9.5 Interaction Models: 2×2 Factorial Designs

9.5.1 Questions of Interest

Factorial designs are those where more than one experimental dimension is being varied and each combination of treatment conditions is observed. Suppose that cells are extracted from wild type and mutant mice and these cells are either stimulated (S) or unstimulated (U). RNA from the treated cells is then extracted and hybridized to a microarray. We will assume for simplicity that the arrays are single-color arrays such as Affymetrix. Consider the following targets frame:

FileName	Strain	Treatment
File1	WT	U
File2	WT	S
File3	Mu	U
File4	Mu	S
File5	Mu	S

The two experimental dimensions or *factors* here are Strain and Treatment. Strain specifies the genotype of the mouse from which the cells are extracted and Treatment specifies whether the cells are stimulated or not. All four combinations of Strain and Treatment are observed, so this is a factorial design. It will be convenient for us to collect the Strain/Treatment combinations into one vector as follows:

```
> TS <- paste(targets$Strain, targets$Treatment, sep=".")
> TS

[1] "WT.U" "WT.S" "Mu.U" "Mu.S" "Mu.S"
```

It is especially important with a factorial design to decide what are the comparisons of interest. We will assume here that the experimenter is interested in

1. which genes respond to stimulation in wild-type cells,

2. which genes respond to stimulation in mutant cells, and

3. which genes respond differently in mutant compared to wild-type cells.

as these are the questions which are most usually relevant in a molecular biology context. The first of these questions relates to the `WT.S` vs `WT.U` comparison and the second to `Mu.S` vs `Mu.U`. The third relates to the difference of differences, i.e., `(Mu.S-Mu.U)-(WT.S-WT.U)`, which is called the *interaction* term.

9.5.2 Analysing as for a Single Factor

We describe first a simple way to analyze this experiment using limma commands in a similar way to that in which two-sample designs were analyzed. Then we will go on to describe the more classical statistical approaches using factorial model formulas. All the approaches considered are equivalent and yield identical bottom-line results. The most basic approach is to fit a model with a coefficient for each of the four factor combinations and then to extract the comparisons of interest as contrasts:

```
> TS <- factor(TS, levels=c("WT.U","WT.S","Mu.U","Mu.S"))
> design <- model.matrix(~0+TS)
> colnames(design) <- levels(TS)
> fit <- lmFit(eset, design)
```

This fits a model with four coefficients corresponding to `WT.U`, `WT.S`, `Mu.U` and `Mu.S` respectively. Our three contrasts of interest can be extracted by

```
> cont.matrix <- makeContrasts(
+     SvsUinWT=WT.S-WT.U,
+     SvsUinMu=Mu.S-Mu.U,
+     Diff=(Mu.S-Mu.U)-(WT.S-WT.U),
+     levels=design)
> fit2 <- contrasts.fit(fit, cont.matrix)
> fit2 <- eBayes(fit2)
```

We can use `topTable()` to look at lists of differentially expressed genes for each of three contrasts, or else

```
> results <- decideTests(fit2)
> vennDiagram(results)
```

to look at all three contrasts simultaneously.

This approach is recommended for most users, because the contrasts that are being tested are formed explicitly.