

## Hypothesis testing

A hypothesis test was used to judge the ability of the technology to discriminate between sample types as follows. Assays were categorized into distinct cohorts based on assay and user metadata (e.g. distinct laboratory or user proficiency level). For each sample type, the sample means were consolidated across assays within a specific cohort. A Wilcoxon signed-rank test was subsequently executed, comparing the sample means of type  $i$  with those of type  $j$  within the given assay cohort. The p-value for this test was determined for every pertinent assay cohort and for each pairing of sample types. The outcomes of these tests were graphically represented in Figs 3A and 5A.