Supplementary figures
Benelli M, Romagnoli D, Demichelis F
Figure 1: AUC and beta scores.
Figure 2: Correlation coefficients and RMSD computed with different number of informative sites in all cancer types.
Figure 3: Comparison between state-of-the-art methods.
Figure 4: PAMES and InfiniumPurify correlation.
Figure 5: Boxplots of the distribution of purity estimates by PAMES (left) and InfiniumPurify (right) across 14 TCGA datasets (n= 5,623 samples).
Figure 6: Scatter plot of purity estimates by a) PAMES, b) PAMES using thresholds calculated by pan-cancer analysis, c) PAMES using thresholds calculated by tumor-type specific analysis versus InfiniumPurify related to 14 datasets from TCGA (n=5,623 samples). R and RMSD refer to the Pearson’s correlation coefficient and Root Mean Square Deviation averaged (mean) across the 14 tumor types.
Figure 7: Box plot of the distribution of the differences between purity estimates by a) PAMES, b) PAMES using thresholds calculated by pan-cancer analysis, c) PAMES using thresholds calculated by tumor-type specific analysis and InfiniumPurify in 14 datasets from TCGA (n=5,623 samples).
**Figure 8:** Box plots of the distribution of purity estimates by PAMES in tumor (red, n=5,623) and normal (green, n=712) tissue samples across 14 datasets from TCGA.
Figure 9: Box plot of the distribution of PAMES purity estimates of tumor, adjacent normal and tumor-free normal breast samples.
Figure 10: Box plot of the distribution of purity estimates of $N=374$ cancer cell lines of 13 different tumor types by PAMES (red, left) and by a modified version of PAMES that takes into consideration a random selection of differentially methylated sites instead of informative sites (blue, right). The difference between the two distributions is statistically significant (WMW p-val < 1e-90).
Figure 11: Purity estimates by PAMES in 6 TCGA PRAD samples showing flat or noisy genomic profiles. Theoretical log2-ratio (Log2(R)) values for different set of SCNA events are represented in the dashed line.
Figure 12: Density plot of the beta values of CpG islands when first quartile (dashed green line), median (continuous red line) or third quartile (dashed blue line) is used as summarization strategy.
Figure 13: CpG islands and Sites correlation.
Figure 14: Box plots of the distribution of purity estimates by PAMES in normal prostate tissues profiled by arrays (TCGA, left) and eRRBS (Lin et al., 2013, right).