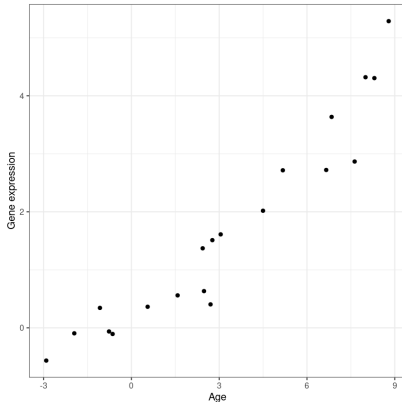
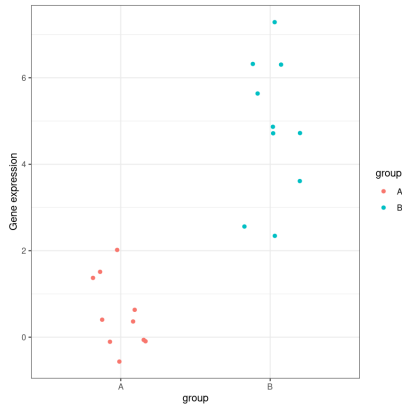


Regression with many features

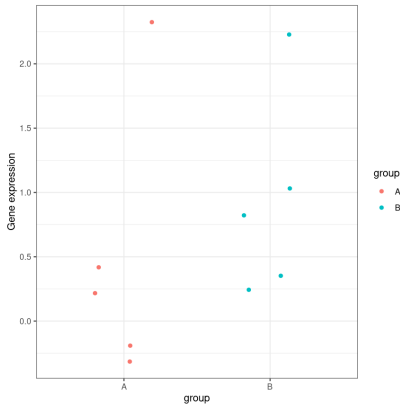
An example of a strong linear association between a continuous phenotype (age) on the x-axis and a feature of interest (gene expression for a given gene) on the y-axis. A strong linear relationship with a positive slope exists between the two.



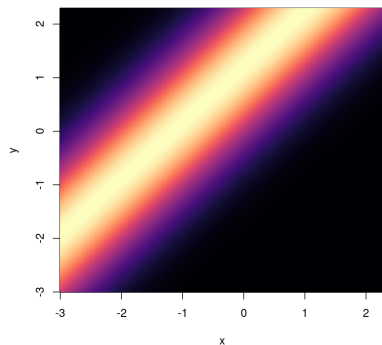
An example of a strong linear association between a discrete phenotype (group) on the x-axis and a feature of interest (gene expression for a given gene) on the y-axis. The two groups clearly differ with respect to gene expression.



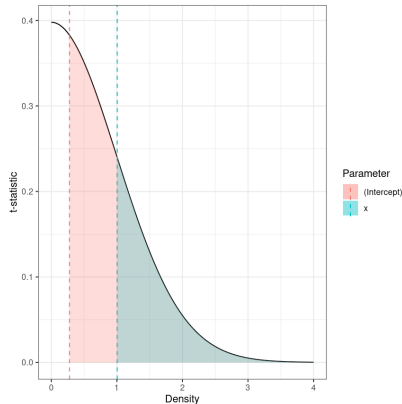
An example of a strong linear association between a discrete phenotype (group) on the x-axis and a feature of interest (gene expression for a given gene) on the y-axis. The two groups seem to differ with respect to gene expression, but the relationship is weak.



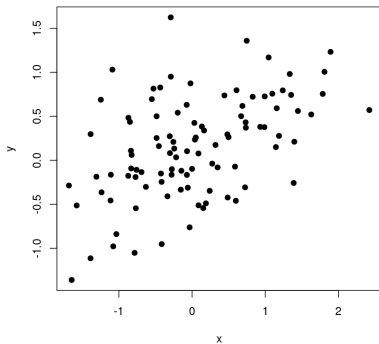
The generative model of a simple linear regression with a fixed slope and intercept. Lightly shaded regions represent regions where observations are probable, and darker regions represent lower probability.



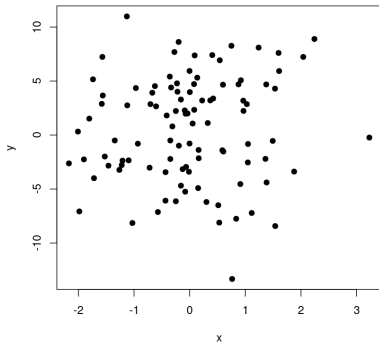
Density plot of a t-distribution showing the observed test statistics (here, t-statistics). The p-values, visualised here with shaded regions, represent the portion of the null distribution that is as extreme or more extreme as the observed test statistics, which are shown as dashed lines.



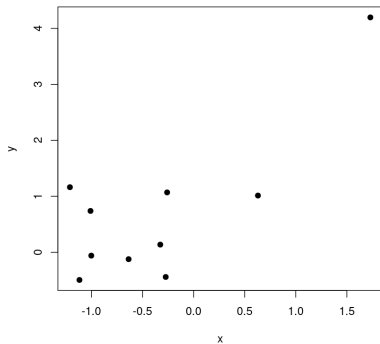
An example of a linear relationship for 100 points with a small amount of noise and small effect sizes that is statistically significant.



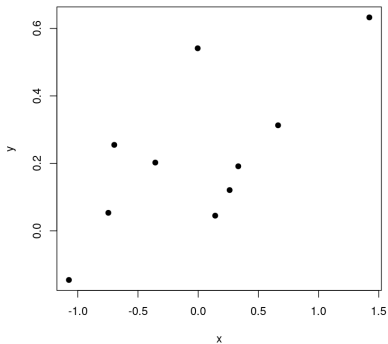
An example of a linear relationship for 100 points with a large amount of noise and large effect sizes that is not statistically significant.



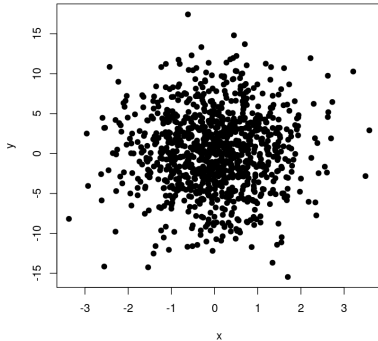
An example of a linear relationship for 10 points with a large amount of noise and large effect sizes that is not statistically significant.



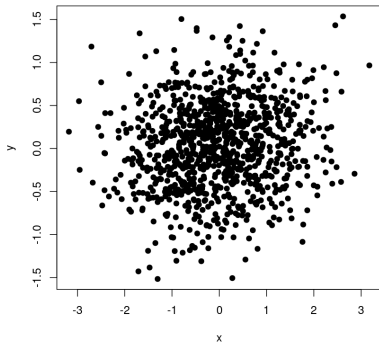
An example of a linear relationship for 10 points with a small amount of noise and small effect sizes that is statistically significant.



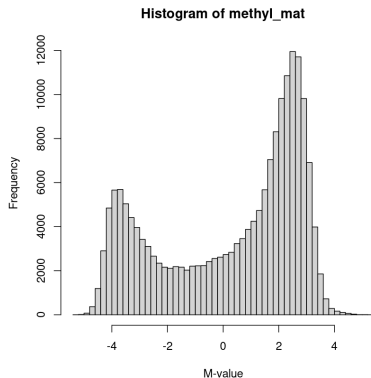
An example of a linear relationship for 1,000 points with a large amount of noise and small effect sizes that is statistically significant.



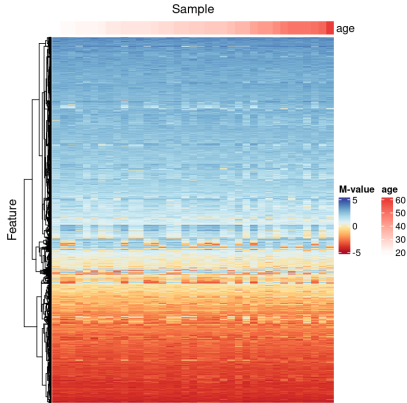
An example of a linear relationship for 1,000 points with a small amount of noise and small effect sizes that is statistically significant.



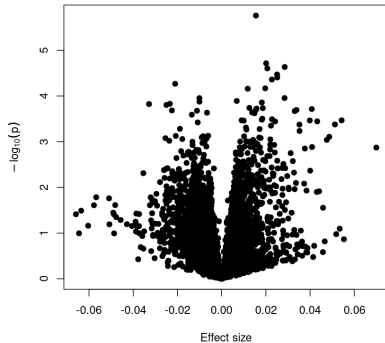
Histogram of M-values for all features. The distribution appears to be bimodal, with a large number of unmethylated features as well as many methylated features, and many intermediate features.



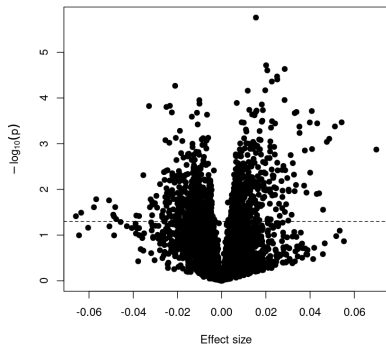
Heatmap of methylation values across all features.
Samples are ordered according to age.



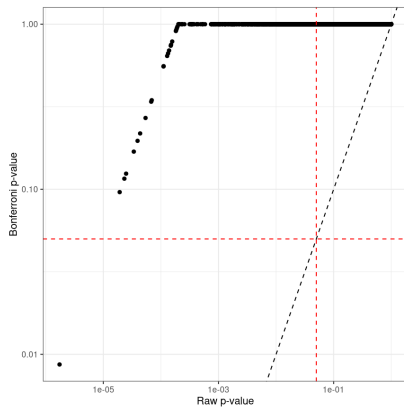
Plot of $-\log_{10}(p)$ against effect size estimates for a regression of age against methylation level for each feature in the data.



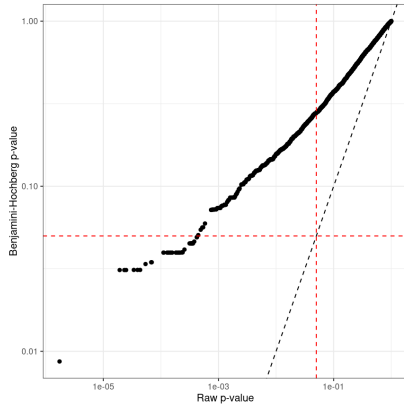
Plot of $-\log_{10}(p)$ against effect size estimates for a regression of a made-up feature against methylation level for each feature in the data. A dashed line represents a 0.05 significance level.



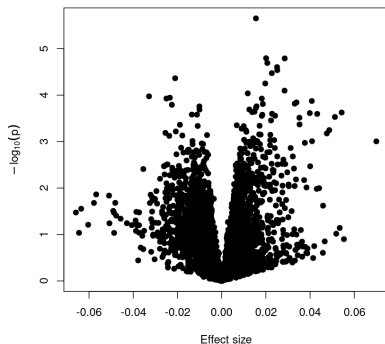
Plot of Bonferroni-adjusted p-values (y) against unadjusted p-values (x). A dashed black line represents the identity (where $x=y$), while dashed red lines represent 0.05 significance thresholds.



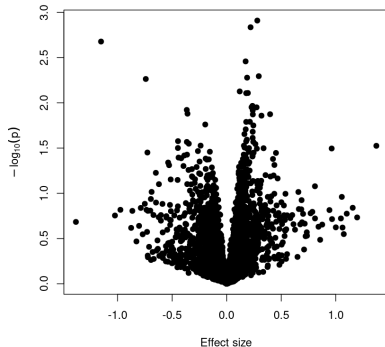
Plot of Benjamini-Hochberg-adjusted p-values (y) against unadjusted p-values (x). A dashed black line represents the identity (where $x=y$), while dashed red lines represent 0.05 significance thresholds.



A plot of $-\log_{10}(p)$ against effect size estimates for a regression of age against methylation using limma.



A plot of $-\log_{10}(p)$ against effect size estimates for a regression of smoking status against methylation using limma.



Alt

