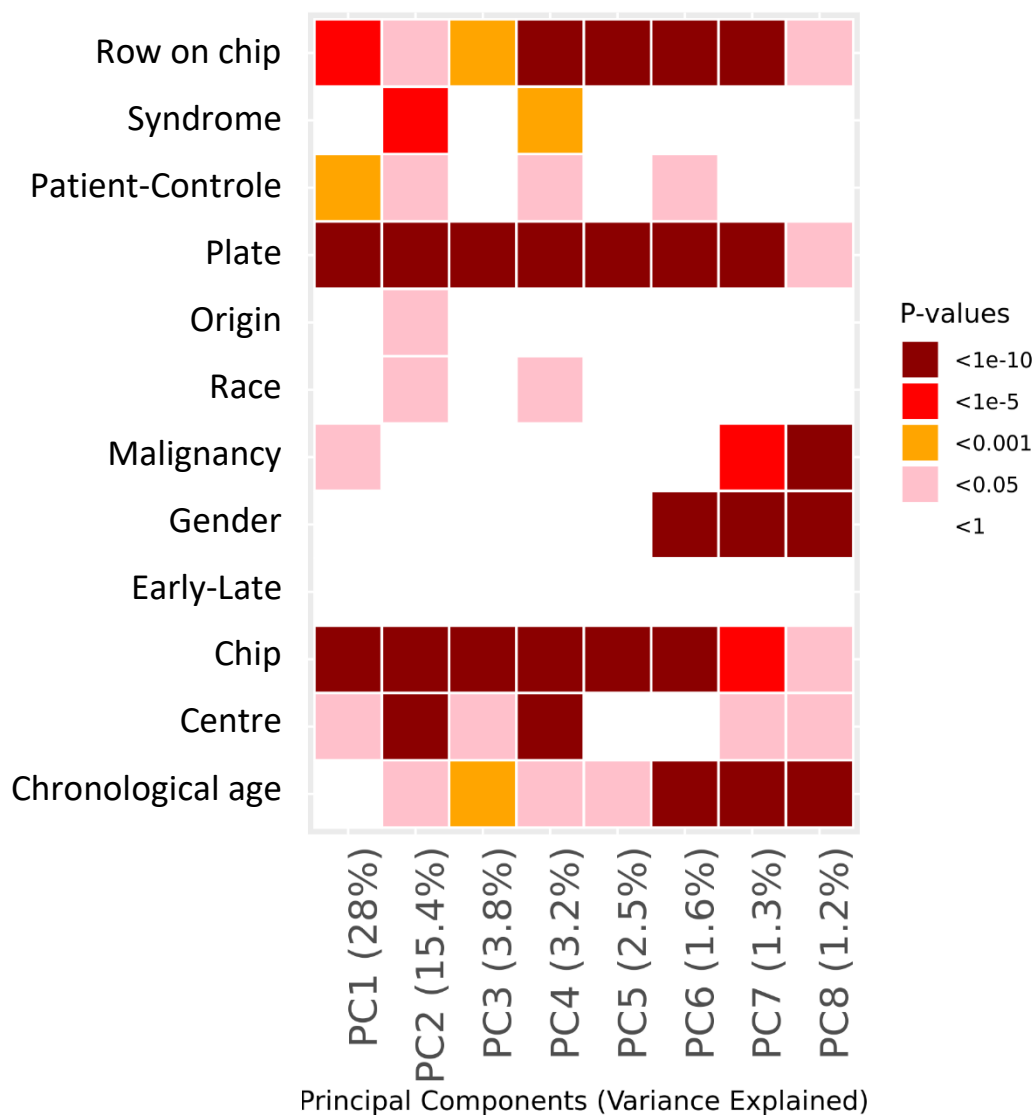


Additional file 2: Quality assessment of the DNA methylation data: principal component analysis



Principal Component Analysis (PCA) is a technique used to bring out strong patterns in a dataset by reducing the dimensionality via creating new uncorrelated variables (principal components or PCs) that successively maximize variance. This p-value heatmap is a visualisation of the PCA, showing the association between the first eight PCs with biological (syndrome, patient vs control, origin, race, malignancy, gender, early-PN vs late-PN, centre, and chronological age) and non-biological factors (row on chip, plate, and chip). An increasing red intensity indicates a stronger association between the two variables, based on the magnitude of the p-value of their association as assessed by ANOVA. The first three PCs explain 47.2% of the variance in the data. The heat map shows strong correlations between various principal components and the microarray chip, the well and plate on which a sample was run, which indicates that technical variation is present, warranting further correction.