Gene Expression Profiling in SMC Samples

Ignat Drozdov, MD PhD

March 27, 2017

Results

Overall, n = 30 expression arrays containing n = 6553600 features were included in the analysis. Robust Multiarray Averaging (RMA) algorithm was used to background correct and normalise all arrays using Affymetrix core level summarisation. Normalised data was reduced to n = 22011 features. Subsequently, all features were mapped to Entrez Gene Identifiers (IDs) and features with missing annotation were removed, yielding a final dataset consisting of n = 15185 annotated transcripts.

Microarray quality control

Boxplots of cumulative signal intensities across all samples (Figure 1) indicate that there are no outlying microarray samples given that there is no significant variability in average signal intensities across all samples.

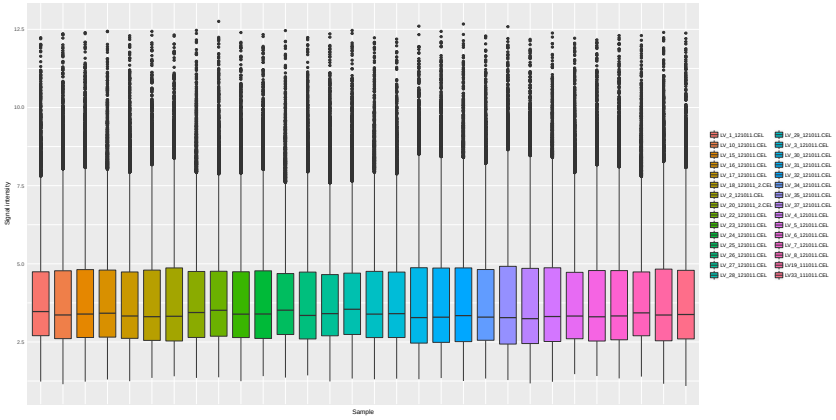

Figure 1: Boxplots visualising signal intensity distributions across all samples

Principal Component Analysis (PCA) was used to visualise genome-wide similarities and differences between patient samples (Figure 2).

Differential Expression Analysis

Differential expression of n = 15185 annotated transcripts was assessed using Empirical Bayes Statistics for Differential Expression. All transcripts with adjuster p-values < 0.05 were considered significant.

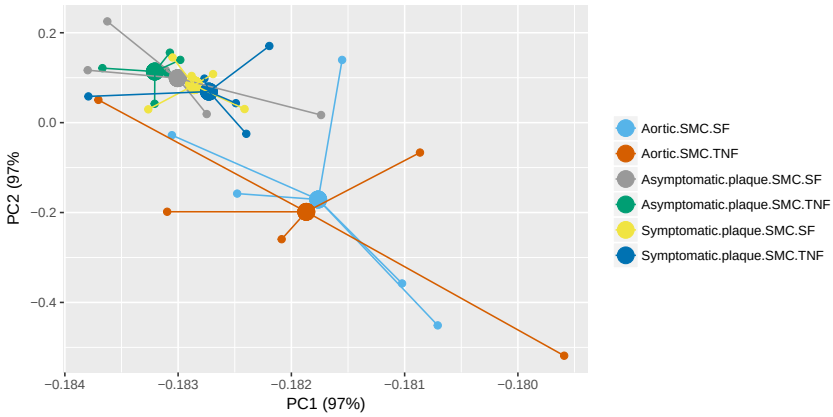

Figure 2: Principal Component Analysis (PCA) scatter plot. Smaller distances between points reflect increased similarity between biological samples on the genome-wide expression level.

There were n = 103 down-regulated and n = 9 up-regulated genes (Figure 3).

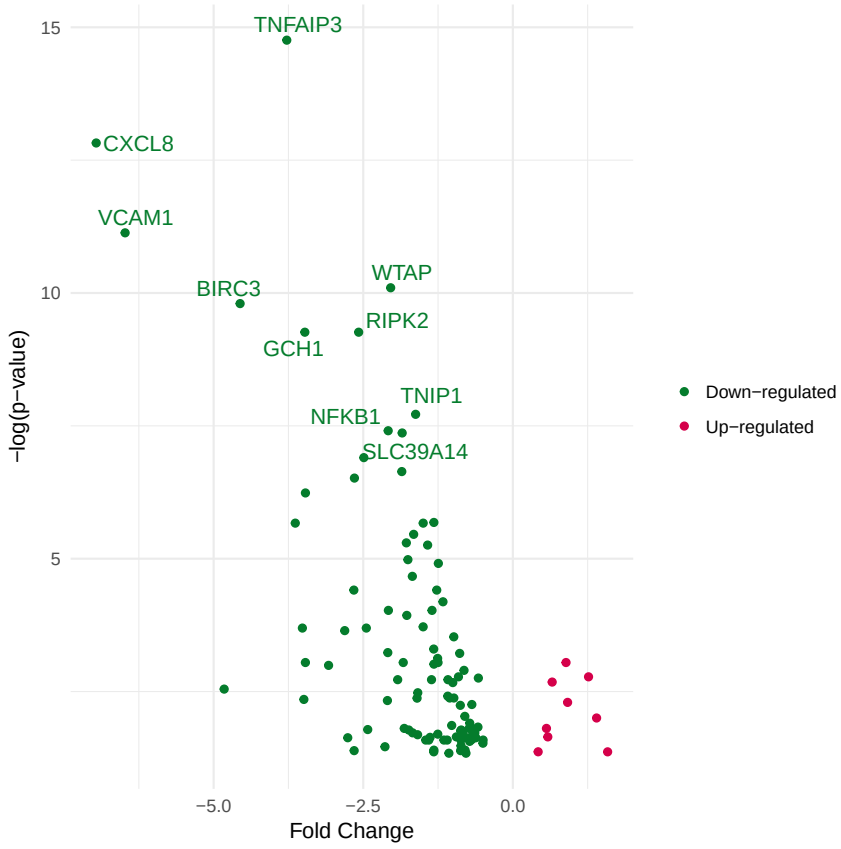

Figure 3: Volcano plot of all differentially expressed genes (adjusted p-value < 0.05). Top 10 differentially expressed genes are labeled.