In [1]:
```
library(limma)
library(affy)
```

```
Loading required package: BiocGenerics
Loading required package: parallel

Attaching package: ‘BiocGenerics’

The following objects are masked from ‘package:parallel’:

    clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
    clusterExport, clusterMap, parApply, parCapply, parLapply,
    parLapplyLB, parRapply, parSapply, parSapplyLB

The following object is masked from ‘package:limma’:

    plotMA

The following object is masked from ‘package:stats’:

    xtabs

The following objects are masked from ‘package:base’:

    anyDuplicated, append, as.data.frame, as.vector, cbind, colnames,
    do.call, duplicated, eval, evalq, Filter, Find, get, intersect,
    is.unsorted, lapply, Map, mapply, match, mget, order, paste, pmax,
    pmax.int, pmin, pmin.int, Position, rank, rbind, Reduce, rep.int,
    rownames, sapply, setdiff, sort, table, tapply, union, unique,
    unlist, unsplit

Loading required package: Biobase
Welcome to Bioconductor

    Vignettes contain introductory material; view with
    'browseVignettes()'. To cite Bioconductor, see
    'citation("Biobase")', and for packages 'citation("pkgname")'.
```

In [2]:
```
celfiles = Sys.glob("/mnt/storage/r0372261/groupwork/CELfiles2/GSM11108*")
#create name to find cel files
#we used this dataset https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45627
```

To determine the genetic interactions of MiR-221 and identify these target genes MiR-221 overexpressing cells were compared to untreated PCa cells. A microarray provides a set of differentially expressed genes which in turn lead to a set of genes and a pathay that is affected by MiR-221. Interaction of miR-221 with potential target genes was analyzed by a genome wide expression profiling. The overexpression leads to activation of apoptosis, growth arrest and reduced invasivness in PCa cells.The transcriptional consequences of MiR-221 are of importance in the tumirogenesis in PCa cells.

In [3]:
```
celfiles
#check if they contain the CEl files
```

```
'/mnt/storage/r0372261/groupwork/CELfiles2/GSM1110879_Nr1_pre221.CEL'
'/mnt/storage/r0372261/groupwork/CELfiles2/GSM1110880_Nr2_pre221.CEL'
'/mnt/storage/r0372261/groupwork/CELfiles2/GSM1110881_Nr3_ctrl.CEL'
'/mnt/storage/r0372261/groupwork/CELfiles2/GSM1110882_Nr4_ctrl.CEL'
```

In [4]:
```
rawcel <- ReadAffy(filenames=celfiles)
#load to work with it in program
```

In [5]:
```
sampleNames (rawcel)
#give own meaningfull names to samples
```

'GSM1110879_Nr1_pre221.CEL' 'GSM1110880_Nr2_pre221.CEL'
'GSM1110881_Nr3_ctrl.CEL' 'GSM1110882_Nr4_ctrl.CEL'

In [6]:
```
treatment = c("stimulated", "stimulated", "unstimulated", "unstimulated")
#create name vector for samples
```

In [7]:
```
pd <- data.frame(treatment = treatment)
rownames(pd) <- sampleNames(rawcel)
pd
#check if vector correlates with the right samples
```

	treatment
GSM1110879_Nr1_pre221.CEL	stimulated
GSM1110880_Nr2_pre221.CEL	stimulated
GSM1110881_Nr3_ctrl.CEL	unstimulated
GSM1110882_Nr4_ctrl.CEL	unstimulated

In [8]:
```
library(Biobase)
#download data.frame that cointains anottated date for affy package

phenoData(rawcel) <- AnnotatedDataFrame(data=pd)
#assign annotation
```

In [9]:
```
library(frma)
```

```
Creating a generic function for ‘nchar’ from package ‘base’ in package ‘S4Vectors’
```

In [10]:
```
eset <- frma(rawcel, verbose=TRUE)
#Verbose is to see messages when running
#normalizes data
save(eset, file="/mnt/storage/r0372261/groupwork2.eset")
```

```
Background Correcting ...


Normalizing ...

Summarizing ...
```

In [11]:
```
eset
#check if worked
```

```
ExpressionSet (storageMode: lockedEnvironment)
assayData: 54675 features, 4 samples
  element names: exprs, se.exprs
protocolData: none
phenoData
  sampleNames: GSM1110879_Nr1_pre221.CEL GSM1110880_Nr2_pre221.CEL
    GSM1110881_Nr3_ctrl.CEL GSM1110882_Nr4_ctrl.CEL
  varLabels: treatment
  varMetadata: labelDescription
featureData: none
experimentData: use 'experimentData(object)'
Annotation: hgu133plus2
```

In [12]:
```
exprs (eset)
```

	GSM1110879_Nr1_pre221.CEL	GSM1110880_Nr2_pre221.CEL	GSM1110881_Nr3_ctr
1007_s_at	7.811605	8.143876	8.321028
1053_at	8.288777	8.404835	8.609546
117_at	6.002085	5.419955	5.849699
121_at	7.475468	7.501379	7.469068
1255_g_at	3.556231	3.466736	3.423771
1294_at	7.005376	5.943170	5.889277
1316_at	5.554376	5.878458	5.663777
1320_at	4.861914	4.736126	4.350290
1405_i_at	5.165580	3.903487	3.707000
1431_at	3.685560	3.583523	3.529870
1438_at	5.966205	6.002227	6.245711
1487_at	7.628095	8.009850	7.737913
1494_f_at	5.364669	5.294262	5.457119
1552256_a_at	8.244695	8.052900	8.405697
1552257_a_at	7.570779	7.862574	8.013220
1552258_at	5.251120	4.651213	5.419008
1552261_at	4.195342	4.198160	4.455030
1552263_at	5.683391	5.908333	5.279475
1552264_a_at	6.418248	6.316203	6.099221
1552266_at	3.402047	3.371037	3.418627
1552269_at	4.457569	4.700752	4.541291
1552271_at	5.883166	5.714249	5.905891
1552272_a_at	5.445157	5.364924	5.360327
1552274_at	6.823570	6.663058	6.594479
1552275_s_at	7.009317	6.760134	6.248437
1552276_a_at	5.363360	5.166106	5.229727
1552277_a_at	9.839086	9.770047	9.762194
1552278_a_at	4.746285	4.990956	4.776170
1552279_a_at	6.385850	6.638732	6.579666
1552280_at	3.565339	3.610838	3.877816
⋮	⋮	⋮	⋮
AFFX-PheX-3_at	6.539190	6.662466	6.660344
AFFX-PheX-5_at	5.906363	5.913123	5.928302

	GSM1110879_Nr1_pre221.CEL	GSM1110880_Nr2_pre221.CEL	GSM1110881_Nr3_ctr
AFFX-PheX-M_at	6.185481	6.440238	6.344328
AFFX-r2-Bs-dap-3_at	9.380118	9.692248	9.712387
AFFX-r2-Bs-dap-5_at	7.603790	7.452639	7.602544
AFFX-r2-Bs-dap-M_at	8.667730	8.792746	8.799369
AFFX-r2-Bs-lys-3_at	5.893447	5.942680	6.130836
AFFX-r2-Bs-lys-5_at	5.546407	5.915175	5.718738
AFFX-r2-Bs-lys-M_at	5.852291	6.021126	5.906801
AFFX-r2-Bs-phe-3_at	7.138527	7.226842	7.187398
AFFX-r2-Bs-phe-5_at	6.416698	6.565569	6.484479
AFFX-r2-Bs-phe-M_at	6.491919	6.740076	6.389071
AFFX-r2-Bs-thr-3_s_at	7.733801	7.840802	7.804803
AFFX-r2-Bs-thr-5_s_at	5.982105	5.637055	5.925948
AFFX-r2-Bs-thr-M_s_at	7.119856	6.881621	7.214284
AFFX-r2-Ec-bioB-3_at	8.299450	8.384834	8.289836
AFFX-r2-Ec-bioB-5_at	8.212420	8.256166	8.244061
AFFX-r2-Ec-bioB-M_at	8.444875	8.487957	8.440258
AFFX-r2-Ec-bioC-3_at	9.782289	9.838125	9.838568
AFFX-r2-Ec-bioC-5_at	9.400611	9.443648	9.381960
AFFX-r2-Ec-bioD-3_at	11.948837	12.056734	12.008104
AFFX-r2-Ec-bioD-5_at	11.439493	11.643706	11.541331
AFFX-r2-P1-cre-3_at	13.659122	13.740519	13.584824

	GSM1110879_Nr1_pre221.CEL	GSM1110880_Nr2_pre221.CEL	GSM1110881_Nr3_ctr
AFFX-r2-P1-cre-5_at	13.503382	13.586569	13.457823
AFFX-ThrX-3_at	7.465558	7.556607	7.570421
AFFX-ThrX-5_at	5.580822	5.133750	5.609562
AFFX-ThrX-M_at	6.365543	6.203833	6.447412
AFFX-TrpnX-3_at	3.049658	2.978359	3.055909
AFFX-TrpnX-5_at	3.653017	3.334958	3.300384
AFFX-TrpnX-M_at	3.247260	3.264249	3.264868

In [13]:
```
eset.labels <- c("stim1", "stim2", "unstim1", "unstim2")
#give the samplenames a meaningful name
sampleNames(eset) <- eset.labels
```

In [14]:
```
sampleNames(eset)
```

'stim1' 'stim2' 'unstim1' 'unstim2'

In [15]:
```
pData(phenoData(eset))
#correlate the samples with a treatment
```

	treatment
stim1	stimulated
stim2	stimulated
unstim1	unstimulated
unstim2	unstimulated

In [16]:
```
probe.counts <- exprs(eset)
head(probe.counts)
#check
```

	stim1	stim2	unstim1	unstim2
1007_s_at	7.811605	8.143876	8.321028	8.353263
1053_at	8.288777	8.404835	8.609546	8.418000
117_at	6.002085	5.419955	5.849699	5.439175
121_at	7.475468	7.501379	7.469068	7.583601
1255_g_at	3.556231	3.466736	3.423771	3.467697
1294_at	7.005376	5.943170	5.889277	5.825513

In [17]:
```
library(annotate)
library(hgu133plus2.db)
```

```
Loading required package: AnnotationDbi
Loading required package: stats4
Loading required package: GenomeInfoDb
Loading required package: S4Vectors
Loading required package: IRanges
Loading required package: XML
Loading required package: org.Hs.eg.db
```

In [18]:
```
probe.names <- rownames(probe.counts)
gene.symbols <- getSYMBOL(probe.names, "hgu133plus2")
head (gene.symbols)
#Give the probe names the corresponding gene symbol
```

1007_s_at	'NA'
1053_at	'RFC2'
117_at	'HSPA6'
121_at	'PAX8'
1255_g_at	'GUCA1A'
1294_at	'NA'

In [19]:
```
probe.counts.genesymbol <- as.data.frame(cbind(as.data.frame(probe.counts), gene.symbols
head(probe.counts.genesymbol)
#add the gene names to the column
```

	stim1	stim2	unstim1	unstim2	gene.symbols
1007_s_at	7.811605	8.143876	8.321028	8.353263	NA
1053_at	8.288777	8.404835	8.609546	8.418000	RFC2
117_at	6.002085	5.419955	5.849699	5.439175	HSPA6
121_at	7.475468	7.501379	7.469068	7.583601	PAX8
1255_g_at	3.556231	3.466736	3.423771	3.467697	GUCA1A
1294_at	7.005376	5.943170	5.889277	5.825513	NA

In [20]:
```
probe.counts.genesymbol = probe.counts.genesymbol[complete.cases(probe.counts.genesymbol
#delete all missing gene symbol data
probe.counts.genesymbol$gene.symbols <- as.character(probe.counts.genesymbol$gene.symbol
#make sure the gene symbols are all strings
```

In [21]:
```
#join multiple probes for 1 gene
library(reshape2)
```

In [22]:
```
m <- melt(probe.counts.genesymbol, id.vars = c("gene.symbols"))
#melt same gene symbol data together
```

In [23]:
```
counts <- dcast(m, gene.symbols ~ variable, fun.aggregate=median)
#take median of same gene symbols
```

In [24]:
```
head(counts)
#check
```

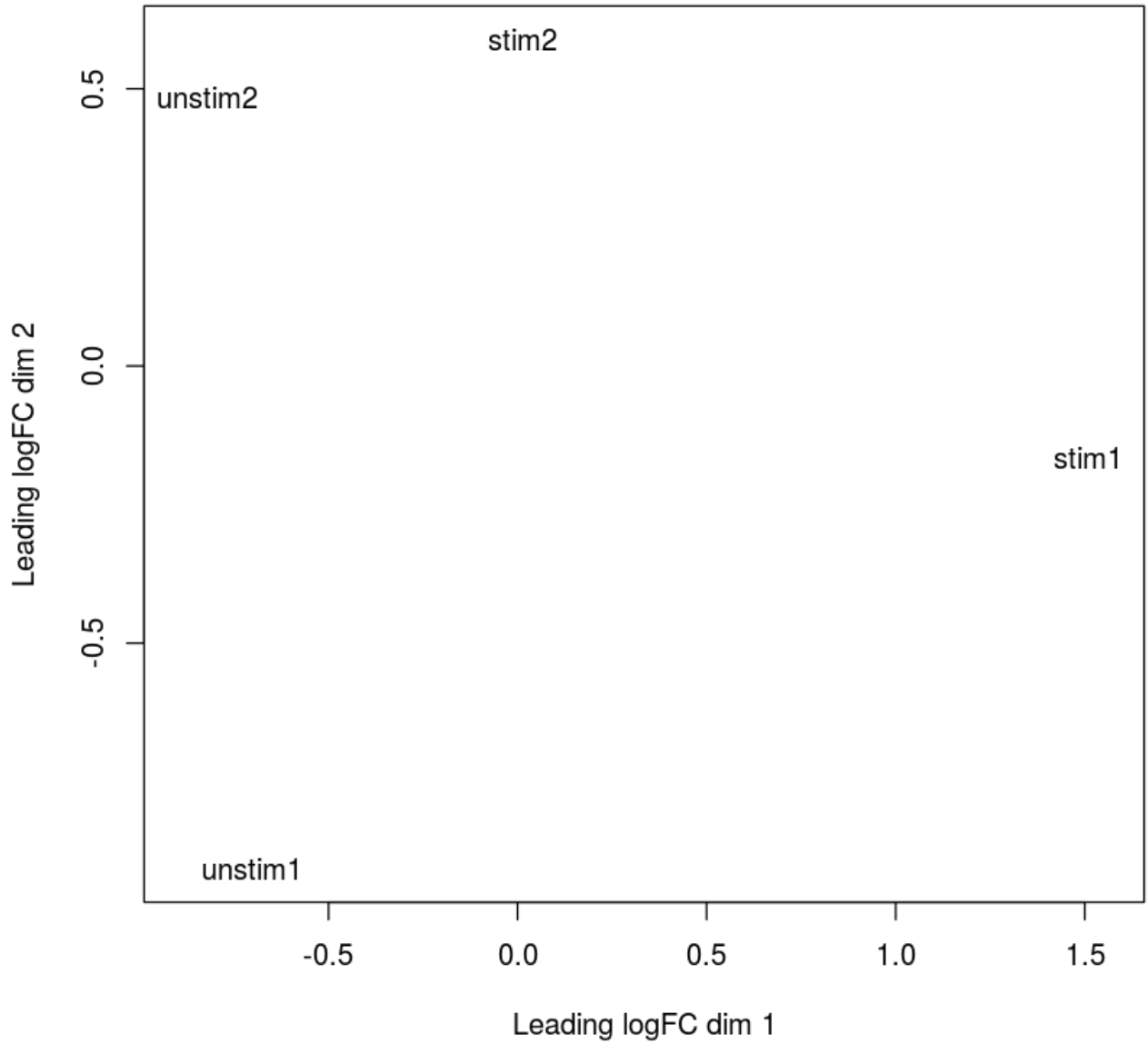
gene.symbols	stim1	stim2	unstim1	unstim2
A1BG	6.008552	5.563582	5.955143	5.561371
A1BG-AS1	5.542202	5.374091	5.340208	5.065887
A1CF	3.854480	3.939221	3.916352	3.947483
A2M	4.482044	4.349491	4.489376	4.275783
A2M-AS1	3.798917	3.840970	4.064091	3.993132
A2ML1	3.622471	3.769957	3.776260	3.815509

In [25]:
```
#make gene symbols into a row name
rownames(counts) <- counts$gene.symbols
#delete name of gene symbols
counts$gene.symbols <- NULL
```

In [26]:
```
head(counts)
#check
```

	stim1	stim2	unstim1	unstim2
A1BG	6.008552	5.563582	5.955143	5.561371
A1BG-AS1	5.542202	5.374091	5.340208	5.065887
A1CF	3.854480	3.939221	3.916352	3.947483
A2M	4.482044	4.349491	4.489376	4.275783
A2M-AS1	3.798917	3.840970	4.064091	3.993132
A2ML1	3.622471	3.769957	3.776260	3.815509

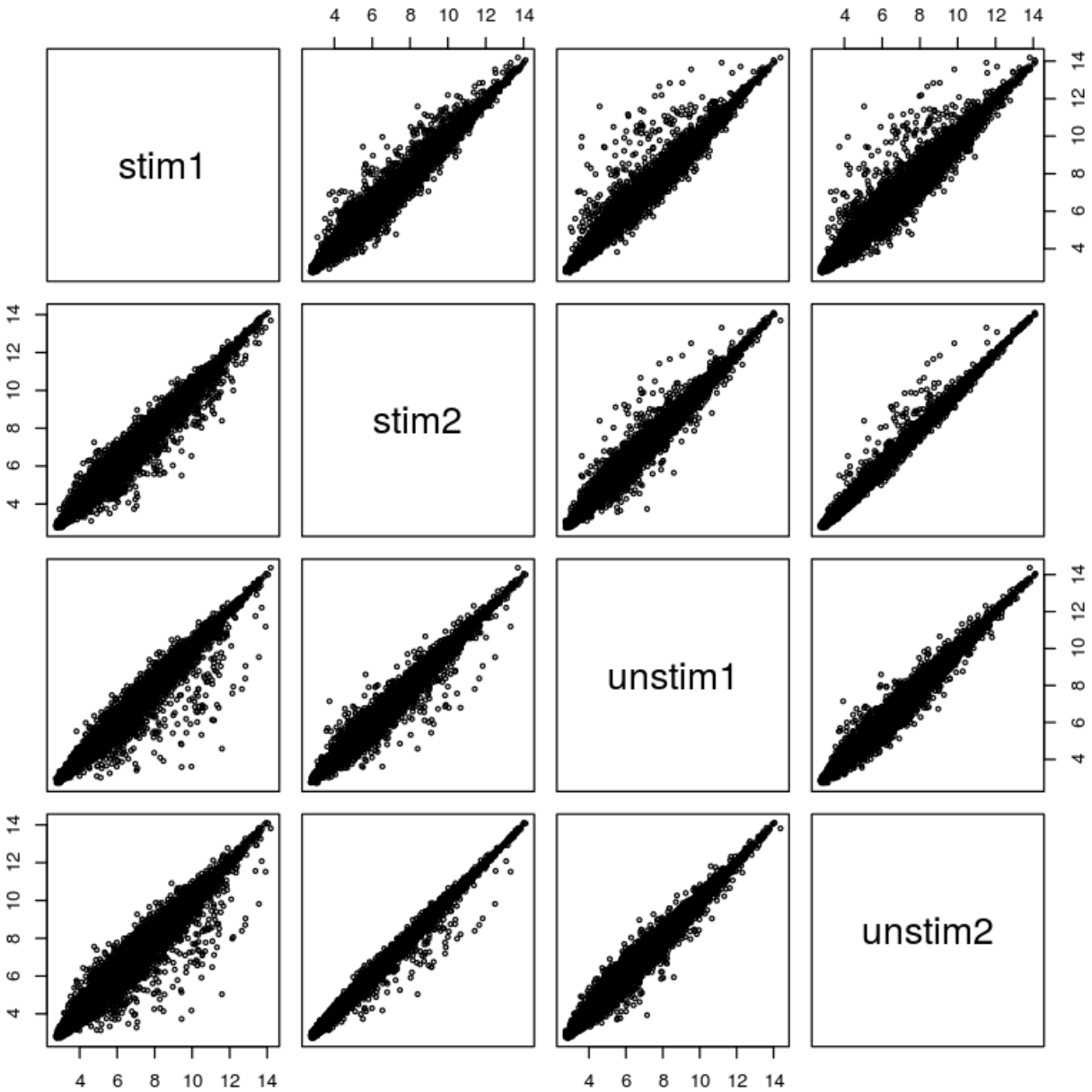
In [27]:
```
plotMDS(eset, ndim=2, dim.plot=c(1,2))
#Check if big difference between treatments and recplicates
#although unstim 1 and 2 are not exactly the same, we can still see them together on one
#This means they are more different from the stimulated than from eachother
```



In [28]:
```
goi <- c("IRF1",  "IRF2", "SOCS3", "STAT1", "STAT2", "STAT3")
counts[goi,]
#predicted difference in the STAT/JAK pathway -> check some important genes
#indeed difference (upregulation in the stimulated samples)
```

	stim1	stim2	unstim1	unstim2
IRF1	8.445419	7.093983	6.631854	6.914710
IRF2	7.587974	6.393101	6.065582	6.257732
SOCS3	5.467248	5.561640	5.448048	5.613179
STAT1	10.724175	9.850468	7.463788	7.775934
STAT2	7.050507	5.685396	5.244096	5.279582
STAT3	4.649420	4.750468	4.809563	4.597772

In [29]:
```
plot (counts, cex=0.5)
#another way to check the differences visualy
```



In [30]:
```
library(limma)
#do differential expression analysis
```

In [31]:
```
treat <- factor(c('stimulated', 'stimulated', 'unstimulated', 'unstimulated'))
treat
#assign a treatment option to the samples
```

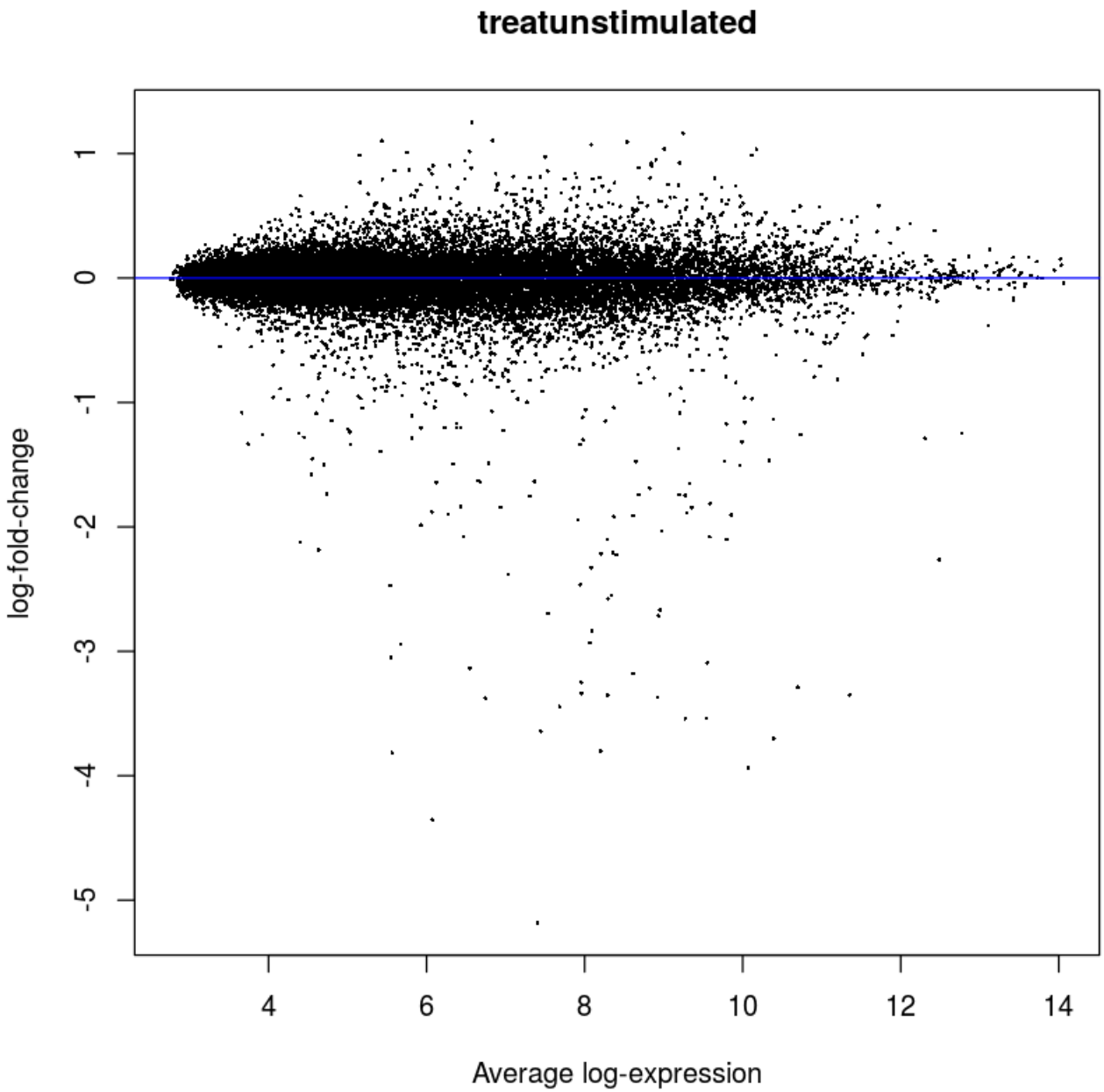
stimulated stimulated unstimulated unstimulated

In [32]:
```
design <- model.matrix(~treat)
design
#matrix needed  for limma (says if untreated or treated)
```

	(Intercept)	treatunstimulated
1	1	0
2	1	0
3	1	1
4	1	1

In [33]:
```
fit <- eBayes(lmFit(counts, design))
#create linear model
```

In [34]:
```
limma::plotMA(fit, 2)
abline(0,0,col='blue')
#gives graph log-fold-change stimulated vs unstimulated
#unstimulated has a lot of genes underexpressed compared to stimulated
```



In [35]:
```
#extract top differential expressed genes
topTable(fit, coef=2)
```

	logFC	AveExpr	t	P.Value	adj.P.Val	B
IFI44L	-3.179707	8.612673	-11.843552	0.0001087339	0.9998485	-3.384600
IFI6	-3.288430	10.697722	-9.683379	0.0002724193	0.9998485	-3.415263
ISG15	-2.262120	12.486394	-9.365291	0.0003168317	0.9998485	-3.421490
IFIT1	-3.699822	10.393635	-9.089225	0.0003625979	0.9998485	-3.427382
IFITM1	-3.349837	11.352938	-9.036628	0.0003721967	0.9998485	-3.428561
DDX60	-3.248647	7.954843	-9.010421	0.0003770935	0.9998485	-3.429155
MX1	-3.934744	10.068944	-8.661755	0.0004502201	0.9998485	-3.437525
IFI44	-3.444251	7.683571	-8.643580	0.0004544788	0.9998485	-3.437987
STAT1	-2.667460	8.953591	-8.469108	0.0004979222	0.9998485	-3.442552
OSTM1	1.008880	5.753688	8.300031	0.0005448630	0.9998485	-3.447222

In [63]:
```
select = p.adjust(fit[,2]$p.value, method="BH") <0.01
select1 = p.adjust(fit[,2]$p.value, method="none") <0.01
#selects all genes still significant after correcting for multiple testing
```

In [64]:
```
summary(select)
summary(select1)
#no genes significant after correction --> original paper probably used another method
#Without correction 104 genes significant expressed
```

```
   Mode   FALSE    NA's
logical   20389       0
```

```
   Mode   FALSE    TRUE    NA's
logical   20285     104       0
```

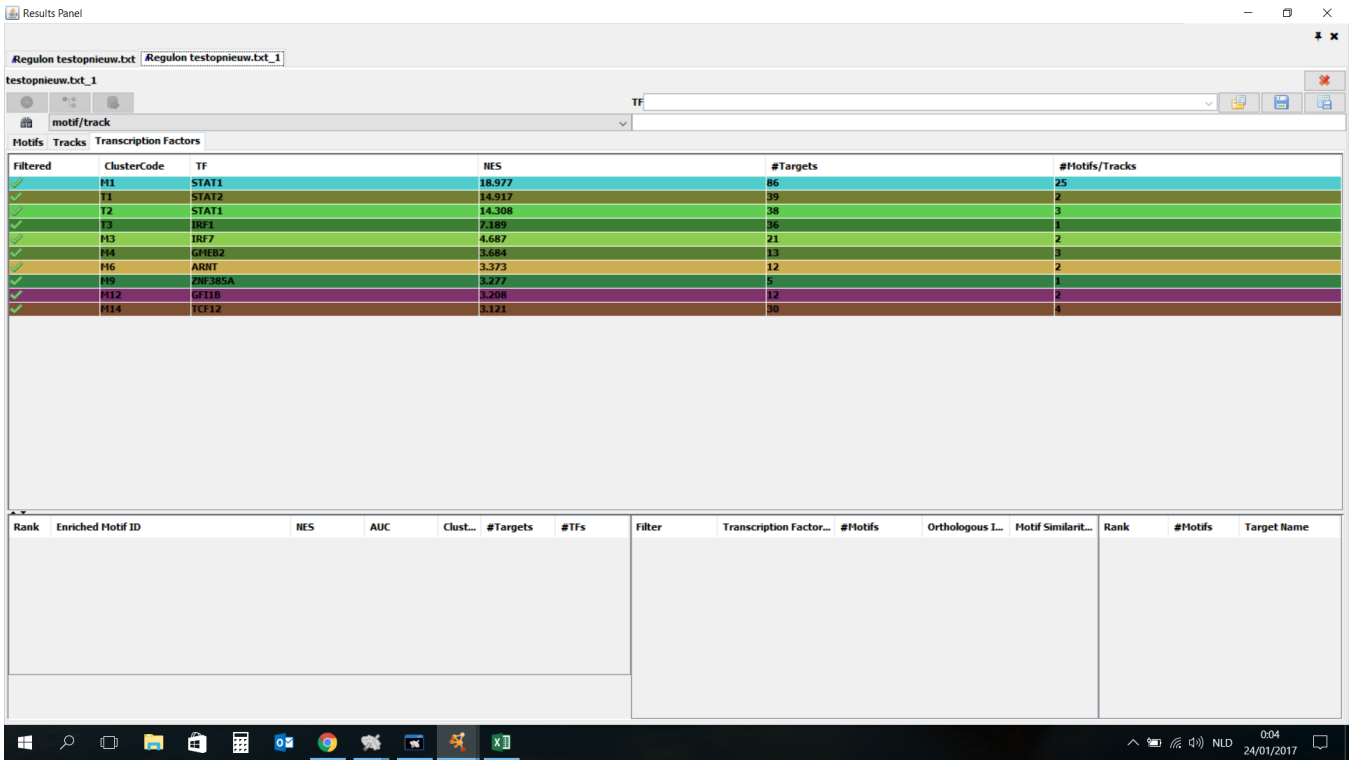
In [38]:
```
top = topTable(fit, coef=2, number=100000)
```

In [39]:
```
write.table(top, "GSM11195.tophits.tsv", sep="\t", quote=FALSE, col.names=NA)
#created excel file
```

In [40]:
```
head(top)
```

	logFC	AveExpr	t	P.Value	adj.P.Val	B
IFI44L	-3.179707	8.612673	-11.843552	0.0001087339	0.9998485	-3.384600
IFI6	-3.288430	10.697722	-9.683379	0.0002724193	0.9998485	-3.415263
ISG15	-2.262120	12.486394	-9.365291	0.0003168317	0.9998485	-3.421490
IFIT1	-3.699822	10.393635	-9.089225	0.0003625979	0.9998485	-3.427382
IFITM1	-3.349837	11.352938	-9.036628	0.0003721967	0.9998485	-3.428561
DDX60	-3.248647	7.954843	-9.010421	0.0003770935	0.9998485	-3.429155

In [ ]:
```
#Take top 104 genes out of exceltable, make it a txt file and analyze with Iregulon
```



After Iregulon analysis we also come to the conclusion that the treatment interferes with the JAK/STAT pathway because the TFs regulating this pathway are in common with the differentially expressed genes (eg STAT, IRF1). This supports the hypothesis of the activation of apoptosis and growth arrest pathways by MiR-221 interaction with its target genes which cause a increase in expression of important TFs of these pathways and its components. MiR-221 overexpression leads to activation of apoptosis, growth arrest and reduced invasivness in PCa cells. This can be used as potential cancer treatment as it reduces the viability and invasiveness of cancer cells.

In [ ]:
```
```