

Supportive information

Combined transcriptomics analysis for classification of adverse effects as a potential endpoint in effect based screening

Tjalf E. de Boer ^{a,d*}, Thierry K.S. Janssens ^b, Juliette Legler ^c, Nico M. van Straalen ^d and Dick Roelofs ^d.

Supportive information for the above mentioned manuscript. The supportive information for this manuscript consists of two tables in Microsoft excel format. Supplementary Table 1 shows information about gene annotation and sequence for the different feature selection sets in Microsoft Excel format. Each excel tab contains information for a single set: RFE_low for the RFE low selection set, RFE_high for the RFE high selection set and LM for the LM selection set. Sequences represent contigs based on multiple ESTs which were submitted to Genbank individually. Supplementary table 2 shows the results from the Gene Ontology analysis in the Biological Process and Molecular Function categories of all three gene sets (significant GO terms only) in Microsoft Excel format. Significance is determined after the weighted adjustment in TopGO. Each excel tab contains results from a single set: RFE_low for the RFE low selection set, RFE_high for the RFE high selection set and LM for the LM selection set. Supplementary figure 1 shows density plots of the normalization efficiency of the combined dataset with the raw data (A), the data normalized within microarray (B) and the data normalized between microarrays (C). As can be seen in figure C most of the systematic variance is removed. The data from figure C were used for the final analysis. Volcano plot depicting fold change (x-axis) and significance (y-axis) of the linear model feature selection analysis between the (A) control and low groups and between the (B) control and high groups. Green dots represent significant down-regulated genes and red dots represent significant up regulated genes while black dots were non-significant genes.