Lab: cDNA microarray Preprocessing and Analysis

Nolwenn Le Meur

April 13^{th} 2006

For this lab, we make use of **Chapter 4 and 23** from *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* by R. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry and S. Dudoit.

There are two alternative, partially overlapping paths through this lab, one for people with a preference for data preprocessing, and one for people interested in using linear models for differential expression analysis of data arising from different experimental design. You will need the packages *beta7*, *marray*, *marrayQuality*, *vsn* and *limma*.

1 Data preprocessing

The last paragraph of Chapter 4 "Preprocessing Two-Color Arrays" by Y.H Yang and A.C. Paquet contains a case study for spotted chip preprocessing and quality control using the *arrayQuality* package. Working through these exercises should take you around 30 min, depending on your experience with R.

1.1 beta7 Dataset

In this case study, they make use of an experiment conducted by the Erle Lab in UC San Francisco. This experiment aims to study the cell adhesion molecule integrin alpha4/beta7 which assists in directing the migration of blood lymphocytes to the intestine and associated lymphoid tissues. The goal of the study is to identify differentially expressed genes between the alpha4/beta7+ and alpha4/beta7- memory T helper cells. The study hypothesizes that differentially expressed genes may play a role in the adhesion or migration of T cells. Further details and results of the experiments can be found in Rodriguez (2004).

The data set given here is a subset from the original dataset consisting of 6 replicated slides from different subjects. Complete information about the array platform and data from each of the individual arrays is available from GEO (accession number GSE 1039). Each hybridization involved beta 7+ cell RNA from a single subject (labeled with one dye) and beta7- cell

RNA from the same subject (labeled with the other dye). Target RNA was hybridized to microarrays containing 23,184 probes including the Operon Human version 2 set of 70-mer oligonucleotide probes and 1760 controls spots (e.g., negative, positive and normalization controls spots). Microarrays were printed using 12x4 print-tips and are thus partitioned into a 12x4 grid matrix. Each grid consists of a 21x23 spot matrix that was printed with a single print-tip.

Each of the arrays were scanned using an Axon GenePix 4000B scanner and images were processed using GenePix 5.0 image processing software. The data comprises 6 GenePix gpr output files. Each gpr file contains 23,184 rows and 56 columns; rows correspond to probes (spots) while columns correspond to different statistics from the image analysis output. The gpr files also contain probe names and IDs.

If you have brought your own data, process it in a similar way.

1.2 Supplement exercice

This exercice is adapted from G.K Smyth's Bioconductor Short Course, Seattle Aug. 2005. Try different background correction using the *backgroundCorrect* from the *limma* package and compare the resulting the MA-plots.

```
> library("limma")
> f <- function(x) as.numeric(x$Flags > -75)
> RG <- read.maimages(TargetInfo@maInfo$FileName, source="genepix", wt.fun=f)
> RG$printer <- getLayout(RG$genes)

> RGsu <- backgroundCorrect(RG, method="subtract")  # the default
> RGno <- backgroundCorrect(RG, method="none")
> RGne <- backgroundCorrect(RG, method="normexp", offset=25)
```

Examine closely the MA-plots from the three background correction methods. Notice that subtracting produces a decreasing fan effect with intensity while not background correcting produces an increasing fan effect. The 'normexp' produces a more balanced stabilisation of the variances. It also preserves all the data. To examine all the MA-plots efficiently, you may find it helpful to use the following commands, which write all the MA-plots to png disk files in compact format:

```
> plotMA3by2(RGsu, prefix="MAsu")
> plotMA3by2(RGno, prefix="MAno")
> plotMA3by2(RGne, prefix="MAne")
```

2 Differential Expression

We will make use of the *limma* package and the linear models case studies presented in Chapter 23 by G.K Smyth.

limma is a package for differential expression analysis of data from microarrays experiments. The package is designed to analyze complex experiements involving comparisons between many RNA targets simultaneously. Working through these exercises will give you a survey of differential expression analysis from the simplest replicated designs to time course experiments.

3 Memento

marray and *limma* are the 2 main packages you will be using to read and preprocess the data. They propose equivalent but different classes, objects and methods to read and analysis the raw data. Table 1 summarize the function, classes and objects handle by *marray* and *limma*.

	Preprocessing			
	***limma* package**		***marray* package**	
Action	**Function**	**Class - Object**	**marray**	**Class - Object**
read target file	readTargets	*dataframe*	read.marrayInfo	marrayInfo
read image file	read.maimages	RGList	read.marrayRaw, read.GenePix, read.Spot, read.SMD, read.Agilent	marrayRaw
read gene list	readGAL	RGList$genes	read.Galfile	marrayInfo, marrayLayout
read spot type	readSpotTypes, controlStatus	RGList$genes$status		
array layout	getLayout	RGList$printer	read.marrayLayout, Layout	marrayLayout
background correction	backgroundCorrect			
one array normalization	normalizeWithinArrays, MA.RG	MAList	maNormMain	marrayNorm
normalization between arrays	normalizeBetweenArrays	MAList		

	Differential expression	
	***limma* package**	
Action	**Function**	**Class - Object**
design matrix	modelMatrix	*matrix*
fitting model	lmFit	MArrayLM
contrast matrix	makeContrasts	*matrix*
pairwise comparaison	contrasts.fit	*list*
moderated t-statistics	ebayes, eBayes	*list*
test summary	toptable, topTable	*dataframe*
test summary	decideTests	TestResults