

GSE14038

Teresa Rozza

7/22/2021

```
# Instalar Bioconductor
if (!requireNamespace("BiocManager", quietly = TRUE)) # install.packages("BiocManager")
BiocManager::install(version = "3.13")
# Instalar y cargar libraries
if (!requireNamespace("GEOquery", quietly = TRUE))
BiocManager::install("GEOquery")
if (!requireNamespace("limma", quietly = TRUE))
BiocManager::install("limma")
if (!requireNamespace("umap", quietly = TRUE))
BiocManager::install("umap")
if (!requireNamespace("pvca", quietly = TRUE))
BiocManager::install("pvca")
if (!requireNamespace("oligo", quietly = TRUE))
BiocManager::install("oligo")
if (!requireNamespace("pd.mogene.2.1.st", quietly = TRUE))
BiocManager::install("pd.mogene.2.1.st")
if (!requireNamespace("Biobase", quietly = TRUE))
BiocManager::install("Biobase")
if (!requireNamespace("genefilter", quietly = TRUE))
BiocManager::install("genefilter")
if (!requireNamespace("mogene21sttranscriptcluster.db", quietly = TRUE))
BiocManager::install("mogene21sttranscriptcluster.db")
if (!requireNamespace("annotate", quietly = TRUE))
BiocManager::install("annotate")
if (!requireNamespace("org.Mm.eg.db", quietly = TRUE))
BiocManager::install("org.Mm.eg.db")
if (!requireNamespace("ReactomePA", quietly = TRUE))
BiocManager::install("ReactomePA")
if (!requireNamespace("reactome.db", quietly = TRUE))
BiocManager::install("reactome.db")
```

```
library(GEOquery)
library(limma)
library(umap)
library(pvca)
library(oligo)
library(pd.mogene.2.1.st)
library(Biobase)
library(genefilter)
library(mogene21sttranscriptcluster.db)
library(annotate)
library(org.Mm.eg.db)
library(ReactomePA)
library(reactome.db)
```

```
datadir <- setwd("~/Desktop/hack4rare/GSE14038_RAW")
```

STEP 1 - Obtain Microarray Data

```
# Store the dataset ids in a vector GEO_DATASETS just in case you want to loop through several GEO ids
GEO_DATASETS <- "GSE14038"
```

```
#loading dataset
gset <- getGEO("GSE14038", destdir=datadir, GSElimits=NULL, GSEMatrix=TRUE, AnnotGPL=FALSE, getGPL=FALSE)
if (length(gset) > 1) idx <- grep("GPL7869", attr(gset, "names")) else idx <- 1
gset <- gset[[idx]]

ex <- exprs(gset)
# log2 transform
qx <- as.numeric(quantile(ex, c(0., 0.25, 0.5, 0.75, 0.99, 1.0), na.rm=T))
LogC <- (qx[5] > 100) ||
        (qx[6]-qx[1] > 50 && qx[2] > 0)
if (LogC) { ex[which(ex <= 0)] <- NaN
  ex <- log2(ex) }
```

```
#quality data
library(arrayQualityMetrics)
arrayQualityMetrics(gset)
```

From the file index.html we can see that only 4 specimen are marked once, so we can keep them in the analysis.

```
#explore structure of data
gset
```

```
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 26183 features, 86 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: GSM352448 GSM352450 ... GSM352553 (86 total)
##   varLabels: title geo_accession ... relation (32 total)
##   varMetadata: labelDescription
## featureData: none
## experimentData: use 'experimentData(object)'
##   pubMedIds: 20049725
##   23685747
##   24832557
## Annotation: GPL7869
```

```
head(exprs(gset))
```

```
##           GSM352448 GSM352450 GSM352451 GSM352453 GSM352454 GSM352455
## AFX-BioB-3_at    12.09108   10.96124   11.01926   11.67409   11.65724   11.99511
## AFX-BioB-5_at    11.84192   10.45048   10.51598   11.19520   11.21031   11.65988
## AFX-BioB-M_at    12.35858   11.30336   11.38033   11.90779   11.91642   12.28062
## AFX-BioC-3_at    12.12781   10.92782   11.00451   11.61707   11.64711   12.09226
## AFX-BioC-5_at    12.72149   11.71685   11.80867   12.28593   12.29666   12.62544
## AFX-BioDn-3_at   13.70425   13.32563   13.35103   13.55919   13.51428   13.65361
##           GSM352457 GSM352458 GSM352460 GSM352461 GSM352463 GSM352464
## AFX-BioB-3_at    11.11827   11.19862   11.07628   10.142407  10.241164  10.190896
## AFX-BioB-5_at    10.75063   10.79751   10.55486   9.593927   9.740788   9.693328
## AFX-BioB-M_at    11.49875   11.50761   11.43046   10.503953  10.614904  10.537533
## AFX-BioC-3_at    11.19257   11.10526   11.07983   10.178316  10.288737  10.226805
## AFX-BioC-5_at    11.90911   11.99042   11.86498   10.943615  11.081559  11.037192
## AFX-BioDn-3_at   13.33304   13.39534   13.39710   12.925375  12.939490  12.923748
##           GSM352466 GSM352467 GSM352469 GSM352470 GSM352471 GSM352473
## AFX-BioB-3_at    10.270593  10.336755  10.072369  10.264592  10.171352  10.247153
## AFX-BioB-5_at     9.794574   9.878099   9.557853   9.777857   9.756054   9.700566
## AFX-BioB-M_at    10.586469  10.686047  10.411490  10.603538  10.546048  10.640611
## AFX-BioC-3_at    10.280467  10.391016  10.103925  10.308279  10.235950  10.296716
## AFX-BioC-5_at    11.095126  11.195630  10.950243  11.105343  11.067910  11.125998
## AFX-BioDn-3_at   12.885573  12.990555  12.971288  12.967510  12.942823  13.006323
##           GSM352475 GSM352476 GSM352477 GSM352479 GSM352480 GSM352481
## AFX-BioB-3_at    10.204934  10.69708   10.514580  10.80426   10.70074   10.73901
## AFX-BioB-5_at     9.663128  10.10854   9.971971  10.31391  10.20304  10.30972
## AFX-BioB-M_at    10.556968  11.00993  10.861583  11.14602  10.96003  11.05600
## AFX-BioC-3_at    10.262977  10.69434  10.515112  10.70249  10.62888  10.68297
## AFX-BioC-5_at    11.025405  11.49397  11.299006  11.53503  11.41374  11.51116
## AFX-BioDn-3_at   12.952490  13.17601  13.070940  13.12476  13.04982  13.04117
##           GSM352483 GSM352484 GSM352486 GSM352487 GSM352489 GSM352490
## AFX-BioB-3_at    10.76171  10.68561  10.57062  10.74651  10.77081  10.70444
## AFX-BioB-5_at    10.28333  10.15720  10.13767  10.18003  10.16653  10.21192
```

##	AFFX-BioB-M_at	11.08337	10.99806	10.96635	11.03466	11.04695	11.03495
##	AFFX-BioC-3_at	10.68880	10.66345	10.54643	10.68763	10.69229	10.68256
##	AFFX-BioC-5_at	11.48147	11.41077	11.35575	11.48866	11.44004	11.43314
##	AFFX-BioDn-3_at	13.10346	13.14335	13.03147	13.10297	13.14417	13.08528
##		GSM352492	GSM352493	GSM352495	GSM352496	GSM352497	GSM352498
##	AFFX-BioB-3_at	10.73014	10.97719	10.124147	10.548898	10.346205	10.288886
##	AFFX-BioB-5_at	10.29522	10.41677	9.602872	9.932916	9.807584	9.852111
##	AFFX-BioB-M_at	11.02916	11.26630	10.367117	10.829853	10.691065	10.654976
##	AFFX-BioC-3_at	10.62826	10.91128	10.242515	10.599081	10.454556	10.396623
##	AFFX-BioC-5_at	11.47562	11.69006	11.018212	11.430983	11.266079	11.195135
##	AFFX-BioDn-3_at	13.10052	13.32706	12.967662	13.226914	13.100389	12.989146
##		GSM352500	GSM352501	GSM352503	GSM352504	GSM352505	GSM352507
##	AFFX-BioB-3_at	10.362456	10.100221	10.297114	11.79192	10.48880	11.09285
##	AFFX-BioB-5_at	9.839513	9.556148	9.697522	11.31840	10.02543	10.56452
##	AFFX-BioB-M_at	10.681215	10.434083	10.617794	12.02467	10.83431	11.39018
##	AFFX-BioC-3_at	10.469075	10.177521	10.411046	11.82722	10.58909	11.14755
##	AFFX-BioC-5_at	11.271082	11.005890	11.104592	12.46614	11.39308	11.90818
##	AFFX-BioDn-3_at	13.168255	13.024496	13.055826	13.60558	13.15203	13.43027
##		GSM352508	GSM352510	GSM352511	GSM352512	GSM352514	GSM352515
##	AFFX-BioB-3_at	10.64064	10.83838	10.148714	10.171001	11.17832	9.910880
##	AFFX-BioB-5_at	10.10279	10.40254	9.763192	9.719322	10.79082	9.439334
##	AFFX-BioB-M_at	10.91986	11.22559	10.575706	10.589205	11.55097	10.375304
##	AFFX-BioC-3_at	10.71651	10.98085	10.352880	10.359459	11.39696	10.129617
##	AFFX-BioC-5_at	11.52784	11.67183	11.104397	11.109166	11.99056	10.857110
##	AFFX-BioDn-3_at	13.27912	13.27298	13.065265	13.002176	13.47218	12.903659
##		GSM352516	GSM352517	GSM352518	GSM352519	GSM352520	GSM352521
##	AFFX-BioB-3_at	10.007633	10.54150	10.43455	10.431128	10.031671	10.653492
##	AFFX-BioB-5_at	9.481095	10.21263	10.09614	9.880242	9.565338	9.971229
##	AFFX-BioB-M_at	10.394014	10.98379	10.85257	10.772572	10.371806	10.889090
##	AFFX-BioC-3_at	10.186420	10.74580	10.62278	10.561206	10.197748	10.663839
##	AFFX-BioC-5_at	10.975450	11.47301	11.34610	11.301851	11.002286	11.464994
##	AFFX-BioDn-3_at	12.995699	13.17990	13.11197	13.164354	12.955005	13.324631
##		GSM352522	GSM352523	GSM352524	GSM352525	GSM352526	GSM352527
##	AFFX-BioB-3_at	11.06603	10.83872	10.160720	10.51893	10.296459	10.37480
##	AFFX-BioB-5_at	10.55669	10.50597	9.741786	10.02490	9.846137	9.91919
##	AFFX-BioB-M_at	11.37943	11.24970	10.607940	10.91114	10.676582	10.74699
##	AFFX-BioC-3_at	11.22273	10.97836	10.368945	10.66553	10.481412	10.56680
##	AFFX-BioC-5_at	11.90733	11.74023	11.117630	11.43670	11.276267	11.37689
##	AFFX-BioDn-3_at	13.48931	13.41473	13.114386	13.26271	13.186896	13.11631
##		GSM352528	GSM352529	GSM352530	GSM352531	GSM352532	GSM352533
##	AFFX-BioB-3_at	10.379987	10.86907	10.75420	10.47950	10.66943	10.264458
##	AFFX-BioB-5_at	9.944453	10.41085	10.32796	10.03740	10.17541	9.742432
##	AFFX-BioB-M_at	10.822718	11.27146	11.20862	10.90100	11.01125	10.676167
##	AFFX-BioC-3_at	10.606905	11.01792	10.95012	10.65976	10.81330	10.651410
##	AFFX-BioC-5_at	11.378044	11.81112	11.75003	11.44999	11.58541	11.444923
##	AFFX-BioDn-3_at	13.281209	13.36919	13.36700	13.24174	13.38147	13.283856
##		GSM352534	GSM352535	GSM352536	GSM352537	GSM352538	GSM352539
##	AFFX-BioB-3_at	10.368794	10.182634	10.68092	10.370056	10.330443	10.69151
##	AFFX-BioB-5_at	9.857719	9.590479	10.17171	9.832423	9.784331	10.11360
##	AFFX-BioB-M_at	10.733956	10.547618	11.08224	10.756083	10.726881	11.04376

```
## AFX-BioC-3_at 10.559524 10.537775 11.02864 10.696841 10.669585 11.04404
## AFX-BioC-5_at 11.417828 11.355207 11.84012 11.522411 11.442056 11.79442
## AFX-BioDn-3_at 13.081369 13.260035 13.39789 13.267628 13.242596 13.45147
## GSM352540 GSM352541 GSM352542 GSM352543 GSM352544 GSM352545
## AFX-BioB-3_at 10.326056 10.62738 10.213533 10.41838 10.203904 10.171635
## AFX-BioB-5_at 9.838021 10.10597 9.698102 10.00251 9.792832 9.774741
## AFX-BioB-M_at 10.780496 10.98064 10.610691 10.83626 10.624502 10.625988
## AFX-BioC-3_at 10.669123 10.92678 10.497910 10.72307 10.520461 10.566844
## AFX-BioC-5_at 11.534022 11.74263 11.359097 11.65307 11.433640 11.387019
## AFX-BioDn-3_at 13.353354 13.33736 13.229074 13.18755 13.157004 13.180407
## GSM352546 GSM352547 GSM352548 GSM352549 GSM352550 GSM352551
## AFX-BioB-3_at 10.81328 11.22788 10.57445 10.304978 10.345810 10.092580
## AFX-BioB-5_at 10.35814 10.76735 10.08430 9.782221 9.883265 9.706487
## AFX-BioB-M_at 11.18876 11.53656 10.98867 10.715034 10.758199 10.558858
## AFX-BioC-3_at 11.10720 11.37077 10.90157 10.616143 10.667104 10.493181
## AFX-BioC-5_at 11.96888 12.21693 11.76804 11.478659 11.504585 11.293226
## AFX-BioDn-3_at 13.35810 13.54472 13.34529 13.283913 13.224406 13.184534
## GSM352552 GSM352553
## AFX-BioB-3_at 10.277888 10.65908
## AFX-BioB-5_at 9.734644 10.16479
## AFX-BioB-M_at 10.721479 11.14311
## AFX-BioC-3_at 10.630465 11.03614
## AFX-BioC-5_at 11.397320 11.91022
## AFX-BioDn-3_at 13.185466 13.42056
```

```
# Column-wise summary statistics,3
round(apply(ex,2, summary),3)
```

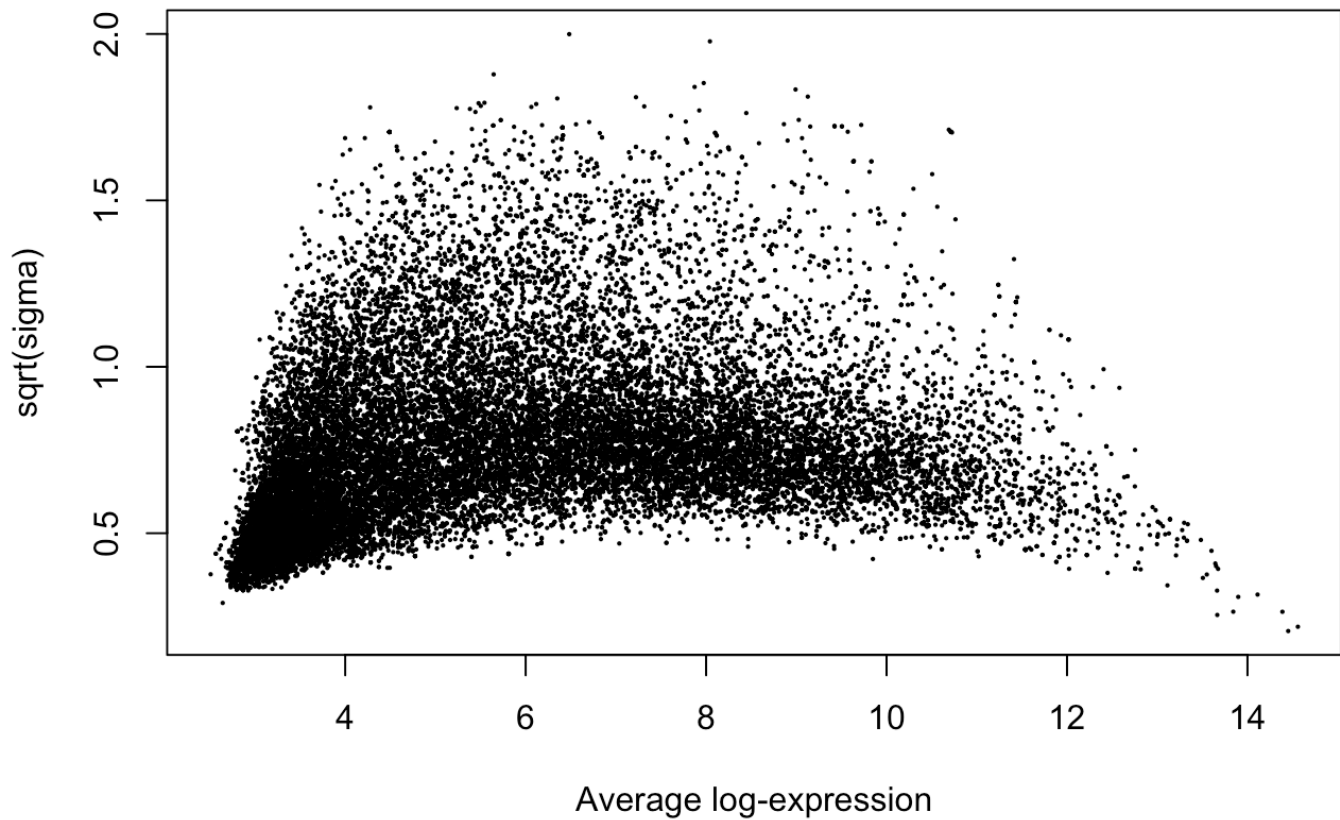
```
## GSM352448 GSM352450 GSM352451 GSM352453 GSM352454 GSM352455 GSM352457
## Min. 2.286 2.342 2.263 2.151 2.376 2.251 2.309
## 1st Qu. 3.852 3.604 3.599 3.797 3.690 3.838 3.654
## Median 5.134 5.263 5.215 5.187 5.133 5.065 5.207
## Mean 5.770 5.760 5.756 5.785 5.762 5.747 5.763
## 3rd Qu. 7.317 7.513 7.479 7.367 7.447 7.248 7.447
## Max. 14.564 14.590 14.486 14.590 14.541 14.525 14.608
## GSM352458 GSM352460 GSM352461 GSM352463 GSM352464 GSM352466 GSM352467
## Min. 2.314 2.345 2.287 2.295 2.312 2.399 2.221
## 1st Qu. 3.671 3.648 3.477 3.446 3.467 3.513 3.497
## Median 5.189 5.208 5.325 5.269 5.175 5.157 5.200
## Mean 5.767 5.797 5.778 5.782 5.761 5.751 5.751
## 3rd Qu. 7.476 7.559 7.621 7.664 7.664 7.589 7.575
## Max. 14.551 14.604 14.522 14.523 14.494 14.570 14.528
## GSM352469 GSM352470 GSM352471 GSM352473 GSM352475 GSM352476 GSM352477
## Min. 2.312 2.309 2.257 2.282 2.307 2.312 2.329
## 1st Qu. 3.485 3.502 3.503 3.468 3.460 3.603 3.509
## Median 5.312 5.307 5.247 5.221 5.225 5.297 5.131
## Mean 5.790 5.804 5.787 5.765 5.770 5.817 5.759
## 3rd Qu. 7.642 7.679 7.651 7.661 7.615 7.603 7.623
## Max. 14.594 14.486 14.486 14.576 14.488 14.527 14.511
```

##	GSM352479	GSM352480	GSM352481	GSM352483	GSM352484	GSM352486	GSM352487
## Min.	2.408	2.368	2.266	2.277	2.259	2.227	2.256
## 1st Qu.	3.571	3.551	3.613	3.593	3.502	3.549	3.546
## Median	5.105	5.118	5.044	5.167	5.189	5.166	5.037
## Mean	5.735	5.734	5.755	5.763	5.749	5.754	5.726
## 3rd Qu.	7.524	7.539	7.530	7.539	7.586	7.564	7.555
## Max.	14.541	14.528	14.528	14.561	14.588	14.532	14.463
##	GSM352489	GSM352490	GSM352492	GSM352493	GSM352495	GSM352496	GSM352497
## Min.	2.382	2.389	2.276	2.208	2.196	2.252	2.377
## 1st Qu.	3.531	3.575	3.551	3.631	3.552	3.606	3.541
## Median	5.049	5.019	5.111	5.256	5.132	5.370	5.198
## Mean	5.733	5.731	5.735	5.808	5.780	5.786	5.766
## 3rd Qu.	7.583	7.564	7.540	7.602	7.657	7.552	7.603
## Max.	14.537	14.494	14.537	14.541	14.427	14.551	14.440
##	GSM352498	GSM352500	GSM352501	GSM352503	GSM352504	GSM352505	GSM352507
## Min.	2.401	2.263	2.227	2.082	2.274	2.316	2.368
## 1st Qu.	3.609	3.566	3.513	3.514	3.887	3.587	3.710
## Median	4.995	5.285	5.108	5.097	5.170	5.223	5.203
## Mean	5.749	5.759	5.765	5.754	5.777	5.761	5.774
## 3rd Qu.	7.602	7.517	7.679	7.645	7.339	7.528	7.474
## Max.	14.563	14.487	14.617	14.537	14.578	14.564	14.577
##	GSM352508	GSM352510	GSM352511	GSM352512	GSM352514	GSM352515	GSM352516
## Min.	2.270	2.354	2.227	2.280	2.208	2.291	2.300
## 1st Qu.	3.659	3.601	3.571	3.549	3.713	3.552	3.503
## Median	5.165	5.397	5.296	5.012	5.302	5.069	5.144
## Mean	5.792	5.794	5.773	5.734	5.755	5.751	5.754
## 3rd Qu.	7.564	7.553	7.551	7.587	7.390	7.612	7.604
## Max.	14.489	14.578	14.601	14.556	14.482	14.537	14.524
##	GSM352517	GSM352518	GSM352519	GSM352520	GSM352521	GSM352522	GSM352523
## Min.	2.276	2.342	2.394	2.346	2.408	2.324	2.389
## 1st Qu.	3.615	3.624	3.583	3.516	3.640	3.626	3.662
## Median	5.212	5.221	5.255	5.067	5.230	5.243	5.258
## Mean	5.748	5.758	5.748	5.749	5.806	5.760	5.770
## 3rd Qu.	7.503	7.524	7.509	7.610	7.592	7.483	7.496
## Max.	14.573	14.537	14.563	14.578	14.591	14.590	14.697
##	GSM352524	GSM352525	GSM352526	GSM352527	GSM352528	GSM352529	GSM352530
## Min.	2.187	2.176	2.282	2.216	2.301	2.238	2.320
## 1st Qu.	3.520	3.569	3.536	3.521	3.510	3.620	3.610
## Median	5.184	5.316	5.076	5.125	5.303	5.409	5.397
## Mean	5.771	5.775	5.761	5.743	5.760	5.793	5.789
## 3rd Qu.	7.582	7.546	7.660	7.566	7.583	7.524	7.528
## Max.	14.538	14.564	14.617	14.522	14.594	14.543	14.632
##	GSM352531	GSM352532	GSM352533	GSM352534	GSM352535	GSM352536	GSM352537
## Min.	2.314	2.290	2.305	2.254	2.170	2.358	2.340
## 1st Qu.	3.525	3.638	3.499	3.548	3.454	3.552	3.541
## Median	5.132	5.225	5.250	5.084	5.138	5.217	5.075
## Mean	5.746	5.793	5.741	5.727	5.738	5.757	5.734
## 3rd Qu.	7.583	7.576	7.566	7.521	7.649	7.569	7.565
## Max.	14.564	14.590	14.591	14.577	14.549	14.601	14.616
##	GSM352538	GSM352539	GSM352540	GSM352541	GSM352542	GSM352543	GSM352544

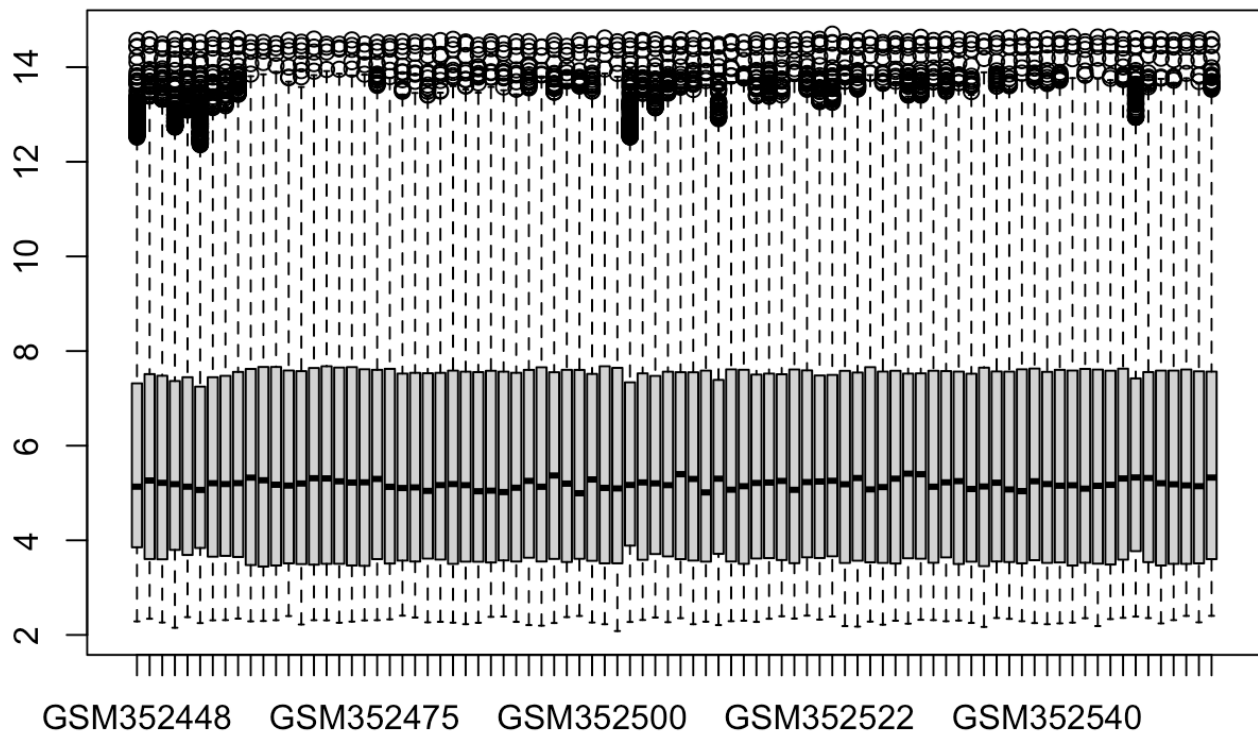
## Min.	2.290	2.251	2.227	2.245	2.260	2.353	2.188
## 1st Qu.	3.513	3.580	3.522	3.532	3.467	3.525	3.506
## Median	5.039	5.248	5.189	5.154	5.163	5.092	5.152
## Mean	5.727	5.799	5.743	5.762	5.731	5.751	5.746
## 3rd Qu.	7.614	7.630	7.560	7.603	7.588	7.622	7.609
## Max.	14.640	14.577	14.601	14.590	14.640	14.537	14.637
##	GSM352545	GSM352546	GSM352547	GSM352548	GSM352549	GSM352550	GSM352551
## Min.	2.335	2.361	2.388	2.356	2.244	2.314	2.402
## 1st Qu.	3.489	3.591	3.771	3.538	3.468	3.502	3.502
## Median	5.172	5.306	5.328	5.316	5.206	5.183	5.161
## Mean	5.734	5.813	5.783	5.770	5.746	5.751	5.753
## 3rd Qu.	7.584	7.627	7.422	7.556	7.587	7.587	7.610
## Max.	14.639	14.590	14.601	14.573	14.590	14.566	14.564
##	GSM352552	GSM352553					
## Min.	2.267	2.403					
## 1st Qu.	3.513	3.602					
## Median	5.143	5.326					
## Mean	5.739	5.804					
## 3rd Qu.	7.571	7.565					
## Max.	14.578	14.590					

```
# mean-variance trend
ex <- na.omit(ex) # eliminate rows with NAs
plotSA(lmFit(ex), main="Mean variance trend, GSE14038")
```

Mean variance trend, GSE14038



```
boxplot(ex)
```

```
colnames(pData(gset))
```

```
## [1] "title" "geo_accession"
## [3] "status" "submission_date"
## [5] "last_update_date" "type"
## [7] "channel_count" "source_name_ch1"
## [9] "organism_ch1" "characteristics_ch1"
## [11] "molecule_ch1" "extract_protocol_ch1"
## [13] "label_ch1" "label_protocol_ch1"
## [15] "taxid_ch1" "hyb_protocol"
## [17] "scan_protocol" "description"
## [19] "data_processing" "platform_id"
## [21] "contact_name" "contact_email"
## [23] "contact_phone" "contact_institute"
## [25] "contact_address" "contact_city"
## [27] "contact_state" "contact_zip/postal_code"
## [29] "contact_country" "supplementary_file"
## [31] "data_row_count" "relation"
```

```
pData(gset)$data_processing[1]
```

```
## [1] "Microarrays were analyzed with Affymetrix Microarray Suite 5.0 to generate
"CEL" files that were processed using Robust Multichip Analysis (RMA) in Bioconduc
tor/R. A custom CDF was used (Brainarray, Hs133P_Hs_REFSEQ_8)."
```

We now have the metadata, phenodata (sample data), and experimental data (microarray intensities) for this experiment. Now we have to download the raw data files to your computer. For Affymetrix microarrays, the raw data format is a CEL file.

```
#downloading raw data
celFiles <- list.celfiles("~/Desktop/hack4rare/GSE14038_RAW", full.names = TRUE)
Data <- read.celfiles(celFiles)
```

```
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352448.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352450.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352451.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352453.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352454.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352455.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352457.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352458.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352460.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352461.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352463.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352464.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352466.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352467.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352469.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352470.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352471.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352473.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352475.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352476.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352477.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352479.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352480.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352481.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352483.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352484.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352486.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352487.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352489.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352490.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352492.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352493.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352495.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352496.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352497.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352498.CEL
## Reading in : /Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352500.CEL
```

[illegible]

```
head(Data)
```

```
## ExpressionFeatureSet (storageMode: lockedEnvironment)
## assayData: 6 features, 86 samples
##   element names: exprs
## protocolData
##   rowNames: GSM352448.CEL GSM352450.CEL ... GSM352553.CEL (86 total)
##   varLabels: exprs dates
##   varMetadata: labelDescription channel
## phenoData
##   rowNames: GSM352448.CEL GSM352450.CEL ... GSM352553.CEL (86 total)
##   varLabels: index
##   varMetadata: labelDescription channel
## featureData: none
## experimentData: use 'experimentData(object)'
## Annotation: pd.hg.u133.plus.2
```

STEP 2 - Processing Microarray Data

The first step to process CEL files with “affy” is to tell the program the directory that it can find the CEL files and the corresponding file with phenoData (sample data).

Preparing the Phenodata

The rownames of the data frame must be the same as the CEL file names

```
head(celFiles)
```

```
## [1] "/Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352448.CEL"
## [2] "/Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352450.CEL"
## [3] "/Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352451.CEL"
## [4] "/Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352453.CEL"
## [5] "/Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352454.CEL"
## [6] "/Users/teresa/Desktop/hack4rare/GSE14038_RAW/GSM352455.CEL"
```

```
#eliminate unseful text
celFiles <- gsub("/Users/teresa/Desktop/hack4rare/GSE14038_RAW/", "", celFiles)
```

```
my.pdata <- as.data.frame(pData(gset), stringsAsFactors=F)
my.pdata <- my.pdata[, c("title", "geo_accession", "description")]
my.pdata <- my.pdata[order(rownames(my.pdata)), ]
head(my.pdata, 10)
```

```
##          title geo_accession
## GSM352448      dNF_AS18      GSM352448
## GSM352450      dNF_AS5       GSM352450
## GSM352451      MPNST_AS10    GSM352451
## GSM352453      MPNST_AS13    GSM352453
## GSM352454      MPNST_AS15    GSM352454
## GSM352455      pNF_AS16     GSM352455
## GSM352457      pNF_AS7      GSM352457
## GSM352458      pNF_AS8      GSM352458
## GSM352460      Universal_STD_1 GSM352460
## GSM352461      88_3         GSM352461
##
##                                     description
## GSM352448      Gene expression data from neurofibroma tumor.
## GSM352450      Gene expression data from neurofibroma tumor.
## GSM352451      Gene expression data from neurofibroma tumor.
## GSM352453      Gene expression data from neurofibroma tumor.
## GSM352454      Gene expression data from neurofibroma tumor.
## GSM352455      Gene expression data from neurofibroma tumor.
## GSM352457      Gene expression data from neurofibroma tumor.
## GSM352458      Gene expression data from neurofibroma tumor.
## GSM352460      Gene expression data from whole mouse.
## GSM352461      Gene expression data from neurofibroma-derived cell culture.
```

```
group <- my.pdata$description
```

```
#eliminate unseful text
group <- gsub("Gene expression data from", "", group)
my.pdata <- cbind(my.pdata, group)
my.pdata
```

```
##          title geo_accession
## GSM352448      dNF_AS18      GSM352448
## GSM352450      dNF_AS5       GSM352450
## GSM352451      MPNST_AS10    GSM352451
## GSM352453      MPNST_AS13    GSM352453
## GSM352454      MPNST_AS15    GSM352454
## GSM352455      pNF_AS16     GSM352455
## GSM352457      pNF_AS7      GSM352457
## GSM352458      pNF_AS8      GSM352458
## GSM352460      Universal_STD_1 GSM352460
## GSM352461      88_3         GSM352461
## GSM352463      90_8         GSM352463
## GSM352464      pNF00.6      GSM352464
## GSM352466      pNF95.3      GSM352466
## GSM352467      pNF95.6      GSM352467
## GSM352469      S462         GSM352469
## GSM352470      S520         GSM352470
## GSM352471      ST8814       GSM352471
```

## GSM352473	STS26T	GSM352473
## GSM352475	T265	GSM352475
## GSM352476	Universal_STD_2	GSM352476
## GSM352477	YST1	GSM352477
## GSM352479	ABC_8N_-1-	GSM352479
## GSM352480	AIBC_2N_-1-	GSM352480
## GSM352481	CLT_6N_+1-	GSM352481
## GSM352483	MGF_33N_+1-	GSM352483
## GSM352484	MPNST_02_2	GSM352484
## GSM352486	MPNST_94_3	GSM352486
## GSM352487	NHSC_303	GSM352487
## GSM352489	NHSC_339	GSM352489
## GSM352490	NHSC_771	GSM352490
## GSM352492	SCC_7N_-1-	GSM352492
## GSM352493	Universal_STD_3	GSM352493
## GSM352495	ADN1N_KO	GSM352495
## GSM352496	dNF.AS3	GSM352496
## GSM352497	dNF.AS46	GSM352497
## GSM352498	MPNST97.6	GSM352498
## GSM352500	MPNST.AS37	GSM352500
## GSM352501	NHSC216	GSM352501
## GSM352503	pNF03.3	GSM352503
## GSM352504	pNF.AS12	GSM352504
## GSM352505	pNF.AS38	GSM352505
## GSM352507	pNF.AS40	GSM352507
## GSM352508	UniversalSTD_1	GSM352508
## GSM352510	dNF.AS22	GSM352510
## GSM352511	dNF.AS4	GSM352511
## GSM352512	MPNST95.3b	GSM352512
## GSM352514	MPNST.AS42	GSM352514
## GSM352515	NHSC323	GSM352515
## GSM352516	pNF04.4	GSM352516
## GSM352517	pNF.AS23	GSM352517
## GSM352518	pNF.AS24	GSM352518
## GSM352519	pNF.AS47	GSM352519
## GSM352520	RMN9N_KO	GSM352520
## GSM352521	UniversalSTD_2	GSM352521
## GSM352522	dNF.AS35	GSM352522
## GSM352523	dNF.AS44	GSM352523
## GSM352524	MPNST96.2	GSM352524
## GSM352525	MPNST.AS45	GSM352525
## GSM352526	NHSC338	GSM352526
## GSM352527	pNF05.3	GSM352527
## GSM352528	pNF.AS32	GSM352528
## GSM352529	pNF.AS33	GSM352529
## GSM352530	pNF.AS34	GSM352530
## GSM352531	SCC5N_KO	GSM352531
## GSM352532	UniversalSTD_3	GSM352532
## GSM352533	dNF_AS55	GSM352533
## GSM352534	dNFSC_ERS	GSM352534

##	GSM352535	NHSC_286	GSM352535
##	GSM352536	pNF_AS48	GSM352536
##	GSM352537	pNFSC_04.7	GSM352537
##	GSM352538	pNFSC_05.4	GSM352538
##	GSM352539	Uni_STD_1	GSM352539
##	GSM352540	dNF_CSG94	GSM352540
##	GSM352541	dNF_CSG95	GSM352541
##	GSM352542	dNFSC_ABB	GSM352542
##	GSM352543	NHSC_J017	GSM352543
##	GSM352544	pNFSC_00.13	GSM352544
##	GSM352545	pNFSC_05.5	GSM352545
##	GSM352546	Uni_STD_2	GSM352546
##	GSM352547	dNF_AS50_HG_U13Plus_2.CEL	GSM352547
##	GSM352548	DNF_LRS_HG_U13Plus_2.CEL	GSM352548
##	GSM352549	dNFSC_JLM_HG_U13Plus_2.CEL	GSM352549
##	GSM352550	NHSC_02.8_HG_U13Plus_2.CEL	GSM352550
##	GSM352551	NHSC_J037_HG_U13Plus_2.CEL	GSM352551
##	GSM352552	pNFSC_97.9_HG_U13Plus_2.CEL	GSM352552
##	GSM352553	Uni_STD_3	GSM352553
##			description
##	GSM352448	Gene expression data from neurofibroma tumor.	
##	GSM352450	Gene expression data from neurofibroma tumor.	
##	GSM352451	Gene expression data from neurofibroma tumor.	
##	GSM352453	Gene expression data from neurofibroma tumor.	
##	GSM352454	Gene expression data from neurofibroma tumor.	
##	GSM352455	Gene expression data from neurofibroma tumor.	
##	GSM352457	Gene expression data from neurofibroma tumor.	
##	GSM352458	Gene expression data from neurofibroma tumor.	
##	GSM352460	Gene expression data from whole mouse.	
##	GSM352461	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352463	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352464	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352466	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352467	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352469	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352470	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352471	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352473	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352475	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352476	Gene expression data from whole mouse.	
##	GSM352477	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352479	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352480	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352481	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352483	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352484	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352486	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352487	Gene expression data from normal human Schwann cell culture.	
##	GSM352489	Gene expression data from normal human Schwann cell culture.	
##	GSM352490	Gene expression data from normal human Schwann cell culture.	

```
## GSM352492 Gene expression data from neurofibroma-derived cell culture.
## GSM352493 Gene expression data from whole mouse.
## GSM352495 Gene expression data from neurofibroma-derived cell culture.
## GSM352496 Gene expression data from neurofibroma tumor.
## GSM352497 Gene expression data from neurofibroma tumor.
## GSM352498 Gene expression data from neurofibroma-derived cell culture.
## GSM352500 Gene expression data from neurofibroma tumor.
## GSM352501 Gene expression data from normal human Schwann cell culture.
## GSM352503 Gene expression data from neurofibroma-derived cell culture.
## GSM352504 Gene expression data from neurofibroma tumor.
## GSM352505 Gene expression data from neurofibroma tumor.
## GSM352507 Gene expression data from neurofibroma tumor.
## GSM352508 Gene expression data from whole mouse.
## GSM352510 Gene expression data from neurofibroma tumor.
## GSM352511 Gene expression data from neurofibroma tumor.
## GSM352512 Gene expression data from neurofibroma-derived cell culture.
## GSM352514 Gene expression data from neurofibroma tumor.
## GSM352515 Gene expression data from normal human Schwann cell culture.
## GSM352516 Gene expression data from neurofibroma-derived cell culture.
## GSM352517 Gene expression data from neurofibroma tumor.
## GSM352518 Gene expression data from neurofibroma tumor.
## GSM352519 Gene expression data from neurofibroma tumor.
## GSM352520 Gene expression data from neurofibroma-derived cell culture.
## GSM352521 Gene expression data from whole mouse.
## GSM352522 Gene expression data from neurofibroma tumor.
## GSM352523 Gene expression data from neurofibroma tumor.
## GSM352524 Gene expression data from neurofibroma-derived cell culture.
## GSM352525 Gene expression data from neurofibroma tumor.
## GSM352526 Gene expression data from normal human Schwann cell culture.
## GSM352527 Gene expression data from neurofibroma-derived cell culture.
## GSM352528 Gene expression data from neurofibroma tumor.
## GSM352529 Gene expression data from neurofibroma tumor.
## GSM352530 Gene expression data from neurofibroma tumor.
## GSM352531 Gene expression data from neurofibroma-derived cell culture.
## GSM352532 Gene expression data from whole mouse.
## GSM352533 Gene expression data from neurofibroma tumor.
## GSM352534 Gene expression data from neurofibroma-derived cell culture.
## GSM352535 Gene expression data from normal human Schwann cell culture.
## GSM352536 Gene expression data from neurofibroma tumor.
## GSM352537 Gene expression data from neurofibroma-derived cell culture.
## GSM352538 Gene expression data from neurofibroma-derived cell culture.
## GSM352539 Gene expression data from whole mouse.
## GSM352540 Gene expression data from neurofibroma tumor.
## GSM352541 Gene expression data from neurofibroma tumor.
## GSM352542 Gene expression data from neurofibroma-derived cell culture.
## GSM352543 Gene expression data from normal human Schwann cell culture.
## GSM352544 Gene expression data from neurofibroma-derived cell culture.
## GSM352545 Gene expression data from neurofibroma-derived cell culture.
## GSM352546 Gene expression data from whole mouse.
## GSM352547 Gene expression data from neurofibroma tumor.
```



```
## GSM352548           Gene expression data from neurofibroma tumor.
## GSM352549 Gene expression data from neurofibroma-derived cell culture.
## GSM352550 Gene expression data from normal human Schwann cell culture.
## GSM352551 Gene expression data from normal human Schwann cell culture.
## GSM352552 Gene expression data from neurofibroma-derived cell culture.
## GSM352553           Gene expression data from whole mouse.
##
##                               group
## GSM352448           neurofibroma tumor.
## GSM352450           neurofibroma tumor.
## GSM352451           neurofibroma tumor.
## GSM352453           neurofibroma tumor.
## GSM352454           neurofibroma tumor.
## GSM352455           neurofibroma tumor.
## GSM352457           neurofibroma tumor.
## GSM352458           neurofibroma tumor.
## GSM352460           whole mouse.
## GSM352461 neurofibroma-derived cell culture.
## GSM352463 neurofibroma-derived cell culture.
## GSM352464 neurofibroma-derived cell culture.
## GSM352466 neurofibroma-derived cell culture.
## GSM352467 neurofibroma-derived cell culture.
## GSM352469 neurofibroma-derived cell culture.
## GSM352470 neurofibroma-derived cell culture.
## GSM352471 neurofibroma-derived cell culture.
## GSM352473 neurofibroma-derived cell culture.
## GSM352475 neurofibroma-derived cell culture.
## GSM352476           whole mouse.
## GSM352477 neurofibroma-derived cell culture.
## GSM352479 neurofibroma-derived cell culture.
## GSM352480 neurofibroma-derived cell culture.
## GSM352481 neurofibroma-derived cell culture.
## GSM352483 neurofibroma-derived cell culture.
## GSM352484 neurofibroma-derived cell culture.
## GSM352486 neurofibroma-derived cell culture.
## GSM352487 normal human Schwann cell culture.
## GSM352489 normal human Schwann cell culture.
## GSM352490 normal human Schwann cell culture.
## GSM352492 neurofibroma-derived cell culture.
## GSM352493           whole mouse.
## GSM352495 neurofibroma-derived cell culture.
## GSM352496           neurofibroma tumor.
## GSM352497           neurofibroma tumor.
## GSM352498 neurofibroma-derived cell culture.
## GSM352500           neurofibroma tumor.
## GSM352501 normal human Schwann cell culture.
## GSM352503 neurofibroma-derived cell culture.
## GSM352504           neurofibroma tumor.
## GSM352505           neurofibroma tumor.
## GSM352507           neurofibroma tumor.
## GSM352508           whole mouse.
```

```
## GSM352510      neurofibroma tumor.
## GSM352511      neurofibroma tumor.
## GSM352512      neurofibroma-derived cell culture.
## GSM352514      neurofibroma tumor.
## GSM352515      normal human Schwann cell culture.
## GSM352516      neurofibroma-derived cell culture.
## GSM352517      neurofibroma tumor.
## GSM352518      neurofibroma tumor.
## GSM352519      neurofibroma tumor.
## GSM352520      neurofibroma-derived cell culture.
## GSM352521      whole mouse.
## GSM352522      neurofibroma tumor.
## GSM352523      neurofibroma tumor.
## GSM352524      neurofibroma-derived cell culture.
## GSM352525      neurofibroma tumor.
## GSM352526      normal human Schwann cell culture.
## GSM352527      neurofibroma-derived cell culture.
## GSM352528      neurofibroma tumor.
## GSM352529      neurofibroma tumor.
## GSM352530      neurofibroma tumor.
## GSM352531      neurofibroma-derived cell culture.
## GSM352532      whole mouse.
## GSM352533      neurofibroma tumor.
## GSM352534      neurofibroma-derived cell culture.
## GSM352535      normal human Schwann cell culture.
## GSM352536      neurofibroma tumor.
## GSM352537      neurofibroma-derived cell culture.
## GSM352538      neurofibroma-derived cell culture.
## GSM352539      whole mouse.
## GSM352540      neurofibroma tumor.
## GSM352541      neurofibroma tumor.
## GSM352542      neurofibroma-derived cell culture.
## GSM352543      normal human Schwann cell culture.
## GSM352544      neurofibroma-derived cell culture.
## GSM352545      neurofibroma-derived cell culture.
## GSM352546      whole mouse.
## GSM352547      neurofibroma tumor.
## GSM352548      neurofibroma tumor.
## GSM352549      neurofibroma-derived cell culture.
## GSM352550      normal human Schwann cell culture.
## GSM352551      normal human Schwann cell culture.
## GSM352552      neurofibroma-derived cell culture.
## GSM352553      whole mouse.
```

```
head(celFiles)
```

```
## [1] "GSM352448.CEL" "GSM352450.CEL" "GSM352451.CEL" "GSM352453.CEL"
## [5] "GSM352454.CEL" "GSM352455.CEL"
```

```
head(rownames(my.pdata))
```

```
## [1] "GSM352448" "GSM352450" "GSM352451" "GSM352453" "GSM352454" "GSM352455"
```

```
table(rownames(my.pdata) == celFiles)
```

```
##  
## FALSE  
##      86
```

```
temp.rownames <- paste(rownames(my.pdata), ".CEL", sep="")  
table(temp.rownames == celFiles)
```

```
##  
## TRUE  
##      86
```

```
rownames(my.pdata) <- temp.rownames  
rm(temp.rownames)  
table(rownames(my.pdata) == celFiles)
```

```
##  
## TRUE  
##      86
```

```
head(my.pdata)
```

```
##          title geo_accession
## GSM352448.CEL    dNF_AS18      GSM352448
## GSM352450.CEL    dNF_AS5       GSM352450
## GSM352451.CEL    MPNST_AS10     GSM352451
## GSM352453.CEL    MPNST_AS13     GSM352453
## GSM352454.CEL    MPNST_AS15     GSM352454
## GSM352455.CEL    pNF_AS16      GSM352455
##
##                                description
## GSM352448.CEL Gene expression data from neurofibroma tumor.
## GSM352450.CEL Gene expression data from neurofibroma tumor.
## GSM352451.CEL Gene expression data from neurofibroma tumor.
## GSM352453.CEL Gene expression data from neurofibroma tumor.
## GSM352454.CEL Gene expression data from neurofibroma tumor.
## GSM352455.CEL Gene expression data from neurofibroma tumor.
##
##                                group
## GSM352448.CEL neurofibroma tumor.
## GSM352450.CEL neurofibroma tumor.
## GSM352451.CEL neurofibroma tumor.
## GSM352453.CEL neurofibroma tumor.
## GSM352454.CEL neurofibroma tumor.
## GSM352455.CEL neurofibroma tumor.
```

Reading the CEL Files

Now that we have directory of CEL files and a corresponding data frame with the phenoData, we can read the CEL files into R

```
list.files("~/Desktop/hack4rare/GSE14038_RAW")
```

```
## [1] "arrayQualityMetrics report for gset"
## [2] "GPL7868.soft"
## [3] "GPL7869_Hs133P_Hs_REFSEQ_8_probe_tab 2.txt"
## [4] "GPL7869_Hs133P_Hs_REFSEQ_8_probe_tab.txt"
## [5] "GPL7869_Hs133P_Hs_REFSEQ_8_probe_tab.txt.gz"
## [6] "GPL7869_Hs133P_Hs_REFSEQ_8.cdf"
## [7] "GPL7869_Hs133P_Hs_REFSEQ_8.cdf 2"
## [8] "GPL7869_Hs133P_Hs_REFSEQ_8.cdf.gz"
## [9] "GPL7869.soft"
## [10] "GSE14038 1.2.Rmd"
## [11] "GSE14038_DataMatrix.csv"
## [12] "GSE14038_series_matrix 2.txt"
## [13] "GSE14038_series_matrix.txt"
## [14] "GSE14038_series_matrix.txt.gz"
## [15] "GSE14038-1.2_files"
## [16] "GSE14038-1.2.html"
## [17] "GSE14038-1.2.Rmd"
## [18] "GSE14038.Rmd"
## [19] "GSM352448.CEL"
```

```
## [20] "GSM352450.CEL"
## [21] "GSM352451.CEL"
## [22] "GSM352453.CEL"
## [23] "GSM352454.CEL"
## [24] "GSM352455.CEL"
## [25] "GSM352457.CEL"
## [26] "GSM352458.CEL"
## [27] "GSM352460.CEL"
## [28] "GSM352461.CEL"
## [29] "GSM352463.CEL"
## [30] "GSM352464.CEL"
## [31] "GSM352466.CEL"
## [32] "GSM352467.CEL"
## [33] "GSM352469.CEL"
## [34] "GSM352470.CEL"
## [35] "GSM352471.CEL"
## [36] "GSM352473.CEL"
## [37] "GSM352475.CEL"
## [38] "GSM352476.CEL"
## [39] "GSM352477.CEL"
## [40] "GSM352479.CEL"
## [41] "GSM352480.CEL"
## [42] "GSM352481.CEL"
## [43] "GSM352483.CEL"
## [44] "GSM352484.CEL"
## [45] "GSM352486.CEL"
## [46] "GSM352487.CEL"
## [47] "GSM352489.CEL"
## [48] "GSM352490.CEL"
## [49] "GSM352492.CEL"
## [50] "GSM352493.CEL"
## [51] "GSM352495.CEL"
## [52] "GSM352496.CEL"
## [53] "GSM352497.CEL"
## [54] "GSM352498.CEL"
## [55] "GSM352500.CEL"
## [56] "GSM352501.CEL"
## [57] "GSM352503.CEL"
## [58] "GSM352504.CEL"
## [59] "GSM352505.CEL"
## [60] "GSM352507.CEL"
## [61] "GSM352508.CEL"
## [62] "GSM352510.CEL"
## [63] "GSM352511.CEL"
## [64] "GSM352512.CEL"
## [65] "GSM352514.CEL"
## [66] "GSM352515.CEL"
## [67] "GSM352516.CEL"
## [68] "GSM352517.CEL"
## [69] "GSM352518.CEL"
```

```
## [70] "GSM352519.CEL"
## [71] "GSM352520.CEL"
## [72] "GSM352521.CEL"
## [73] "GSM352522.CEL"
## [74] "GSM352523.CEL"
## [75] "GSM352524.CEL"
## [76] "GSM352525.CEL"
## [77] "GSM352526.CEL"
## [78] "GSM352527.CEL"
## [79] "GSM352528.CEL"
## [80] "GSM352529.CEL"
## [81] "GSM352530.CEL"
## [82] "GSM352531.CEL"
## [83] "GSM352532.CEL"
## [84] "GSM352533.CEL"
## [85] "GSM352534.CEL"
## [86] "GSM352535.CEL"
## [87] "GSM352536.CEL"
## [88] "GSM352537.CEL"
## [89] "GSM352538.CEL"
## [90] "GSM352539.CEL"
## [91] "GSM352540.CEL"
## [92] "GSM352541.CEL"
## [93] "GSM352542.CEL"
## [94] "GSM352543.CEL"
## [95] "GSM352544.CEL"
## [96] "GSM352545.CEL"
## [97] "GSM352546.CEL"
## [98] "GSM352547.CEL"
## [99] "GSM352548.CEL"
## [100] "GSM352549.CEL"
## [101] "GSM352550.CEL"
## [102] "GSM352551.CEL"
## [103] "GSM352552.CEL"
## [104] "GSM352553.CEL"
```

```
##Perform affy normalization
```

```
library(affy)
```

```
my.affy <- ReadAffy(celfile.path = "~/Desktop/hack4rare/GSE14038_RAW", phenoData =  
my.pdata)
```

```
show(my.affy)
```

```
## AffyBatch object
## size of arrays=1164x1164 features (62 kb)
## cdf=HG-U133_Plus_2 (54675 affyids)
## number of samples=86
## number of genes=54675
## annotation=hgu133plus2
## notes=
```

```
head(exprs(my.affy),5)
```

```
## GSM352448.CEL GSM352450.CEL GSM352451.CEL GSM352453.CEL GSM352454.CEL
## 1          77          97          157          76          110
## 2          10794         13473         13152         11364         13024
## 3           86          110          161          115          117
## 4          11445         13744         13810         12189         13780
## 5           55           67          112           77           81
## GSM352455.CEL GSM352457.CEL GSM352458.CEL GSM352460.CEL GSM352461.CEL
## 1           89           87           67           87           91
## 2          13151         11932         11417         13548         13075
## 3           110          116          116          128          170
## 4          13531         12260         12037         14220         13555
## 5           92          126           85          104           81
## GSM352463.CEL GSM352464.CEL GSM352466.CEL GSM352467.CEL GSM352469.CEL
## 1           111          200          103          103          111
## 2          13131         12224         13331         10974         12978
## 3           147          171          112           85          116
## 4          13557         12582         13849         12639         13374
## 5           113          219          144          105          123
## GSM352470.CEL GSM352471.CEL GSM352473.CEL GSM352475.CEL GSM352476.CEL
## 1           132          119           92          127           85
## 2          9474         10799         12723         13463         11241
## 3           144          144          125          134          134
## 4          10609         12182         12639         13989         12435
## 5           129          159          120          164           78
## GSM352477.CEL GSM352479.CEL GSM352480.CEL GSM352481.CEL GSM352483.CEL
## 1           138           76           67          180           73
## 2          12251         9577          8044         10979         7126
## 3           148           97           85          197           73
## 4          12563         10331         9942          11976         8080
## 5           98           85           84          242           75
## GSM352484.CEL GSM352486.CEL GSM352487.CEL GSM352489.CEL GSM352490.CEL
## 1           66           77           98           92           70
## 2          11570         9151         10446         10177         9784
## 3           87          125          107          107           97
## 4          11954         9954         10904         10355         10606
## 5           94          123          126           86           60
## GSM352492.CEL GSM352493.CEL GSM352495.CEL GSM352496.CEL GSM352497.CEL
## 1           122           77           77           60           85
```

##	2	9059	10870	10156	10239	11895
##	3	129	102	88	75	99
##	4	8897	11312	10730	11409	11809
##	5	105	100	82	86	63
##	GSM352498.CEL GSM352500.CEL GSM352501.CEL GSM352503.CEL GSM352504.CEL					
##	1	121	81	83	91	101
##	2	9347	10446	10204	9583	10479
##	3	132	82	86	82	134
##	4	10124	10768	10584	10032	11208
##	5	141	87	64	78	110
##	GSM352505.CEL GSM352507.CEL GSM352508.CEL GSM352510.CEL GSM352511.CEL					
##	1	80	90	79	68	95
##	2	9870	10514	10441	3070	2958
##	3	105	134	113	94	103
##	4	10836	11013	11657	3227	3195
##	5	86	94	88	76	118
##	GSM352512.CEL GSM352514.CEL GSM352515.CEL GSM352516.CEL GSM352517.CEL					
##	1	91	67	80	152	67
##	2	2908	3395	4129	3744	4873
##	3	75	63	69	104	97
##	4	2958	3546	4260	3700	5188
##	5	60	54	95	109	80
##	GSM352518.CEL GSM352519.CEL GSM352520.CEL GSM352521.CEL GSM352522.CEL					
##	1	74	73	67	54	107
##	2	2623	3016	3066	5713	9324
##	3	78	80	87	62	114
##	4	2734	3224	3175	5729	9635
##	5	82	69	74	67	85
##	GSM352523.CEL GSM352524.CEL GSM352525.CEL GSM352526.CEL GSM352527.CEL					
##	1	93	126	127	204	92
##	2	10923	11975	7063	7765	6592
##	3	140	154	121	172	134
##	4	11295	12273	7005	7709	6635
##	5	110	154	135	198	88
##	GSM352528.CEL GSM352529.CEL GSM352530.CEL GSM352531.CEL GSM352532.CEL					
##	1	99	91	91	94	100
##	2	6608	6709	5976	6631	8434
##	3	129	142	113	130	149
##	4	6754	6983	6147	6541	8589
##	5	106	86	112	77	83
##	GSM352533.CEL GSM352534.CEL GSM352535.CEL GSM352536.CEL GSM352537.CEL					
##	1	93	131	95	100	123
##	2	11501	11101	12329	13757	12825
##	3	125	136	118	150	107
##	4	11881	11318	12848	14571	13930
##	5	122	378	121	134	112
##	GSM352538.CEL GSM352539.CEL GSM352540.CEL GSM352541.CEL GSM352542.CEL					
##	1	91	87	95	93	74
##	2	13645	14099	8735	9150	12095
##	3	115	144	125	139	137


```
## 4      13975      15013      9267      9634      12171
## 5      112      124      71      91      84
## GSM352543.CEL GSM352544.CEL GSM352545.CEL GSM352546.CEL GSM352547.CEL
## 1      90      80      92      113      88
## 2      9869      12648      8273      8942      12901
## 3      108      117      126      122      125
## 4      10532      12728      8388      8913      13516
## 5      86      89      75      76      95
## GSM352548.CEL GSM352549.CEL GSM352550.CEL GSM352551.CEL GSM352552.CEL
## 1      88      101      81      98      86
## 2      8542      9958      10037      9275      11068
## 3      128      125      107      134      106
## 4      8946      9904      9679      9396      11483
## 5      87      84      87      79      93
## GSM352553.CEL
## 1      164
## 2      7686
## 3      131
## 4      7809
## 5      59
```

```
dim(exprs(my.affy))
```

```
## [1] 1354896      86
```

```
colnames(pData(my.affy))
```

```
## [1] "title"      "geo_accession" "description"  "group"
```

Calculating Gene Expression Measurements

the expression data consists of 1354896 individual probe intensities for each of our 86 samples. We need to combine the individual probe intensities to probeset-level (gene-level) measurements. Each probeset typically consists of 11 - 20 individual probes. We will use the `rma()` function to combine the individual probe intensities to a probeset intensity.

```
##Calculate gene level expression measures
my.rma <- rma(my.affy, normalize=F, background=F)
```

```
## Calculating Expression
```

```
head(exprs(my.rma))
```

```
##      GSM352448.CEL GSM352450.CEL GSM352451.CEL GSM352453.CEL GSM352454.CEL
```

##	1007_s_at	6.645129	8.003747	7.030529	6.848707	7.073694
##	1053_at	6.222420	6.831108	7.432634	6.829680	7.280242
##	117_at	6.281286	6.573014	6.470188	6.534854	6.469390
##	121_at	6.209114	6.323868	6.494820	6.444614	6.627807
##	1255_g_at	5.752690	5.546719	5.949862	5.905944	5.971686
##	1294_at	6.240355	6.884619	6.444142	6.495858	6.392872
##	GSM352455.CEL	GSM352457.CEL	GSM352458.CEL	GSM352460.CEL	GSM352461.CEL	
##	1007_s_at	6.818725	7.079061	6.776019	7.720002	7.641538
##	1053_at	6.509329	6.313031	6.571129	7.672797	7.994916
##	117_at	6.871921	6.334965	7.320243	6.361891	6.662739
##	121_at	6.567756	6.405573	6.359454	6.617352	6.497731
##	1255_g_at	6.062997	5.853547	5.735549	5.949417	5.951687
##	1294_at	6.264290	6.306418	6.550735	6.453806	6.539290
##	GSM352463.CEL	GSM352464.CEL	GSM352466.CEL	GSM352467.CEL	GSM352469.CEL	
##	1007_s_at	7.921209	9.290329	7.589048	7.750189	7.748852
##	1053_at	8.133146	8.153098	7.417355	7.177170	8.454968
##	117_at	6.276630	7.354559	6.707741	6.298000	6.487925
##	121_at	6.459792	7.473902	7.103586	6.739787	6.664184
##	1255_g_at	5.927471	7.111800	6.464445	5.983947	6.040500
##	1294_at	6.306901	8.508872	6.906198	6.754694	6.407852
##	GSM352470.CEL	GSM352471.CEL	GSM352473.CEL	GSM352475.CEL	GSM352476.CEL	
##	1007_s_at	7.771032	8.034823	8.958499	7.968376	7.977549
##	1053_at	8.249130	8.572068	8.107228	9.035843	7.802335
##	117_at	6.936081	6.989006	6.519315	6.769918	6.454150
##	121_at	7.059947	7.165212	6.912320	6.946965	6.792367
##	1255_g_at	6.719671	6.857779	6.146290	6.165148	6.019507
##	1294_at	6.685760	6.802081	7.339145	6.550020	6.487362
##	GSM352477.CEL	GSM352479.CEL	GSM352480.CEL	GSM352481.CEL	GSM352483.CEL	
##	1007_s_at	8.227648	7.403657	8.068526	8.138220	7.142426
##	1053_at	7.835912	6.873022	6.434335	7.805853	6.908183
##	117_at	6.713865	6.261742	6.238260	7.536656	6.205019
##	121_at	7.123788	6.622881	6.450621	7.721149	6.632728
##	1255_g_at	6.343581	5.911373	5.803533	7.433398	5.899312
##	1294_at	6.536249	6.387346	6.264494	7.669891	6.318476
##	GSM352484.CEL	GSM352486.CEL	GSM352487.CEL	GSM352489.CEL	GSM352490.CEL	
##	1007_s_at	7.245419	7.636185	9.125605	8.475540	8.311684
##	1053_at	7.092769	6.928027	6.709723	7.245066	6.403728
##	117_at	6.082043	6.711082	6.875713	6.816054	6.320089
##	121_at	6.480821	6.856872	6.948154	6.719729	6.683543
##	1255_g_at	5.934069	6.527539	6.354755	6.028280	5.944587
##	1294_at	7.118295	7.125098	6.680354	6.586945	6.375934
##	GSM352492.CEL	GSM352493.CEL	GSM352495.CEL	GSM352496.CEL	GSM352497.CEL	
##	1007_s_at	8.182208	7.629818	9.013702	8.163313	6.894054
##	1053_at	6.861831	7.741052	7.438290	6.871596	6.674174
##	117_at	6.631474	6.541081	6.248296	6.502712	6.774071
##	121_at	6.830427	6.781940	6.418290	6.318206	6.516229
##	1255_g_at	6.420570	6.170181	5.866474	5.477623	5.878111
##	1294_at	6.696148	6.519700	6.229796	6.817969	6.654143
##	GSM352498.CEL	GSM352500.CEL	GSM352501.CEL	GSM352503.CEL	GSM352504.CEL	
##	1007_s_at	7.285636	7.454866	7.945298	8.687405	7.281667

##	1053_at	7.229825	7.178534	7.446685	6.751229	6.790326
##	117_at	6.938080	6.617383	6.441652	6.282731	6.645201
##	121_at	7.365002	6.618096	6.715268	6.665874	6.833650
##	1255_g_at	6.812840	5.908899	5.807080	6.013531	6.509419
##	1294_at	7.051025	6.698450	6.303214	6.337043	6.611672
##		GSM352505.CEL	GSM352507.CEL	GSM352508.CEL	GSM352510.CEL	GSM352511.CEL
##	1007_s_at	7.018756	7.751848	7.726523	7.451825	8.137630
##	1053_at	6.562222	6.988176	7.681359	6.638262	7.407614
##	117_at	6.248943	7.351726	6.427966	6.375535	7.025243
##	121_at	6.630285	6.771284	7.011080	6.138961	7.079782
##	1255_g_at	5.994791	6.293232	6.222723	5.677513	6.449601
##	1294_at	6.452136	7.100028	6.627653	6.993424	7.254659
##		GSM352512.CEL	GSM352514.CEL	GSM352515.CEL	GSM352516.CEL	GSM352517.CEL
##	1007_s_at	7.447951	7.159903	8.574014	8.712049	6.829682
##	1053_at	6.758306	6.772804	7.023637	6.751118	6.644379
##	117_at	6.209885	6.373048	6.760195	6.819295	7.218066
##	121_at	6.539673	6.256735	6.798216	7.038408	6.372411
##	1255_g_at	5.912142	5.803277	6.042098	5.993918	5.775654
##	1294_at	6.436974	6.339656	6.671482	6.799305	6.851862
##		GSM352518.CEL	GSM352519.CEL	GSM352520.CEL	GSM352521.CEL	GSM352522.CEL
##	1007_s_at	7.503470	8.497162	9.392036	7.506394	7.757849
##	1053_at	6.857802	6.667792	7.359660	7.752035	7.001174
##	117_at	6.554985	7.195814	6.468873	6.297862	6.554782
##	121_at	6.574942	6.559441	6.804469	6.667615	6.680877
##	1255_g_at	5.992336	5.850618	5.862542	5.780411	5.926216
##	1294_at	6.767416	6.712888	6.218267	6.351230	7.029779
##		GSM352523.CEL	GSM352524.CEL	GSM352525.CEL	GSM352526.CEL	GSM352527.CEL
##	1007_s_at	8.079809	8.655242	7.762511	8.995563	9.047172
##	1053_at	7.216944	8.172837	8.007611	8.127804	6.939659
##	117_at	7.031888	7.072273	6.815111	7.636112	6.642382
##	121_at	7.044903	7.245955	6.926368	7.718840	7.046476
##	1255_g_at	6.141287	6.745550	6.760109	7.244485	6.129765
##	1294_at	7.179129	7.174502	6.953802	7.309114	6.662335
##		GSM352528.CEL	GSM352529.CEL	GSM352530.CEL	GSM352531.CEL	GSM352532.CEL
##	1007_s_at	8.102235	8.533405	8.628142	8.434376	8.299966
##	1053_at	7.155674	6.987416	7.106430	7.385380	8.263295
##	117_at	7.398066	6.646422	6.855366	6.590185	6.694443
##	121_at	6.729210	6.492497	6.728288	6.707729	7.120801
##	1255_g_at	6.052113	5.943828	5.907426	6.038848	6.402550
##	1294_at	7.384888	6.822857	7.021457	6.598470	6.764588
##		GSM352533.CEL	GSM352534.CEL	GSM352535.CEL	GSM352536.CEL	GSM352537.CEL
##	1007_s_at	7.538695	7.853292	9.080991	6.887027	9.140697
##	1053_at	6.789270	6.842478	7.663704	7.061965	6.622780
##	117_at	6.709125	6.529273	6.701540	6.722086	6.683826
##	121_at	6.719729	6.975551	6.766659	6.927070	6.558701
##	1255_g_at	6.029791	6.223586	6.051860	6.136102	5.985064
##	1294_at	7.009445	6.687462	6.690813	6.927198	6.462549
##		GSM352538.CEL	GSM352539.CEL	GSM352540.CEL	GSM352541.CEL	GSM352542.CEL
##	1007_s_at	8.851641	8.011003	8.555544	7.458113	8.430927
##	1053_at	6.671870	7.902435	6.729841	6.863300	7.018831

```
## 117_at      6.617887      6.521827      7.876035      7.944232      6.307716
## 121_at      6.673026      6.951794      6.587205      6.757520      6.879985
## 1255_g_at   5.937564      6.197283      5.888136      5.955142      5.945492
## 1294_at     6.534941      6.590691      6.831102      6.379924      6.504903
##           GSM352543.CEL GSM352544.CEL GSM352545.CEL GSM352546.CEL GSM352547.CEL
## 1007_s_at   7.998070      8.632102      7.962358      7.750605      7.093068
## 1053_at     6.818899      6.833088      6.698797      7.686014      6.554363
## 117_at      6.341375      6.406989      6.152459      6.151424      6.684646
## 121_at      6.451785      6.546678      6.848540      6.560663      6.665838
## 1255_g_at   5.698942      5.658207      5.744861      5.773758      5.695541
## 1294_at     6.129259      6.342191      6.303261      6.247991      6.909530
##           GSM352548.CEL GSM352549.CEL GSM352550.CEL GSM352551.CEL GSM352552.CEL
## 1007_s_at   7.992850      9.039365      9.852244      9.062332      8.069734
## 1053_at     6.293961      6.867591      6.264639      6.873733      6.524271
## 117_at      6.322633      6.360169      6.731429      6.542245      6.121543
## 121_at      6.677064      6.608692      6.443608      6.799230      6.732459
## 1255_g_at   5.610013      5.994178      5.559502      5.796617      5.788245
## 1294_at     6.797498      6.512081      6.477204      6.468143      6.454790
##           GSM352553.CEL
## 1007_s_at   8.153857
## 1053_at     7.368427
## 117_at      6.140521
## 121_at      6.627416
## 1255_g_at   5.936214
## 1294_at     6.376053
```

```
pData(my.rma)
```

```
##           title geo_accession
## GSM352448.CEL      dNF_AS18      GSM352448
## GSM352450.CEL      dNF_AS5       GSM352450
## GSM352451.CEL      MPNST_AS10    GSM352451
## GSM352453.CEL      MPNST_AS13    GSM352453
## GSM352454.CEL      MPNST_AS15    GSM352454
## GSM352455.CEL      pNF_AS16     GSM352455
## GSM352457.CEL      pNF_AS7      GSM352457
## GSM352458.CEL      pNF_AS8      GSM352458
## GSM352460.CEL      Universal_STD_1 GSM352460
## GSM352461.CEL      88_3         GSM352461
## GSM352463.CEL      90_8         GSM352463
## GSM352464.CEL      pNF00.6      GSM352464
## GSM352466.CEL      pNF95.3      GSM352466
## GSM352467.CEL      pNF95.6      GSM352467
## GSM352469.CEL      S462         GSM352469
## GSM352470.CEL      S520         GSM352470
## GSM352471.CEL      ST8814       GSM352471
## GSM352473.CEL      STS26T       GSM352473
## GSM352475.CEL      T265         GSM352475
## GSM352476.CEL      Universal_STD_2 GSM352476
```

## GSM352477.CEL	YST1	GSM352477
## GSM352479.CEL	ABC_8N_-1-	GSM352479
## GSM352480.CEL	AIBC_2N_-1-	GSM352480
## GSM352481.CEL	CLT_6N_+1-	GSM352481
## GSM352483.CEL	MGF_33N_+1-	GSM352483
## GSM352484.CEL	MPNST_02_2	GSM352484
## GSM352486.CEL	MPNST_94_3	GSM352486
## GSM352487.CEL	NHSC_303	GSM352487
## GSM352489.CEL	NHSC_339	GSM352489
## GSM352490.CEL	NHSC_771	GSM352490
## GSM352492.CEL	SCC_7N_-1-	GSM352492
## GSM352493.CEL	Universal_STD_3	GSM352493
## GSM352495.CEL	ADN1N_KO	GSM352495
## GSM352496.CEL	dNF.AS3	GSM352496
## GSM352497.CEL	dNF.AS46	GSM352497
## GSM352498.CEL	MPNST97.6	GSM352498
## GSM352500.CEL	MPNST.AS37	GSM352500
## GSM352501.CEL	NHSC216	GSM352501
## GSM352503.CEL	pNF03.3	GSM352503
## GSM352504.CEL	pNF.AS12	GSM352504
## GSM352505.CEL	pNF.AS38	GSM352505
## GSM352507.CEL	pNF.AS40	GSM352507
## GSM352508.CEL	UniversalSTD_1	GSM352508
## GSM352510.CEL	dNF.AS22	GSM352510
## GSM352511.CEL	dNF.AS4	GSM352511
## GSM352512.CEL	MPNST95.3b	GSM352512
## GSM352514.CEL	MPNST.AS42	GSM352514
## GSM352515.CEL	NHSC323	GSM352515
## GSM352516.CEL	pNF04.4	GSM352516
## GSM352517.CEL	pNF.AS23	GSM352517
## GSM352518.CEL	pNF.AS24	GSM352518
## GSM352519.CEL	pNF.AS47	GSM352519
## GSM352520.CEL	RMN9N_KO	GSM352520
## GSM352521.CEL	UniversalSTD_2	GSM352521
## GSM352522.CEL	dNF.AS35	GSM352522
## GSM352523.CEL	dNF.AS44	GSM352523
## GSM352524.CEL	MPNST96.2	GSM352524
## GSM352525.CEL	MPNST.AS45	GSM352525
## GSM352526.CEL	NHSC338	GSM352526
## GSM352527.CEL	pNF05.3	GSM352527
## GSM352528.CEL	pNF.AS32	GSM352528
## GSM352529.CEL	pNF.AS33	GSM352529
## GSM352530.CEL	pNF.AS34	GSM352530
## GSM352531.CEL	SCC5N_KO	GSM352531
## GSM352532.CEL	UniversalSTD_3	GSM352532
## GSM352533.CEL	dNF_AS55	GSM352533
## GSM352534.CEL	dNFSC_ERS	GSM352534
## GSM352535.CEL	NHSC_286	GSM352535
## GSM352536.CEL	pNF_AS48	GSM352536
## GSM352537.CEL	pNFSC_04.7	GSM352537

##	GSM352538.CEL	pNFSC_05.4	GSM352538
##	GSM352539.CEL	Uni_STD_1	GSM352539
##	GSM352540.CEL	dNF_CSG94	GSM352540
##	GSM352541.CEL	dNF_CSG95	GSM352541
##	GSM352542.CEL	dNFSC_ABB	GSM352542
##	GSM352543.CEL	NHSC_J017	GSM352543
##	GSM352544.CEL	pNFSC_00.13	GSM352544
##	GSM352545.CEL	pNFSC_05.5	GSM352545
##	GSM352546.CEL	Uni_STD_2	GSM352546
##	GSM352547.CEL	dNF_AS50_HG_U13Plus_2.CEL	GSM352547
##	GSM352548.CEL	DNF_LRS_HG_U13Plus_2.CEL	GSM352548
##	GSM352549.CEL	dNFSC_JLM_HG_U13Plus_2.CEL	GSM352549
##	GSM352550.CEL	NHSC_02.8_HG_U13Plus_2.CEL	GSM352550
##	GSM352551.CEL	NHSC_J037_HG_U13Plus_2.CEL	GSM352551
##	GSM352552.CEL	pNFSC_97.9_HG_U13Plus_2.CEL	GSM352552
##	GSM352553.CEL	Uni_STD_3	GSM352553
##			description
##	GSM352448.CEL	Gene expression data from neurofibroma tumor.	
##	GSM352450.CEL	Gene expression data from neurofibroma tumor.	
##	GSM352451.CEL	Gene expression data from neurofibroma tumor.	
##	GSM352453.CEL	Gene expression data from neurofibroma tumor.	
##	GSM352454.CEL	Gene expression data from neurofibroma tumor.	
##	GSM352455.CEL	Gene expression data from neurofibroma tumor.	
##	GSM352457.CEL	Gene expression data from neurofibroma tumor.	
##	GSM352458.CEL	Gene expression data from neurofibroma tumor.	
##	GSM352460.CEL	Gene expression data from whole mouse.	
##	GSM352461.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352463.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352464.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352466.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352467.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352469.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352470.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352471.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352473.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352475.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352476.CEL	Gene expression data from whole mouse.	
##	GSM352477.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352479.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352480.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352481.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352483.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352484.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352486.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352487.CEL	Gene expression data from normal human Schwann cell culture.	
##	GSM352489.CEL	Gene expression data from normal human Schwann cell culture.	
##	GSM352490.CEL	Gene expression data from normal human Schwann cell culture.	
##	GSM352492.CEL	Gene expression data from neurofibroma-derived cell culture.	
##	GSM352493.CEL	Gene expression data from whole mouse.	
##	GSM352495.CEL	Gene expression data from neurofibroma-derived cell culture.	

```
## GSM352496.CEL      Gene expression data from neurofibroma tumor.
## GSM352497.CEL      Gene expression data from neurofibroma tumor.
## GSM352498.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352500.CEL      Gene expression data from neurofibroma tumor.
## GSM352501.CEL      Gene expression data from normal human Schwann cell culture.
## GSM352503.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352504.CEL      Gene expression data from neurofibroma tumor.
## GSM352505.CEL      Gene expression data from neurofibroma tumor.
## GSM352507.CEL      Gene expression data from neurofibroma tumor.
## GSM352508.CEL      Gene expression data from whole mouse.
## GSM352510.CEL      Gene expression data from neurofibroma tumor.
## GSM352511.CEL      Gene expression data from neurofibroma tumor.
## GSM352512.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352514.CEL      Gene expression data from neurofibroma tumor.
## GSM352515.CEL      Gene expression data from normal human Schwann cell culture.
## GSM352516.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352517.CEL      Gene expression data from neurofibroma tumor.
## GSM352518.CEL      Gene expression data from neurofibroma tumor.
## GSM352519.CEL      Gene expression data from neurofibroma tumor.
## GSM352520.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352521.CEL      Gene expression data from whole mouse.
## GSM352522.CEL      Gene expression data from neurofibroma tumor.
## GSM352523.CEL      Gene expression data from neurofibroma tumor.
## GSM352524.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352525.CEL      Gene expression data from neurofibroma tumor.
## GSM352526.CEL      Gene expression data from normal human Schwann cell culture.
## GSM352527.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352528.CEL      Gene expression data from neurofibroma tumor.
## GSM352529.CEL      Gene expression data from neurofibroma tumor.
## GSM352530.CEL      Gene expression data from neurofibroma tumor.
## GSM352531.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352532.CEL      Gene expression data from whole mouse.
## GSM352533.CEL      Gene expression data from neurofibroma tumor.
## GSM352534.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352535.CEL      Gene expression data from normal human Schwann cell culture.
## GSM352536.CEL      Gene expression data from neurofibroma tumor.
## GSM352537.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352538.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352539.CEL      Gene expression data from whole mouse.
## GSM352540.CEL      Gene expression data from neurofibroma tumor.
## GSM352541.CEL      Gene expression data from neurofibroma tumor.
## GSM352542.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352543.CEL      Gene expression data from normal human Schwann cell culture.
## GSM352544.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352545.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352546.CEL      Gene expression data from whole mouse.
## GSM352547.CEL      Gene expression data from neurofibroma tumor.
## GSM352548.CEL      Gene expression data from neurofibroma tumor.
## GSM352549.CEL      Gene expression data from neurofibroma-derived cell culture.
## GSM352550.CEL      Gene expression data from normal human Schwann cell culture.
```

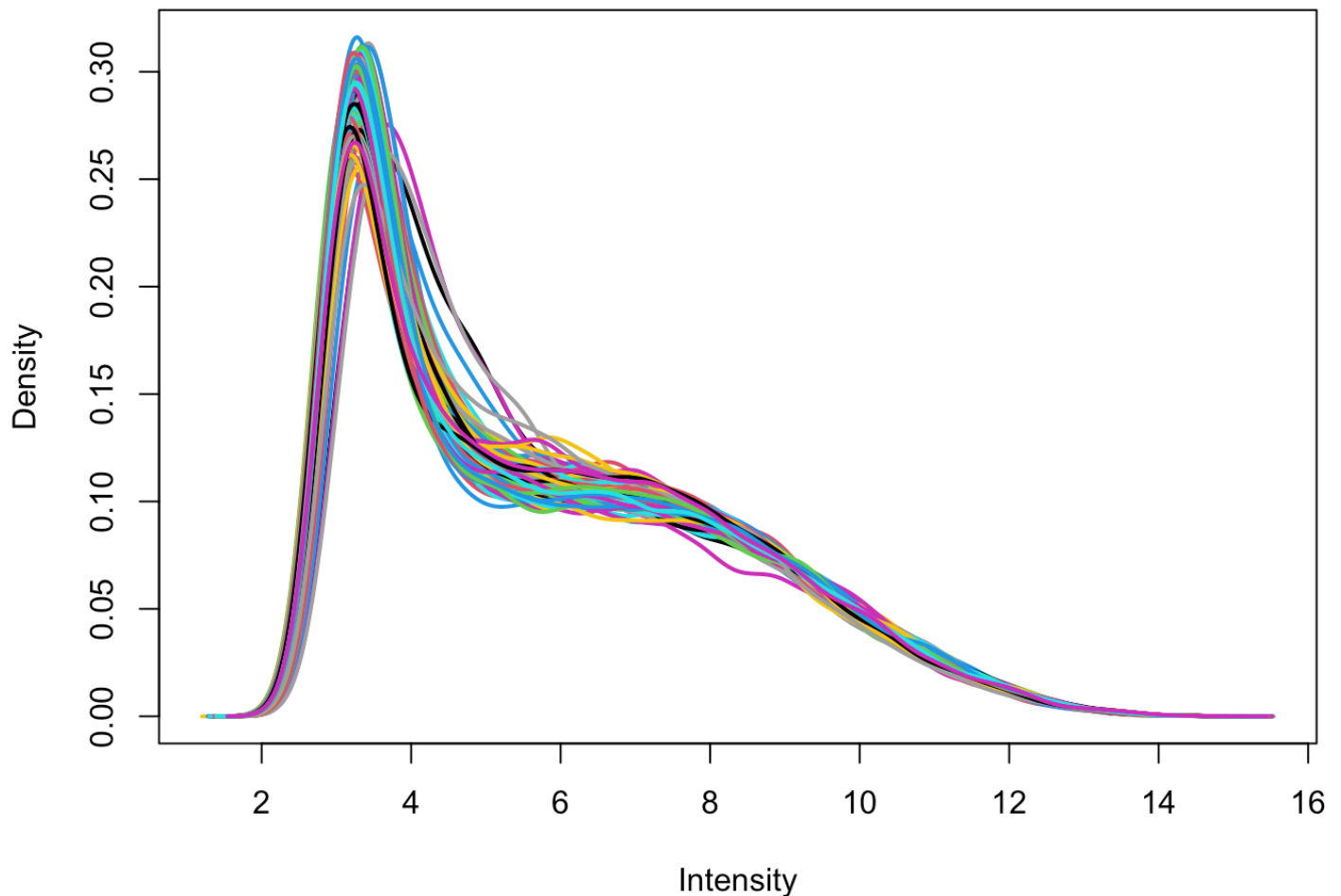
```
## GSM352551.CEL Gene expression data from normal human Schwann cell culture.
## GSM352552.CEL Gene expression data from neurofibroma-derived cell culture.
## GSM352553.CEL Gene expression data from whole mouse.
##
## group
## GSM352448.CEL neurofibroma tumor.
## GSM352450.CEL neurofibroma tumor.
## GSM352451.CEL neurofibroma tumor.
## GSM352453.CEL neurofibroma tumor.
## GSM352454.CEL neurofibroma tumor.
## GSM352455.CEL neurofibroma tumor.
## GSM352457.CEL neurofibroma tumor.
## GSM352458.CEL neurofibroma tumor.
## GSM352460.CEL whole mouse.
## GSM352461.CEL neurofibroma-derived cell culture.
## GSM352463.CEL neurofibroma-derived cell culture.
## GSM352464.CEL neurofibroma-derived cell culture.
## GSM352466.CEL neurofibroma-derived cell culture.
## GSM352467.CEL neurofibroma-derived cell culture.
## GSM352469.CEL neurofibroma-derived cell culture.
## GSM352470.CEL neurofibroma-derived cell culture.
## GSM352471.CEL neurofibroma-derived cell culture.
## GSM352473.CEL neurofibroma-derived cell culture.
## GSM352475.CEL neurofibroma-derived cell culture.
## GSM352476.CEL whole mouse.
## GSM352477.CEL neurofibroma-derived cell culture.
## GSM352479.CEL neurofibroma-derived cell culture.
## GSM352480.CEL neurofibroma-derived cell culture.
## GSM352481.CEL neurofibroma-derived cell culture.
## GSM352483.CEL neurofibroma-derived cell culture.
## GSM352484.CEL neurofibroma-derived cell culture.
## GSM352486.CEL neurofibroma-derived cell culture.
## GSM352487.CEL normal human Schwann cell culture.
## GSM352489.CEL normal human Schwann cell culture.
## GSM352490.CEL normal human Schwann cell culture.
## GSM352492.CEL neurofibroma-derived cell culture.
## GSM352493.CEL whole mouse.
## GSM352495.CEL neurofibroma-derived cell culture.
## GSM352496.CEL neurofibroma tumor.
## GSM352497.CEL neurofibroma tumor.
## GSM352498.CEL neurofibroma-derived cell culture.
## GSM352500.CEL neurofibroma tumor.
## GSM352501.CEL normal human Schwann cell culture.
## GSM352503.CEL neurofibroma-derived cell culture.
## GSM352504.CEL neurofibroma tumor.
## GSM352505.CEL neurofibroma tumor.
## GSM352507.CEL neurofibroma tumor.
## GSM352508.CEL whole mouse.
## GSM352510.CEL neurofibroma tumor.
## GSM352511.CEL neurofibroma tumor.
## GSM352512.CEL neurofibroma-derived cell culture.
```



```
## GSM352514.CEL          neurofibroma tumor.
## GSM352515.CEL normal human Schwann cell culture.
## GSM352516.CEL neurofibroma-derived cell culture.
## GSM352517.CEL          neurofibroma tumor.
## GSM352518.CEL          neurofibroma tumor.
## GSM352519.CEL          neurofibroma tumor.
## GSM352520.CEL neurofibroma-derived cell culture.
## GSM352521.CEL          whole mouse.
## GSM352522.CEL          neurofibroma tumor.
## GSM352523.CEL          neurofibroma tumor.
## GSM352524.CEL neurofibroma-derived cell culture.
## GSM352525.CEL          neurofibroma tumor.
## GSM352526.CEL normal human Schwann cell culture.
## GSM352527.CEL neurofibroma-derived cell culture.
## GSM352528.CEL          neurofibroma tumor.
## GSM352529.CEL          neurofibroma tumor.
## GSM352530.CEL          neurofibroma tumor.
## GSM352531.CEL neurofibroma-derived cell culture.
## GSM352532.CEL          whole mouse.
## GSM352533.CEL          neurofibroma tumor.
## GSM352534.CEL neurofibroma-derived cell culture.
## GSM352535.CEL normal human Schwann cell culture.
## GSM352536.CEL          neurofibroma tumor.
## GSM352537.CEL neurofibroma-derived cell culture.
## GSM352538.CEL neurofibroma-derived cell culture.
## GSM352539.CEL          whole mouse.
## GSM352540.CEL          neurofibroma tumor.
## GSM352541.CEL          neurofibroma tumor.
## GSM352542.CEL neurofibroma-derived cell culture.
## GSM352543.CEL normal human Schwann cell culture.
## GSM352544.CEL neurofibroma-derived cell culture.
## GSM352545.CEL neurofibroma-derived cell culture.
## GSM352546.CEL          whole mouse.
## GSM352547.CEL          neurofibroma tumor.
## GSM352548.CEL          neurofibroma tumor.
## GSM352549.CEL neurofibroma-derived cell culture.
## GSM352550.CEL normal human Schwann cell culture.
## GSM352551.CEL normal human Schwann cell culture.
## GSM352552.CEL neurofibroma-derived cell culture.
## GSM352553.CEL          whole mouse.
```

```
# expression value distribution plot
par(mar=c(4,4,2,1))
title <- paste ("GSE14038", "/", annotation(gset), " value distribution", sep = "")
plotDensities(ex, main=title, legend=F)
```

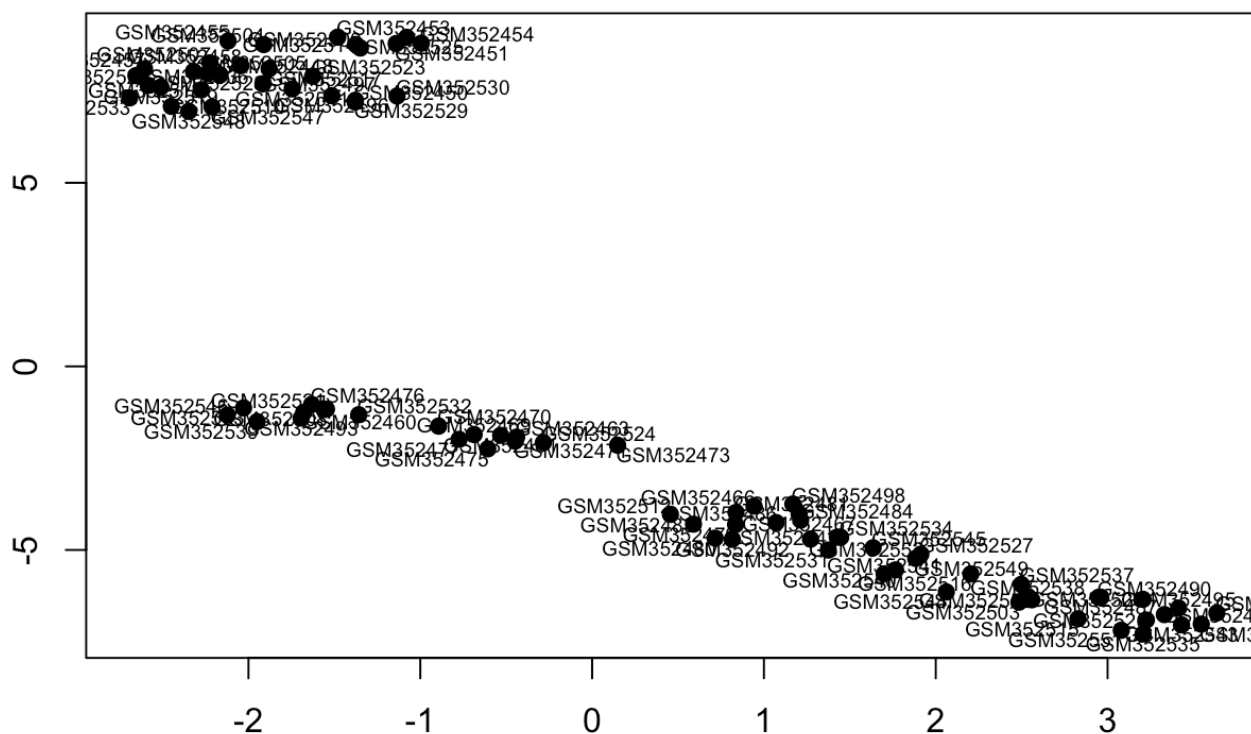
GSE14038/GPL7869 value distribution



STEP 3 - EXPLORATORY ANALYSIS

```
# UMAP plot (multi-dimensional scaling)
ex <- ex[!duplicated(ex), ] # remove duplicates
ump <- umap(t(ex), n_neighbors = 15, random_state = 123)
plot(ump$layout, main="UMAP plot, nbrs=15", xlab="", ylab="", pch=20, cex=1.5)
library("maptools") # point labels without overlaps
pointLabel(ump$layout, labels = rownames(ump$layout), method="SANN", cex=0.6)
```

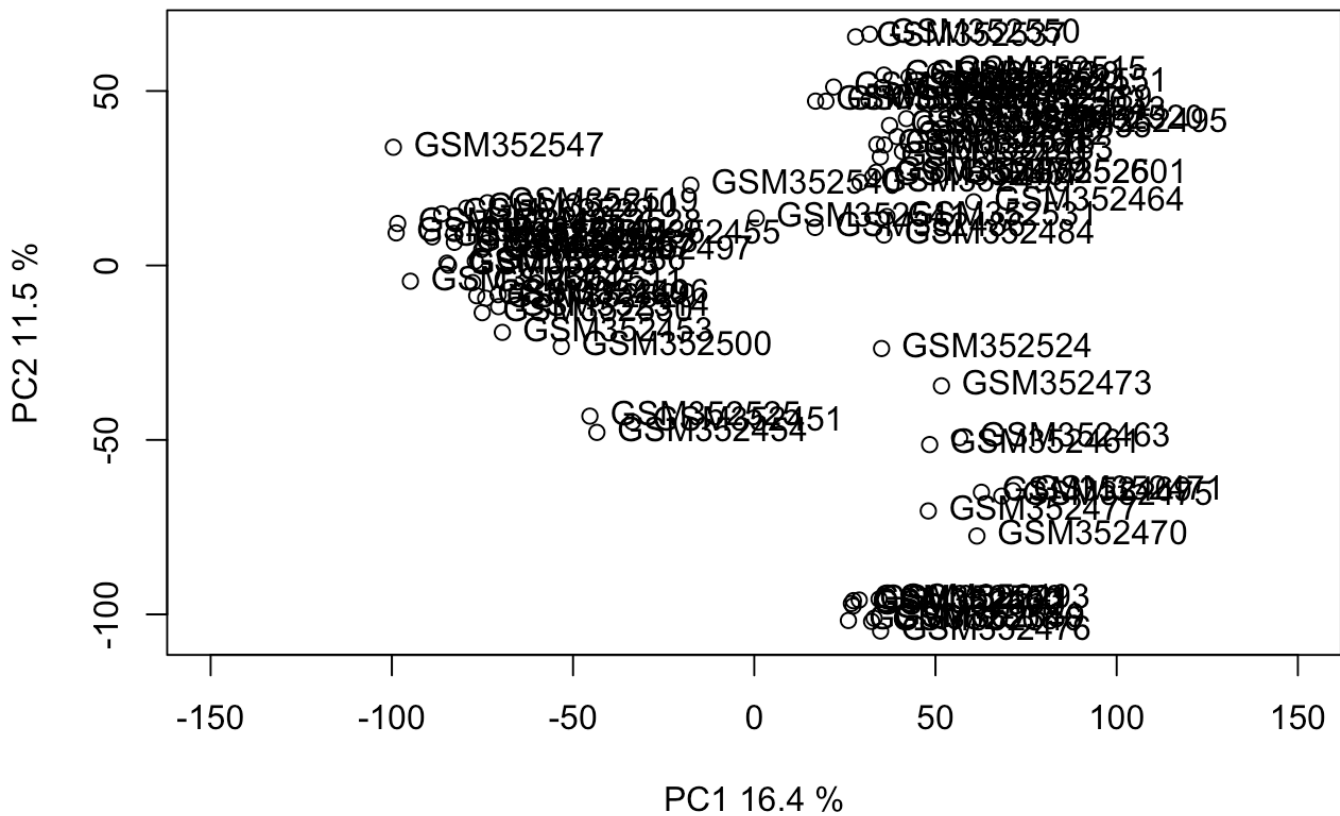
UMAP plot, nbrs=15



```
#computing principal components and loadings.
pcX<-prcomp(t(ex), scale=TRUE)
loads<- round(pcX$sdev^2/sum(pcX$sdev^2)*100,1)
```

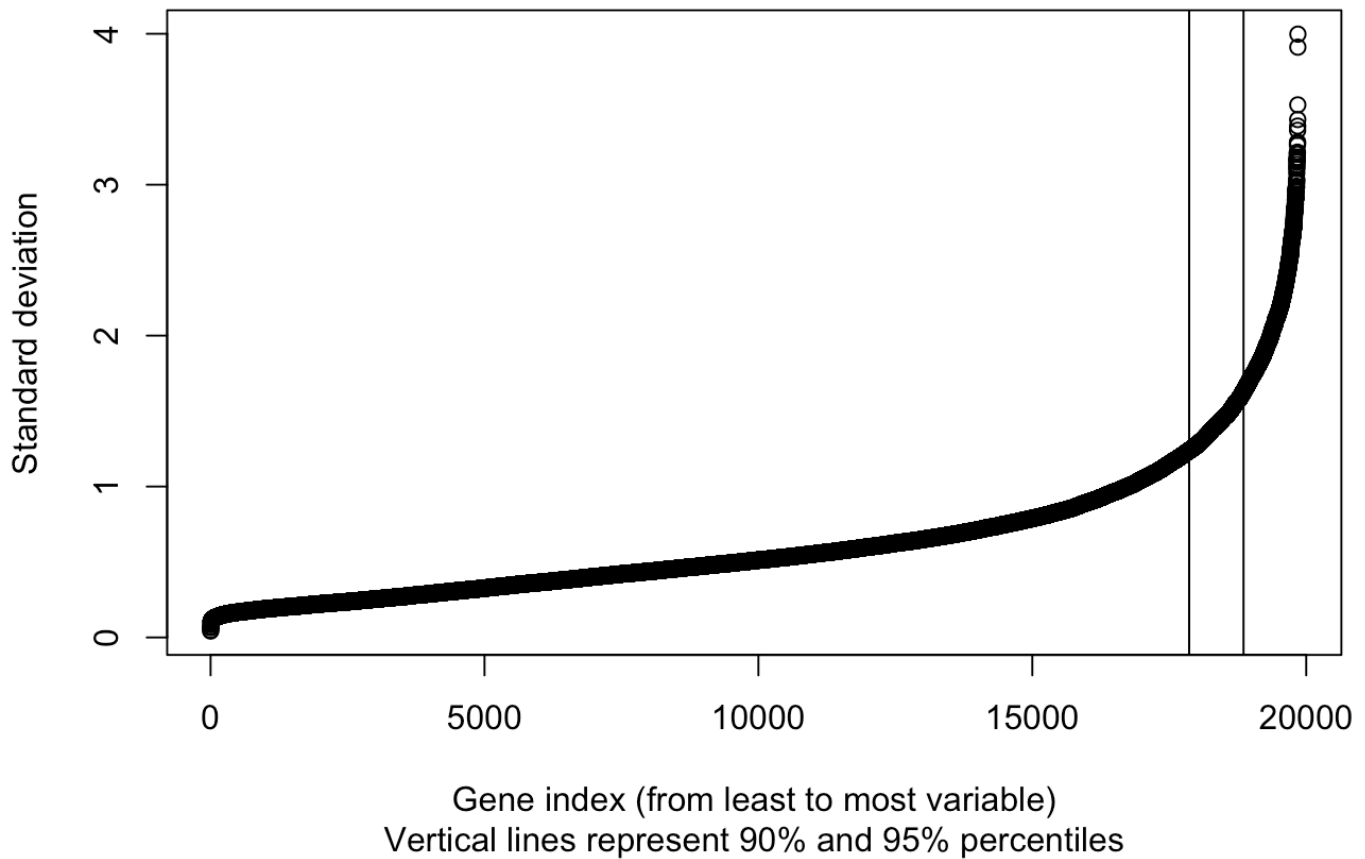
```
xlab<-c(paste("PC1",loads[1],"%"))
ylab<-c(paste("PC2",loads[2],"%"))
plot(pcX$x[,1:2],xlab=xlab,ylab=ylab, xlim=c(-150, 150))
title("Principal components (PCA)")
text(pcX$x[,1],pcX$x[,2], colnames(ex), pos=4)
```

Principal components (PCA)



```
sds <- apply(ex, 1, sd)
sds0<- sort(sds)
plot(1:length(sds0), sds0, main="Distribution of variability for all genes",
     sub="Vertical lines represent 90% and 95% percentiles",
     xlab="Gene index (from least to most variable)", ylab="Standard deviation")
abline(v=length(sds)*c(0.9,0.95))
```

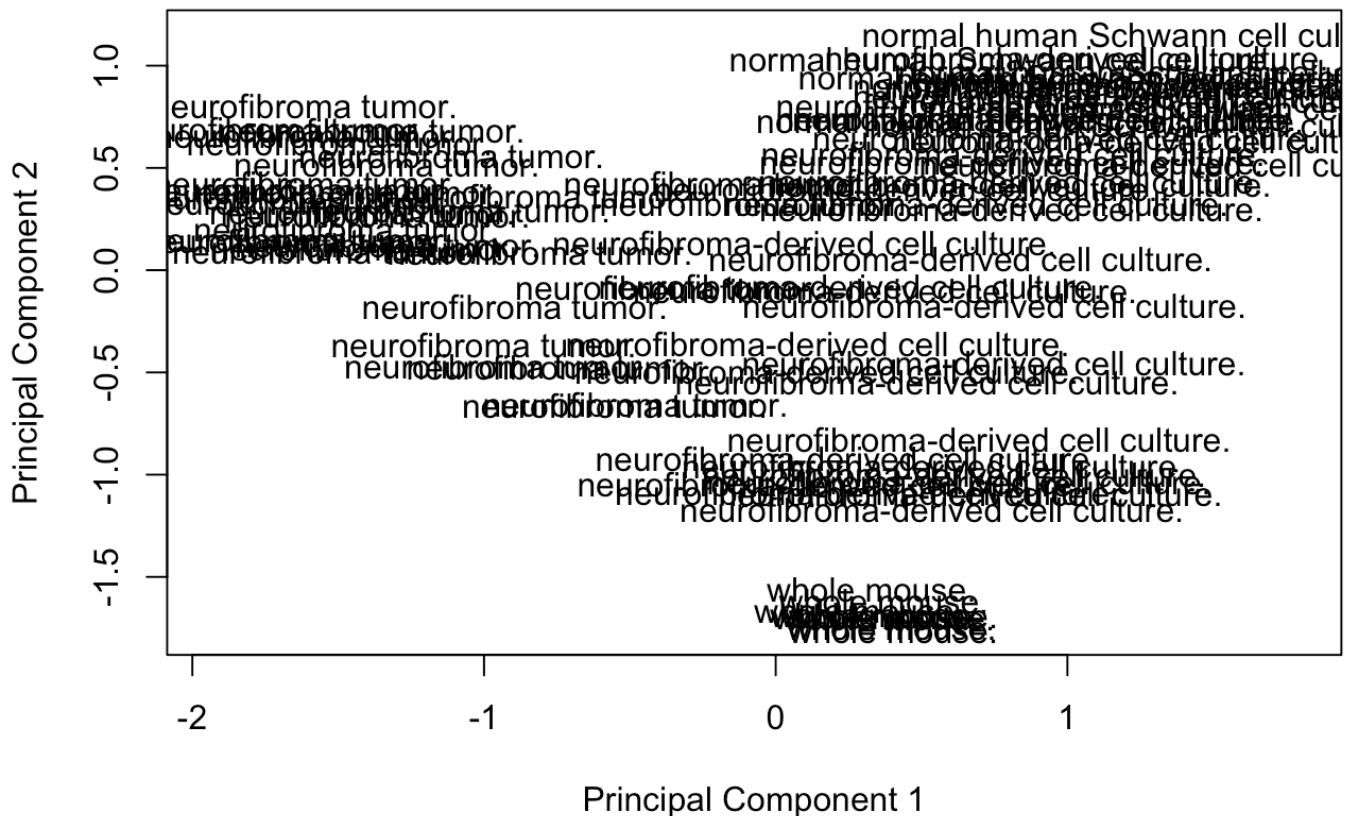
Distribution of variability for all genes



STEP 4 - Differential gene Expression

```
plotMDS(exprs(my.rma), labels=pData(my.rma)$group, top=500, gene.selection="common", main="MDS Plot to Compare Replicates")
```

MDS Plot to Compare Replicates



Alternatively a hierarchical clustering can be applied to detect any expected (or unexpected) grouping of the samples).

```
cluster.dat <- exprs(my.rma)
gene.mean <- apply(cluster.dat, 1, mean)
gene.sd <- apply(cluster.dat, 1, sd)
cluster.dat <- sweep(cluster.dat, 1, gene.mean, "-")
cluster.dat <- sweep(cluster.dat, 1, gene.sd, "/")
my.dist <- dist(t(cluster.dat), method="euclidean")
my.hclust <- hclust(my.dist, method="average")
my.hclust$labels <- pData(my.rma)$group
plot(my.hclust, cex=0.75, main="Comparison of Biological Replicates", xlab="Euclidean Distance")
```

[illegible]

Euclidean Distance
hclust (*, "average")

Design of the experiment

```
##determine the average effect (coefficient) for each condition
library(limma)
designMat<- model.matrix(~0+group, pData(my.rma))
colnames(designMat) <- c("neuro_whole", "neuro_cell", "schwann_cell", "mouse_whole")
print(designMat)
```

##	neuro_whole	neuro_cell	schwann_cell	mouse_whole
## GSM352448.CEL	1	0	0	0
## GSM352450.CEL	1	0	0	0
## GSM352451.CEL	1	0	0	0
## GSM352453.CEL	1	0	0	0
## GSM352454.CEL	1	0	0	0
## GSM352455.CEL	1	0	0	0
## GSM352457.CEL	1	0	0	0
## GSM352458.CEL	1	0	0	0
## GSM352460.CEL	0	0	0	1
## GSM352461.CEL	0	1	0	0
## GSM352463.CEL	0	1	0	0

## GSM352464.CEL	0	1	0	0
## GSM352466.CEL	0	1	0	0
## GSM352467.CEL	0	1	0	0
## GSM352469.CEL	0	1	0	0
## GSM352470.CEL	0	1	0	0
## GSM352471.CEL	0	1	0	0
## GSM352473.CEL	0	1	0	0
## GSM352475.CEL	0	1	0	0
## GSM352476.CEL	0	0	0	1
## GSM352477.CEL	0	1	0	0
## GSM352479.CEL	0	1	0	0
## GSM352480.CEL	0	1	0	0
## GSM352481.CEL	0	1	0	0
## GSM352483.CEL	0	1	0	0
## GSM352484.CEL	0	1	0	0
## GSM352486.CEL	0	1	0	0
## GSM352487.CEL	0	0	1	0
## GSM352489.CEL	0	0	1	0
## GSM352490.CEL	0	0	1	0
## GSM352492.CEL	0	1	0	0
## GSM352493.CEL	0	0	0	1
## GSM352495.CEL	0	1	0	0
## GSM352496.CEL	1	0	0	0
## GSM352497.CEL	1	0	0	0
## GSM352498.CEL	0	1	0	0
## GSM352500.CEL	1	0	0	0
## GSM352501.CEL	0	0	1	0
## GSM352503.CEL	0	1	0	0
## GSM352504.CEL	1	0	0	0
## GSM352505.CEL	1	0	0	0
## GSM352507.CEL	1	0	0	0
## GSM352508.CEL	0	0	0	1
## GSM352510.CEL	1	0	0	0
## GSM352511.CEL	1	0	0	0
## GSM352512.CEL	0	1	0	0
## GSM352514.CEL	1	0	0	0
## GSM352515.CEL	0	0	1	0
## GSM352516.CEL	0	1	0	0
## GSM352517.CEL	1	0	0	0
## GSM352518.CEL	1	0	0	0
## GSM352519.CEL	1	0	0	0
## GSM352520.CEL	0	1	0	0
## GSM352521.CEL	0	0	0	1
## GSM352522.CEL	1	0	0	0
## GSM352523.CEL	1	0	0	0
## GSM352524.CEL	0	1	0	0
## GSM352525.CEL	1	0	0	0
## GSM352526.CEL	0	0	1	0
## GSM352527.CEL	0	1	0	0
## GSM352528.CEL	1	0	0	0


```
## GSM352529.CEL      1      0      0      0
## GSM352530.CEL      1      0      0      0
## GSM352531.CEL      0      1      0      0
## GSM352532.CEL      0      0      0      1
## GSM352533.CEL      1      0      0      0
## GSM352534.CEL      0      1      0      0
## GSM352535.CEL      0      0      1      0
## GSM352536.CEL      1      0      0      0
## GSM352537.CEL      0      1      0      0
## GSM352538.CEL      0      1      0      0
## GSM352539.CEL      0      0      0      1
## GSM352540.CEL      1      0      0      0
## GSM352541.CEL      1      0      0      0
## GSM352542.CEL      0      1      0      0
## GSM352543.CEL      0      0      1      0
## GSM352544.CEL      0      1      0      0
## GSM352545.CEL      0      1      0      0
## GSM352546.CEL      0      0      0      1
## GSM352547.CEL      1      0      0      0
## GSM352548.CEL      1      0      0      0
## GSM352549.CEL      0      1      0      0
## GSM352550.CEL      0      0      1      0
## GSM352551.CEL      0      0      1      0
## GSM352552.CEL      0      1      0      0
## GSM352553.CEL      0      0      0      1
## attr(,"assign")
## [1] 1 1 1 1
## attr(,"contrasts")
## attr(,"contrasts")$group
## [1] "contr.treatment"
```

Fitting coefficients

“neuro_whole”, “neuro_cell”, “schwann_cell”, “mouse_whole”

```
##specify the contrast of interest using the levels from the design matrix
cont.matrix <- makeContrasts (neuroVsschwann_cell = neuro_cell-schwann_cell,
                             neuroVsmouse_whole = neuro_whole-mouse_whole,
                             neurowVsschwann = neuro_whole-schwann_cell,
                             INT = (neuro_cell-schwann_cell) - (neuro_whole-schwann_cell),
                             levels=designMat)
print(cont.matrix)
```

```
##           Contrasts
## Levels      neuroVSSchawnn_cell neuroVSmouse_whole neurowVSSchwann INT
##   neuro_whole                0                1                1 -1
##   neuro_cell                 1                0                0  1
##   schwann_cell               -1                0               -1  0
##   mouse_whole                0               -1                0  0
```

Now that we have a design matrix, we need to estimate the coefficients. For this design, we will essentially average the replicate arrays for each sample level. In addition, we will calculate standard deviations for each gene, and the average intensity for the genes across all microarrays.

```
##determine the average effect (coefficient) for each treatment
my.fit <- lmFit(my.rma, designMat)
my.fit
```

```
## An object of class "MArrayLM"
## $coefficients
##           neuro_whole neuro_cell schwann_cell mouse_whole
## 1007_s_at      7.525282   8.210682    8.742134    7.863969
## 1053_at       6.869100   7.340025    7.057762    7.763305
## 117_at        6.790593   6.566683    6.716640    6.399018
## 121_at        6.594987   6.839949    6.804503    6.792337
## 1255_g_at     5.951118   6.179855    6.052821    6.050227
## 54670 more rows ...
##
## $rank
## [1] 4
##
## $assign
## [1] 1 1 1 1
##
## $qr
## $qr
##           neuro_whole neuro_cell schwann_cell mouse_whole
## GSM352448.CEL -5.6568542   0.00000    0.000000        0
## GSM352450.CEL  0.1767767  -5.91608    0.000000        0
## GSM352451.CEL  0.1767767   0.00000   -3.162278        0
## GSM352453.CEL  0.1767767   0.00000    0.000000       -3
## GSM352454.CEL  0.1767767   0.00000    0.000000        0
## 81 more rows ...
##
## $qraux
## [1] 1.176777 1.000000 1.000000 1.000000
##
## $pivot
## [1] 1 2 3 4
##
## $tol
## [1] 1e-07
```

```
##
## $rank
## [1] 4
##
##
## $df.residual
## [1] 82 82 82 82 82
## 54670 more elements ...
##
## $sigma
## 1007_s_at 1053_at 117_at 121_at 1255_g_at
## 0.5889262 0.5336611 0.3738674 0.2793913 0.3581521
## 54670 more elements ...
##
## $cov.coefficients
##          neuro_whole neuro_cell schwann_cell mouse_whole
## neuro_whole      0.03125 0.000000000      0.0  0.0000000
## neuro_cell       0.00000 0.02857143      0.0  0.0000000
## schwann_cell     0.00000 0.00000000      0.1  0.0000000
## mouse_whole      0.00000 0.00000000      0.0  0.1111111
##
## $stdev.unscaled
##          neuro_whole neuro_cell schwann_cell mouse_whole
## 1007_s_at  0.1767767 0.1690309  0.3162278  0.3333333
## 1053_at    0.1767767 0.1690309  0.3162278  0.3333333
## 117_at     0.1767767 0.1690309  0.3162278  0.3333333
## 121_at     0.1767767 0.1690309  0.3162278  0.3333333
## 1255_g_at  0.1767767 0.1690309  0.3162278  0.3333333
## 54670 more rows ...
##
## $pivot
## [1] 1 2 3 4
##
## $Amean
## 1007_s_at 1053_at 117_at 121_at 1255_g_at
## 7.981162 7.176273 6.649889 6.739696 6.066406
## 54670 more elements ...
##
## $method
## [1] "ls"
##
## $design
##          neuro_whole neuro_cell schwann_cell mouse_whole
## GSM352448.CEL      1      0      0      0
## GSM352450.CEL      1      0      0      0
## GSM352451.CEL      1      0      0      0
## GSM352453.CEL      1      0      0      0
## GSM352454.CEL      1      0      0      0
## 81 more rows ...
```

Obtaining lists of differentially expressed genes and performing statistical analysis

The eBayes function performs the tests, and there are several parameters (arguments) that can be changed. We are going to change only the proportion of genes that we expect to be differentially expressed.

```
#linear model fit
fit.main<-contrasts.fit(my.fit, cont.matrix)
names(fit.main)
```

```
## [1] "coefficients"      "rank"              "assign"            "qr"
## [5] "df.residual"       "sigma"             "cov.coefficients"  "stdev.unscaled"
## [9] "pivot"            "Amean"            "method"            "design"
## [13] "contrasts"
```

```
fit.Bayes<-eBayes(fit.main, proportion=0.1, trend=FALSE, robust=FALSE)
```

The limma package implements function `topTable` which contains, for a given contrast a list of genes ordered from smallest to biggest p-value which can be considered to be most to least differential expressed. For each gene the following statistics are provided:

- `logFC` : Mean difference between groups.
- `AveExpr` : Average expression of all genes in the comparison.
- `t` : Moderated t-statistic (t-test-like statistic for the comparison).
- `P.Value` : Test p-value.
- `adj.P.Val` : Adjusted p-value
- `B` : B-statistic: Posterior log odds of the gene of being vs non being differential expressed. `topTable` will adjust the p-values and return the top genes that meet the cutoffs.

```
#first comparison
topTab_neuroVSSchawnn_cell <- topTable (fit.Bayes, number=nrow(fit.Bayes), coef="n
euroVSSchawnn_cell", adjust="fdr")
head(topTab_neuroVSSchawnn_cell)
```

##		logFC	AveExpr	t	P.Value	adj.P.Val	B
##	224323_s_at	-2.098916	6.616958	-12.388855	4.775796e-21	2.611166e-16	35.87612
##	231358_at	-3.605856	7.884132	-11.601476	1.792024e-19	4.898946e-15	32.89800
##	201438_at	2.953648	11.153228	10.623735	1.739439e-17	3.170127e-13	29.08226
##	205722_s_at	-1.664342	6.703654	-10.466409	3.654413e-17	4.995126e-13	28.45753
##	205721_at	-2.401388	7.116952	-9.846701	6.885888e-16	7.529719e-12	25.97230
##	231227_at	3.362307	8.920307	9.651881	1.738179e-15	1.583915e-11	25.18396

```
#second comparison
topTab_neurowVSSchwann <- topTable (fit.Bayes, number=nrow(fit.Bayes), coef="neuro
wVSSchwann", adjust="fdr")
head(topTab_neurowVSSchwann)
```

```
##           logFC AveExpr      t      P.Value    adj.P.Val      B
## 200795_at      5.062173 8.237421 20.74241 8.854338e-36 4.841109e-31 70.22812
## 201525_at      6.142178 9.055918 17.26950 4.121088e-30 1.126603e-25 58.52026
## 202878_s_at     4.215110 7.960574 16.25408 2.479726e-28 4.519300e-24 54.77390
## 1554474_a_at   -4.279320 7.820559 -16.00380 6.945796e-28 9.494035e-24 53.82743
## 204232_at      3.356503 7.904532 14.73757 1.438117e-25 1.572580e-21 48.89907
## 208982_at      4.331440 8.452634 14.44233 5.133396e-25 4.677807e-21 47.71667
```

```
#third comparison
topTab_neuroVSmouse_whole <- topTable (fit.Bayes, number=nrow(fit.Bayes), coef="ne
uroVSmouse_whole", adjust="fdr")
head(topTab_neuroVSmouse_whole)
```

```
##           logFC AveExpr      t      P.Value    adj.P.Val      B
## 219465_at     -6.311018 6.734726 -51.31397 1.008734e-67 5.515255e-63 125.6058
## 219466_s_at   -6.299173 6.774839 -48.73690 8.333101e-66 2.278062e-61 122.9922
## 204694_at     -5.671861 6.817432 -48.24587 1.980376e-65 3.609236e-61 122.4671
## 211298_s_at   -6.424858 6.778037 -45.18520 5.285019e-63 6.809869e-59 118.9806
## 204419_x_at   -6.313646 6.815767 -45.09805 6.227590e-63 6.809869e-59 118.8756
## 210929_s_at   -6.103459 6.744407 -43.90618 6.053741e-62 5.516471e-58 117.4067
```

```
#third comparison: Genes that behave differently between comparison 1 and 2
topTab_INT <- topTable(fit.Bayes, number=nrow(fit.Bayes), coef="INT", adjust="fdr"
)
head(topTab_INT)
```

```
##           logFC AveExpr      t      P.Value    adj.P.Val      B
## 200795_at     -4.950210 8.237421 -30.04465 2.906456e-48 1.589105e-43 98.44634
## 201525_at     -5.304254 9.055918 -22.09037 8.156669e-38 2.229829e-33 76.55718
## 202878_s_at   -3.821647 7.960574 -21.82857 1.995667e-37 3.637103e-33 75.72270
## 202613_at      3.026465 9.606935 20.67075 1.142442e-35 1.561576e-31 71.93294
## 200756_x_at    2.403127 9.373701 19.92969 1.649892e-34 1.804157e-30 69.42005
## 214845_s_at    2.493555 9.575478 19.59854 5.554312e-34 5.061366e-30 68.27451
```

Finally, `decideTests` will make calls for DEGs by adjusting the p-values and applying a logFC cutoff similar to `topTable`.

```
contrast.tests <- decideTests(fit.Bayes, method="separate", adjust.method="BH", p.
value=0.05, lfc=0)
head(contrast.tests)
```

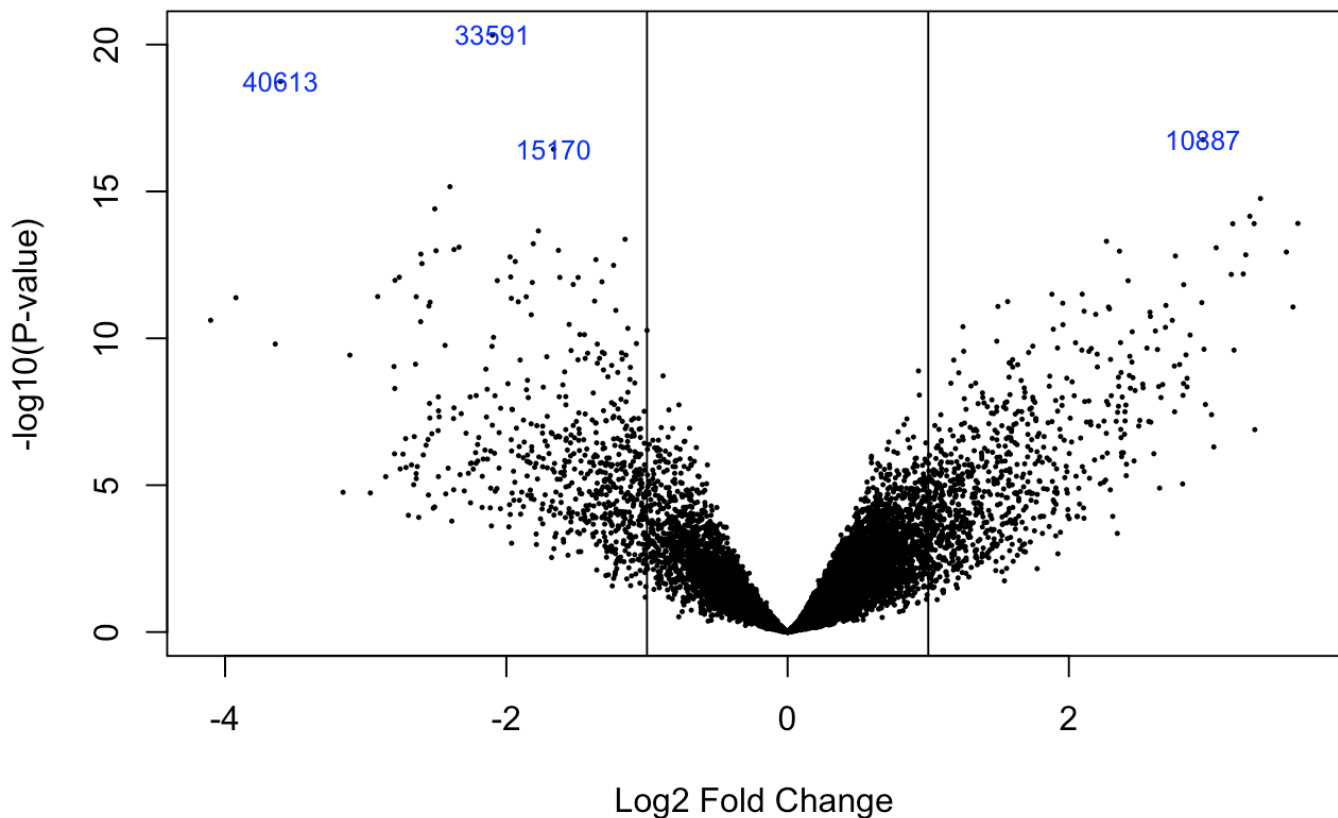
```
## TestResults matrix
##           Contrasts
##           neuroVSschawnn_cell neuroVSmouse_whole neurowVSschwann INT
## 1007_s_at           0           0           -1    1
## 1053_at             0          -1            0    1
## 117_at              0            1            0   -1
## 121_at              0            0            0    1
## 1255_g_at           0            0            0    1
## 1294_at             0            0            0    0
```

Assessing the Results

There are several ways to visualize the statistical results from the DGE analysis. Limma has a `volcanoplot` function.

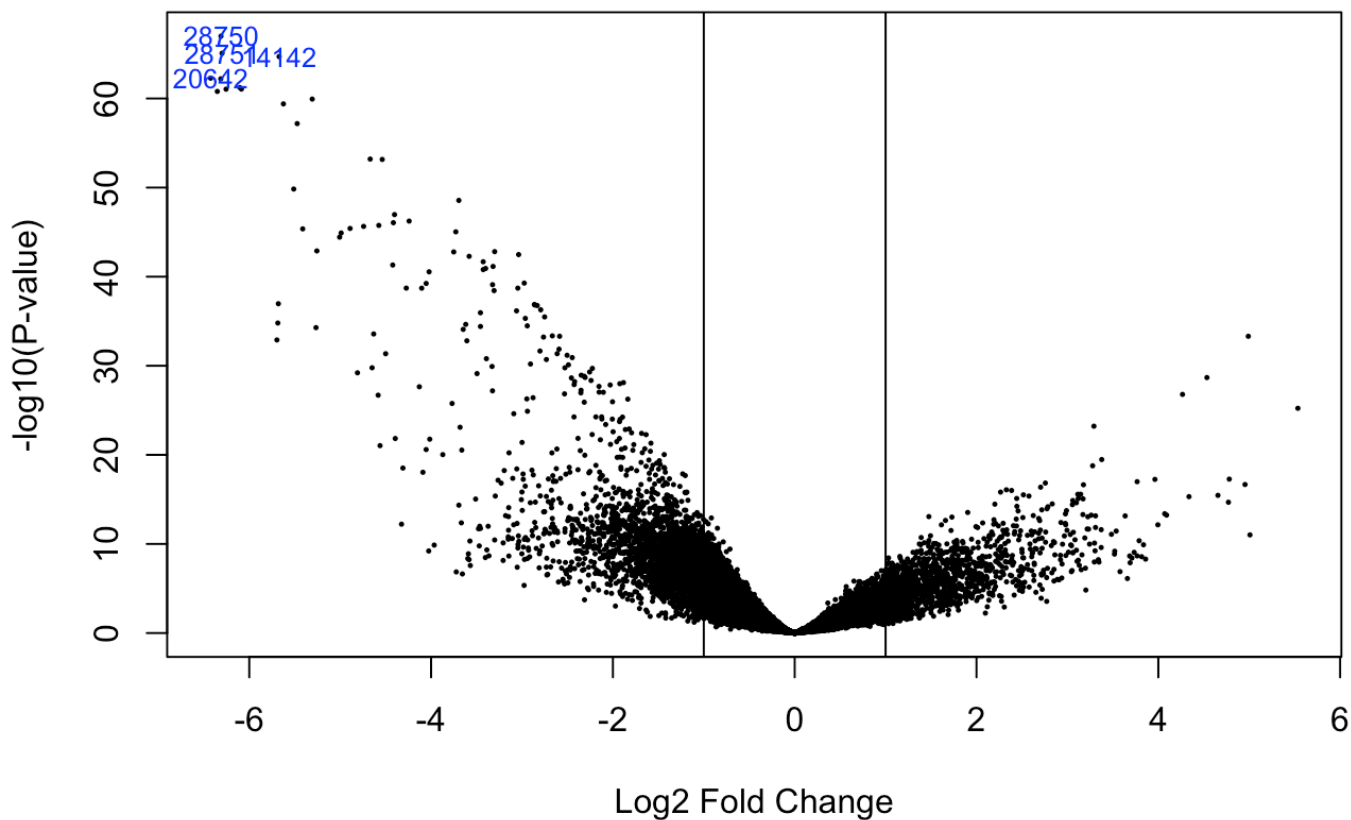
```
volcanoplot(fit.Bayes, coef=1, highlight=4,
            main=paste("Differentially expressed genes", colnames(cont.matrix)[1],
            sep="\n"))
            abline(v=c(-1,1))
```

Differentially expressed genes neuroVSschawnn_cell



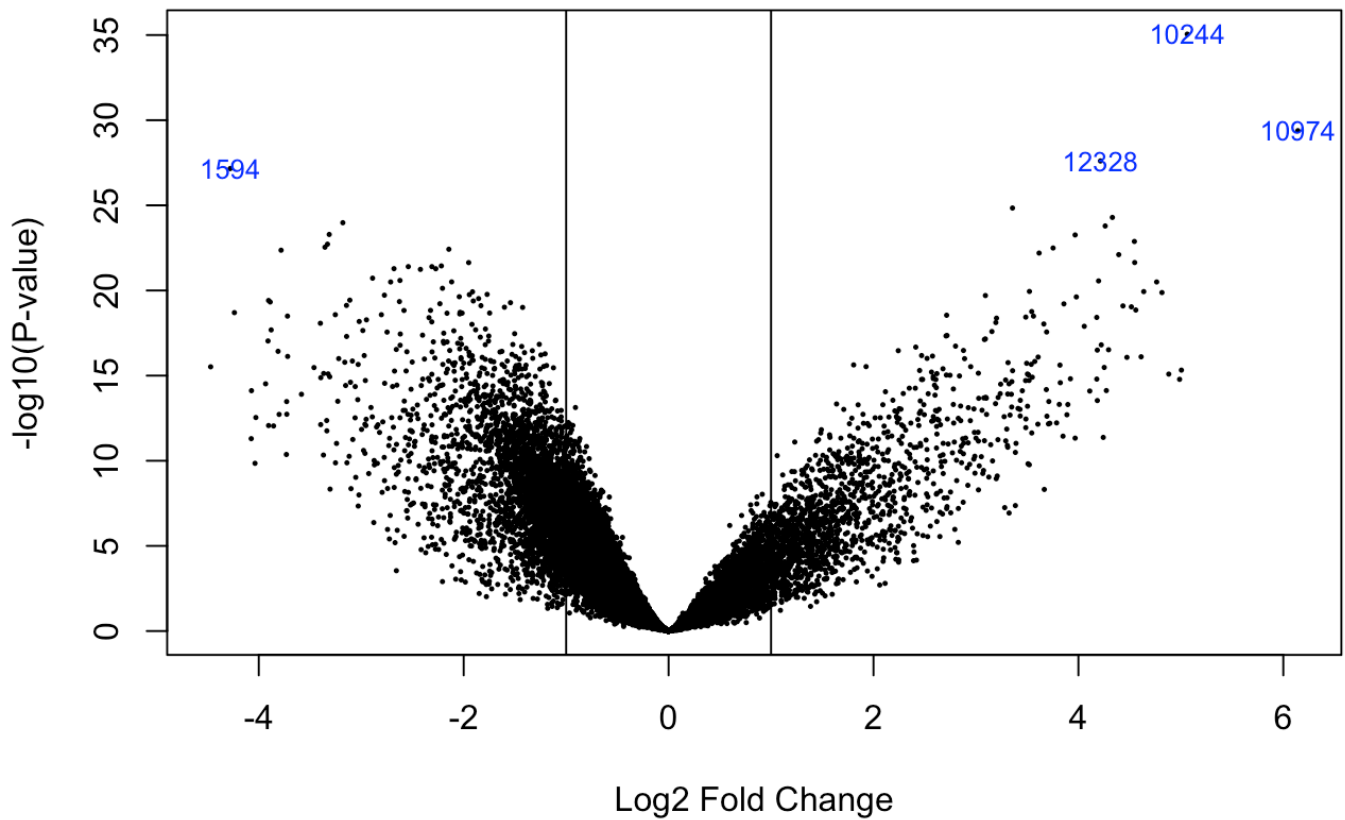
```
volcanoplot(fit.Bayes, coef=2, highlight=4,
            main=paste("Differentially expressed genes", colnames(cont.matrix)[2],
            sep="\n"))
            abline(v=c(-1,1))
```

Differentially expressed genes neuroVSmouse_whole



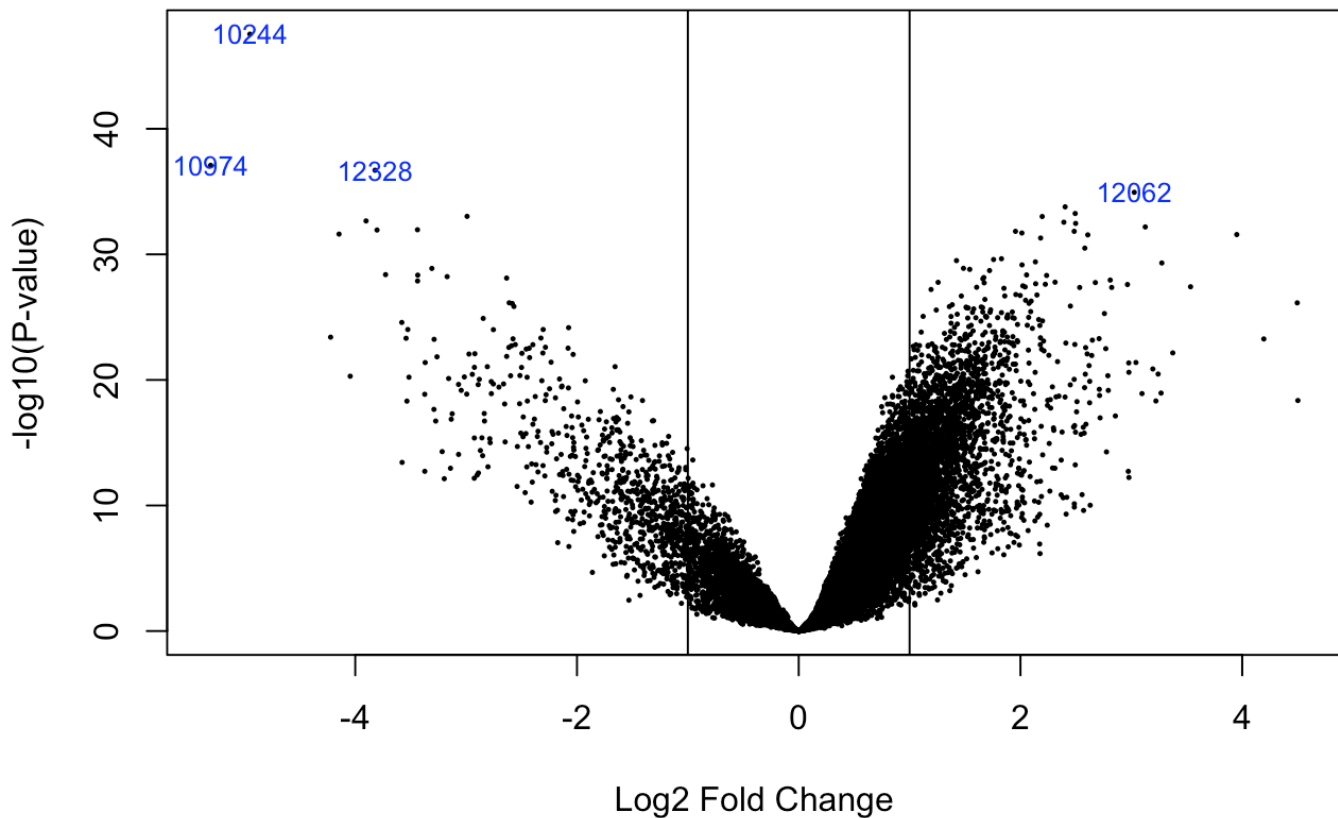
```
volcanoplot(fit.Bayes, coef=3, highlight=4,
            main=paste("Differentially expressed genes", colnames(cont.matrix)[3],
            sep="\n"))
            abline(v=c(-1,1))
```

Differentially expressed genes neurowVSschwann



```
volcanoplot(fit.Bayes, coef=4, highlight=4,  
            main=paste("Differentially expressed genes", colnames(cont.matrix)[4],  
            sep="\n"))  
abline(v=c(-1,1))
```


Differentially expressed genes INT



Biological significance

```
library(hgu133plus2.db)
columns(hgu133plus2.db)
```

```
## [1] "ACCNUM"      "ALIAS"       "ENSEMBL"     "ENSEMBLPROT" "ENSEMBLTRANS"
## [6] "ENTREZID"    "ENZYME"      "EVIDENCE"     "EVIDENCEALL"  "GENENAME"
## [11] "GO"          "GOALL"       "IPI"          "MAP"           "OMIM"
## [16] "ONTOLOGY"    "ONTOLOGYALL" "PATH"         "PFAM"          "PMID"
## [21] "PROBEID"     "PROSITE"     "REFSEQ"       "SYMBOL"        "UCSCKG"
## [26] "UNIGENE"     "UNIPROT"
```

```
gene.data1 <- select(hgu133plus2.db, keys=rownames(topTab_neuroVSSchawnn_cell), ke
ytype="PROBEID", columns=c("ENTREZID", "GENENAME", "SYMBOL"))
head(gene.data1)
```

```
##          PROBEID ENTREZID          GENENAME SYMBOL
## 1 224323_s_at      83876          maestro      MRO
## 2  231358_at      83876          maestro      MRO
## 3  201438_at      1293 collagen type VI alpha 3 chain COL6A3
## 4 205722_s_at      2675  GDNF family receptor alpha 2  GFRA2
## 5  205721_at      2675  GDNF family receptor alpha 2  GFRA2
## 6  231227_at      <NA>          <NA>      <NA>
```

```
gene.data2 <- select(hgu133plus2.db, keys=rownames(topTab_neurowVSschwann), keytype="PROBEID", columns=c("ENTREZID", "GENENAME", "SYMBOL"))
head(gene.data2)
```

```
##          PROBEID ENTREZID          GENENAME
## 1  200795_at      8404          SPARC like 1
## 2  201525_at      347          apolipoprotein D
## 3 202878_s_at      22918          CD93 molecule
## 4 1554474_a_at      26002          monooxygenase DBH like 1
## 5  204232_at      2207          Fc fragment of IgE receptor Ig
## 6  208982_at      5175 platelet and endothelial cell adhesion molecule 1
##          SYMBOL
## 1 SPARCL1
## 2  APOD
## 3  CD93
## 4  MOXD1
## 5 FCER1G
## 6 PECAM1
```

```
gene.data3 <- select(hgu133plus2.db, keys=rownames(topTab_neuroVSmouse_whole), keytype="PROBEID", columns=c("ENTREZID", "GENENAME", "SYMBOL"))
head(gene.data3)
```

```
##          PROBEID ENTREZID          GENENAME SYMBOL
## 1  219465_at      336          apolipoprotein A2  APOA2
## 2 219466_s_at      336          apolipoprotein A2  APOA2
## 3  204694_at      174          alpha fetoprotein  AFP
## 4 211298_s_at      213          albumin          ALB
## 5 204419_x_at      3048 hemoglobin subunit gamma 2  HBG2
## 6 204419_x_at      3047 hemoglobin subunit gamma 1  HBG1
```

```
gene.data4 <- select(hgu133plus2.db, keys=rownames(topTab_INT), keytype="PROBEID", columns=c("ENTREZID", "GENENAME", "SYMBOL"))
head(gene.data4)
```

##	PROBEID	ENTREZID	GENENAME	SYMBOL
## 1	200795_at	8404	SPARC like 1	SPARCL1
## 2	201525_at	347	apolipoprotein D	APOD
## 3	202878_s_at	22918	CD93 molecule	CD93
## 4	202613_at	1503	CTP synthase 1	CTPS1
## 5	200756_x_at	813	calumenin	CALU
## 6	214845_s_at	813	calumenin	CALU