Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression.

Varambally et al. (2005) doi:10.1016/j.ccr.2005.10.001

In this document, I describe how the GEO data entry for the Varambally data 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("GSE3325")[[1]]
> geoData
```

The phenoData contains lots of metadata from GEO that probably won't be useful for our analysis. We just take the sample name and group colums. Groups are re-ordered according to cancer progression. We re-name the feature data so that the gene symbol column doesn't contain a space character (which will cause problems later-on).

```
> pData(geoData) <- pData(geoData)[,c("geo_accession","title","description")]
> Sample_Group <- NULL
> Sample_Group[grep("Benign", pData(geoData)$title)] <- "Benign"
> Sample_Group[grep("prostate cancer",pData(geoData)$title)] <- "Tumour"
> Sample_Group[grep("Metastatic", pData(geoData)$title)] <- "Metastatic"
> pData(geoData)$Sample_Group <- factor(Sample_Group,levels=c("Benign","Tumour","Metastatic'
> colnames(fData(geoData))[which(colnames(fData(geoData)) == "Gene Symbol")] <- "Symbol"
> varambally <- geoData
> save(varambally, file="data/varambally.rda",compress="xz")
```