

# Group 1: Result comparison

Luca Visser

2022

## Contents

Load data	1
Comparison of RNA-seq and expression micro-array	1
Venn diagram	3
Gene set analysis	4

## Load data

```
results_files <- list.files(pattern = ".*.txt")
for (i in 1:length(results_files)) { #loop through .txt files, assign them to their respective variable
  assign(sub("\\..*", "", results_files[i]), read.table(results_files[i]))
}
```

## Comparison of RNA-seq and expression micro-array

```
head(RNAseq_results)
```

##		logFC	logCPM	F	PValue	FDR
##	ENSG00000092850	-1.0310797	1.390662	37.34583	8.346117e-07	0.00123596
##	ENSG00000174132	0.7477023	4.103482	36.94319	9.175015e-07	0.00123596
##	ENSG00000231889	-0.5807385	4.691997	37.83752	7.440575e-07	0.00123596
##	ENSG00000105877	-1.2620152	2.564962	43.31467	2.185691e-07	0.00123596
##	ENSG00000136933	0.5893419	3.701181	44.27497	1.780250e-07	0.00123596
##	ENSG00000166295	-0.5812867	4.828000	36.89729	9.274916e-07	0.00123596

```
head(limmaExprsArray_results)
```

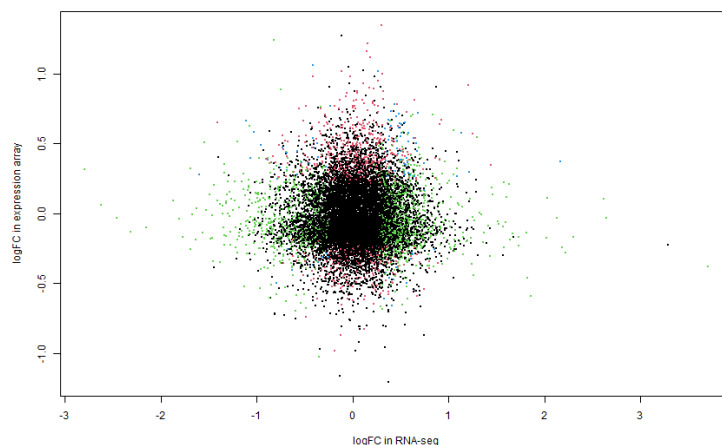
```
##          logFC    AveExpr      t      P.Value    adj.P.Val
## 33250.ENSG00000242779  0.6161607 11.814335  8.428341 1.675760e-07 0.005579778
## 14868.ENSG00000087008  0.9353745  8.488672  7.613016 6.759165e-07 0.009510968
## 20726.ENSG00000156313  0.6073496  7.709785  7.456015 8.930147e-07 0.009510968
## 9948.ENSG00000160472   0.5512948  9.784034  7.246011 1.302871e-06 0.009510968
## 20772.ENSG00000274588 -0.4732591  7.186132 -7.046497 1.875616e-06 0.009510968
## 18268.ENSG00000127399  0.5039090 10.022263  6.974057 2.143775e-06 0.009510968
##          B
## 33250.ENSG00000242779  7.174066
## 14868.ENSG00000087008  5.980536
## 20726.ENSG00000156313  5.738539
## 9948.ENSG00000160472   5.408578
## 20772.ENSG00000274588  5.088416
## 18268.ENSG00000127399  4.970554
```

```
ENS <- gsub(".*\\."," ",rownames(limmaExprsArray_results))
ENS_uniq <- ENS[!(duplicated(ENS) | is.na(ENS))]
Arr_filtered <- limmaExprsArray_results[!(duplicated(ENS) | is.na(ENS)),]
rownames(Arr_filtered) <- ENS_uniq
present_in_both <- intersect(ENS_uniq, rownames(RNAseq_results))
length(present_in_both)
```

```
## [1] 14238
```

```
RNAseq_filtered <- RNAseq_results[which(rownames(RNAseq_results) %in% present_in_both), ]
Arr_filtered <- Arr_filtered[which(gsub(".*\\."," ",rownames(Arr_filtered)) %in% present_in_both), ]
RNAseq_filtered <- RNAseq_filtered[order(rownames(RNAseq_filtered)), ]
Arr_filtered <- Arr_filtered[order(rownames(Arr_filtered)), ]
sign_RNA <- as.factor(RNAseq_filtered$FDR < 0.05)
sign_arr <- as.factor(Arr_filtered$adj.P.Val < 0.05)
sign_combined <- as.factor(paste(as.double(sign_RNA),as.double(sign_arr), sep="."))

plot(RNAseq_filtered$logFC, Arr_filtered$logFC, pch=20, xlab="logFC in RNA-seq", ylab="logFC in expression array")
```



```

print("Percentage of genes with same sign in their logFCs:")

## [1] "Percentage of genes with same sign in their logFCs:"

paste(substr(as.character(sum(sign(RNAseq_filtered$logFC) == sign(Arr_filtered$logFC))/
length(RNAseq_filtered$logFC)*100),1,5), "%", sep="")

## [1] "48.85%"

print("Percentage of genes which are significantly DE for both analyses (blue in plot) with same sign in both analyses:")

## [1] "Percentage of genes which are significantly DE for both analyses (blue in plot) with same sign in both analyses:"

paste(substr(as.character(sum((sign(RNAseq_filtered$logFC) == sign(Arr_filtered$logFC))
[as.character(sign_combined)=="2.2"])/sum(as.character(sign_combined)=="2.2")*100),1,5),
"%", sep="")

## [1] "62.62%"

```

## Venn diagram

```

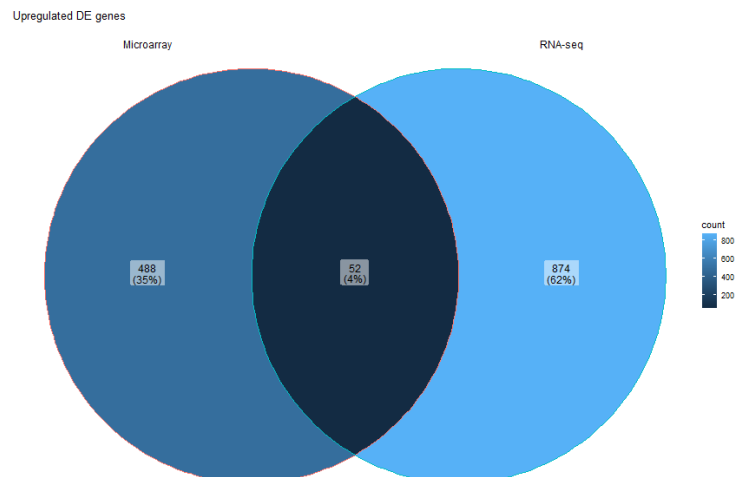
library(ggVennDiagram)
library(ggplot2)

Arr.up <- rownames(Arr_filtered[Arr_filtered$logFC > 0 & Arr_filtered$adj.P.Val < 0.05,])
Arr.down <- rownames(Arr_filtered[Arr_filtered$logFC < 0 & Arr_filtered$adj.P.Val < 0.05,])

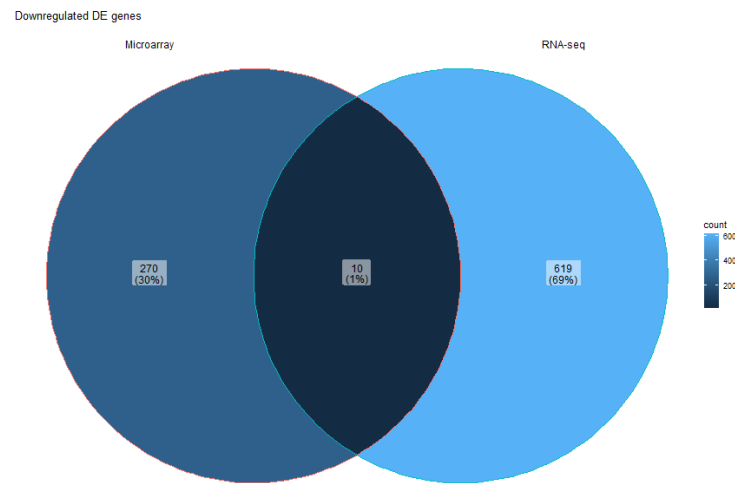
Seq.up <- rownames(RNAseq_filtered[RNAseq_filtered$logFC > 0 & RNAseq_filtered$FDR < 0.05,])
Seq.down <- rownames(RNAseq_filtered[RNAseq_filtered$logFC < 0 & RNAseq_filtered$FDR < 0.05,])

venn_up <- ggVennDiagram(list("Microarray" = Arr.up, "RNA-seq" = Seq.up)) +
  labs(title="Upregulated DE genes")
print(venn_up)

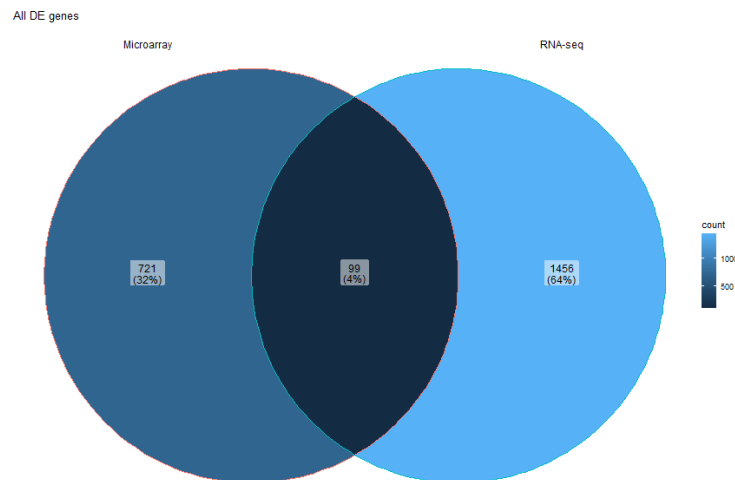
```



```
venn_down <- ggVennDiagram(list("Microarray" = Arr.down, "RNA-seq" = Seq.down)) +
  labs(title="Downregulated DE genes")
print(venn_down)
```



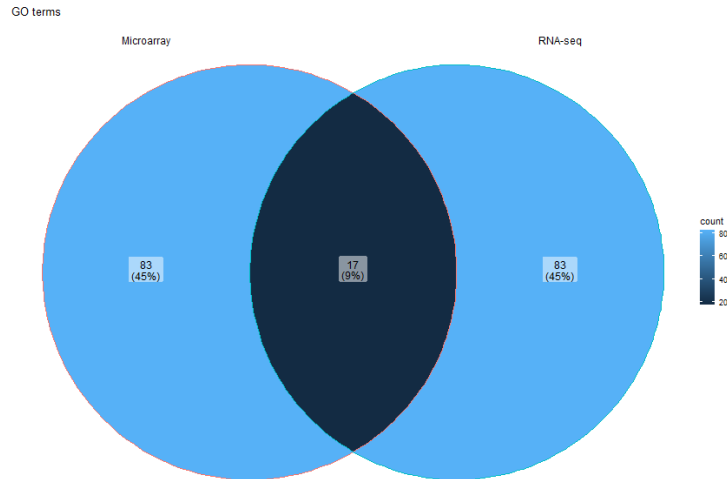
```
venn_all <- ggVennDiagram(list("Microarray" = append(Arr.up, Arr.down), "RNA-seq" = append(Seq.up, Seq.down))) +
  labs(title="All DE genes")
print(venn_all)
```



## Gene set analysis

```
RNATERMS <- rownames(RNAseq_GSA_results)
Arrterms <- rownames(array_GSA_results)

venn_GO <- ggVennDiagram(list("Microarray" = Arrterms, "RNA-seq" = RNATERMS)) +
  labs(title="GO terms")
print(venn_GO)
```



```
library(dplyr)

common_arr <- array_GSA_results[Arrrterms %in% intersect(RNATERms, Arrrterms), ] %>%
  rename("DE.Array" = DE) %>%
  rename("P.Array" = P.DE) %>%
  rename("FDR.Array" = FDR.DE)
common_seq <- RNAseq_GSA_results[RNATERms %in% intersect(RNATERms, Arrrterms), ] %>%
  rename("DE.RNAseq" = DE) %>%
  rename("P.RNAseq" = P.DE) %>%
  rename("FDR.RNAseq" = FDR.DE)

print(common_arr)
```

```
##                               Term Ont      N DE.Array
## GO:0009987                    cellular process BP 17266    2557
## GO:1901564 organonitrogen compound metabolic process BP 6535    1019
## GO:0044237                    cellular metabolic process BP 10271    1554
## GO:0006807                    nitrogen compound metabolic process BP 10310    1555
## GO:0044238                    primary metabolic process BP 10825    1621
## GO:0044260                    cellular macromolecule metabolic process BP 3266    524
## GO:0009058                    biosynthetic process BP 5962    918
## GO:1901576                    organic substance biosynthetic process BP 5877    904
## GO:0051171 regulation of nitrogen compound metabolic process BP 5792    891
## GO:0044249                    cellular biosynthetic process BP 5807    893
## GO:0071840                    cellular component organization or biogenesis BP 6601    1008
## GO:0044085                    cellular component biogenesis BP 3273    517
## GO:0034641                    cellular nitrogen compound metabolic process BP 7053    1068
## GO:0043933                    protein-containing complex organization BP 1845    302
## GO:0044271                    cellular nitrogen compound biosynthetic process BP 4865    748
## GO:0016043                    cellular component organization BP 6395    969
## GO:0051641                    cellular localization BP 3351    524
##                               P.Array    FDR.Array
## GO:0009987 7.884158e-09 4.252359e-05
## GO:1901564 6.408401e-05 1.051115e-01
## GO:0044237 8.867290e-05 1.357464e-01
## GO:0006807 1.794420e-04 2.168698e-01
## GO:0044238 5.032971e-04 3.873451e-01
```

```
## G0:0044260 6.695712e-04 4.393718e-01
## G0:0009058 1.044661e-03 5.578733e-01
## G0:1901576 1.322674e-03 6.074514e-01
## G0:0051171 1.447683e-03 6.392911e-01
## G0:0044249 1.485936e-03 6.428746e-01
## G0:0071840 1.537623e-03 6.428746e-01
## G0:0044085 2.722484e-03 7.612893e-01
## G0:0034641 3.099615e-03 7.700151e-01
## G0:0043933 3.381217e-03 8.172936e-01
## G0:0044271 4.328794e-03 9.123703e-01
## G0:0016043 5.087989e-03 9.655826e-01
## G0:0051641 5.571582e-03 1.000000e+00
```

```
print(common_seq)
```

##		Term	Ont	N
## G0:0044237		cellular metabolic process	BP	10271
## G0:0006807		nitrogen compound metabolic process	BP	10310
## G0:0044238		primary metabolic process	BP	10825
## G0:0034641		cellular nitrogen compound metabolic process	BP	7053
## G0:0009987		cellular process	BP	17266
## G0:1901564		organonitrogen compound metabolic process	BP	6535
## G0:0044260		cellular macromolecule metabolic process	BP	3266
## G0:1901576		organic substance biosynthetic process	BP	5877
## G0:0009058		biosynthetic process	BP	5962
## G0:0044085		cellular component biogenesis	BP	3273
## G0:0071840		cellular component organization or biogenesis	BP	6601
## G0:0043933		protein-containing complex organization	BP	1845
## G0:0044271		cellular nitrogen compound biosynthetic process	BP	4865
## G0:0016043		cellular component organization	BP	6395
## G0:0044249		cellular biosynthetic process	BP	5807
## G0:0051641		cellular localization	BP	3351
## G0:0051171		regulation of nitrogen compound metabolic process	BP	5792
##	DE.RNAseq	P.RNAseq	FDR.RNAseq	
## G0:0044237	965	4.271072e-20	6.538442e-17	
## G0:0006807	947	6.055583e-16	5.039696e-13	
## G0:0044238	986	6.145168e-16	5.039696e-13	
## G0:0034641	681	4.234797e-14	2.946777e-11	
## G0:0009987	1421	7.990243e-12	3.335999e-09	
## G0:1901564	623	2.049681e-11	7.757276e-09	
## G0:0044260	344	1.127885e-10	3.597170e-08	
## G0:1901576	563	1.995961e-10	5.952370e-08	
## G0:0009058	569	2.809006e-10	8.062899e-08	
## G0:0044085	342	3.424318e-10	9.707731e-08	
## G0:0071840	620	3.954915e-10	1.094177e-07	
## G0:0043933	213	5.339682e-10	1.442531e-07	
## G0:0044271	473	1.909195e-09	4.566755e-07	
## G0:0016043	597	3.285398e-09	7.254096e-07	
## G0:0044249	547	6.712226e-09	1.401208e-06	
## G0:0051641	331	3.009259e-07	4.669028e-05	
## G0:0051171	531	6.160069e-07	8.896457e-05	