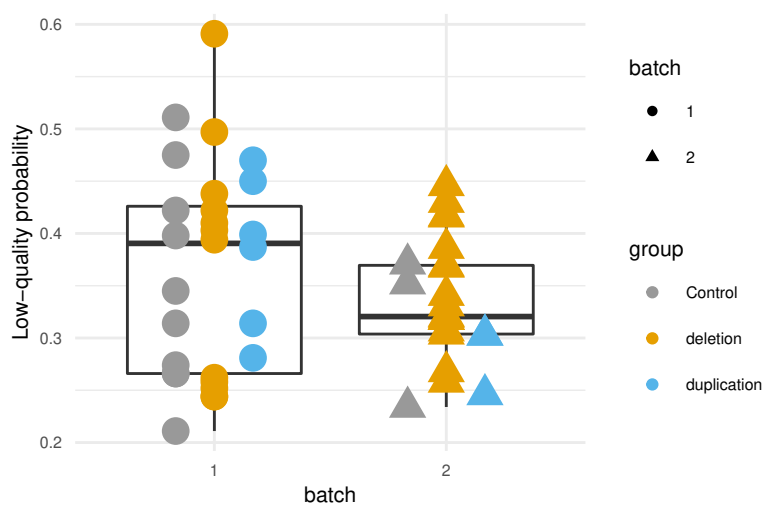


PCA corrected Batch and Plow and no outliers: GSE144736

DesignBias=0.49, Gamma=0.05, Dunn1=0.02, WbRatio=0.87

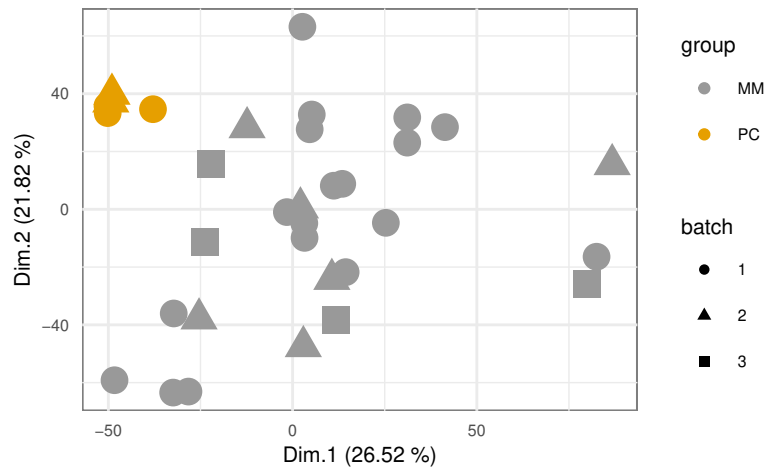


GSE144736: Kruskal-Wallis' p-value: 3.63e-01



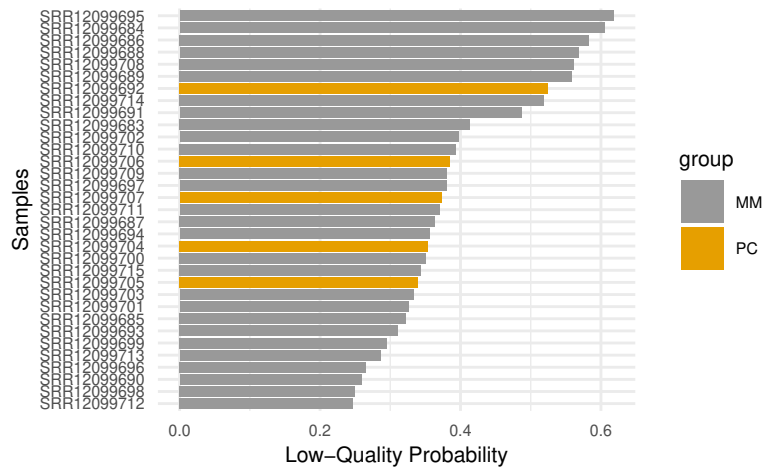
PCA Abundance: GSE153380

DesignBias=0.44, Gamma=0.3, Dunn1=0.16, WbRatio=0.64, DEGs=398



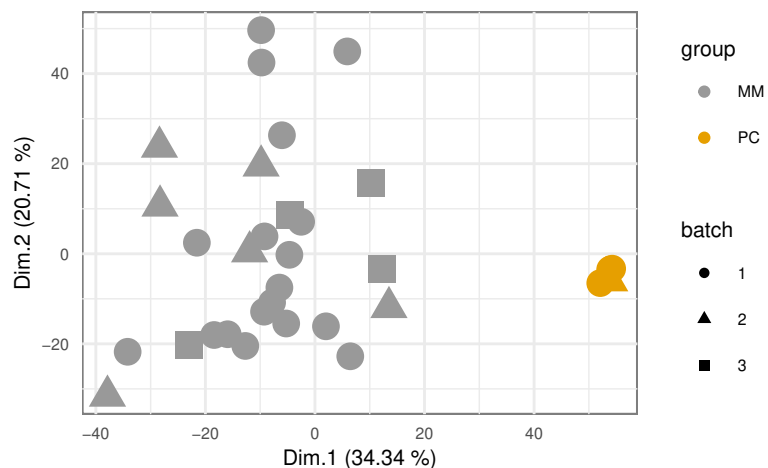
Low Quality Probabilities in GSE153380

Correlation coefficient of P_low vs Group: 0.4385



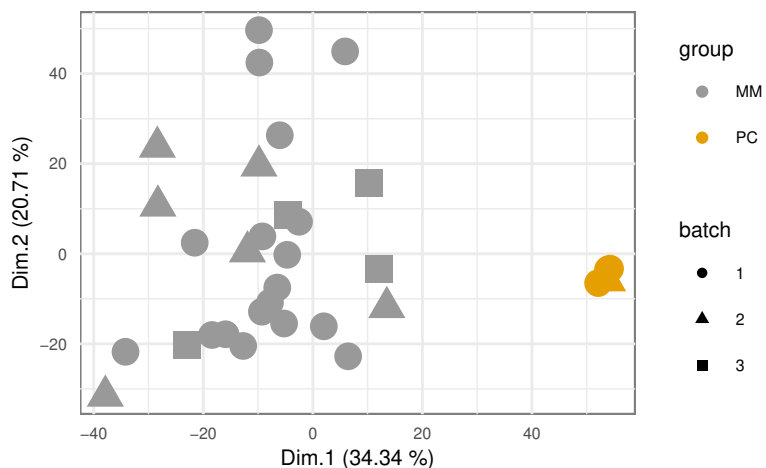
PCA corrected Batch: GSE153380

DesignBias=0.44, Gamma=0.69, Dunn1=0.44, WbRatio=0.41



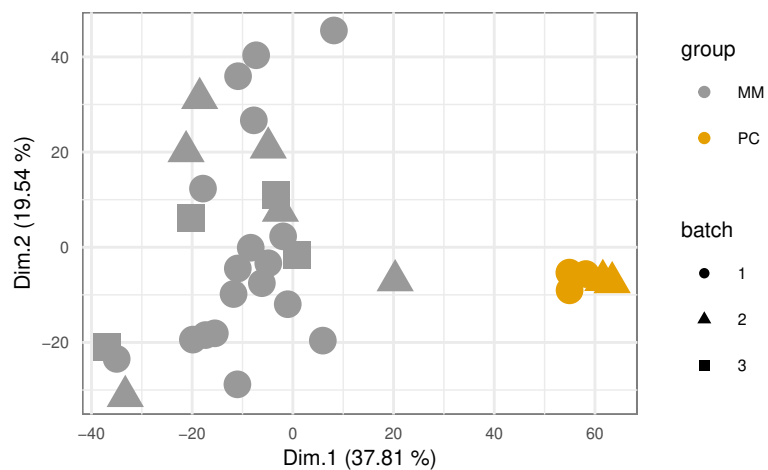
PCA corrected Batch and no Outliers: GSE153380

DesignBias=0.44, Gamma=0.69, Dunn1=0.44, WbRatio=0.41



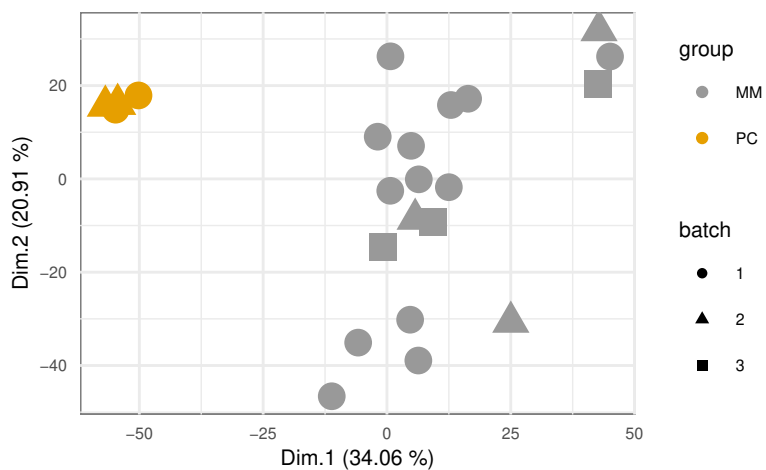
PCA corrected Plow: GSE153380

DesignBias=0.44, Gamma=0.75, Dunn1=0.4, WbRatio=0.37



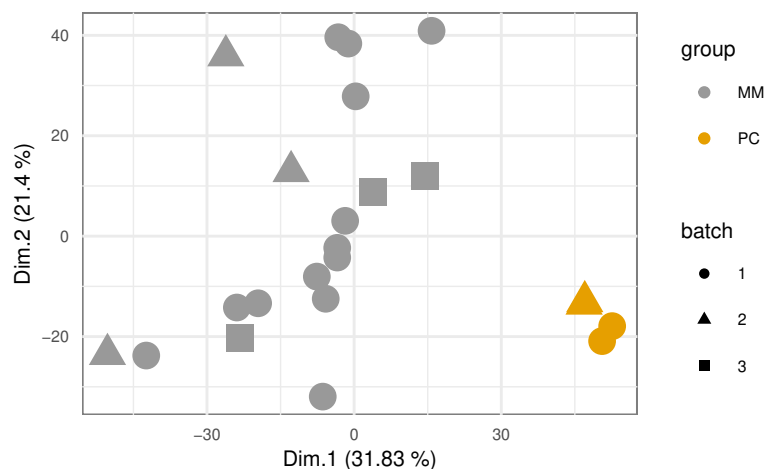
PCA corrected Plow and no outliers: GSE153380

DesignBias=0.44, Gamma=0.67, Dunn1=0.52, WbRatio=0.42

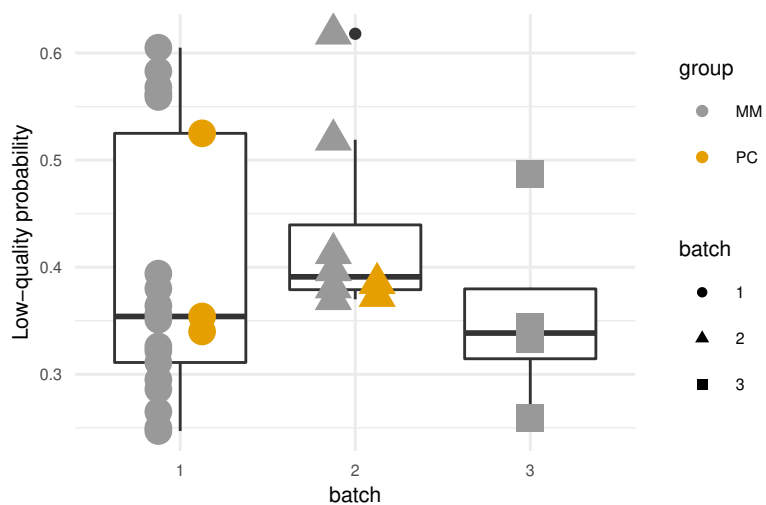


PCA corrected Batch and Plow and no outliers: GSE153380

DesignBias=0.44, Gamma=0.62, Dunn1=0.45, WbRatio=0.46

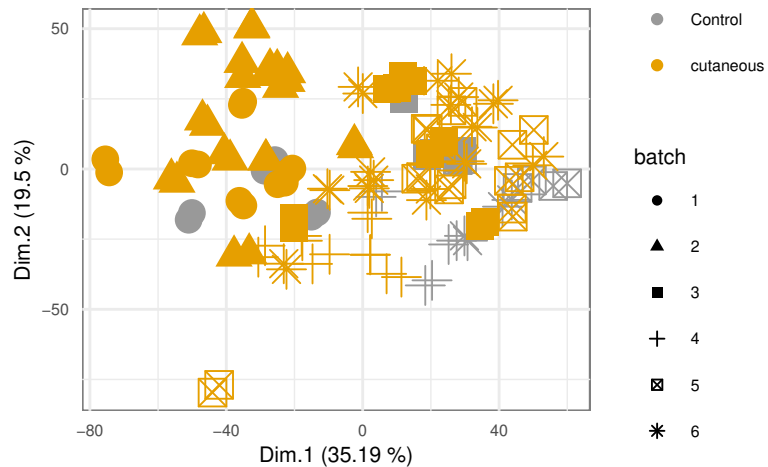


GSE153380: Kruskal-Wallis' p-value: 1.29e-01



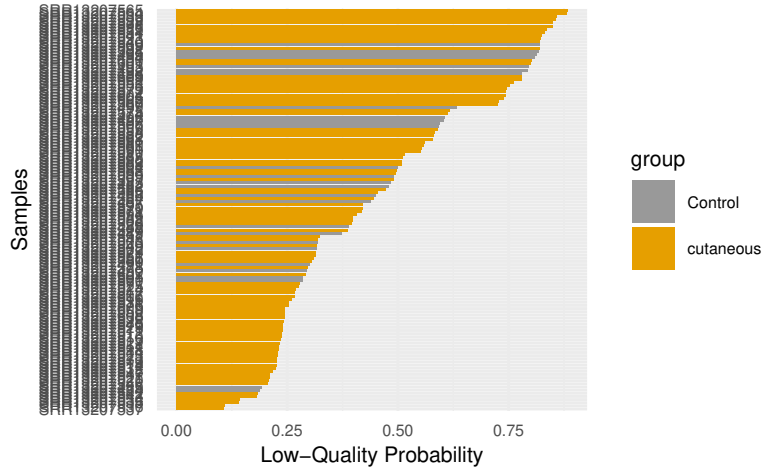
PCA Abundance: GSE162760

DesignBias=0.5, Gamma=0.01, Dunn1=0.01, WbRatio=0.97, PEGs=223



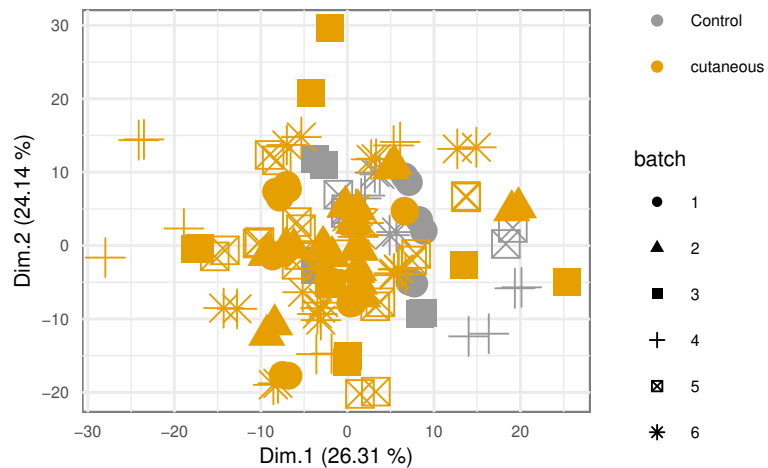
Low Quality Probabilities in GSE162760

Correlation coefficient of P_low vs Group: 0.4964



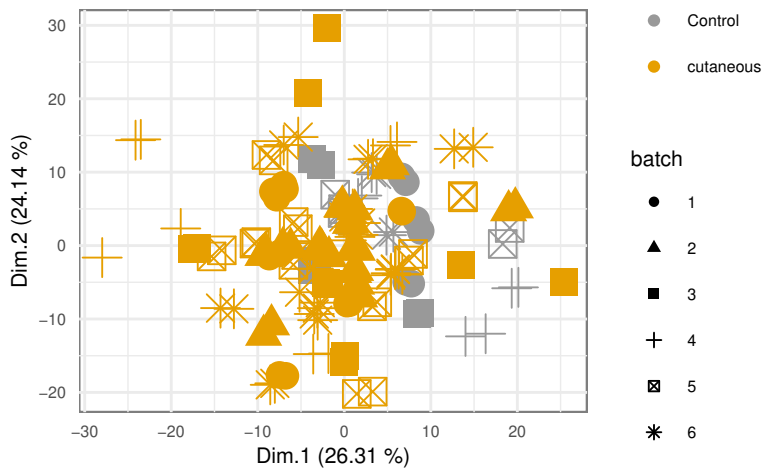
PCA corrected Batch: GSE162760

DesignBias=0.5, Gamma=0.02, Dunn1=0, WbRatio=0.94



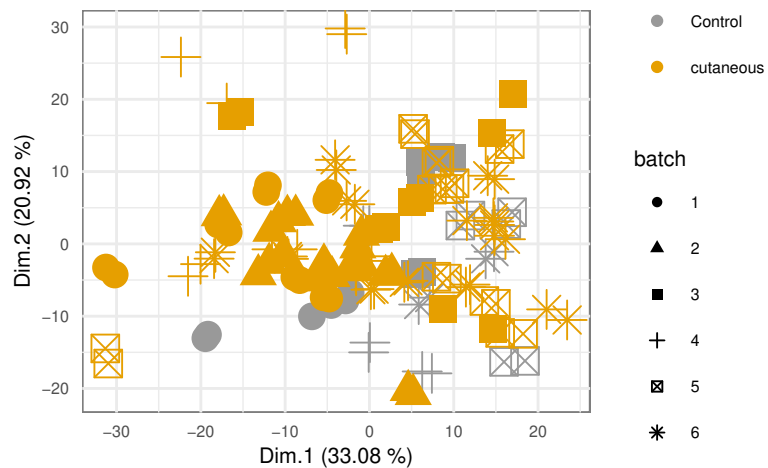
PCA corrected Batch and no Outliers: GSE162760

DesignBias=0.5, Gamma=0.02, Dunn1=0, WbRatio=0.94



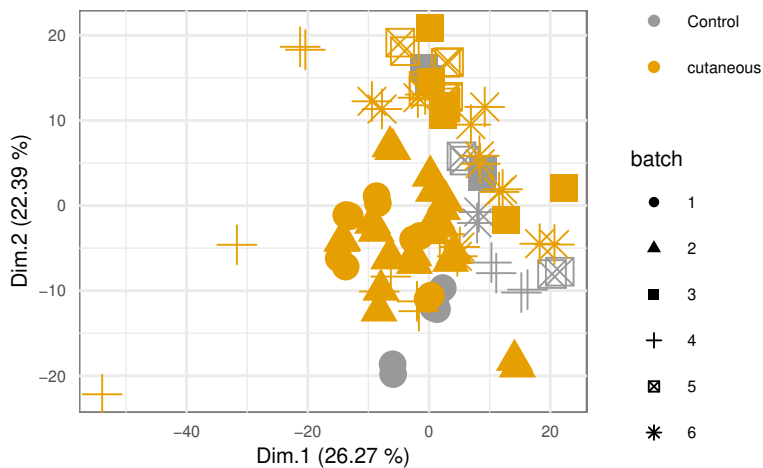
PCA corrected Plow: GSE162760

DesignBias=0.5, Gamma=0.01, Dunn1=0.01, WbRatio=0.97



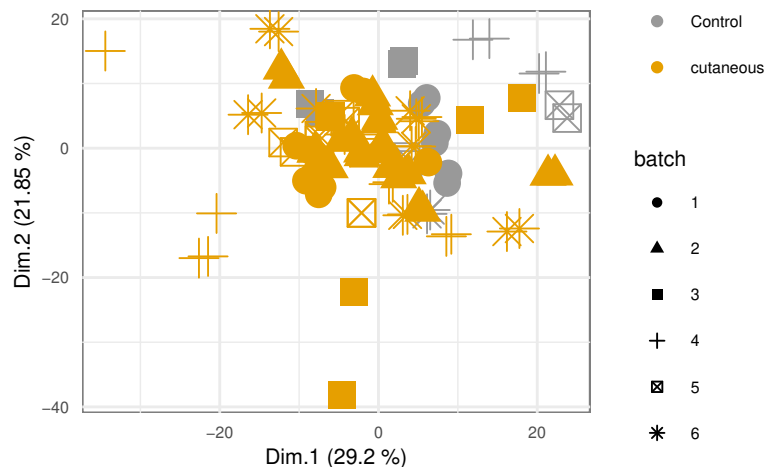
PCA corrected Plow and no outliers: GSE162760

DesignBias=0.5, Gamma=0.05, Dunn1=0, WbRatio=0.92

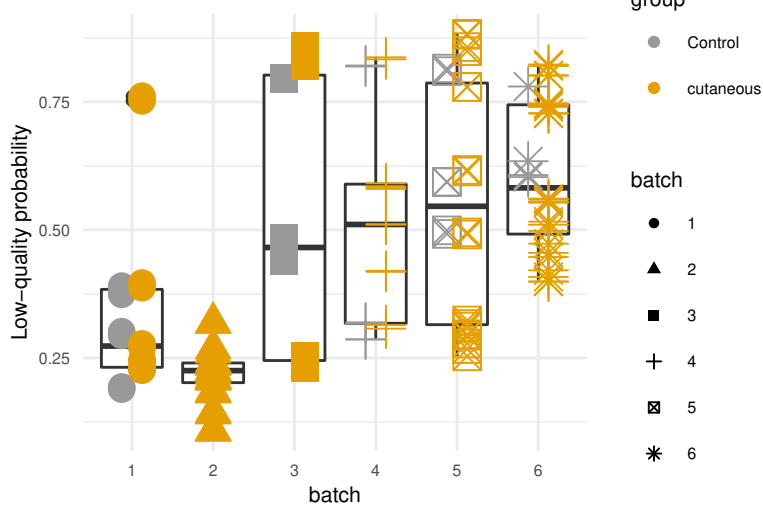


PCA corrected Batch and Plow and no outliers: GSE162760

DesignBias=0.5, Gamma=0.13, Dunn1=0.01, WbRatio=0.84

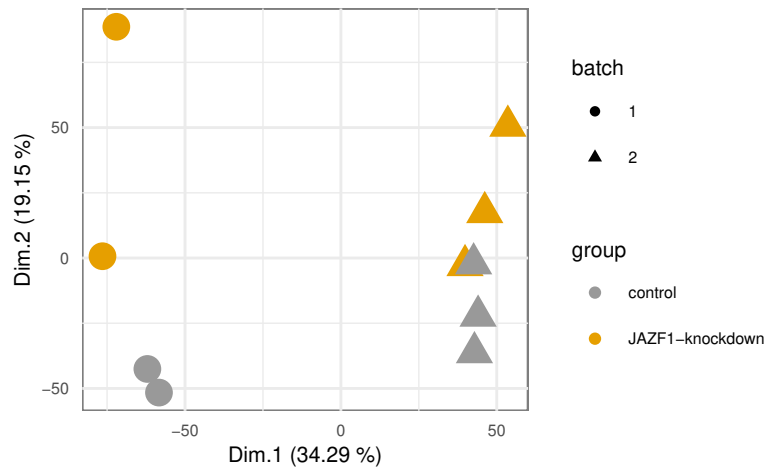


GSE162760: Kruskal-Wallis' p-value: 2.76e-12



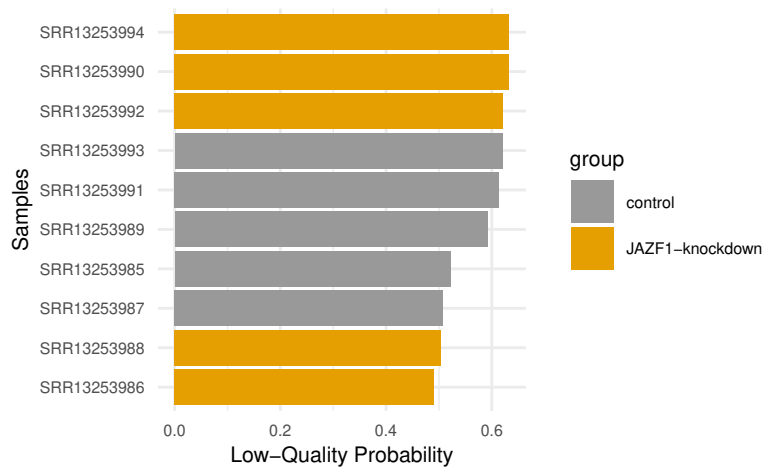
PCA Abundance: GSE163214

DesignBias=0.46, Gamma=0.09, Dunn1=0.02, WbRatio=0.91, DEGs=3



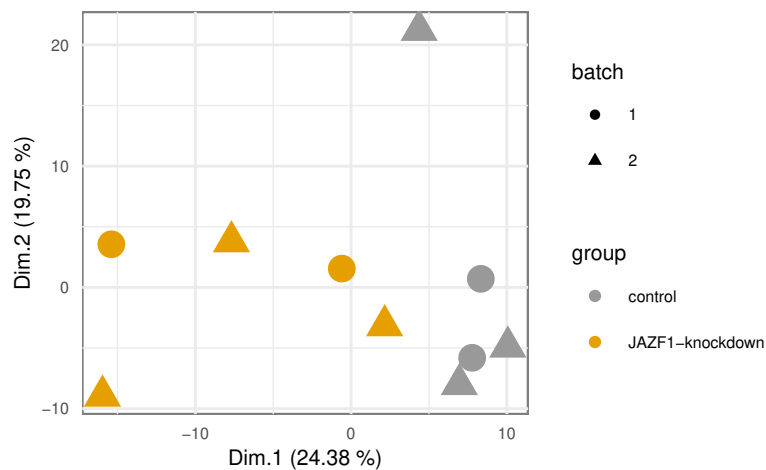
Low Quality Probabilities in GSE163214

Correlation coefficient of P_low vs Group: 0.4571



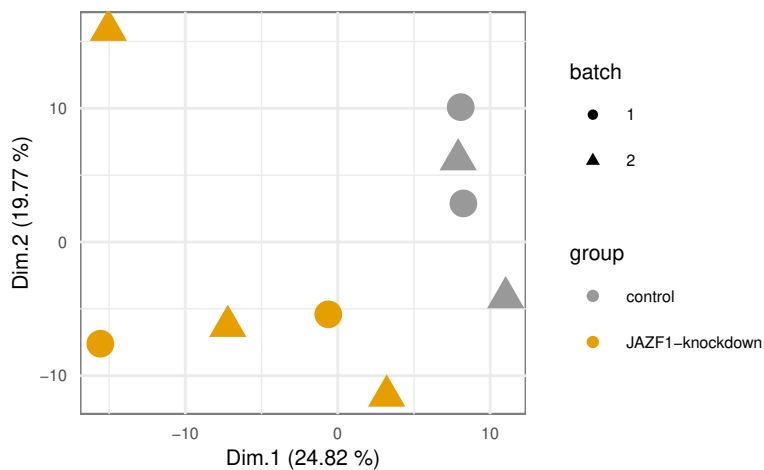
PCA corrected Batch: GSE163214

DesignBias=0.46, Gamma=0.33, Dunn1=0.21, WbRatio=0.7



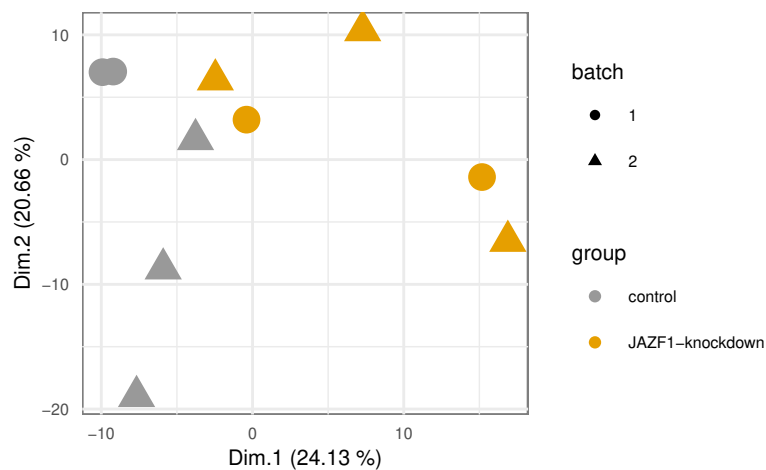
PCA corrected Batch and no Outliers: GSE163214

DesignBias=0.46, Gamma=0.44, Dunn1=0.33, WbRatio=0.63



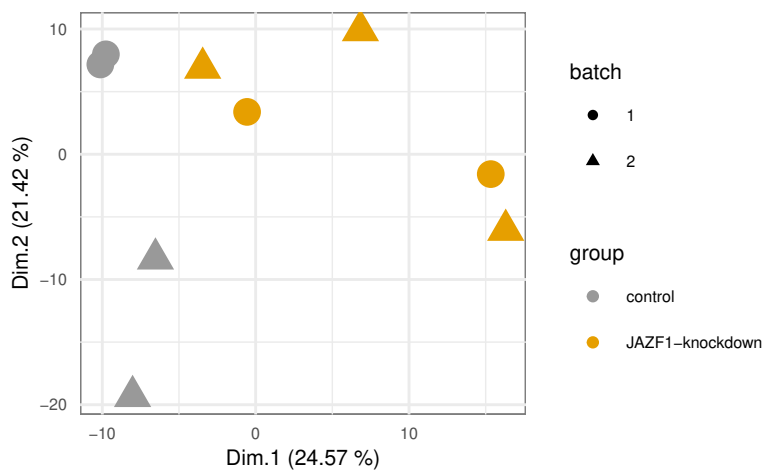
PCA corrected Plow: GSE163214

DesignBias=0.46, Gamma=0.3, Dunn1=0.14, WbRatio=0.74



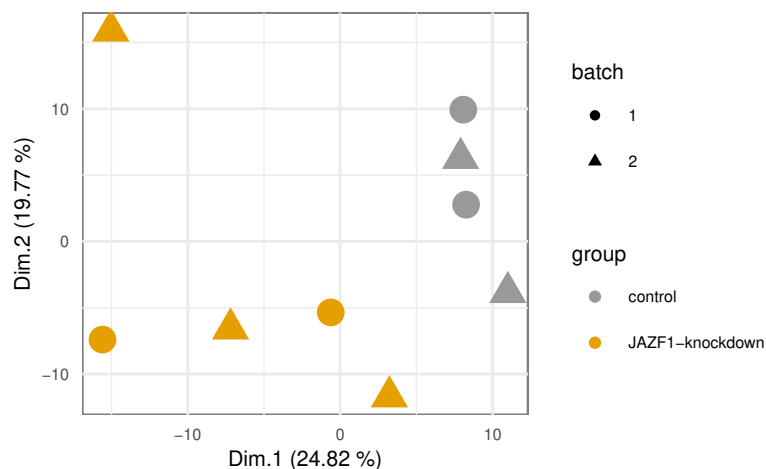
PCA corrected Plow and no outliers: GSE163214

DesignBias=0.46, Gamma=0.35, Dunn1=0.23, WbRatio=0.73

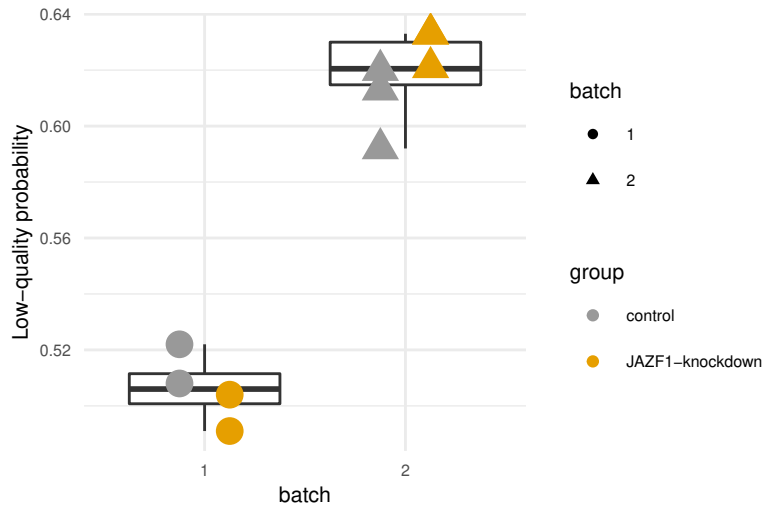


PCA corrected Batch and Plow and no outliers: GSE163214

DesignBias=0.46, Gamma=0.44, Dunn1=0.33, WbRatio=0.63

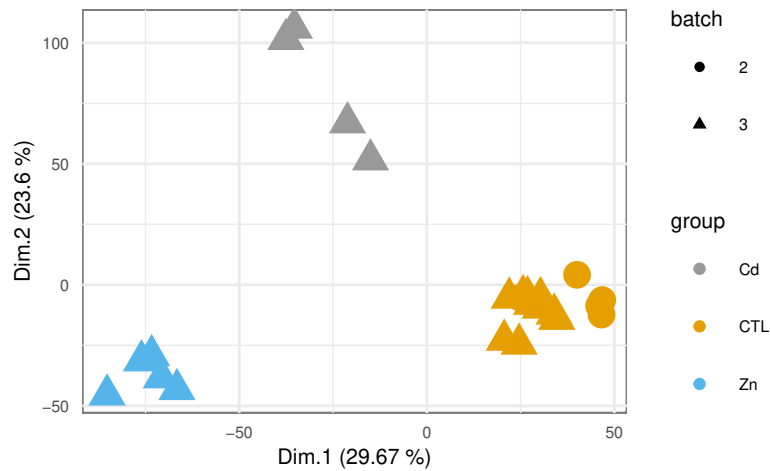


GSE163214: Kruskal-Wallis' p-value: 1.03e-02



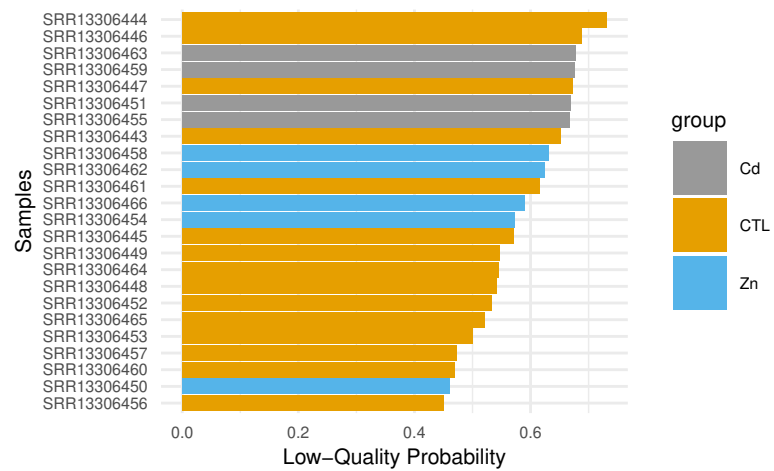
PCA Abundance: GSE163857

DesignBias=0.52, Gamma=0.95, Dunn1=1.17, WbRatio=0.16, DEGs=1726



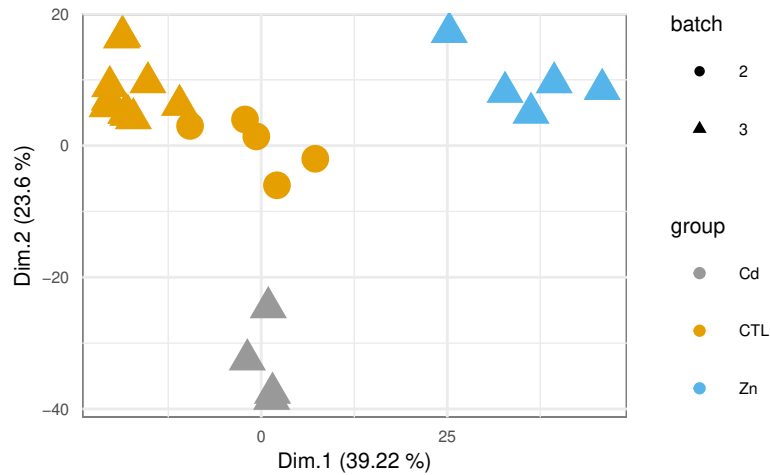
Low Quality Probabilities in GSE163857

Correlation coefficient of P_low vs Group: 0.5222



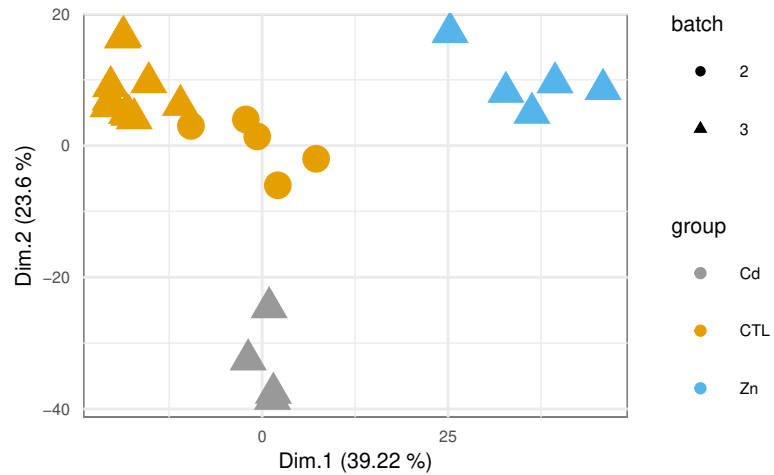
PCA corrected Batch: GSE163857

DesignBias=0.52, Gamma=0.87, Dunn1=0.58, WbRatio=0.25



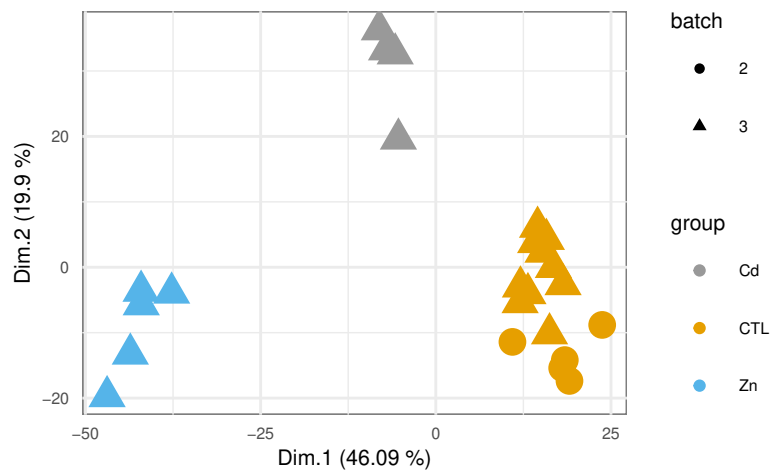
PCA corrected Batch and no Outliers: GSE163857

DesignBias=0.52, Gamma=0.87, Dunn1=0.58, WbRatio=0.25



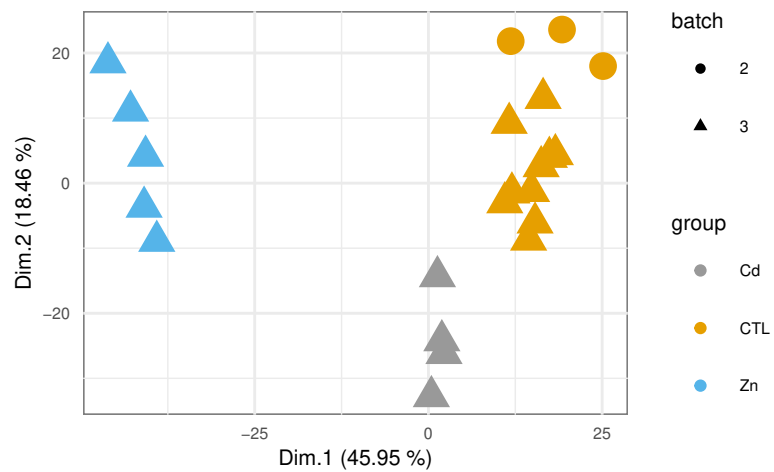
PCA corrected Plow: GSE163857

DesignBias=0.52, Gamma=0.92, Dunn1=1, WbRatio=0.19



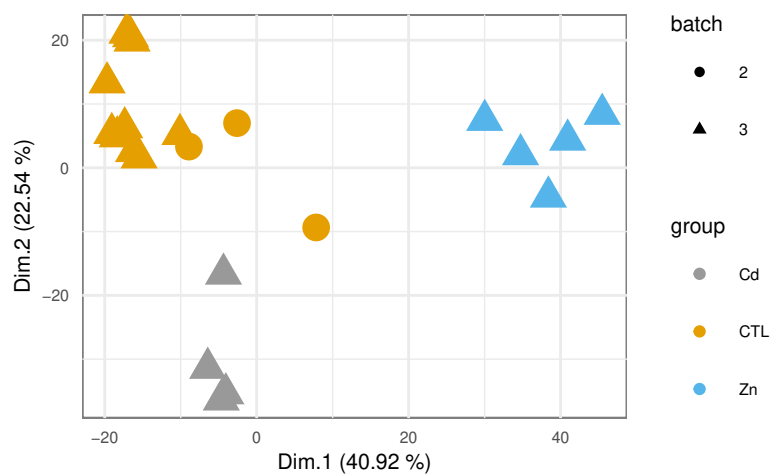
PCA corrected Plow and no outliers: GSE163857

DesignBias=0.52, Gamma=0.82, Dunn1=0.44, WbRatio=0.27

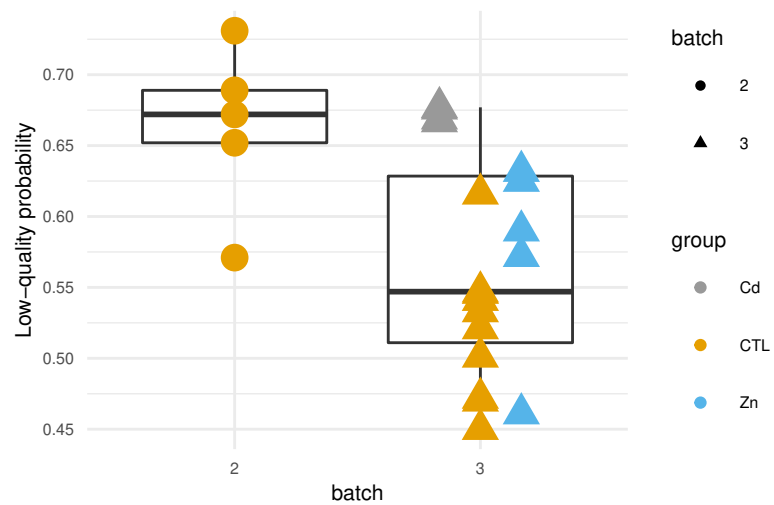


PCA corrected Batch and Plow and no outliers: GSE163857

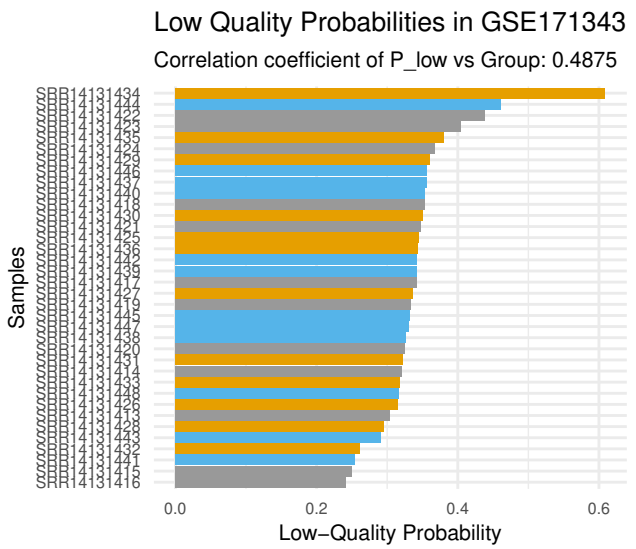
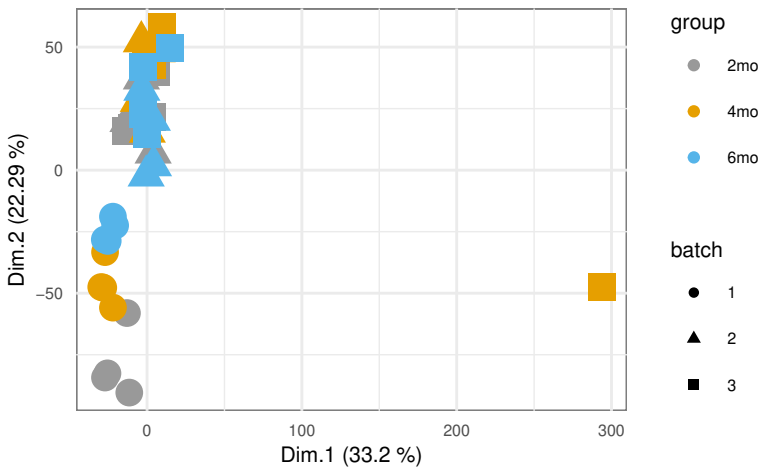
DesignBias=0.52, Gamma=0.85, Dunn1=0.36, WbRatio=0.26



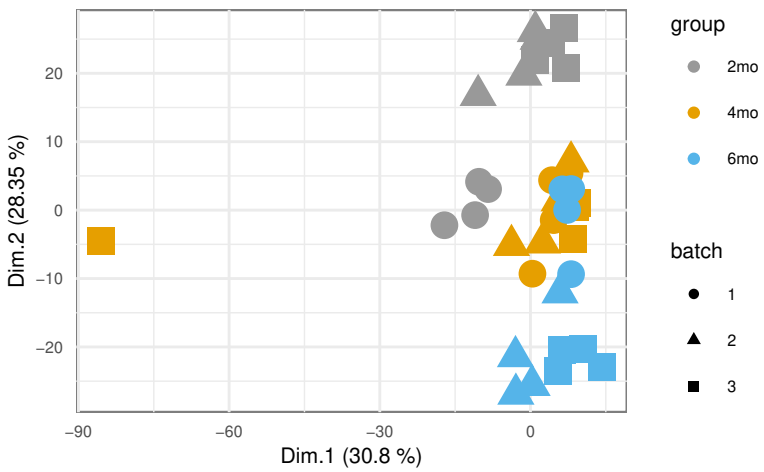
GSE163857: Kruskal-Wallis' p-value: 2.09e-02



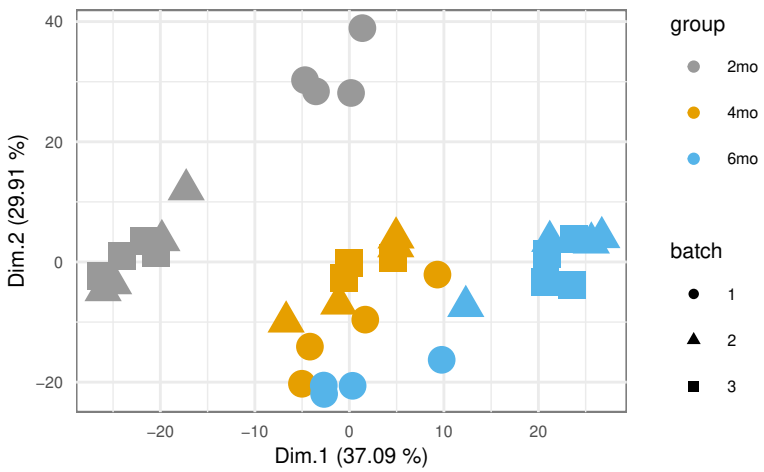
PCA Abundance: GSE171343
 DesignBias=0.49, Gamma=0.01, Dunn1=0, WbRatio=0.99, DEGs=1111



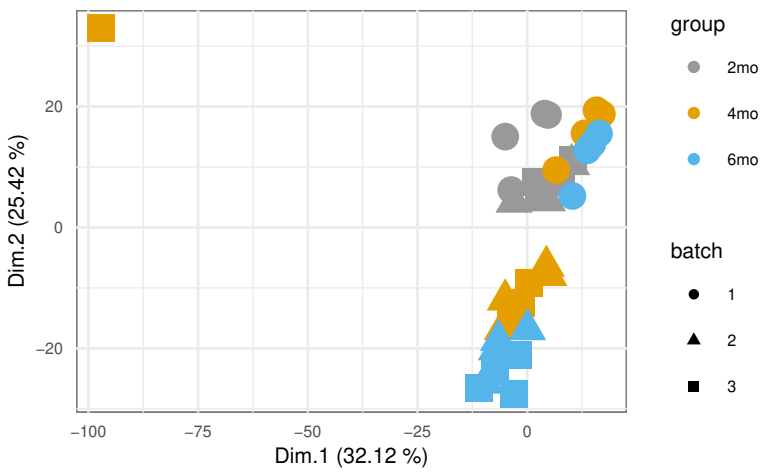
PCA corrected Batch: GSE171343
 DesignBias=0.49, Gamma=0.25, Dunn1=0.02, WbRatio=0.63



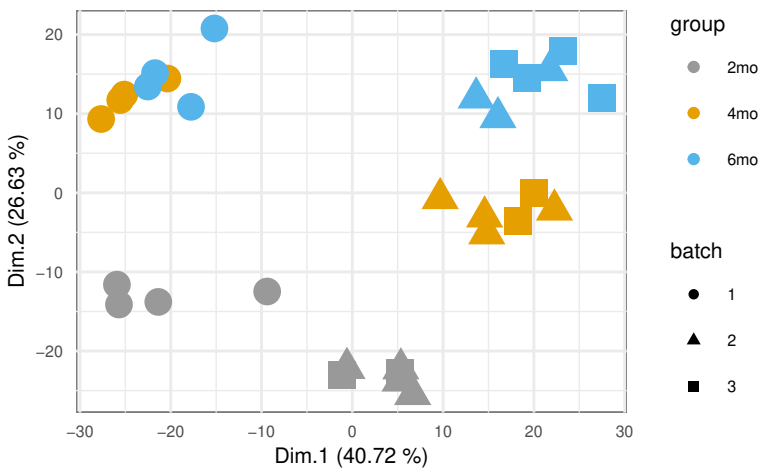
PCA corrected Batch and no Outliers: GSE171343
 DesignBias=0.49, Gamma=0.47, Dunn1=0.05, WbRatio=0.55



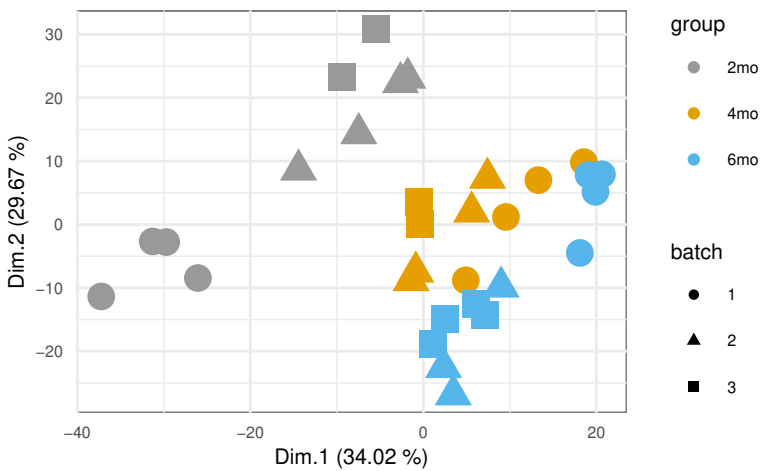
PCA corrected Plow: GSE171343
 DesignBias=0.49, Gamma=0.11, Dunn1=0.01, WbRatio=0.8



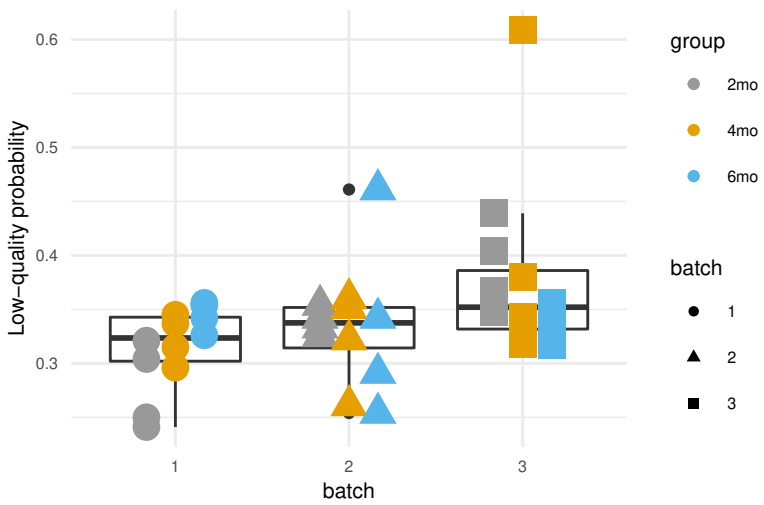
PCA corrected Plow and no outliers: GSE171343
 DesignBias=0.49, Gamma=0.37, Dunn1=0.03, WbRatio=0.64



PCA corrected Batch and Plow and no outliers: GSE171343
 DesignBias=0.49, Gamma=0.4, Dunn1=0.04, WbRatio=0.61

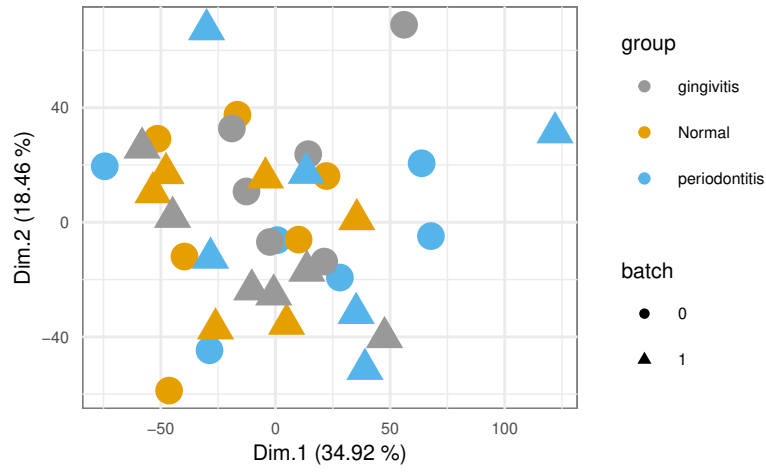


GSE171343: Kruskal-Wallis' p-value: 4.82e-02



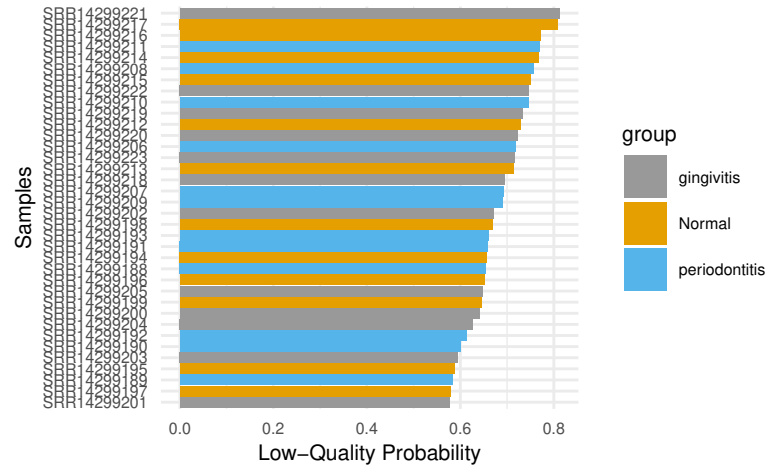
PCA Abundance: GSE173078

DesignBias=0.48, Gamma=0.01, Dunn1=0.02, WbRatio=0.99, DEGs=100



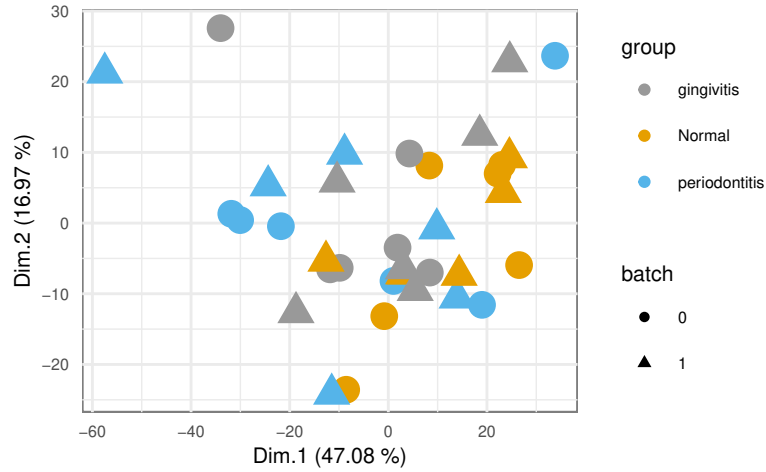
Low Quality Probabilities in GSE173078

Correlation coefficient of P_low vs P: 0.4788



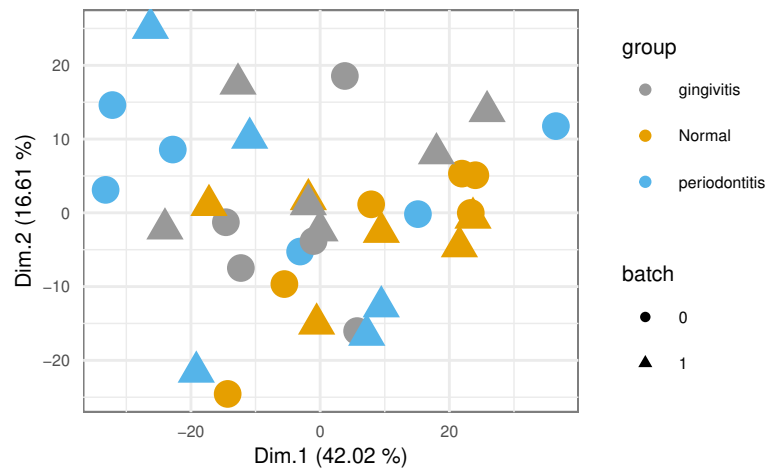
PCA corrected Batch: GSE173078

DesignBias=0.48, Gamma=0.05, Dunn1=0.01, WbRatio=0.94



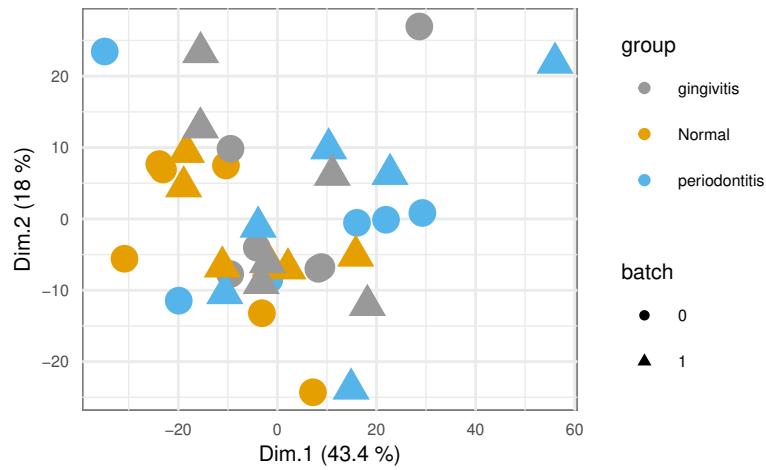
PCA corrected Batch and no Outliers: GSE173078

DesignBias=0.48, Gamma=0.05, Dunn1=0.01, WbRatio=0.95



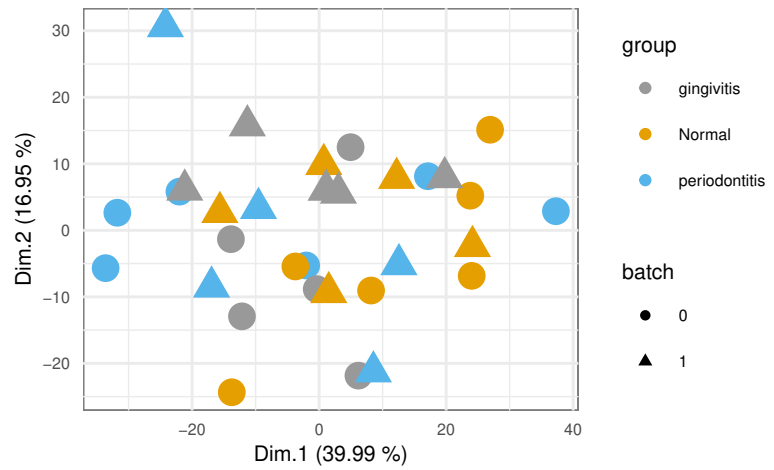
PCA corrected Plow: GSE173078

DesignBias=0.48, Gamma=0.04, Dunn1=0.01, WbRatio=0.95



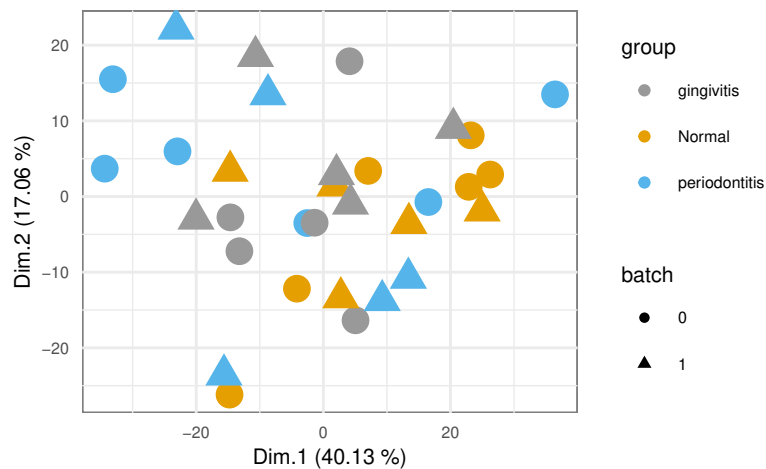
PCA corrected Plow and no outliers: GSE173078

DesignBias=0.48, Gamma=0.01, Dunn1=0.01, WbRatio=0.99

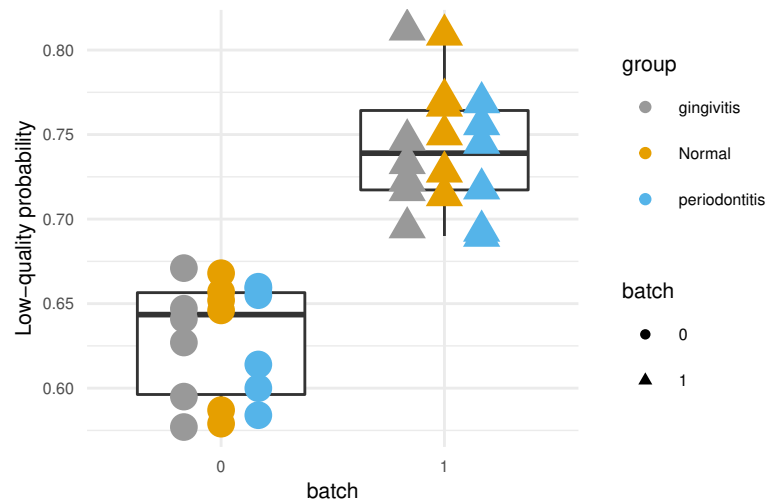


PCA corrected Batch and Plow and no outliers: GSE173078

DesignBias=0.48, Gamma=0.03, Dunn1=0.02, WbRatio=0.96

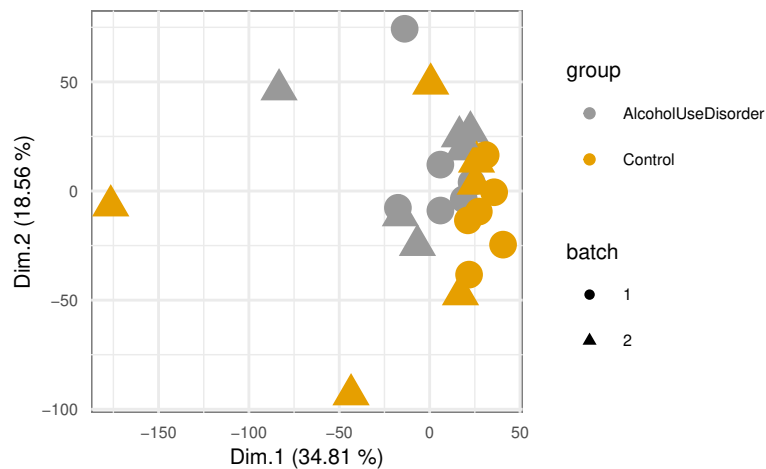


GSE173078: Kruskal-Wallis' p-value: 2.97e-07



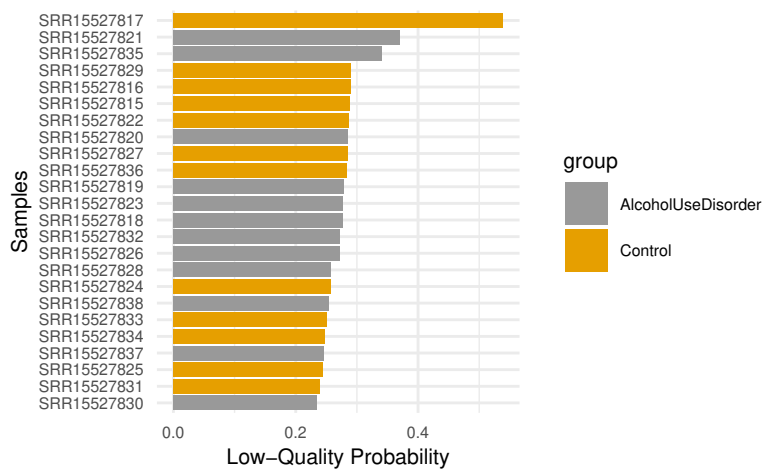
PCA Abundance: GSE182440

DesignBias=0.49, Gamma=0.03, Dunn1=0.01, WbRatio=0.95, DEGs=3



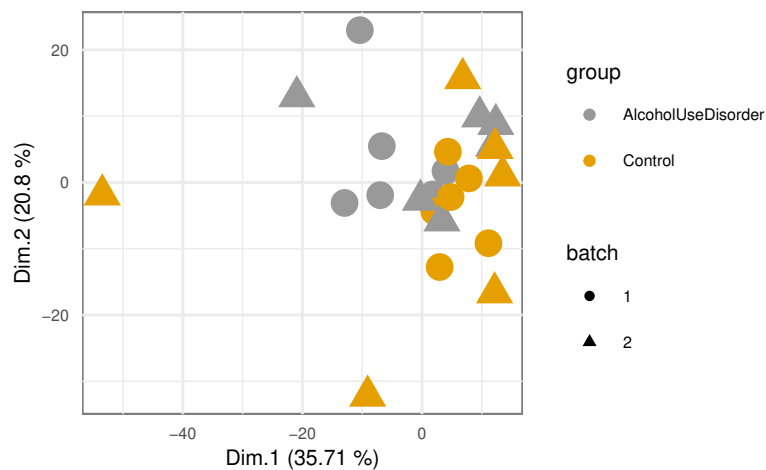
Low Quality Probabilities in GSE182440

Correlation coefficient of P_low vs Group: 0.4939



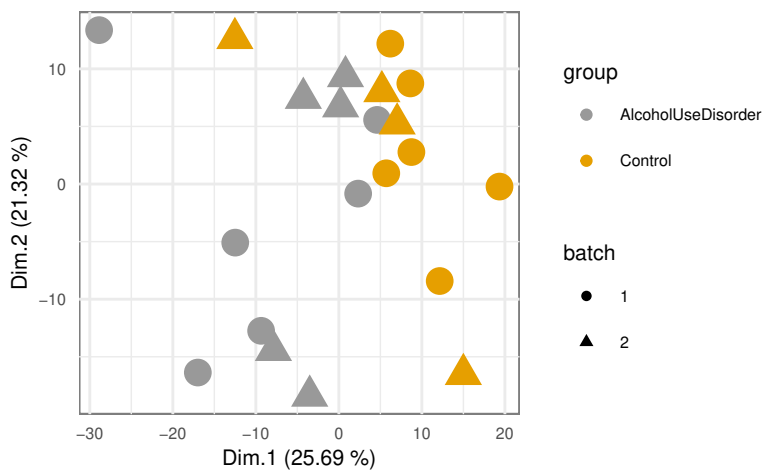
PCA corrected Batch: GSE182440

DesignBias=0.49, Gamma=0.03, Dunn1=0.01, WbRatio=0.95



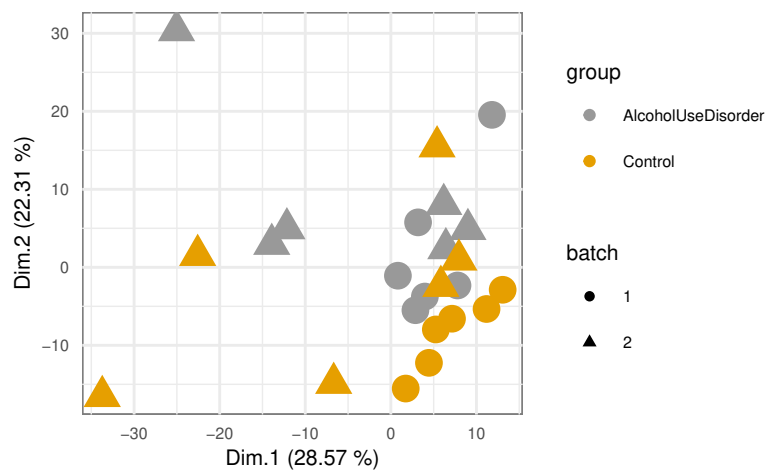
PCA corrected Batch and no Outliers: GSE182440

DesignBias=0.49, Gamma=0.2, Dunn1=0.06, WbRatio=0.8



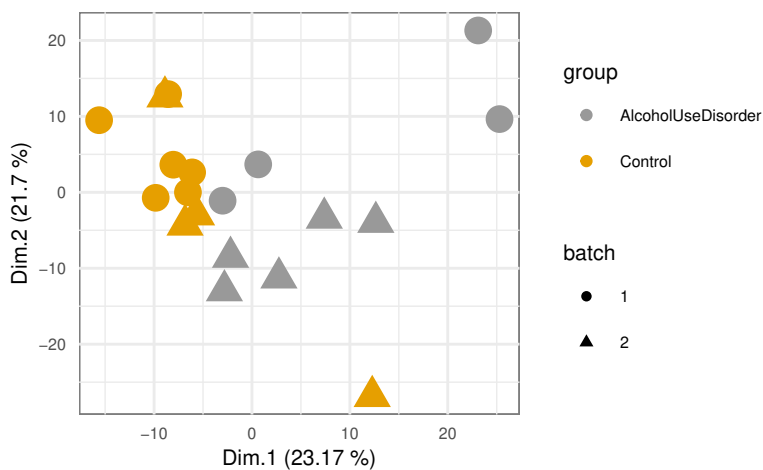
PCA corrected Plow: GSE182440

DesignBias=0.49, Gamma=0.06, Dunn1=0.04, WbRatio=0.92



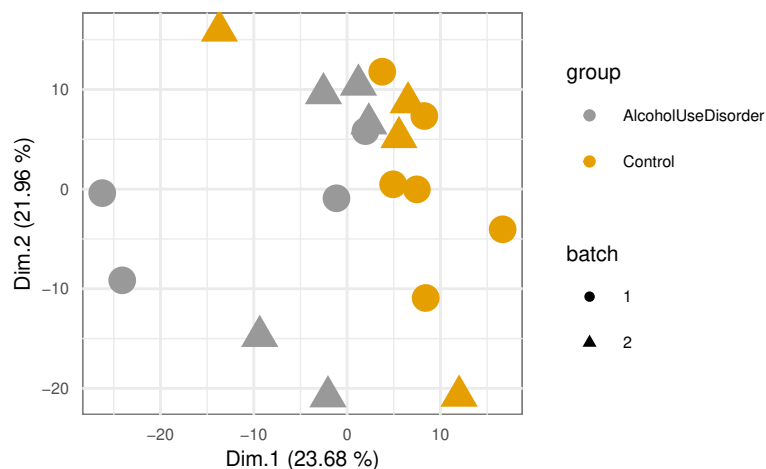
PCA corrected Plow and no outliers: GSE182440

DesignBias=0.49, Gamma=0.2, Dunn1=0.07, WbRatio=0.78



PCA corrected Batch and Plow and no outliers: GSE182440

DesignBias=0.49, Gamma=0.13, Dunn1=0.06, WbRatio=0.87



GSE182440: Kruskal-Wallis' p-value: 3.50e-02

