Using microarray data to find link between gene expression and phenotype is .

Finding a key set of genes where changes in expression indicate a the presence or absence of a condition, such as disease state or drug response.

Chronic Myeloid Leukaemia is a disease that is successfully treated by the drug imatinib . Unfortunately 20% of patients do not respond and others become resistant to the drug after a long period of successful treatment. Being able to predict the non-responders early gives the clinicians the opportunity to try alternative treatment options early.

In this assignment I attempt to use two supervised learning classifiers, logistic regression and random forest ,to produce models that can predict drug response from micro array data.

The Data.

The dataset GSE14671was downloaded from GEO at the NCBI websites deposited by McWeeney et al(1). It consists 59 samples of microarray data each of 54675 probes, the metadata includes information on drug response.

```{r echo=FALSE}

Data Exploration

References

[1] Shannon K. McWeeney, Lucy C. Pemberton, Marc M. Loriaux, Kristina

Vartanian, Stephanie G. Willis, Gregory Yochum, Beth Wilmot, Yaron Tur-

paz, Raji Pillai, Brian J. Druker, Jennifer L. Snead, Mary MacPartlin,

Stephen G. O'Brien, Junia V. Melo, Thoralf Lange, Christina A. Harring-

ton, and Michael W. N. Deininger. A gene expression signature of cd34+

cells to predict major cytogenetic response in chronic-phase chronic myeloid

leukemia patients treated with imatinib. Blood, 115(2):315{325, 2010.

2

Correlation