Figures and figure supplements

Infrared molecular fingerprinting of blood-based liquid biopsies for the detection of cancer
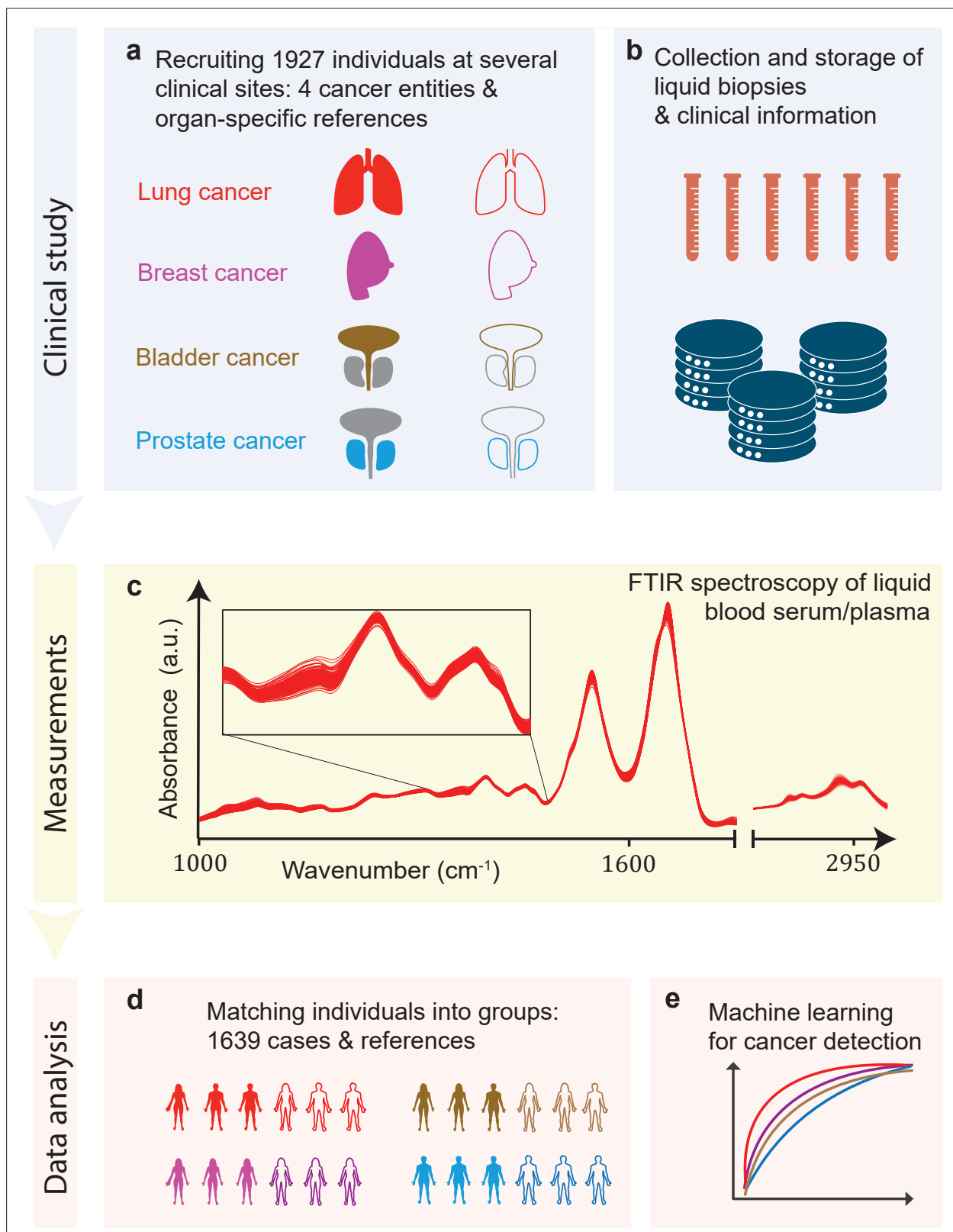
Marinus Huber *et al*

Figure 1. Infrared molecular fingerprinting workflow and clinical study design. (**a**) Cohorts of therapy-naïve, lung, breast, prostate, and bladder cancer patients (cases), and organ-specific symptomatic references as well as non-symptomatic reference individuals were recruited at three different clinical sites – in total, 1927 individuals. (**b**) Blood samples from all individuals were drawn, and sera and plasma were prepared according to well-defined standard operating procedures. (**c**) Automated Fourier-transform infrared spectroscopy of liquid bulk sera and plasma were used to obtain IMFs. The

Figure 1 continued on next page

Figure 1 continued

displayed IMFs were pre-processed using water correction and normalization (see Methods). (**d**) For each clinical question studied, the characteristics of the case and the reference cohorts were matched for age, gender, and body mass index (BMI) to avoid patient selection bias. This resulted in total number of 1639 individuals upon matching. (**e**) Machine learning models were built on training datasets and evaluated on test datasets to separately evaluate the efficiency of classification for each of the four cancer entities.
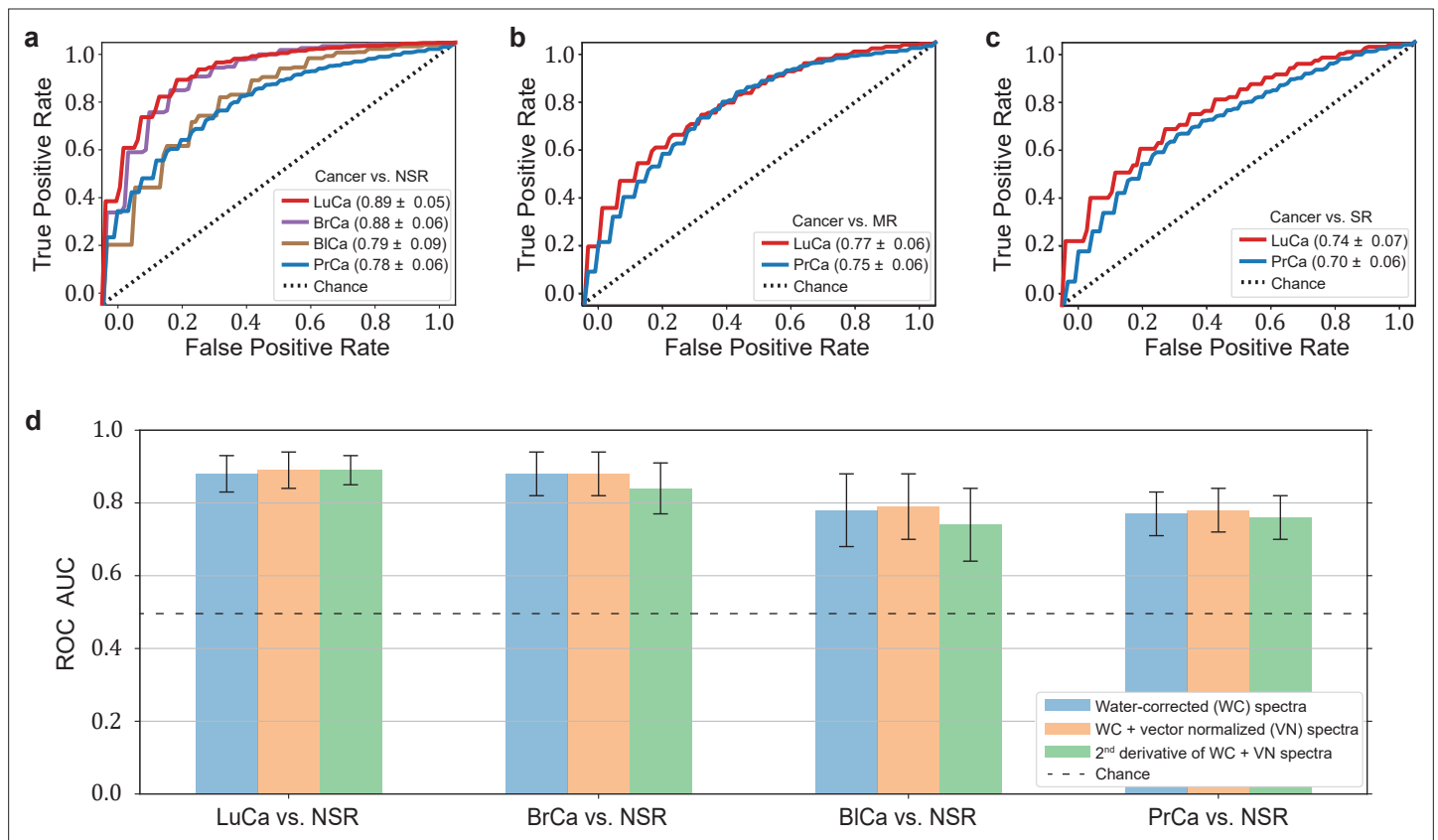
Figure 2. Diagnostic performance of lung, prostate, bladder, and breast cancer detection based on infrared molecular fingerprints (IMFs) of blood sera. Receiver operating characteristic (ROC) curves for the binary classification of the test set with support vector machine (SVM) models trained on water-corrected and vector-normalized IMFs. The different cancer entities were tested against (**a**) non-symptomatic references, (**b**) mixed references that also include organ-specific symptomatic references, and (**c**) organ-specific symptomatic references only. Detailed cohort characteristics can be found in *Figure 2—source data 1*. (**d**) Area under the receiver operating characteristic curve (AUC) for the test sets according to different spectral pre-processing of the IMFs. The error bars show the standard deviation of the individual results of the cross-validation (LuCa: lung cancer; PrCa: prostate cancer; BrCa: breast cancer; BlCa: bladder cancer; NSR: non-symptomatic references; MR: mixed references; SR: symptomatic references).
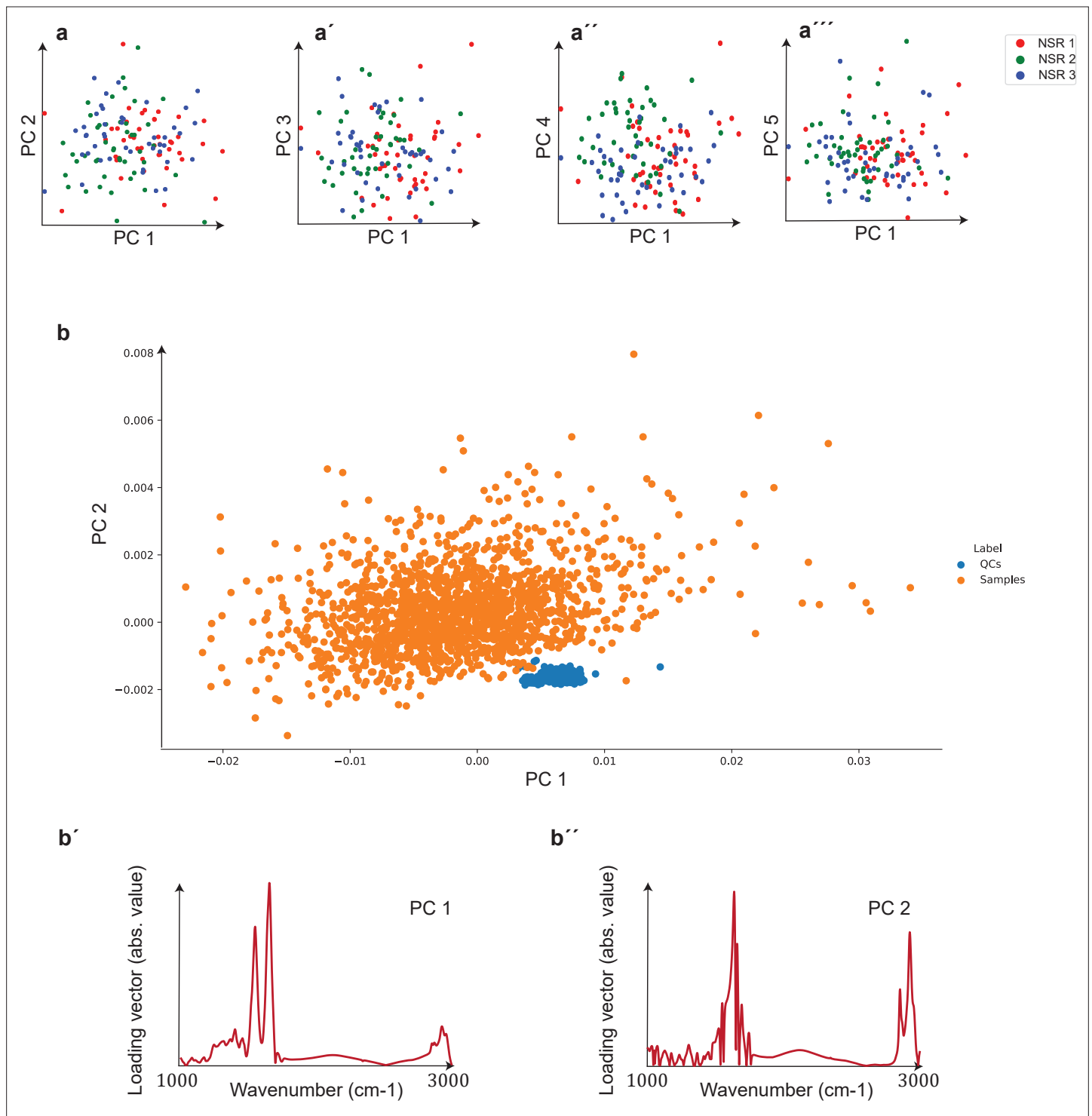
Figure 2—figure supplement 1. Unsupervised comparison between data from the three clinical sites as well as quality control (QC) analysis of measurements. (**a–a'''**) Principal component analysis (PCA) of samples of non-symptomatic healthy individuals collected from three different clinical sites. Plots depict the first five principal components, which correspond to 95% of the explained variance. The three groups are statistically matched in terms of age, gender, and body mass index (BMI). Cohort characteristics are given in *Figure 2—figure supplement 1—source data 1*. (**b**) PCA plot of biological samples and QCs. The two first principal components included in the plot correspond to 93% of the explained variance. (**b', b''**) Loading vectors for the two principal components shown in (**b**).
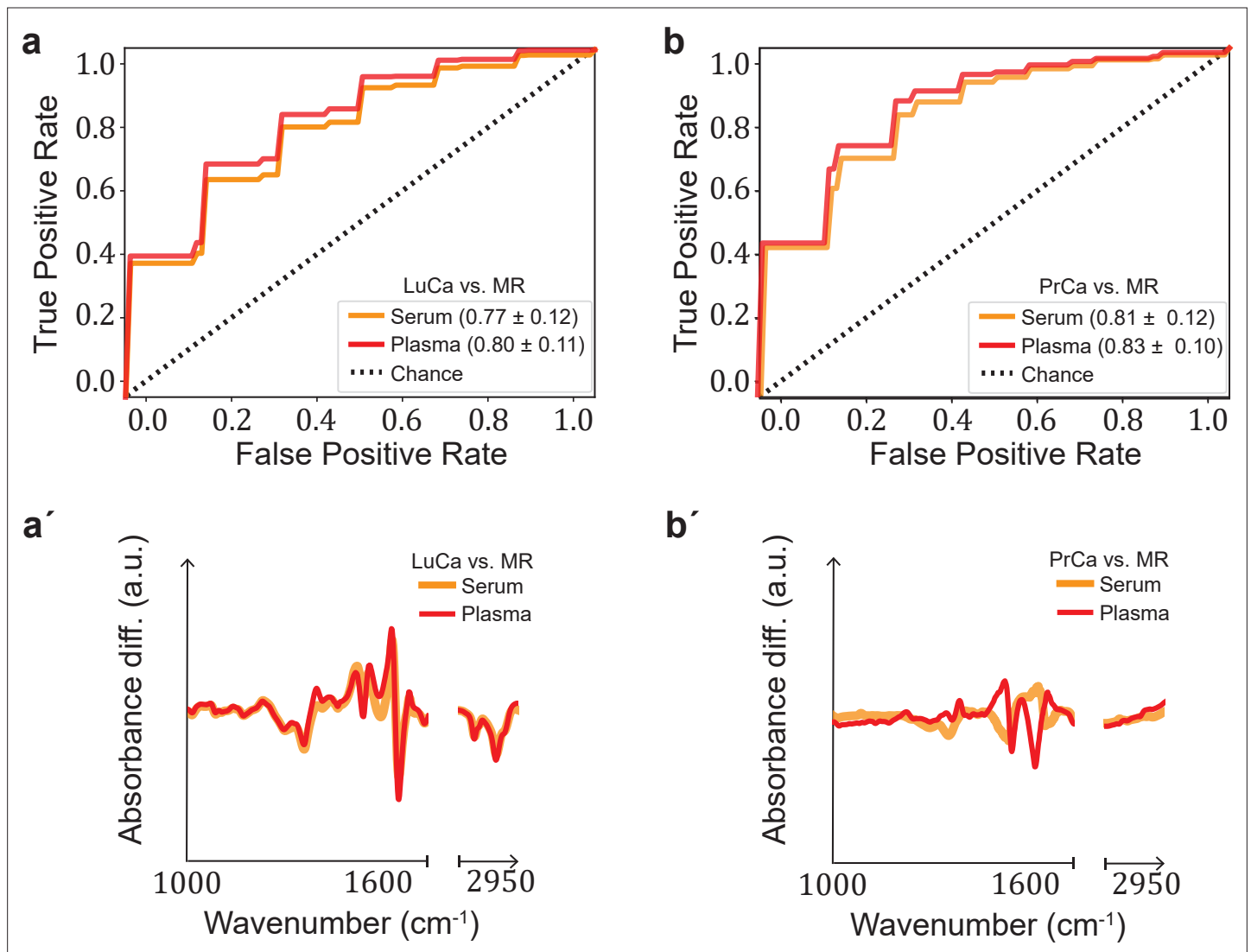
Figure 2—figure supplement 2. Performance comparison of serum- and plasma-based fingerprints for cancer detection. Receiver operating characteristic (ROC) curves for (**a**) lung cancer (LuCa) and (**b**) prostate cancer (PrCa) vs. mixed references (MR). Differential fingerprints (**a′, b′**) for the same comparisons as above. The characteristics of the cohort used for this analysis are given in *Figure 2—figure supplement 2—source data 1*.
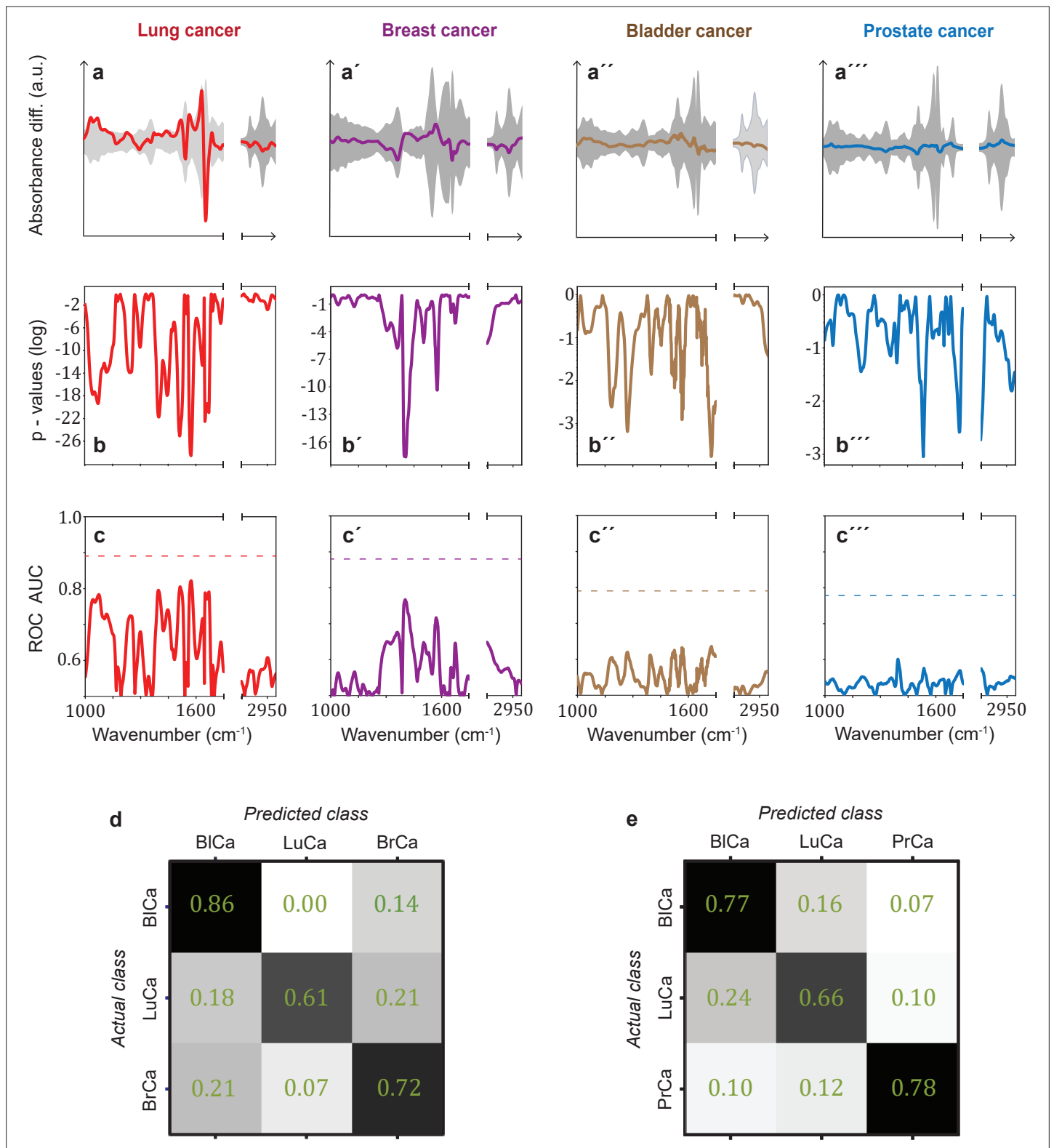
Figure 3. Infrared spectral signatures of lung, prostate, bladder, and breast cancer. (**a-a'''**) Differential fingerprints (standard deviations of the reference cohorts are displayed as grey areas), (**b-b'''**) two-tailed p-value of Student's t-test, and (**c-c'''**) area under the receiver operating characteristic curve (AUC) per wavenumber (extracted by application of Mann–Whitney U test) compared to the AUC of the combined model (dashed horizontal lines). Confusion matrix summarizing the per-class accuracies of multiclass classification of (**d**) lung, bladder, and breast cancer (matched female cohort) with overall model accuracy of 0.73 ± 0.11, and (**e**) lung, bladder, and prostate cancer (matched male cohort) with overall model accuracy of 0.74 ± 0.13.

Figure 3 continued on next page

Figure 3 continued

Detailed cohort characteristics can be found in ***Figure 3—source data 1***. Chance level for the three-class classification corresponds to 0.33 (LuCa: lung cancer; PrCa: prostate cancer; BrCa: breast cancer; BlCa: bladder cancer).
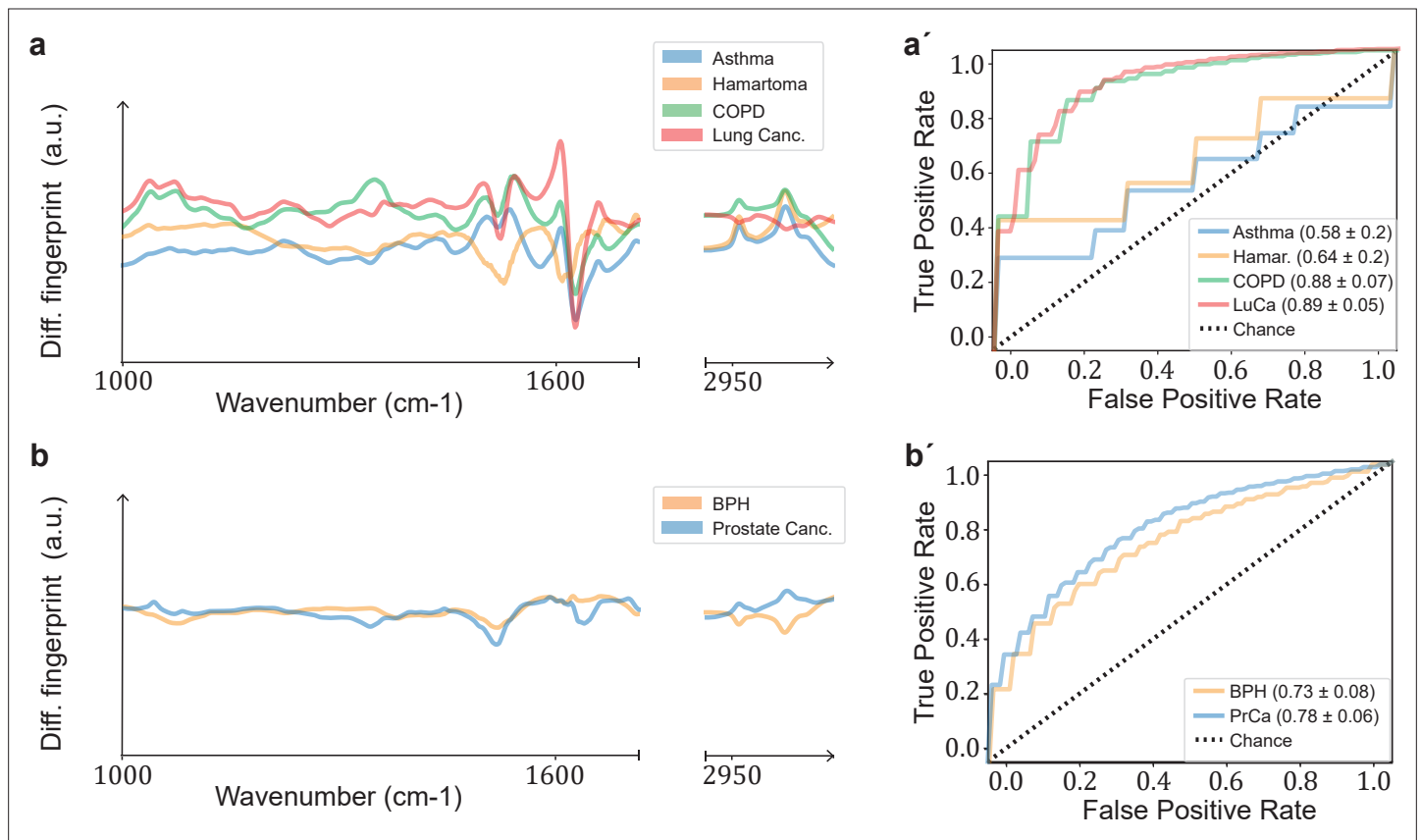
Figure 3—figure supplement 1. Comparison of signatures from different organ-specific pathologies. Differential fingerprints for (**a**) lung-related conditions (asthma, lung hamartoma, chronic obstructive pulmonary disease [COPD], lung cancer) and (**b**) prostate-related pathologies (benign prostate hyperplasia [BPH], prostate cancer). Receiver operating characteristic (ROC) curves for (**a'**) lung and (**b'**) prostate pathologies. All comparisons are against non-symptomatic references. The characteristics of the cohort used for this analysis are given in *Figure 3—figure supplement 1—source data 1*.
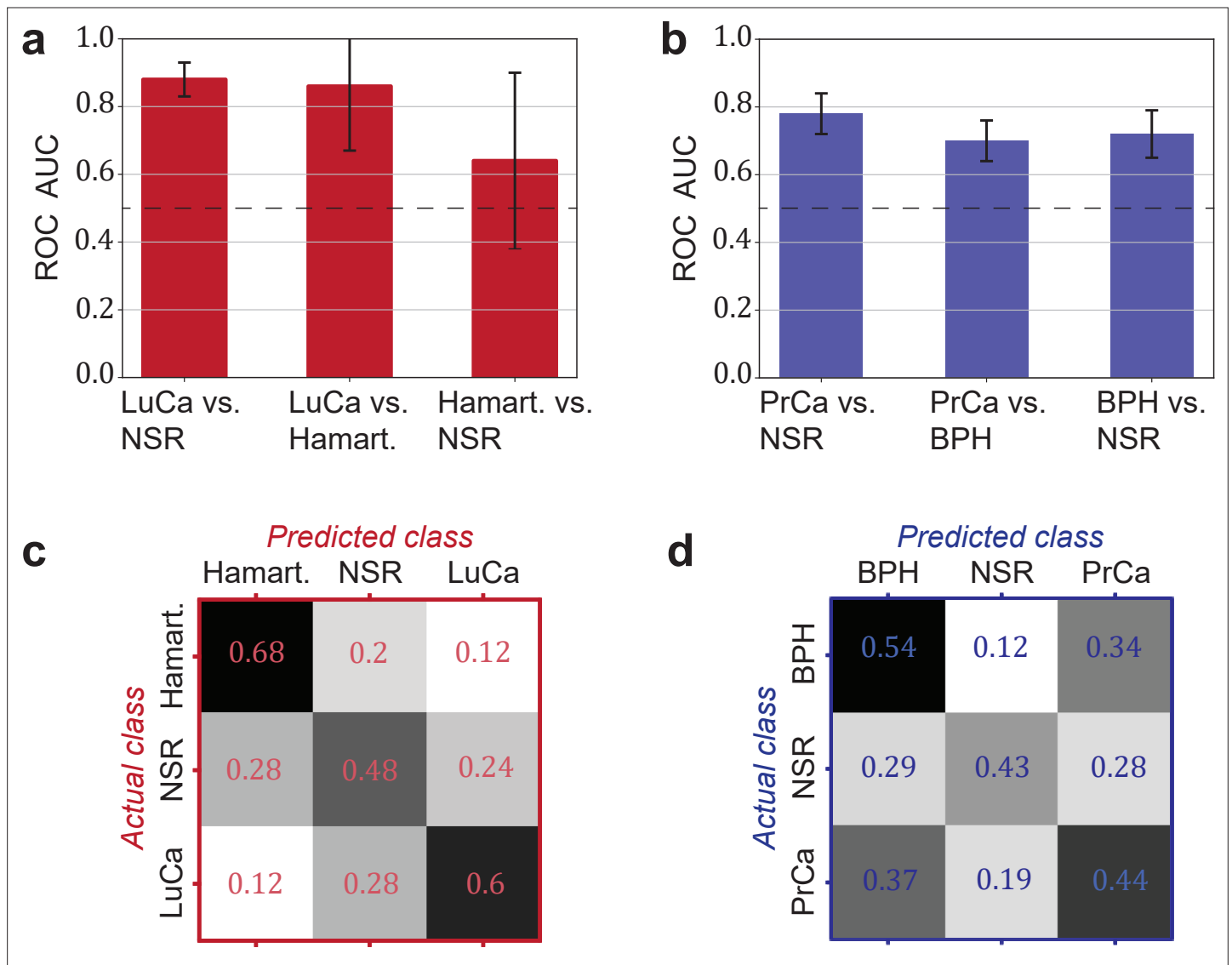
Figure 4. Detection efficiency of benign conditions and multiclass classification. (**a**) Pairwise classification performance results between lung cancer (LuCa), hamartoma (Hamart.) and non-symptomatic reference group (NSR) with overall model accuracy of 0.46 ± 0.18, and (**b**) pairwise classification performance between prostate cancer (PrCa), benign prostate hyperplasia (BPH), and NSR with overall model accuracy of 0.43 ± 0.06. The error bars show the standard deviation of the individual results of the cross-validation. Confusion matrix summarizing the per-class accuracies of multiclass classification in (**c**) the LuCa cohort and (**d**) the PrCa cohort. The characteristics of the cohort used for this analysis are given in *Figure 4—source data 1*. Chance level for the three-class classification corresponds to 0.33.
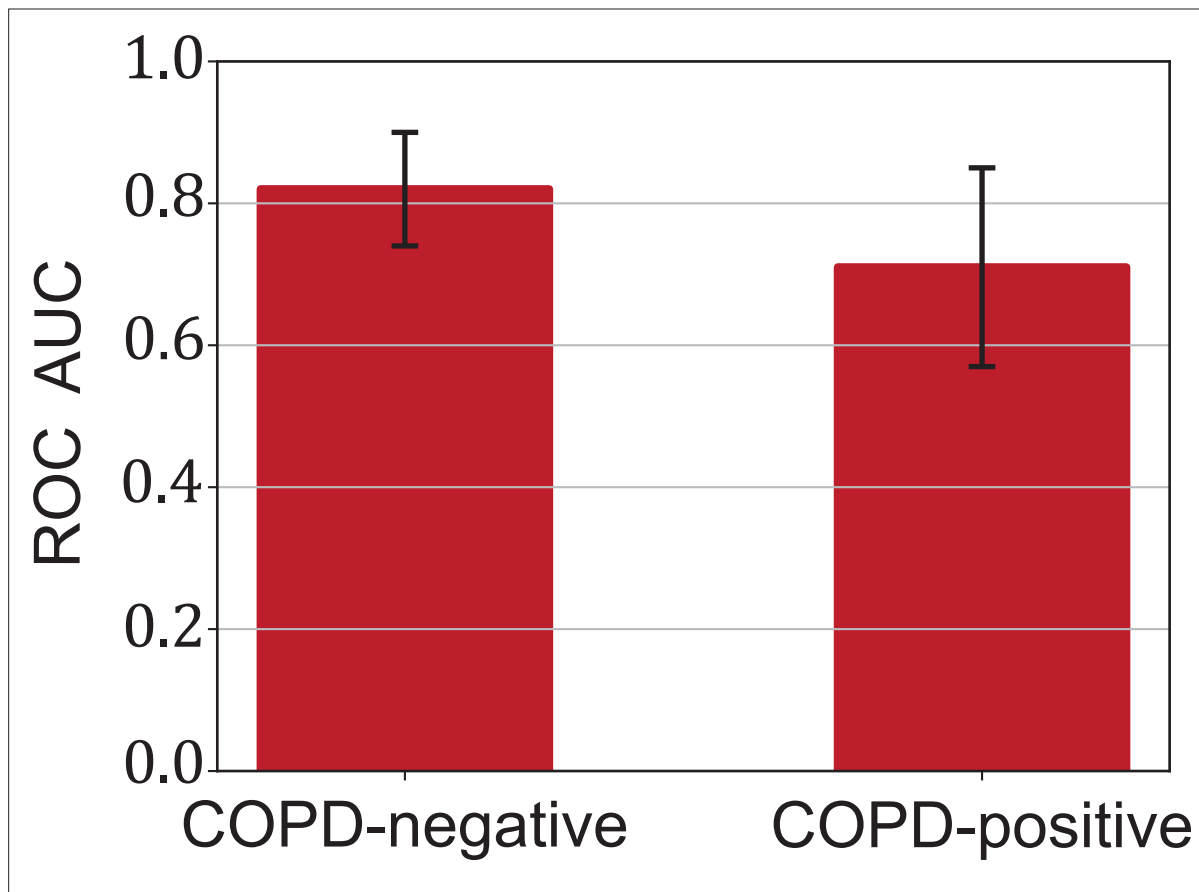
Figure 4—figure supplement 1. Influence of chronic obstructive pulmonary disease (COPD) in lung cancer (LuCa) detection. Classification performance of LuCa vs. mixed references as a function of the COPD status. The characteristics of the cohort used for this analysis are given in *Figure 4—figure supplement 1—source data 1*.
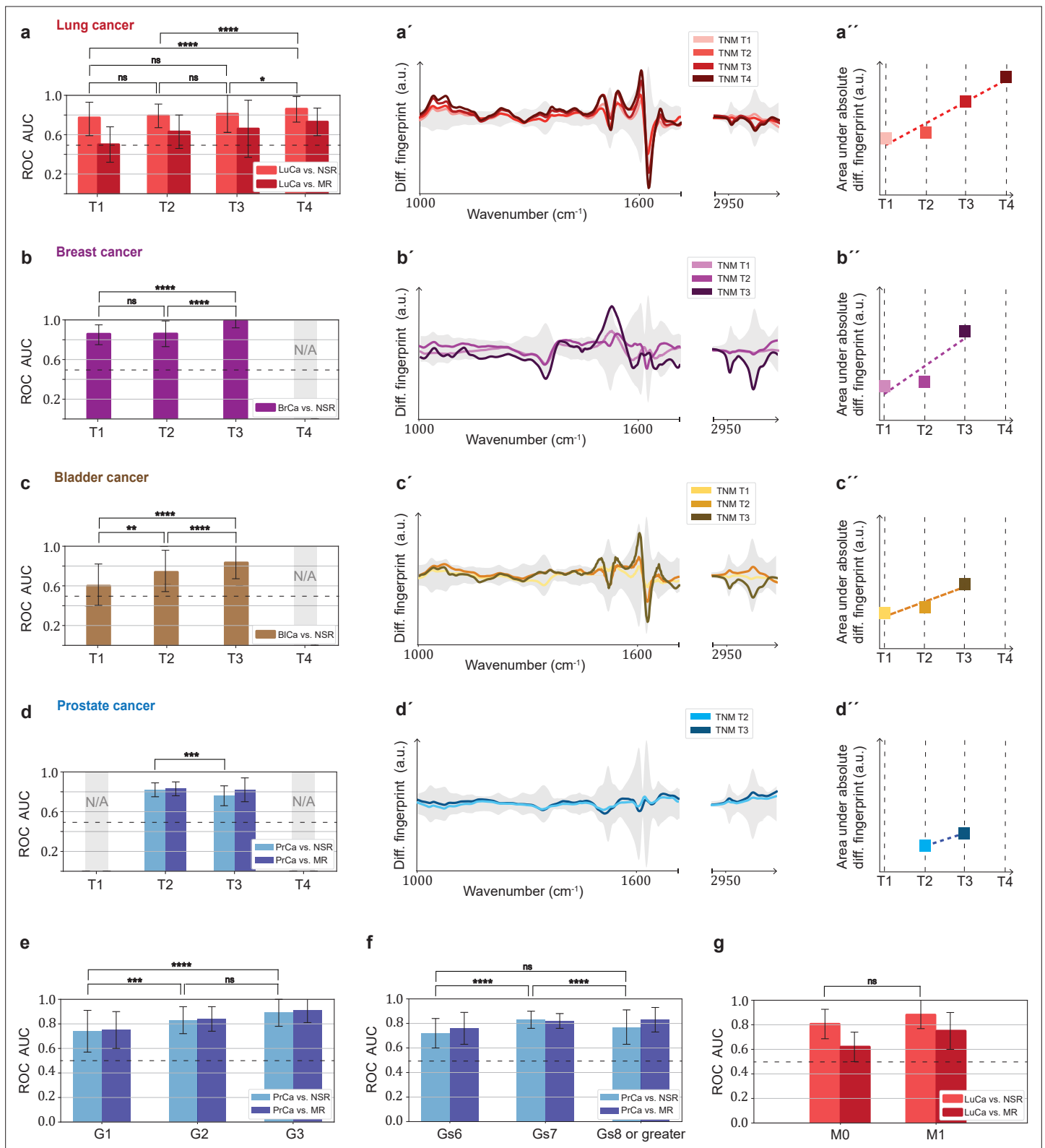
Figure 5. Efficiency of binary classification and infrared spectral changes in dependence of tumour progression. (**a–d**) Binary classification performance of lung, breast, bladder, and prostate cancer against references as a function of T-classification (of TNM-staging). (**a'–d'**) Differential fingerprints in relation with the tumour size (TNM class T) for all four cancer entities. (**a''–d''**) Area under the absolute differential fingerprints in relation with the tumour size for all dour cancer entities. The y-axes of the diagrams in the panels (**a'–d'**) and (**a''–d''**) each have the same linear scaling, thus directly comparable. (**e**) Classification performance of prostate cancer versus references as a function of tumour grade score. (**f**) Classification performance of

Figure 5 continued on next page

Figure 5 continued

prostate cancer as a function of the Gleason score (Gs). (**g**) Classification performance of lung cancer versus references as a function of the metastasis status. The detailed cohort breakdown and classification results are given as *Figure 5—source data 1*, *Figure 5—source data 2*, *Figure 5—source data 3*, *Figure 5—source data 4*. Some cohorts did not include sufficient number of participants so that a reliable machine learning model could not be built and were therefore not evaluated. LuCa: lung cancer; PrCa: prostate cancer; BrCa: breast cancer; BlCa: bladder cancer; NSR: non-symptomatic references; MR: mixed references; n.s.: not significant; *p$<10^{-2}$; **p$<10^{-3}$; ***p$<10^{-4}$; ****p$<10^{-5}$; The error bars show the standard deviation of the individual results of the cross-validation.
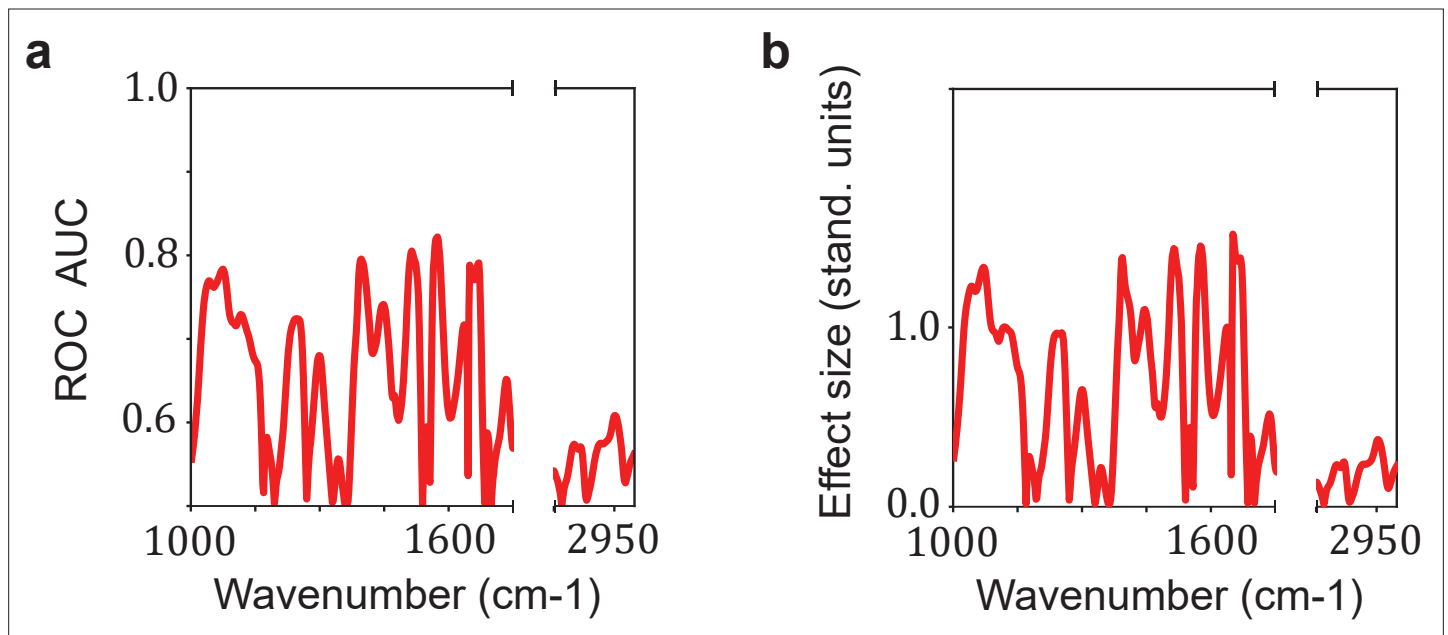
Figure 5—figure supplement 1. Relation between the effect size and the area under the receiver operating characteristic curve (AUC) per wavenumber. Comparison between (**a**) the AUC per wavenumber and (**b**) the effect size per wavenumber. The effect size is defined as the standardized difference between the sample means of cases and references, also known as Cohen's d. The AUC per wavenumber is calculated using the U statistic of Mann–Whitney U test by the relation AUC = $U/(n1 * n2)$. This example was performed for the comparison lung cancer (LuCa) vs. non-symptomatic references (NSR).