The Mutational Landscape of Lethal Castrate Resistant Prostate Cancer Grasso et al. (2012) doi:10.1038/nature11125

In this document, I describe how the Agilent expression data for this datast was processed and saved into an object for analysis in Bioconductor. First of all load the relevant libraries for grabbing and manipulaing the data

```
> library(GEOquery)
```

Now use the 'getGEO' function with the correct ID

```
> geoData <- getGEO("GSE35988")
> geoData
```

We notice that the samples are spread across two different datasets, each of which is a different list entry in the object we just created. We create a new pheno data object that has the sample names and analysis groups. We also cut-down the feature matrix to just ID, Entrez and Symbol.

```
> pd <- rbind(pData(geoData[[1]]),pData(geoData[[2]]))
> groups = NULL
> groups[grep('benign', pd$characteristics_ch2, perl=TRUE)] = "Benign"
> groups[grep('localized', pd$characteristics_ch2, perl=TRUE)] = "HormomeDependant"
> groups[grep('metastatic', pd$characteristics_ch2, perl=TRUE)] = "CastrateResistant"
> pd <- data.frame(geo_accession=pd$geo_accession, Sample.Name= pd$title,Group=groups)
> fd <- fData(geoData[[1]])[,c("ID","GENE","GENE_SYMBOL")]
> exp <- cbind(exprs(geoData[[1]]), exprs(geoData[[2]]))
> geoData <- geoData[[1]]
> exprs(geoData) <- exp
> pData(geoData) <- pd
> fData(geoData) <- fd
> grasso <- geoData
> save(grasso, file="data/grasso.rda",compress="xz")
```