

Microarray2 E-GEO-D-64750

Seoyeon Oh, Tobias Heyman

12/8/2021

```
setwd("/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Datasets")
dir <- "/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Datasets"
```

Microarray2: E-GEO-D-64750

General info

The array used for this dataset is A-AFFY-45 - Affymetrix GeneChip Mouse Genome 430 2.0 [Mouse430_2]. In this experiment, susceptible mice were infected with H5N1 influenza. After 72h RNA was extracted from the lungs of the mice. We used 9 samples of this experiment (susceptible mice).

Intensity values

Read in the microarray data and examine dimensionality of the intensity value matrix.

```
id <- "E-GEO-D-64750"
exonCELs <- list.celfiles("../Datasets/Microarray2/")
data.raw_2 <- read.celfiles(paste(rep("../Datasets/Microarray2/", length(exonCELs)), exonCELs, sep=""))

## Reading in : ../Datasets/Microarray2/GSM1579245_jac013-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579246_jac021-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579247_jac044-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579248_jac045-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579249_jac046-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579250_jac007-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579251_jac008-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579252_jac036-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579253_jac047-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579254_jac014-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579255_jac024-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579256_jac025-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579257_jac011-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579258_jac012-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579259_jac037-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579260_jac048-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579261_jac015-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579262_jac016-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579263_jac017-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579264_jac018-430v2.CEL
```

```

## Reading in : ../Datasets/Microarray2/GSM1579265_jac019-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579266_jac020-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579267_jac027-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579268_jac028-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579269_jac034-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579270_jac033-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579271_jac029-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579272_jac035-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579273_jac041-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579274_jac042-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579275_jac043-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579276_jac030-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579277_jac031-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579278_jac032-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579279_jac038-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579280_jac039-430v2.CEL
## Reading in : ../Datasets/Microarray2/GSM1579281_jac040-430v2.CEL

```

```
dim(exprs(data.raw_2))
```

```
## [1] 1004004      37
```

```
head(exprs(data.raw_2))
```

	GSM1579245_jac013-430v2.CEL	GSM1579246_jac021-430v2.CEL
## 1	62	191
## 2	5707	16272
## 3	69	265
## 4	5913	16107
## 5	57	143
## 6	54	143
## GSM1579247_jac044-430v2.CEL	GSM1579248_jac045-430v2.CEL	
## 1	198	134
## 2	7485	5697
## 3	200	144
## 4	7664	6184
## 5	154	121
## 6	162	128
## GSM1579249_jac046-430v2.CEL	GSM1579250_jac007-430v2.CEL	
## 1	136	123
## 2	6036	7786
## 3	134	162
## 4	6147	7832
## 5	126	111
## 6	113	105
## GSM1579251_jac008-430v2.CEL	GSM1579252_jac036-430v2.CEL	
## 1	139	80
## 2	8743	6310
## 3	154	102
## 4	8940	6335
## 5	102	70
## 6	96	88
## GSM1579253_jac047-430v2.CEL	GSM1579254_jac014-430v2.CEL	

## 1	92	105
## 2	4988	8341
## 3	87	159
## 4	5275	8721
## 5	87	109
## 6	79	96
## GSM1579255_jac024-430v2.CEL	GSM1579256_jac025-430v2.CEL	
## 1	125	153
## 2	9656	8775
## 3	200	180
## 4	10228	8859
## 5	94	101
## 6	118	121
## GSM1579257_jac011-430v2.CEL	GSM1579258_jac012-430v2.CEL	
## 1	182	130
## 2	13154	9498
## 3	218	195
## 4	13689	9705
## 5	104	108
## 6	132	128
## GSM1579259_jac037-430v2.CEL	GSM1579260_jac048-430v2.CEL	
## 1	176	92
## 2	15363	4973
## 3	251	72
## 4	15737	5058
## 5	142	72
## 6	173	63
## GSM1579261_jac015-430v2.CEL	GSM1579262_jac016-430v2.CEL	
## 1	59	69
## 2	5959	6331
## 3	73	72
## 4	6140	6388
## 5	66	65
## 6	51	71
## GSM1579263_jac017-430v2.CEL	GSM1579264_jac018-430v2.CEL	
## 1	58	69
## 2	5856	6670
## 3	78	82
## 4	5730	6455
## 5	61	92
## 6	59	47
## GSM1579265_jac019-430v2.CEL	GSM1579266_jac020-430v2.CEL	
## 1	87	202
## 2	6029	6455
## 3	78	74
## 4	6409	6648
## 5	65	102
## 6	67	58
## GSM1579267_jac027-430v2.CEL	GSM1579268_jac028-430v2.CEL	
## 1	161	123
## 2	7684	8069
## 3	230	156
## 4	7162	8064
## 5	123	97

```

## 6          139          146
##   GSM1579269_jac034-430v2.CEL GSM1579270_jac033-430v2.CEL
## 1          193          150
## 2          14103         7298
## 3          279          231
## 4          14199         7447
## 5          156          145
## 6          197          139
##   GSM1579271_jac029-430v2.CEL GSM1579272_jac035-430v2.CEL
## 1          242          149
## 2          18215         7908
## 3          283          150
## 4          17438         8798
## 5          160          137
## 6          151          141
##   GSM1579273_jac041-430v2.CEL GSM1579274_jac042-430v2.CEL
## 1          132          184
## 2          8993         8146
## 3          185          169
## 4          9202         8344
## 5          153          180
## 6          117          152
##   GSM1579275_jac043-430v2.CEL GSM1579276_jac030-430v2.CEL
## 1          187          101
## 2          6867         6737
## 3          223           92
## 4          7165         6788
## 5          101           61
## 6          138           83
##   GSM1579277_jac031-430v2.CEL GSM1579278_jac032-430v2.CEL
## 1          74           165
## 2          7379         7416
## 3          67           201
## 4          7312         7929
## 5          65           157
## 6          77           141
##   GSM1579279_jac038-430v2.CEL GSM1579280_jac039-430v2.CEL
## 1          72           72
## 2          4883         4924
## 3          68           74
## 4          4814         4988
## 5          68           65
## 6          71           57
##   GSM1579281_jac040-430v2.CEL
## 1          81
## 2          4475
## 3          78
## 4          4646
## 5          59
## 6          67

```

Annotation

Here we provide basic sample annotation, including the phenotype of interest and relevant other features (e.g. confounders). This dataset contains array data (A-AFFY-45) of different mice strains (BXD98, BXD97, BXD83, BXD73, BXD68, BXD67 ,BXD43, C57BL/6J, DBA/2J) infected with influenza virus H5N1.

```
sdrf <- read.delim("./Microarray2/E-GEO-D-64750.sdrf.txt")
print(sdrf[,c("Source.Name", "Comment..Sample_source_name.", "Array.Design.REF", "Characteristics..strain")]
```

##	Source.Name	Comment..Sample_source_name.	
## 1	GSM1579281	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 2	GSM1579280	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 3	GSM1579279	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 4	GSM1579278	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 5	GSM1579277	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 6	GSM1579276	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 7	GSM1579275	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 8	GSM1579274	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 9	GSM1579273	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 10	GSM1579272	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 11	GSM1579271	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 12	GSM1579270	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 13	GSM1579269	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 14	GSM1579268	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 15	GSM1579267	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 16	GSM1579266	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 17	GSM1579265	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 18	GSM1579264	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 19	GSM1579263	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 20	GSM1579262	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 21	GSM1579261	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 22	GSM1579260	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 23	GSM1579259	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 24	GSM1579258	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 25	GSM1579257	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 26	GSM1579256	1	Uninfected control
## 27	GSM1579255	1	Uninfected control
## 28	GSM1579254	1	Uninfected control
## 29	GSM1579253	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 30	GSM1579252	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 31	GSM1579251	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 32	GSM1579250	1 Highly Pathogenic H5N1 Influenza A virus infected; 72 hours	
## 33	GSM1579249	1	Uninfected control
## 34	GSM1579248	1	Uninfected control
## 35	GSM1579247	1	Uninfected control
## 36	GSM1579246	1	Uninfected control
## 37	GSM1579245	1	Uninfected control
##	Array.Design.REF	Characteristics..strain.	
## 1	A-AFFY-45	BXD98	
## 2	A-AFFY-45	BXD98	
## 3	A-AFFY-45	BXD98	
## 4	A-AFFY-45	BXD97	
## 5	A-AFFY-45	BXD97	
## 6	A-AFFY-45	BXD97	

```

## 7      A-AFFY-45          BXD83
## 8      A-AFFY-45          BXD83
## 9      A-AFFY-45          BXD83
## 10     A-AFFY-45          BXD73
## 11     A-AFFY-45          BXD73
## 12     A-AFFY-45          BXD73
## 13     A-AFFY-45          BXD68
## 14     A-AFFY-45          BXD68
## 15     A-AFFY-45          BXD68
## 16     A-AFFY-45          BXD67
## 17     A-AFFY-45          BXD67
## 18     A-AFFY-45          BXD67
## 19     A-AFFY-45          BXD43
## 20     A-AFFY-45          BXD43
## 21     A-AFFY-45          BXD43
## 22     A-AFFY-45          C57BL/6J
## 23     A-AFFY-45          C57BL/6J
## 24     A-AFFY-45          C57BL/6J
## 25     A-AFFY-45          C57BL/6J
## 26     A-AFFY-45          C57BL/6J
## 27     A-AFFY-45          C57BL/6J
## 28     A-AFFY-45          C57BL/6J
## 29     A-AFFY-45          DBA/2J
## 30     A-AFFY-45          DBA/2J
## 31     A-AFFY-45          DBA/2J
## 32     A-AFFY-45          DBA/2J
## 33     A-AFFY-45          DBA/2J
## 34     A-AFFY-45          DBA/2J
## 35     A-AFFY-45          DBA/2J
## 36     A-AFFY-45          DBA/2J
## 37     A-AFFY-45          DBA/2J

```

Comment..Sample_description.

```

## 1 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 2 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 3 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 4 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 5 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 6 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 7 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 8 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 9 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 10 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 11 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 12 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 13 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 14 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 15 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 16 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 17 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 18 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 19 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 20 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 21 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 22 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice

```

```

## 23 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 24 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 25 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 26                                     Gene expression data from lungs of uninfected mice
## 27                                     Gene expression data from lungs of uninfected mice
## 28                                     Gene expression data from lungs of uninfected mice
## 29 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 30 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 31 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 32 Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected mice
## 33                                     Gene expression data from lungs of uninfected mice
## 34                                     Gene expression data from lungs of uninfected mice
## 35                                     Gene expression data from lungs of uninfected mice
## 36                                     Gene expression data from lungs of uninfected mice
## 37                                     Gene expression data from lungs of uninfected mice

```

Which samples we are using, and not using: We will be using samples involving susceptible and resistant mouse strain, DBA/2J (GSM1579245 - GSM1579253) and C57BL/6J (GSM1579254 - GSM1579260) respectively. Each strain was inoculated with H5N1 influenza A virus. We are not using the data from other strains (BXD98, BXD97, BXD83, BXD73, BXD68, BXD67, BXD43) which do not contain non-infected control samples.

```

#Load in the ExpressionFeatureSet object
setwd("/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Database")
MouseExp_AE2 <- ArrayExpress("E-GEO-64750")

```

```
## Copying raw data files
```

```
## Unpacking data files
```

```
## ArrayExpress: Reading pheno data from SDRF
```

```
## ArrayExpress: Reading data files
```

```
## Platform design info loaded.
```

```

## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579281_jac040-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579280_jac039-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579279_jac038-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579278_jac032-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579277_jac031-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579276_jac030-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579275_jac043-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579274_jac042-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579273_jac041-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579272_jac035-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579271_jac029-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579270_jac033-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579269_jac034-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579268_jac028-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579267_jac027-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579266_jac020-430v2.CEL

```

```

## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579265_jac019-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579264_jac018-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579263_jac017-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579262_jac016-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579261_jac015-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579260_jac048-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579259_jac037-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579258_jac012-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579257_jac011-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579256_jac025-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579255_jac024-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579254_jac014-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579253_jac047-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579252_jac036-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579251_jac008-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579250_jac007-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579249_jac046-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579248_jac045-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579247_jac044-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579246_jac021-430v2.CEL
## Reading in : /var/folders/my/9t5kkz3n4x15jqxc9cf6hy740000gn/T//RtmpBy1Wqu/GSM1579245_jac013-430v2.CEL

##
## E-GEOID-64750 was successfully loaded into ExpressionFeatureSet

#Load in the Affybatch object
setwd("/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Database")
MouseExp_pheno2 <- ReadAffy(phenoData=pData(MouseExp_AE2))

#load first lines of output from the object
head(exprs(MouseExp_pheno2))

##      GSM1579281_jac040-430v2.CEL GSM1579280_jac039-430v2.CEL
## 1          81                  72
## 2         4475                 4924
## 3          78                  74
## 4         4646                 4988
## 5          59                  65
## 6          67                  57
##      GSM1579279_jac038-430v2.CEL GSM1579278_jac032-430v2.CEL
## 1          72                  165
## 2         4883                 7416
## 3          68                  201
## 4         4814                 7929
## 5          68                  157
## 6          71                  141
##      GSM1579277_jac031-430v2.CEL GSM1579276_jac030-430v2.CEL
## 1          74                  101
## 2         7379                 6737
## 3          67                  92
## 4         7312                 6788
## 5          65                  61
## 6          77                  83

```

```

##  GSM1579275_jac043-430v2.CEL  GSM1579274_jac042-430v2.CEL
## 1          187          184
## 2          6867         8146
## 3          223          169
## 4          7165         8344
## 5          101          180
## 6          138          152
##  GSM1579273_jac041-430v2.CEL  GSM1579272_jac035-430v2.CEL
## 1          132          149
## 2          8993         7908
## 3          185          150
## 4          9202         8798
## 5          153          137
## 6          117          141
##  GSM1579271_jac029-430v2.CEL  GSM1579270_jac033-430v2.CEL
## 1          242          150
## 2          18215         7298
## 3          283          231
## 4          17438         7447
## 5          160          145
## 6          151          139
##  GSM1579269_jac034-430v2.CEL  GSM1579268_jac028-430v2.CEL
## 1          193          123
## 2          14103         8069
## 3          279          156
## 4          14199         8064
## 5          156          97
## 6          197          146
##  GSM1579267_jac027-430v2.CEL  GSM1579266_jac020-430v2.CEL
## 1          161          202
## 2          7684          6455
## 3          230           74
## 4          7162          6648
## 5          123          102
## 6          139           58
##  GSM1579265_jac019-430v2.CEL  GSM1579264_jac018-430v2.CEL
## 1          87            69
## 2          6029          6670
## 3          78            82
## 4          6409          6455
## 5          65            92
## 6          67            47
##  GSM1579263_jac017-430v2.CEL  GSM1579262_jac016-430v2.CEL
## 1          58            69
## 2          5856          6331
## 3          78            72
## 4          5730          6388
## 5          61            65
## 6          59            71
##  GSM1579261_jac015-430v2.CEL  GSM1579260_jac048-430v2.CEL
## 1          59            92
## 2          5959          4973
## 3          73            72
## 4          6140          5058

```

## 5	66	72
## 6	51	63
## GSM1579259_jac037-430v2.CEL	GSM1579258_jac012-430v2.CEL	
## 1	176	130
## 2	15363	9498
## 3	251	195
## 4	15737	9705
## 5	142	108
## 6	173	128
## GSM1579257_jac011-430v2.CEL	GSM1579256_jac025-430v2.CEL	
## 1	182	153
## 2	13154	8775
## 3	218	180
## 4	13689	8859
## 5	104	101
## 6	132	121
## GSM1579255_jac024-430v2.CEL	GSM1579254_jac014-430v2.CEL	
## 1	125	105
## 2	9656	8341
## 3	200	159
## 4	10228	8721
## 5	94	109
## 6	118	96
## GSM1579253_jac047-430v2.CEL	GSM1579252_jac036-430v2.CEL	
## 1	92	80
## 2	4988	6310
## 3	87	102
## 4	5275	6335
## 5	87	70
## 6	79	88
## GSM1579251_jac008-430v2.CEL	GSM1579250_jac007-430v2.CEL	
## 1	139	123
## 2	8743	7786
## 3	154	162
## 4	8940	7832
## 5	102	111
## 6	96	105
## GSM1579249_jac046-430v2.CEL	GSM1579248_jac045-430v2.CEL	
## 1	136	134
## 2	6036	5697
## 3	134	144
## 4	6147	6184
## 5	126	121
## 6	113	128
## GSM1579247_jac044-430v2.CEL	GSM1579246_jac021-430v2.CEL	
## 1	198	191
## 2	7485	16272
## 3	200	265
## 4	7664	16107
## 5	154	143
## 6	162	143
## GSM1579245_jac013-430v2.CEL		
## 1	62	
## 2	5707	

```

## 3          69
## 4          5913
## 5          57
## 6          54

head(pData(MouseExp_pheno2)) #source name, comment sample description, sample source name, sample titl

##                                         Source.Name
## GSM1579281_jac040-430v2.CEL GSM1579281 1
## GSM1579280_jac039-430v2.CEL GSM1579280 1
## GSM1579279_jac038-430v2.CEL GSM1579279 1
## GSM1579278_jac032-430v2.CEL GSM1579278 1
## GSM1579277_jac031-430v2.CEL GSM1579277 1
## GSM1579276_jac030-430v2.CEL GSM1579276 1
##                                         Comment..Sample_name
## GSM1579281_jac040-430v2.CEL Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected; 72 hours
## GSM1579280_jac039-430v2.CEL Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected; 72 hours
## GSM1579279_jac038-430v2.CEL Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected; 72 hours
## GSM1579278_jac032-430v2.CEL Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected; 72 hours
## GSM1579277_jac031-430v2.CEL Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected; 72 hours
## GSM1579276_jac030-430v2.CEL Gene expression data from lungs of highly pathogenic H5N1 influenza A virus infected; 72 hours
##                                         Comment..Sample_source_name.
## GSM1579281_jac040-430v2.CEL Highly Pathogenic H5N1 Influenza A virus infected; 72 hours
## GSM1579280_jac039-430v2.CEL Highly Pathogenic H5N1 Influenza A virus infected; 72 hours
## GSM1579279_jac038-430v2.CEL Highly Pathogenic H5N1 Influenza A virus infected; 72 hours
## GSM1579278_jac032-430v2.CEL Highly Pathogenic H5N1 Influenza A virus infected; 72 hours
## GSM1579277_jac031-430v2.CEL Highly Pathogenic H5N1 Influenza A virus infected; 72 hours
## GSM1579276_jac030-430v2.CEL Highly Pathogenic H5N1 Influenza A virus infected; 72 hours
##                                         Characteristics..age. Characteristics..organism.
## GSM1579281_jac040-430v2.CEL          6-8wk old           Mus musculus
## GSM1579280_jac039-430v2.CEL          6-8wk old           Mus musculus
## GSM1579279_jac038-430v2.CEL          6-8wk old           Mus musculus
## GSM1579278_jac032-430v2.CEL          6-8wk old           Mus musculus
## GSM1579277_jac031-430v2.CEL          6-8wk old           Mus musculus
## GSM1579276_jac030-430v2.CEL          6-8wk old           Mus musculus
##                                         Term.Source.REF
## GSM1579281_jac040-430v2.CEL          EFO
## GSM1579280_jac039-430v2.CEL          EFO
## GSM1579279_jac038-430v2.CEL          EFO
## GSM1579278_jac032-430v2.CEL          EFO
## GSM1579277_jac031-430v2.CEL          EFO
## GSM1579276_jac030-430v2.CEL          EFO
##                                         Term.Accession.Number
## GSM1579281_jac040-430v2.CEL http://purl.obolibrary.org/obo/NCBITaxon_10090
## GSM1579280_jac039-430v2.CEL http://purl.obolibrary.org/obo/NCBITaxon_10090
## GSM1579279_jac038-430v2.CEL http://purl.obolibrary.org/obo/NCBITaxon_10090
## GSM1579278_jac032-430v2.CEL http://purl.obolibrary.org/obo/NCBITaxon_10090

```

```

## GSM1579277_jac031-430v2.CEL http://purl.obolibrary.org/obo/NCBITaxon_10090
## GSM1579276_jac030-430v2.CEL http://purl.obolibrary.org/obo/NCBITaxon_10090
## Characteristics..sex. Term.Source.REF.1
## GSM1579281_jac040-430v2.CEL female EFO
## GSM1579280_jac039-430v2.CEL female EFO
## GSM1579279_jac038-430v2.CEL female EFO
## GSM1579278_jac032-430v2.CEL female EFO
## GSM1579277_jac031-430v2.CEL female EFO
## GSM1579276_jac030-430v2.CEL female EFO
## Term.Accession.Number.1 Characteristics..strain.
## GSM1579281_jac040-430v2.CEL EFO_0001265 BXD98
## GSM1579280_jac039-430v2.CEL EFO_0001265 BXD98
## GSM1579279_jac038-430v2.CEL EFO_0001265 BXD98
## GSM1579278_jac032-430v2.CEL EFO_0001265 BXD97
## GSM1579277_jac031-430v2.CEL EFO_0001265 BXD97
## GSM1579276_jac030-430v2.CEL EFO_0001265 BXD97
## Term.Source.REF.2 Term.Accession.Number.2
## GSM1579281_jac040-430v2.CEL
## GSM1579280_jac039-430v2.CEL
## GSM1579279_jac038-430v2.CEL
## GSM1579278_jac032-430v2.CEL
## GSM1579277_jac031-430v2.CEL
## GSM1579276_jac030-430v2.CEL
## Protocol.REF Term.Source.REF.3 Protocol.REF.1
## GSM1579281_jac040-430v2.CEL P-GSE64750-2 ArrayExpress P-GSE64750-3
## GSM1579280_jac039-430v2.CEL P-GSE64750-2 ArrayExpress P-GSE64750-3
## GSM1579279_jac038-430v2.CEL P-GSE64750-2 ArrayExpress P-GSE64750-3
## GSM1579278_jac032-430v2.CEL P-GSE64750-2 ArrayExpress P-GSE64750-3
## GSM1579277_jac031-430v2.CEL P-GSE64750-2 ArrayExpress P-GSE64750-3
## GSM1579276_jac030-430v2.CEL P-GSE64750-2 ArrayExpress P-GSE64750-3
## Term.Source.REF.4 Extract.Name
## GSM1579281_jac040-430v2.CEL ArrayExpress GSM1579281 extract 1
## GSM1579280_jac039-430v2.CEL ArrayExpress GSM1579280 extract 1
## GSM1579279_jac038-430v2.CEL ArrayExpress GSM1579279 extract 1
## GSM1579278_jac032-430v2.CEL ArrayExpress GSM1579278 extract 1
## GSM1579277_jac031-430v2.CEL ArrayExpress GSM1579277 extract 1
## GSM1579276_jac030-430v2.CEL ArrayExpress GSM1579276 extract 1
## Material.Type Protocol.REF.2 Term.Source.REF.5
## GSM1579281_jac040-430v2.CEL total RNA P-GSE64750-4 ArrayExpress
## GSM1579280_jac039-430v2.CEL total RNA P-GSE64750-4 ArrayExpress
## GSM1579279_jac038-430v2.CEL total RNA P-GSE64750-4 ArrayExpress
## GSM1579278_jac032-430v2.CEL total RNA P-GSE64750-4 ArrayExpress
## GSM1579277_jac031-430v2.CEL total RNA P-GSE64750-4 ArrayExpress
## GSM1579276_jac030-430v2.CEL total RNA P-GSE64750-4 ArrayExpress
## Labeled.Extract.Name Label Protocol.REF.3
## GSM1579281_jac040-430v2.CEL GSM1579281 LE 1 biotin P-GSE64750-5
## GSM1579280_jac039-430v2.CEL GSM1579280 LE 1 biotin P-GSE64750-5
## GSM1579279_jac038-430v2.CEL GSM1579279 LE 1 biotin P-GSE64750-5
## GSM1579278_jac032-430v2.CEL GSM1579278 LE 1 biotin P-GSE64750-5
## GSM1579277_jac031-430v2.CEL GSM1579277 LE 1 biotin P-GSE64750-5
## GSM1579276_jac030-430v2.CEL GSM1579276 LE 1 biotin P-GSE64750-5
## Term.Source.REF.6 Assay.Name Array.Design.REF
## GSM1579281_jac040-430v2.CEL ArrayExpress GSM1579281 A-AFFY-45
## GSM1579280_jac039-430v2.CEL ArrayExpress GSM1579280 A-AFFY-45

```

```

## GSM1579279_jac038-430v2.CEL      ArrayExpress GSM1579279      A-AFFY-45
## GSM1579278_jac032-430v2.CEL      ArrayExpress GSM1579278      A-AFFY-45
## GSM1579277_jac031-430v2.CEL      ArrayExpress GSM1579277      A-AFFY-45
## GSM1579276_jac030-430v2.CEL      ArrayExpress GSM1579276      A-AFFY-45
##                                         Term.Source.REF.7 Technology.Type Protocol.REF.4
## GSM1579281_jac040-430v2.CEL      ArrayExpress      array assay    P-GSE64750-6
## GSM1579280_jac039-430v2.CEL      ArrayExpress      array assay    P-GSE64750-6
## GSM1579279_jac038-430v2.CEL      ArrayExpress      array assay    P-GSE64750-6
## GSM1579278_jac032-430v2.CEL      ArrayExpress      array assay    P-GSE64750-6
## GSM1579277_jac031-430v2.CEL      ArrayExpress      array assay    P-GSE64750-6
## GSM1579276_jac030-430v2.CEL      ArrayExpress      array assay    P-GSE64750-6
##                                         Term.Source.REF.8          Array.Data.File
## GSM1579281_jac040-430v2.CEL      ArrayExpress GSM1579281_jac040-430v2.CEL
## GSM1579280_jac039-430v2.CEL      ArrayExpress GSM1579280_jac039-430v2.CEL
## GSM1579279_jac038-430v2.CEL      ArrayExpress GSM1579279_jac038-430v2.CEL
## GSM1579278_jac032-430v2.CEL      ArrayExpress GSM1579278_jac032-430v2.CEL
## GSM1579277_jac031-430v2.CEL      ArrayExpress GSM1579277_jac031-430v2.CEL
## GSM1579276_jac030-430v2.CEL      ArrayExpress GSM1579276_jac030-430v2.CEL
##                                         Term.Source.REF.9
## GSM1579281_jac040-430v2.CEL      P-GSE64750-1      ArrayExpress
## GSM1579280_jac039-430v2.CEL      P-GSE64750-1      ArrayExpress
## GSM1579279_jac038-430v2.CEL      P-GSE64750-1      ArrayExpress
## GSM1579278_jac032-430v2.CEL      P-GSE64750-1      ArrayExpress
## GSM1579277_jac031-430v2.CEL      P-GSE64750-1      ArrayExpress
## GSM1579276_jac030-430v2.CEL      P-GSE64750-1      ArrayExpress
##                                         Normalization.Name
## GSM1579281_jac040-430v2.CEL      GSM1579281_sample_table.txt norm
## GSM1579280_jac039-430v2.CEL      GSM1579280_sample_table.txt norm
## GSM1579279_jac038-430v2.CEL      GSM1579279_sample_table.txt norm
## GSM1579278_jac032-430v2.CEL      GSM1579278_sample_table.txt norm
## GSM1579277_jac031-430v2.CEL      GSM1579277_sample_table.txt norm
## GSM1579276_jac030-430v2.CEL      GSM1579276_sample_table.txt norm
##                                         Derived.Array.Data.File
## GSM1579281_jac040-430v2.CEL      GSM1579281_sample_table.txt
## GSM1579280_jac039-430v2.CEL      GSM1579280_sample_table.txt
## GSM1579279_jac038-430v2.CEL      GSM1579279_sample_table.txt
## GSM1579278_jac032-430v2.CEL      GSM1579278_sample_table.txt
## GSM1579277_jac031-430v2.CEL      GSM1579277_sample_table.txt
## GSM1579276_jac030-430v2.CEL      GSM1579276_sample_table.txt
##                                         FactorValue..strain. Term.Source.REF.10
## GSM1579281_jac040-430v2.CEL      ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/GEOD/E-GEOID
## GSM1579280_jac039-430v2.CEL      ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/GEOD/E-GEOID
## GSM1579279_jac038-430v2.CEL      ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/GEOD/E-GEOID
## GSM1579278_jac032-430v2.CEL      ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/GEOD/E-GEOID
## GSM1579277_jac031-430v2.CEL      ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/GEOD/E-GEOID
## GSM1579276_jac030-430v2.CEL      ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/GEOD/E-GEOID

```

```

## GSM1579281_jac040-430v2.CEL           BXD98
## GSM1579280_jac039-430v2.CEL           BXD98
## GSM1579279_jac038-430v2.CEL           BXD98
## GSM1579278_jac032-430v2.CEL           BXD97
## GSM1579277_jac031-430v2.CEL           BXD97
## GSM1579276_jac030-430v2.CEL           BXD97
##                                         Term.Accession.Number.3
## GSM1579281_jac040-430v2.CEL
## GSM1579280_jac039-430v2.CEL
## GSM1579279_jac038-430v2.CEL
## GSM1579278_jac032-430v2.CEL
## GSM1579277_jac031-430v2.CEL
## GSM1579276_jac030-430v2.CEL

filter2 <- colnames(data.raw_2)[data.raw_2@phenoData@data$index <= 16]
filter2

## [1] "GSM1579245_jac013-430v2.CEL" "GSM1579246_jac021-430v2.CEL"
## [3] "GSM1579247_jac044-430v2.CEL" "GSM1579248_jac045-430v2.CEL"
## [5] "GSM1579249_jac046-430v2.CEL" "GSM1579250_jac007-430v2.CEL"
## [7] "GSM1579251_jac008-430v2.CEL" "GSM1579252_jac036-430v2.CEL"
## [9] "GSM1579253_jac047-430v2.CEL" "GSM1579254_jac014-430v2.CEL"
## [11] "GSM1579255_jac024-430v2.CEL" "GSM1579256_jac025-430v2.CEL"
## [13] "GSM1579257_jac011-430v2.CEL" "GSM1579258_jac012-430v2.CEL"
## [15] "GSM1579259_jac037-430v2.CEL" "GSM1579260_jac048-430v2.CEL"

filtered2 <- data.raw_2[,filter2]
filtered2

## ExpressionFeatureSet (storageMode: lockedEnvironment)
## assayData: 1004004 features, 16 samples
##   element names: exprs
## protocolData
##   rowNames: GSM1579245_jac013-430v2.CEL GSM1579246_jac021-430v2.CEL ...
##             GSM1579260_jac048-430v2.CEL (16 total)
##   varLabels: exprs dates
##   varMetadata: labelDescription channel
## phenoData
##   rowNames: GSM1579245_jac013-430v2.CEL GSM1579246_jac021-430v2.CEL ...
##             GSM1579260_jac048-430v2.CEL (16 total)
##   varLabels: index
##   varMetadata: labelDescription channel
## featureData: none
## experimentData: use 'experimentData(object)'
## Annotation: pd.mouse430.2

dim(exprs(filtered)) #1004004 features      9 samples

## Error in h(simpleError(msg, call)): error in evaluating the argument 'object' in selecting a method :

```

```

head(exprs(filtered2))

##   GSM1579245_jac013-430v2.CEL GSM1579246_jac021-430v2.CEL
## 1                      62                      191
## 2                     5707                   16272
## 3                      69                      265
## 4                     5913                  16107
## 5                      57                      143
## 6                      54                      143
##   GSM1579247_jac044-430v2.CEL GSM1579248_jac045-430v2.CEL
## 1                      198                     134
## 2                     7485                   5697
## 3                      200                     144
## 4                     7664                  6184
## 5                      154                     121
## 6                      162                     128
##   GSM1579249_jac046-430v2.CEL GSM1579250_jac007-430v2.CEL
## 1                      136                     123
## 2                     6036                   7786
## 3                      134                     162
## 4                     6147                  7832
## 5                      126                     111
## 6                      113                     105
##   GSM1579251_jac008-430v2.CEL GSM1579252_jac036-430v2.CEL
## 1                      139                      80
## 2                     8743                  6310
## 3                      154                     102
## 4                     8940                  6335
## 5                      102                      70
## 6                      96                      88
##   GSM1579253_jac047-430v2.CEL GSM1579254_jac014-430v2.CEL
## 1                      92                      105
## 2                     4988                 8341
## 3                      87                      159
## 4                     5275                  8721
## 5                      87                      109
## 6                      79                      96
##   GSM1579255_jac024-430v2.CEL GSM1579256_jac025-430v2.CEL
## 1                      125                     153
## 2                     9656                   8775
## 3                      200                     180
## 4                     10228                  8859
## 5                      94                      101
## 6                      118                     121
##   GSM1579257_jac011-430v2.CEL GSM1579258_jac012-430v2.CEL
## 1                      182                     130
## 2                     13154                  9498
## 3                      218                     195
## 4                     13689                  9705
## 5                      104                     108
## 6                      132                     128
##   GSM1579259_jac037-430v2.CEL GSM1579260_jac048-430v2.CEL
## 1                      176                      92

```

```

## 2          15363          4973
## 3          251            72
## 4          15737          5058
## 5          142            72
## 6          173            63

#arrayQualityMetrics(filtered2,outdir=".~/raw2",force=T)
#arrayQualityMetrics(filtered2,outdir=".~/rawlog2",force=T,do.logtransform=T)

miceRMA <- oligo:::rma(filtered2, background=T)

## Background correcting
## Normalizing
## Calculating Expression

head(miceRMA)

## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 6 features, 16 samples
##   element names: exprs
## protocolData
##   rowNames: GSM1579245_jac013-430v2.CEL GSM1579246_jac021-430v2.CEL ...
##   GSM1579260_jac048-430v2.CEL (16 total)
##   varLabels: exprs dates
##   varMetadata: labelDescription channel
## phenoData
##   rowNames: GSM1579245_jac013-430v2.CEL GSM1579246_jac021-430v2.CEL ...
##   GSM1579260_jac048-430v2.CEL (16 total)
##   varLabels: index
##   varMetadata: labelDescription channel
## featureData: none
## experimentData: use 'experimentData(object)'
## Annotation: pd.mouse430.2

#arrayQualityMetrics(miceRMA, outdir=".~/rma2", force=TRUE)

## Differential expression analysis with RMA preprocessed data
#####
## Additional preprocessing
samples <- c(replicate(5, "DBA/2J control"), replicate(4, "DBA/2J infected"), replicate(3, "C57/BL6J control"))
samples

## [1] "DBA/2J control"      "DBA/2J control"      "DBA/2J control"
## [4] "DBA/2J control"      "DBA/2J control"      "DBA/2J infected"
## [7] "DBA/2J infected"     "DBA/2J infected"     "DBA/2J infected"
## [10] "C57/BL6J control"    "C57/BL6J control"    "C57/BL6J control"
## [13] "C57/BL6J infected"   "C57/BL6J infected"   "C57/BL6J infected"
## [16] "C57/BL6J infected"

```

```

condition <- c(replicate(5, "control"), replicate(4, "infected"), replicate(3, "control"), replicate(4, "control"))

pData(miceRMA)[,2] <- condition
pData(miceRMA)[,3] <- c(replicate(9, "DBA/2J"), replicate(7, "C57/BL6J"))
pData(miceRMA)[,4] <- samples

colnames(pData(miceRMA)) <- c("index", "condition", "strain", "samples")
pData(miceRMA)

```

	index	condition	strain	samples
## GSM1579245_jac013-430v2.CEL	1	control	DBA/2J	DBA/2J control
## GSM1579246_jac021-430v2.CEL	2	control	DBA/2J	DBA/2J control
## GSM1579247_jac044-430v2.CEL	3	control	DBA/2J	DBA/2J control
## GSM1579248_jac045-430v2.CEL	4	control	DBA/2J	DBA/2J control
## GSM1579249_jac046-430v2.CEL	5	control	DBA/2J	DBA/2J control
## GSM1579250_jac007-430v2.CEL	6	infected	DBA/2J	DBA/2J infected
## GSM1579251_jac008-430v2.CEL	7	infected	DBA/2J	DBA/2J infected
## GSM1579252_jac036-430v2.CEL	8	infected	DBA/2J	DBA/2J infected
## GSM1579253_jac047-430v2.CEL	9	infected	DBA/2J	DBA/2J infected
## GSM1579254_jac014-430v2.CEL	10	control	C57/BL6J	C57/BL6J control
## GSM1579255_jac024-430v2.CEL	11	control	C57/BL6J	C57/BL6J control
## GSM1579256_jac025-430v2.CEL	12	control	C57/BL6J	C57/BL6J control
## GSM1579257_jac011-430v2.CEL	13	infected	C57/BL6J	C57/BL6J infected
## GSM1579258_jac012-430v2.CEL	14	infected	C57/BL6J	C57/BL6J infected
## GSM1579259_jac037-430v2.CEL	15	infected	C57/BL6J	C57/BL6J infected
## GSM1579260_jac048-430v2.CEL	16	infected	C57/BL6J	C57/BL6J infected

The variability of the strain is encompassed in the model but you do not test for it

```

condition <- factor(pData(miceRMA)[,2])
strain <- factor(pData(miceRMA)[,3])
condition

## [1] control control control control control infected infected infected
## [9] infected control control control infected infected infected
## Levels: control infected

strain

## [1] DBA/2J   DBA/2J   DBA/2J   DBA/2J   DBA/2J   DBA/2J   DBA/2J
## [9] DBA/2J   C57/BL6J C57/BL6J C57/BL6J C57/BL6J C57/BL6J C57/BL6J C57/BL6J
## Levels: C57/BL6J DBA/2J

design <- model.matrix(~0+condition*strain)
colnames(design)<-c("Control", "Infected", "strain", "interaction")
# for strain, 1 represents DBA/2J, 0 represents c57/BL6
design

##      Control Infected strain interaction
## 1          1        0       1         0
## 2          1        0       1         0

```

```

## 3      1      0      1      0
## 4      1      0      1      0
## 5      1      0      1      0
## 6      0      1      1      1
## 7      0      1      1      1
## 8      0      1      1      1
## 9      0      1      1      1
## 10     1      0      0      0
## 11     1      0      0      0
## 12     1      0      0      0
## 13     0      1      0      0
## 14     0      1      0      0
## 15     0      1      0      0
## 16     0      1      0      0

## attr(),"assign")
## [1] 1 1 2 3
## attr(),"contrasts")
## attr(),"contrasts")$condition
## [1] "contr.treatment"
##
## attr(),"contrasts")$strain
## [1] "contr.treatment"

fit_m2 <- lmFit(miceRMA, design)
cont.matrix <- makeContrasts(InfectedvsControl="Infected-Control", levels=design)
cont.matrix

##           Contrasts
## Levels      InfectedvsControl
##   Control          -1
##   Infected          1
##   strain            0
##   interaction       0

fit2_m2 <- contrasts.fit(fit_m2,cont.matrix)
fit2_m2 <- eBayes(fit2_m2)
fit2_m2

## An object of class "MArrayLM"
## $coefficients
##           Contrasts
##           InfectedvsControl
## 1415670_at    0.295615220
## 1415671_at    0.108222965
## 1415672_at   -0.001215109
## 1415673_at   -0.023467868
## 1415674_a_at  0.173622489
## 45096 more rows ...
##
## $rank
## [1] 4
##
## $assign

```

```

## [1] 1 1 2 3
##
## $qr
## $qr
##      Control Infected      strain interaction
## 1 -2.8284271 0.000000 -1.7677670 0.0000000
## 2  0.3535534 -2.828427 -1.4142136 -1.4142136
## 3  0.3535534 0.000000 -1.9685020 -1.0160010
## 4  0.3535534 0.000000  0.1407408  0.9837388
## 5  0.3535534 0.000000  0.1407408  0.1274229
## 11 more rows ...
##
## $qraux
## [1] 1.353553 1.000000 1.140741 1.127423
##
## $pivot
## [1] 1 2 3 4
##
## $tol
## [1] 1e-07
##
## $rank
## [1] 4
##
## $df.residual
## [1] 12 12 12 12 12
## 45096 more elements ...
##
## $sigma
##    1415670_at   1415671_at   1415672_at   1415673_at 1415674_a_at
##    0.3165328    0.2033128    0.1071823    0.3166509    0.1363164
## 45096 more elements ...
##
## $cov.coefficients
##          Contrasts
## Contrasts           InfectedvsControl
## InfectedvsControl       0.5833333
##
## $stdev.unscaled
##          Contrasts
##          InfectedvsControl
##    1415670_at        0.7637626
##    1415671_at        0.7637626
##    1415672_at        0.7637626
##    1415673_at        0.7637626
##    1415674_a_at      0.7637626
## 45096 more rows ...
##
## $pivot
## [1] 1 2 3 4
##
## $Amean
##    1415670_at   1415671_at   1415672_at   1415673_at 1415674_a_at

```

```

##      8.995360   10.365534   10.828212    7.422040   9.235163
## 45096 more elements ...
##
## $method
## [1] "ls"
##
## $design
##   Control Infected strain interaction
## 1       1       0       1       0
## 2       1       0       1       0
## 3       1       0       1       0
## 4       1       0       1       0
## 5       1       0       1       0
## 11 more rows ...
##
## $contrasts
##           Contrasts
## Levels          InfectedvsControl
##   Control                  -1
##   Infected                  1
##   strain                   0
##   interaction                0
##
## $df.prior
## [1] 3.078919
##
## $s2.prior
## [1] 0.03390424
##
## $var.prior
## [1] 61.79253
##
## $proportion
## [1] 0.01
##
## $s2.post
##   1415670_at   1415671_at   1415672_at   1415673_at 1415674_a_at
## 0.08665770   0.03981860   0.01606514   0.08671723   0.02171072
## 45096 more elements ...
##
## $t
##           Contrasts
##           InfectedvsControl
##   1415670_at      1.31481549
##   1415671_at      0.71009763
##   1415672_at     -0.01255205
##   1415673_at     -0.10434281
##   1415674_a_at     1.54280288
## 45096 more rows ...
##
## $df.total
## [1] 15.07892 15.07892 15.07892 15.07892 15.07892
## 45096 more elements ...
##

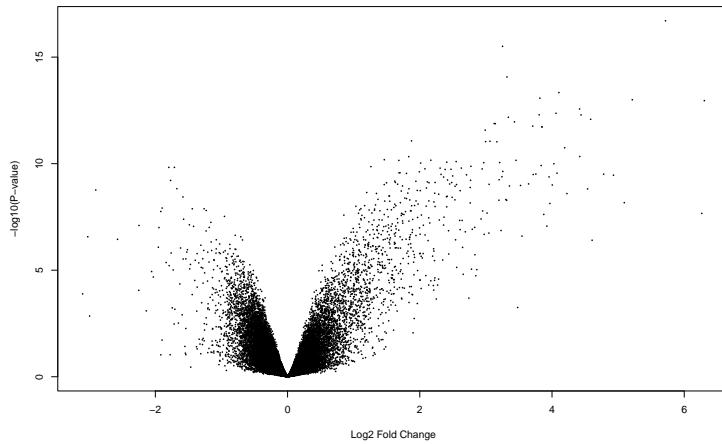
```

```

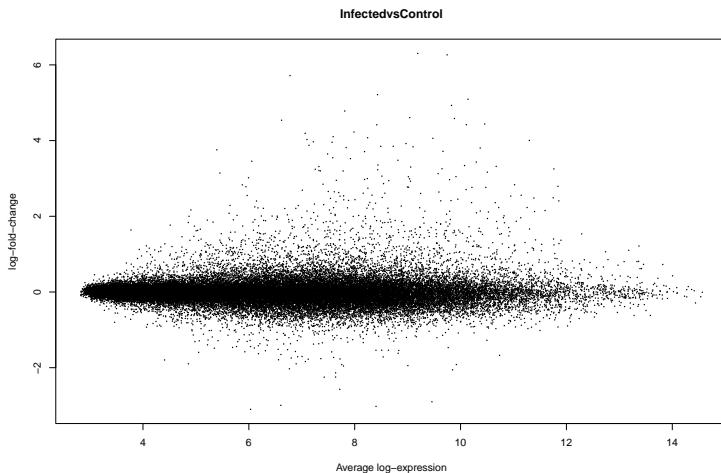
## $p.value
##           Contrasts
##           InfectedvsControl
## 1415670_at      0.2082185
## 1415671_at      0.4884808
## 1415672_at      0.9901497
## 1415673_at      0.9182725
## 1415674_a_at    0.1436020
## 45096 more rows ...
##
## $lods
##           Contrasts
##           InfectedvsControl
## 1415670_at     -6.067244
## 1415671_at     -6.669279
## 1415672_at     -6.931124
## 1415673_at     -6.925459
## 1415674_a_at   -5.764749
## 45096 more rows ...
##
## $F
## [1] 1.7287397757 0.5042386432 0.0001575539 0.0108874228 2.3802407312
## 45096 more elements ...
##
## $F.p.value
## [1] 0.2082185 0.4884808 0.9901497 0.9182725 0.1436020
## 45096 more elements ...

```

```
volcanoplot(fit2_m2)
```



```
limma::plotMA(fit2_m2)
```



```

# DE results
LIMMAout2 <- topTable(fit2_m2, adjust="BH", number=nrow(exprs(miceRMA)))
#head(LIMMAout)

## Check intensity values for top results
head(exprs(miceRMA)[rownames(exprs(miceRMA))%in%rownames(head(LIMMAout2)),])

##          GSM1579245_jac013-430v2.CEL GSM1579246_jac021-430v2.CEL
## 1418293_at           8.850220           8.697405
## 1419697_at           4.873087           5.196294
## 1424518_at           5.610642           5.255470
## 1449025_at           9.824438           9.841855
## 1450297_at           3.450449           3.409701
## 1451905_a_at         5.271445           5.483409
##          GSM1579247_jac044-430v2.CEL GSM1579248_jac045-430v2.CEL
## 1418293_at           8.628802           8.583339
## 1419697_at           4.944273           4.957930
## 1424518_at           5.033368           5.491428
## 1449025_at           9.930787           9.733168
## 1450297_at           3.628502           3.614535
## 1451905_a_at         5.814011           5.916976
##          GSM1579249_jac046-430v2.CEL GSM1579250_jac007-430v2.CEL
## 1418293_at           8.979707           13.280513
## 1419697_at           5.156156           12.484048
## 1424518_at           5.391127           9.926284
## 1449025_at           10.059654          13.751370
## 1450297_at           3.661825           10.544846
## 1451905_a_at         6.290288           12.039855
##          GSM1579251_jac008-430v2.CEL GSM1579252_jac036-430v2.CEL
## 1418293_at           13.17569            13.17675
## 1419697_at           12.58070            12.54747
## 1424518_at           10.07358            10.38063
## 1449025_at           13.69643            13.85968
## 1450297_at           10.77543            10.84424
## 1451905_a_at         11.65031            12.11285
##          GSM1579253_jac047-430v2.CEL GSM1579254_jac014-430v2.CEL
## 1418293_at           13.07904            8.738460

```

```

## 1419697_at           12.59532          4.815025
## 1424518_at           10.27056          5.208682
## 1449025_at           13.77391          10.119994
## 1450297_at           10.40172          3.770888
## 1451905_a_at         11.95163          5.430919
##                 GSM1579255_jac024-430v2.CEL GSM1579256_jac025-430v2.CEL
## 1418293_at           8.723118          8.699255
## 1419697_at           5.156243          4.948877
## 1424518_at           5.235854          5.688553
## 1449025_at           10.063244          10.129423
## 1450297_at           3.496160          3.556766
## 1451905_a_at         5.635768          5.014256
##                 GSM1579257_jac011-430v2.CEL GSM1579258_jac012-430v2.CEL
## 1418293_at           12.021105          12.276407
## 1419697_at           8.623987          9.287277
## 1424518_at           9.305259          9.485391
## 1449025_at           13.289408          13.458878
## 1450297_at           9.421028          9.491732
## 1451905_a_at         10.501785          10.709535
##                 GSM1579259_jac037-430v2.CEL GSM1579260_jac048-430v2.CEL
## 1418293_at           11.864963          11.991672
## 1419697_at           8.614776          8.632264
## 1424518_at           9.530186          9.606000
## 1449025_at           13.312026          13.363678
## 1450297_at           9.073683          9.311915
## 1451905_a_at         10.413056          10.672276

```

```

#mean expression of control/DBA2J
rowMeans(exprs(miceRMA)[rownames(exprs(miceRMA))%in%rownames(head(LIMMAout2)), 1:5])

```

```

## 1418293_at   1419697_at   1424518_at   1449025_at   1450297_at   1451905_a_at
##     8.747894    5.025548    5.356407    9.877981    3.553002    5.755226

```

```

#mean expression of infected/DBA2J
rowMeans(exprs(miceRMA)[rownames(exprs(miceRMA))%in%rownames(head(LIMMAout2)), 6:9])

```

```

## 1418293_at   1419697_at   1424518_at   1449025_at   1450297_at   1451905_a_at
##     13.17800   12.55189   10.16277   13.77035   10.64156   11.93866

```

```

#mean expression of control/C57BL6
rowMeans(exprs(miceRMA)[rownames(exprs(miceRMA))%in%rownames(head(LIMMAout2)), 10:12])

```

```

## 1418293_at   1419697_at   1424518_at   1449025_at   1450297_at   1451905_a_at
##     8.720277   4.973382   5.377696   10.104220   3.607938   5.360314

```

```

#mean expression of infected/C57BL6
rowMeans(exprs(miceRMA)[rownames(exprs(miceRMA))%in%rownames(head(LIMMAout2)), 13:16])

```

```

## 1418293_at   1419697_at   1424518_at   1449025_at   1450297_at   1451905_a_at
##     12.038537   8.789576   9.481709   13.355997   9.324589   10.574163

```

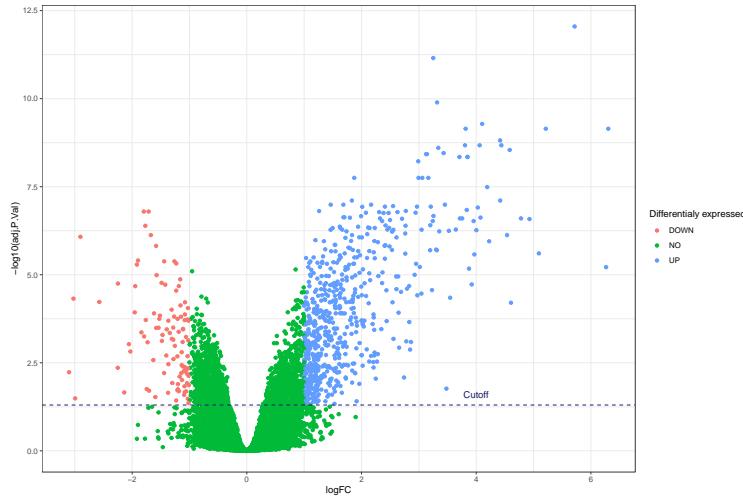
```

#Adjustments on p values using Benjamini-Hochberg
LIMMAout2$diffexpressed <- "NO"
LIMMAout2$diffexpressed[LIMMAout2$logFC > 1 & LIMMAout2$adj.P.Val < 0.05] <- "UP"
LIMMAout2$diffexpressed[LIMMAout2$logFC < -1 & LIMMAout2$adj.P.Val < 0.05] <- "DOWN"

#No adjustments on pvalues
LIMMAout2$diffexpressed_no_BH <- "NO"
LIMMAout2$diffexpressed_no_BH[LIMMAout2$logFC > 1 & LIMMAout2$P.Value < 0.05] <- "UP"
LIMMAout2$diffexpressed_no_BH[LIMMAout2$logFC < -1 & LIMMAout2$P.Value < 0.05] <- "DOWN"

#jpeg("microarray2_volcanoplot.jpg")
ggplot(data = LIMMAout2, aes(x= logFC, y = -log10(adj.P.Val), colour = diffexpressed)) +
  geom_point()+
  theme_bw()+
  geom_hline(yintercept = -log10(0.05), linetype="dashed", color="midnightblue")+
  annotate("text", min(4), 1.3, vjust = -1, label = "Cutoff", color="midnightblue")+
  #ggtitle("Differentiall (unadjusted P-value)") +
  labs(colour = "Differentially expressed")

```

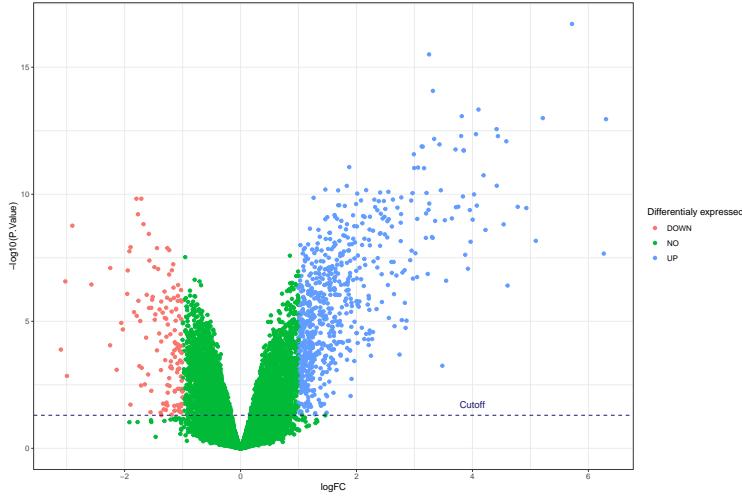


```

#theme(plot.title = element_text(hjust = 0.5, face = "bold.italic"))dev.off()

ggplot(data = LIMMAout2, aes(x= logFC, y = -log10(P.Value), colour = diffexpressed_no_BH)) +
  geom_point()+
  theme_bw()+
  geom_hline(yintercept = -log10(0.05), linetype="dashed", color="midnightblue")+
  annotate("text", min(4), 1.3, vjust = -1, label = "Cutoff", color="midnightblue")+
  #ggtitle("Differentiall (unadjusted P-value)") +
  labs(colour = "Differentially expressed")

```



```
length(which(LIMMAout2$diffexpressed=="UP")) #641 upregulated
```

```
## [1] 641
```

```
length(which(LIMMAout2$diffexpressed=="DOWN")) #100 downregulated
```

```
## [1] 100
```

```
length(which(LIMMAout2$diffexpressed_no_BH=="UP")) #701 upregulated
```

```
## [1] 701
```

```
length(which(LIMMAout2$diffexpressed_no_BH=="DOWN")) #136 downregulated
```

```
## [1] 136
```

```
## Load annotation and sort alphabetically on probe name
setwd("../Datasets/Microarray2/")
annotation_MA2 <- read.table("A-AFFY-45.adf.txt", header=T, sep="\t", skip=17, fill=T)
print(head(annotation_MA2))
```

```
## Composite.Element.Name Composite.Element.Database.Entry.interpro.
## 1 AFFX-BioB-5_at IPR007197
## 2 AFFX-BioB-M_at IPR007197
## 3 AFFX-BioB-3_at IPR007197
## 4 AFFX-BioC-5_at
## 5 AFFX-BioC-3_at
## 6 AFFX-BioDn-5_at IPR002586
## Composite.Element.Database.Entry.embl.
## 1 AFFX-BioB-5
## 2 AFFX-BioB-M
## 3 AFFX-BioB-3
## 4 AFFX-BioC-5
```

```

## 5                      AFFX-BioC-3
## 6                      AFFX-BioDn-5
##   Composite.Element.Database.Entry.affymetrix_netaffx.
## 1                      AFFX-BioB-5_at
## 2                      AFFX-BioB-M_at
## 3                      AFFX-BioB-3_at
## 4                      AFFX-BioC-5_at
## 5                      AFFX-BioC-3_at
## 6                      AFFX-BioDn-5_at
##   Composite.Element.Database.Entry.genbank.
## 1
## 2
## 3
## 4
## 5
## 6                      NP_752788
##   Composite.Element.Database.Entry.ec. Composite.Element.Database.Entry.refseq.
## 1
## 2
## 3
## 4
## 5
## 6
##   Composite.Element.Database.Entry.swall.
## 1
## 2
## 3
## 4
## 5
## 6
##   Composite.Element.Database.Entry.ensembl.
## 1
## 2
## 3
## 4
## 5
## 6
##   Composite.Element.Database.Entry.go.
## 1
## 2
## 3
## 4
## 5
## 6
##   Composite.Element.Database.Entry.unigene.
## 1
## 2
## 3
## 4
## 5
## 6
##   Composite.Element.Database.Entry.mgd. Composite.Element.Database.Entry.locus.
## 1                      NA                      NA
## 2                      NA                      NA

```

```

## 3 NA NA
## 4 NA NA
## 5 NA NA
## 6 NA NA
##   Composite.Element.Database.Entry.pkr_hanks.
## 1
## 2
## 3
## 4
## 5
## 6
##   Composite.Element.Database.Entry.scop.
## 1
## 2
## 3
## 4
## 5
## 6
##   Composite.Element.Database.Entry.cp450.
## 1
## 2
## 3
## 4
## 5
## 6

annotation_MA2 <- annotation_MA2[sort(annotation_MA2$Composite.Element.Name, index.return=T)$ix,]

## Check if all probes are present in both sets
dim(annotation_MA2)

## [1] 45101    16

dim(LIMMAout2)

## [1] 45101     8

## Double check => "Assumption is the mother of all fuck up's ;)"
sum(annotation_MA2$Composite.Element.Name==sort(rownames(LIMMAout2)))

## [1] 45101

## Sort LIMMA output alphabetically on probe name
LIMMAout_sorted2 <- LIMMAout2[sort(rownames(LIMMAout2), index.return=T)$ix,]

## Add gene names to LIMMA output
LIMMAout_sorted2$gene <- annotation_MA2$Composite.Element.Database.Entry.ensembl.

LIMMAout_annot2 <- LIMMAout_sorted2[sort(LIMMAout_sorted2$adj.P.Val, index.return=T)$ix,]

#sort by adjusted p value from most significant to least
LIMMAout_sorted2 <- LIMMAout_sorted2[order(LIMMAout_sorted2$adj.P.Val, decreasing= F),]

```

```

#extract top 50 significant DE genes
LIMMAout_sorted2[1:50,]$gene

## [1] "ENSMUSG00000025746" "ENSMUSG00000045303" "ENSMUSG00000045932"
## [4] "ENSMUSG00000051925" "ENSMUSG00000034855" "ENSMUSG00000029419"
## [7] "ENSMUSG00000023341" "ENSMUSG00000041827" "ENSMUSG00000022548"
## [10] "" "ENSMUSG00000047610" "ENSMUSG00000054261"
## [13] "" "ENSMUSG00000030107" "ENSMUSG00000025165"
## [16] "ENSMUSG00000020638" "ENSMUSG00000015947" "ENSMUSG00000022586"
## [19] "ENSMUSG00000027514" "" "ENSMUSG00000035152"
## [22] "ENSMUSG00000017830" "ENSMUSG00000029561" ""
## [25] "ENSMUSG00000048806" "ENSMUSG00000025498" "ENSMUSG00000035208"
## [28] "ENSMUSG00000022906" "ENSMUSG00000029379" ""
## [31] "ENSMUSG00000010358" "ENSMUSG00000039364" "ENSMUSG00000046031"
## [34] "" "ENSMUSG00000030921" "ENSMUSG00000019910"
## [37] "" "ENSMUSG00000001131" ""
## [40] "ENSMUSG00000003617" "ENSMUSG00000009670" "ENSMUSG00000028957"
## [43] "ENSMUSG00000023341" "ENSMUSG00000055116" "ENSMUSG00000015947"
## [46] "" ""
## [49] "ENSMUSG00000024371" ""

# Have a look at the results and search for other probesets for your DE genes
head(LIMMAout_annot2)

```

```

##          logFC      AveExpr       t     P.Value    adj.P.Val      B
## 1450297_at 5.716651  6.778338 44.60100 1.977975e-17 8.920867e-13 25.91879
## 1449025_at 3.251777 11.762997 37.08774 3.118352e-16 7.032040e-12 24.48160
## 1418293_at 3.318259 10.672902 29.69841 8.518266e-15 1.280608e-10 22.40893
## 1424518_at 4.104013  7.593314 26.49470 4.619705e-14 5.208833e-10 21.21036
## 1418930_at 6.303782  9.190682 24.97055 1.107804e-13 7.137579e-10 20.55617
## 1419697_at 3.816194  7.838359 25.43495 8.441607e-14 7.137579e-10 20.76185
##          diffexpressed diffexpressed_no_BH      gene
## 1450297_at           UP      ENSMUSG00000025746
## 1449025_at           UP      ENSMUSG00000045303
## 1418293_at           UP      ENSMUSG00000045932
## 1424518_at           UP      ENSMUSG00000051925
## 1418930_at           UP      ENSMUSG00000034855
## 1419697_at           UP      ENSMUSG00000029419

```

```
LIMMAout_annot2[LIMMAout_annot2$gene==" ENSMUSG00000025746",]
```

```

## [1] logFC          AveExpr        t
## [4] P.Value        adj.P.Val      B
## [7] diffexpressed diffexpressed_no_BH gene
## <0 rows> (or 0-length row.names)

```

```

ensembl <- useEnsembl(biomart = "genes")
#listDatasets(ensembl)
searchDatasets(mart = ensembl, pattern = "musculus")

```

```

##          dataset      description      version
## 18  bmusculus_gene_ensembl  Blue whale genes (mBalMus1.v2) mBalMus1.v2
## 107 mmusculus_gene_ensembl           Mouse genes (GRCm39)      GRCm39

#mmusculus_gene_ensembl

library('biomaRt')
mart <- useMart("ENSEMBL_MART_ENSEMBL")
mart <- useDataset("mmusculus_gene_ensembl", mart)

ensLookup <- gsub("\\. [0-9]*$", "", c(LIMMAout_sorted2$gene))

annotLookup <- getBM(
  mart=mart,
  attributes=c("ensembl_transcript_id", "ensembl_gene_id",
    "gene_biotype", "external_gene_name"),
  filter="ensembl_gene_id",
  values=ensLookup,
  uniqueRows=TRUE)

#retrieved external gene names
#unique(annotLookup$external_gene_name) #11449 genes

#retrieve the list of gene names from the limma output genes
gene_list <- unique(annotLookup$ensembl_gene_id[annotLookup$ensembl_gene_id %in% LIMMAout_sorted2$gene])
length(gene_list)

## [1] 11449

#extract DE genes
DEgeneIDs <- LIMMAout_sorted2$gene[LIMMAout_sorted2$adj.P.Val <= 0.05]
DEgeneIDs <- DEgeneIDs[DEgeneIDs != ""]
DEgeneIDs <- DEgeneIDs[DEgeneIDs %in% unique(annotLookup$ensembl_gene_id)]
length(DEgeneIDs)

## [1] 1382

#get gene symbols from ENSEMBL gene ids
DEgene_symbols2 <- NULL
for (gene in DEgeneIDs){
  n <- which(gene == gene_list)
  DEgene_symbols2 <- c(DEgene_symbols2, unique(annotLookup$external_gene_name)[n])
}

#top 10 DE genes
DEgene_symbol_and_ID <- cbind(DEgene_symbols2, DEgeneIDs)
head(DEgene_symbol_and_ID , 10)

##          DEgene_symbols2 DEgeneIDs
##  [1,] "I16"              "ENSMUSG00000025746"
##  [2,] "Ifit2"             "ENSMUSG00000045932"
##  [3,] "Cxcl10"            "ENSMUSG00000034855"

```

```

## [4,] "Ajm1"           "ENSMUSG00000029419"
## [5,] "Mx2"             "ENSMUSG00000023341"
## [6,] "Oasl1"           "ENSMUSG00000041827"
## [7,] "Apod"            "ENSMUSG00000022548"
## [8,] "Usp18"           "ENSMUSG00000030107"
## [9,] "Sectm1a"         "ENSMUSG00000025165"
## [10,] "Cmpk2"          "ENSMUSG00000020638"

# transpose the data before Pca as this function requires the variables to b columns
data <- t(as.data.frame(miceRMA@assayData$exprs))
pca <- prcomp(data, center = T, scale. = T)

summary(pca)

## Importance of components:
##                               PC1      PC2      PC3      PC4      PC5      PC6
## Standard deviation     109.5792 94.5875 70.5787 58.29125 49.1200 45.97059
## Proportion of Variance 0.2662  0.1984  0.1105  0.07534  0.0535  0.04686
## Cumulative Proportion  0.2662  0.4646  0.5751  0.65040  0.7039  0.75075
##                               PC7      PC8      PC9      PC10     PC11     PC12
## Standard deviation     42.54130 41.32187 37.95470 35.78304 34.98692 32.96254
## Proportion of Variance 0.04013  0.03786  0.03194  0.02839  0.02714  0.02409
## Cumulative Proportion  0.79088  0.82874  0.86068  0.88907  0.91621  0.94030
##                               PC13     PC14     PC15     PC16
## Standard deviation     31.05439 30.60996 28.12686 4.935e-13
## Proportion of Variance 0.02138  0.02077  0.01754 0.0000e+00
## Cumulative Proportion  0.96168  0.98246  1.00000 1.0000e+00

# save as dataframe and add treatment variable
pca_out <- as.data.frame(pca$x)

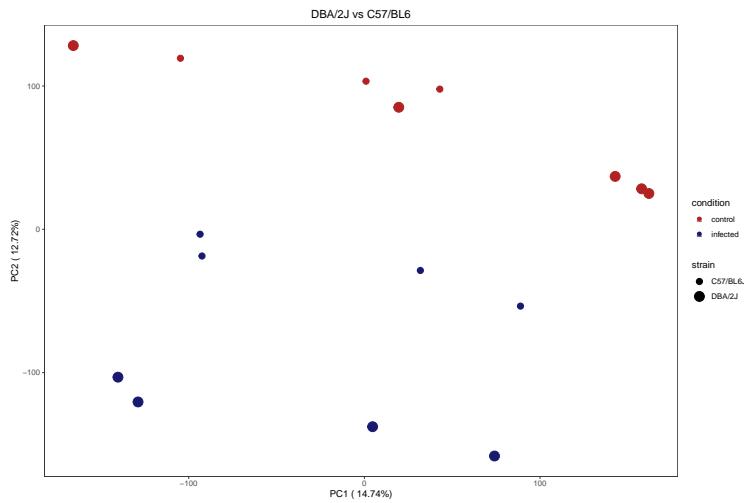
pca_out$condition <- as.character(miceRMA@phenoData$data$condition)

# get labels
percentage <- round(pca$sdev / sum(pca$sdev) * 100, 2)
percentage <- paste( colnames(pca_out), "(", paste( as.character(percentage), "%", ") ", sep="") ) 

ggplot(data = pca_out)+ 
  ggtitle("DBA/2J vs C57/BL6")+
  geom_point(aes(x = PC1, y = PC2, colour = condition, label='', size=strain))+ 
  geom_text(aes(x = PC1, y = PC2, colour = condition, label=''), hjust=0.5, vjust=1.15)+ 
  theme_bw()+
  xlab(percentage[1])+ 
  ylab(percentage[2])+ 
  labs(colour = "condition")+
  theme(plot.title = element_text(hjust = 0.5))+ 
  scale_size_manual(values = c(3, 5))+ 
  scale_colour_manual(values = c("firebrick", "midnightblue"))+ 
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())

## Warning: Ignoring unknown aesthetics: label

```



```
#ggsave("PCA_array2_bothstrains.png", dpi=750, width=8, height = 5)
```