

## ADW lab 5 raport

## 1. Załadowanie bibliotek i wczytanie wyników mapowania odczytów do sekwencji miRNA

```

1 library(limma)
2 library(ggplot2)
3 library(edgeR)
4 library(dplyr)
5 library(edgeR)
6 library(Biobase)
7
8 ### ODCZYTY I FILTROWANIE DANYCH ###
9 micro = read.table('c:\\Users\\marce\\Desktop\\biurowe 5\\Analiza danych wysokoprzepust\\z5\\miHomo.txt')
10 dim(micro)
11
12 n_samples <- dim(micro)[2]
13 row_sums <- apply(micro, 1, sum)
14 avgs <- row_sums / n_samples
15
16 micro <- micro[avgs >= 3.0,]
17 dim(micro)      ###----> brak zmian, dane przefiltrowane na wstepie
18 micro
19

```

Po przefiltrowaniu liczba genów nie uległa zmianie, być może były to już przefiltrowane dane. Nadal tabela posiada 450 miRNA i 32 próbki

Kolejno wczytano dane kliniczne dotyczące próbek:

```

pData=read.table("C:\\Users\\marce\\Desktop\\biurowe 5\\Analiza danych wysokoprzepust\\z5\\phenoData.txt",row.names=1,
                header=TRUE, sep="\t")

pData[pData$subtype %in% c('normalBM', 'normalPB'), 'subtype'] = 'CTRL'

pData$age1 <- case_when(
  pData$age1 == 'CTRL' ~ 'CTRL',
  pData$age1 < 40 ~ 'young',
  pData$age1 >= 40 & pData$age1 < 59 ~ 'middle',
  pData$age1 >= 59 ~ 'old',
  TRUE ~ as.character(pData$age1) # Keep the original value for other cases
)
pData

metaData <- data.frame(labelDescription=
  c("type_of_cancer","Mutation_type", "Sex", "Age"),row.names=
  c("type", "virus", "timepoint", "individual"))

metaData

phenoData <- new("AnnotatedDataFrame",data=pData, varMetadata=metaData)
phenoData

```

Następnym krokiem było stworzenie obiektu DGEList na podstawie uprzednio wczytanych danych. Taki obiekt zawiera w sobie dane dotyczące ekspresji miRNA w każdej próbce oraz powiązane dane kliniczne:

```

### DGE ###
Homo_eSet <- ExpressionSet(assayData = as.matrix(micro), phenoData = phenoData)
Counts1 <- exprs(Homo_eSet)
y1=DGEList(counts=Counts1, genes=rownames(Counts1))

```

Kolejnym etapem było zaprojektowanie macierzy dla naszego eksperymentu, która zawierała logikę porównań próbek.

```

### Design i contrast metrics ###
types <- Homo_eSet$type
design <- model.matrix(~types,ref="AML")
colnames(design) <- c("AML","CTRL")
contrast.matrix <- makeContrasts(AML-CTRL, levels=design)

```

\*\*\* Na tym etapie wydają mi się, że zauważyłem błąd w skrypcie PDF z poleceniami co do projektu. Mianowicie tak konstruując design eksperymentu nie musimy tworzyć contrast matrix (wykorzystujemy podejście z coefficient, a contrast matrix wymagana jest w drugim podejściu str. 54 Limma Userguide). Kontynuując podejście wskazane w PDF w każdym porównaniu wychodziło, że wszystkie geny zawsze ulegają nadekspresji nie zachowując przy tym logiki, tzn w pierwszym porównaniu AML vs CTRL wszystkie geny AML były UP regulated, również w porównaniu M1 a M2 wszystkie geny M1 były UP regulated, natomiast w M1 a kontroli wychodziło, że kontrola ma wszystkie geny UP regulated.

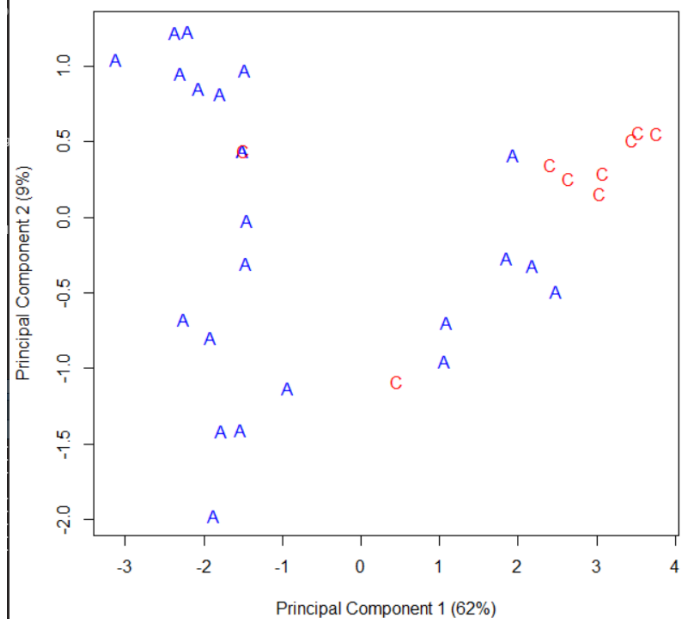
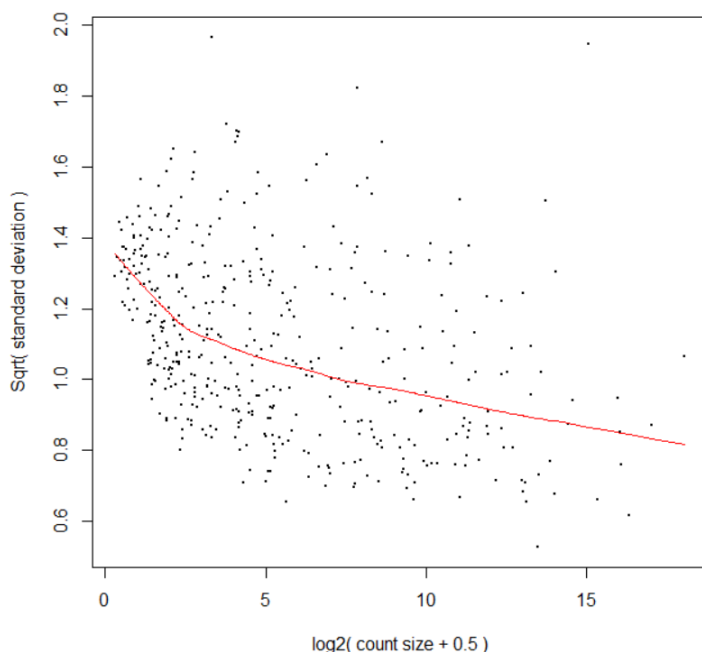
Z powyższego powodu zmodyfikowano projekt eksperymentu i trzymano się tego podejścia dla wszystkich porównań:

```
### Design i contrast matrices ###
Group <- factor(Homo_eSet$type, levels=c("CTRL","AML"))
design <- model.matrix(~Group)
colnames(design) <- c("CTRL","AMLvsCTRL")
design
```

Ostatnim etapem przygotowania danych było obliczenie wektora normalnego dla próbek i normalizacja danych poprzez funkcję voom, oraz wizualizacja w postaci wykresów:

```
### Normalizacja ###
y1 <- calcNormFactors(y1)
v1 <- voom(y1, design, plot=TRUE)
plotMDS(v1, top=50, labels=substring(types,1,1), col=ifelse(types=="AML", "blue", "red"), gene.selection= "common")
```

voom: Mean-variance trend



Wykres po lewej opisuje związek między średnim poziomem ekspresji dla genów (czerwona linia) a tym jak niektóre geny (miRNA) odbiegają od tej średniej (czarne punkty). Na osi Y mamy wariancję, a na osi x logarytm ze zliczeń genów.

Wykres po prawej stosując PCA obrazuje ułożenie próbek w przestrzeni dwuwymiarowej. Widać wówczas czy są jakieś różnice między próbkami nowotworowymi a kontrolnymi. Próbki kontrolne skupione są raczej w jednym miejscu i w większości z dala od próbek nowotworowych, zatem można już założyć, że występują różnice w ekspresji. Co ciekawe dwie próbki C i A nakładają się na siebie, być może doszło do błędu i próbka była powielona ale ze zmienioną kategorią?

Po takim przygotowaniu danych można było znaleźć liczbę miRNA ulegających różnicowej ekspresji przy odpowiednim progu p-value (tutaj 0.01):

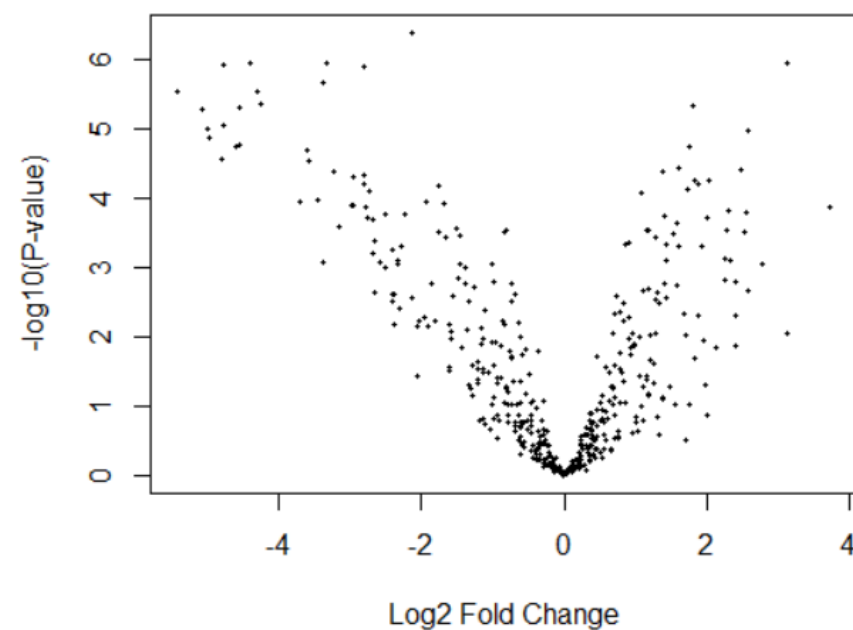
```
### wyniki ###
fit1 = lmFit(v1,design) |
fit1 <- eBayes(fit1)
top = topTable(fit1,coef="AMLvsCTRL",n=Inf)
top1 = top[top$adj.P.Val<=0.01,]
top1
```

Otrzymano 101 miRNA, które ulegają ekspresji różnicowej w próbkach nowotworowych względem próbek kontrolnych. Ponad połowa z nich ulegała obniżonej ekspresji miRNA względem próbek kontrolnych. Fragment tabeli i wyniki przedstawiono poniżej.

	genes	logFC	AveExpr	t	P.Value	adj.P.Val	B
hsa-miR-141-3p	hsa-miR-141-3p	-2.1307166	6.28389533	-6.246144	4.142167e-07	9.261642e-05	6.43800008
hsa-miR-199b-5p	hsa-miR-199b-5p	3.1110937	9.08457121	5.919428	1.097713e-06	9.261642e-05	5.51177335
hsa-miR-150-5p	hsa-miR-150-5p	-3.3256713	12.29935585	-5.919162	1.098586e-06	9.261642e-05	5.45380775
hsa-miR-369-3p	hsa-miR-369-3p	-4.3944396	1.19994232	-5.904583	1.147528e-06	9.261642e-05	5.40489769
hsa-miR-889	hsa-miR-889	-4.7654477	1.97680970	-5.889011	1.202226e-06	9.261642e-05	5.39539184
hsa-miR-342-3p	hsa-miR-342-3p	-2.7969738	10.78073626	-5.880048	1.234886e-06	9.261642e-05	5.36462281
hsa-miR-382-5p	hsa-miR-382-5p	-3.3660449	0.23220729	-5.689731	2.182768e-06	1.403208e-04	4.76513052
hsa-miR-370	hsa-miR-370	-4.2986310	1.30878722	-5.602114	2.837822e-06	1.473703e-04	4.58094843
hsa-miR-410	hsa-miR-410	-5.4104565	3.34229932	-5.589467	2.947406e-06	1.473703e-04	4.58078213
hsa-miR-493-5p	hsa-miR-493-5p	-4.2277979	1.61953937	-5.462784	4.307988e-06	1.773426e-04	4.20691767
hsa-miR-136-3p	hsa-miR-136-3p	-4.5505426	2.01642320	-5.426840	4.797777e-06	1.773426e-04	4.11564917

Showing 1 to 18 of 101 entries, 7 total columns

Dodatkowe zobrazowanie wyników:



```
results <- decideTests(fit1, p.value=0.01)
summary(results)
volcanoplot(fit1, coef=2)
```

```
> summary(results)
      CTRL AMLvsCTRL
Down      0         61
NotSig    82        349
Up       368         40
```

Wykres volcano plot pokazuje czy genu ulegają różnicowej ekspresji (górne wysunięte najbardziej bokach). Oś X opisuje logarytm zmiany, czyli jak duża jest różnica, a oś Y istotność statystyczną obserwacji.

Następnie zbadano inne rodzaje porównań.

- Porównanie typów białaczki M1 i M2

Aby to zrobić należało odpowiednio zaprojektować design matrix wydobywając podtypy, oraz okroić naszą macierz zliczeń usuwając próbki (kolumny) z kontrolami.

```
#porównanie typów białaczki M1 i M2
y1=DGEList(counts=Counts1, genes=rownames(Counts1))

subtypes <- Homo_eSet$subtype
treated_idx = which(subtypes!='CTRL')
y1 <- y1[, treated_idx]
subtypes <- subtypes[subtypes!='CTRL']

Group <- factor(subtypes, levels=c("M1","M2"))
design <- model.matrix(~Group)
colnames(design) <- c("M1","M2vsM1")
design
```

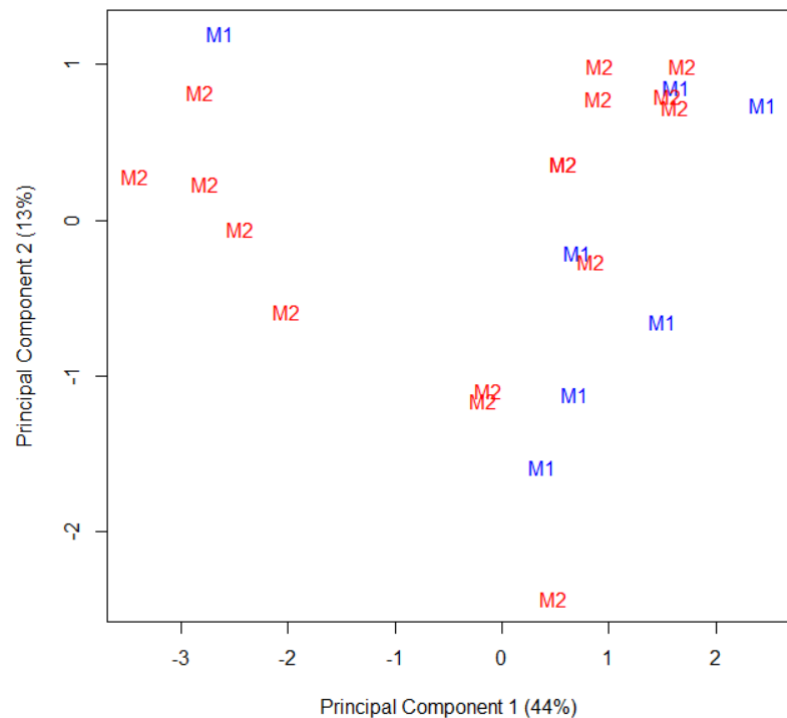
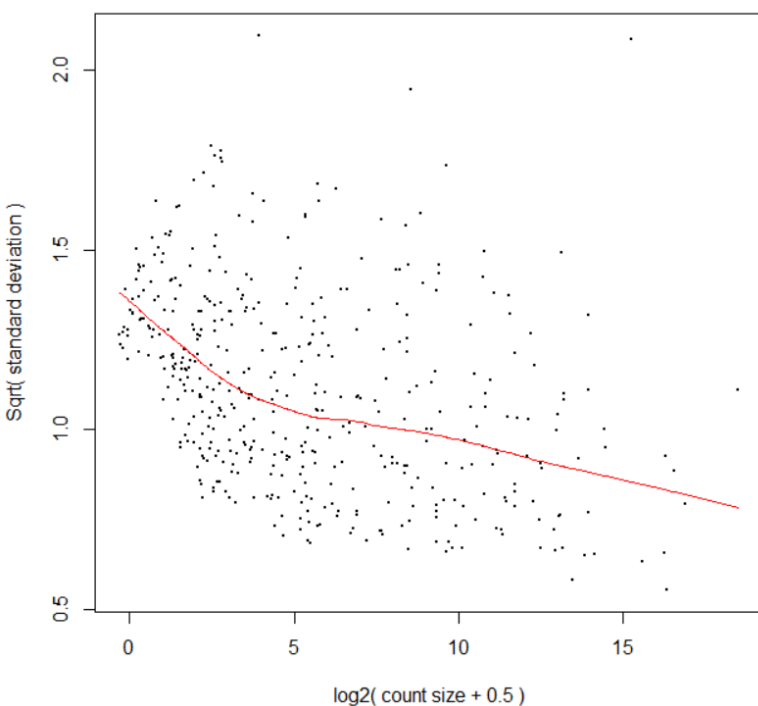
Analogicznie przeprowadzono normalizację oraz wyliczenie wyników i wizualizację:

```
y1 <- calcNormFactors(y1)
v1 <- voom(y1,design,plot=TRUE)
plotMDS(v1,top=50,labels=subtypes, col=ifelse(subtypes=="M1", "blue","red"), gene.selection= "common")

### wyniki ###
fit1 = lmFit(v1,design)
fit1 <- eBayes(fit1)
top = topTable(fit1,coef=2,n=Inf)
top
top1 = top[top$adj.p.val<=0.05,]
top1

results <- decideTests(fit1, p.value=0.01)
summary(results)
volcanoplot(fit1, coef=2)
```

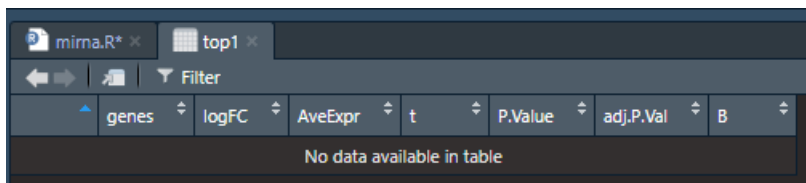
voom: Mean-variance trend



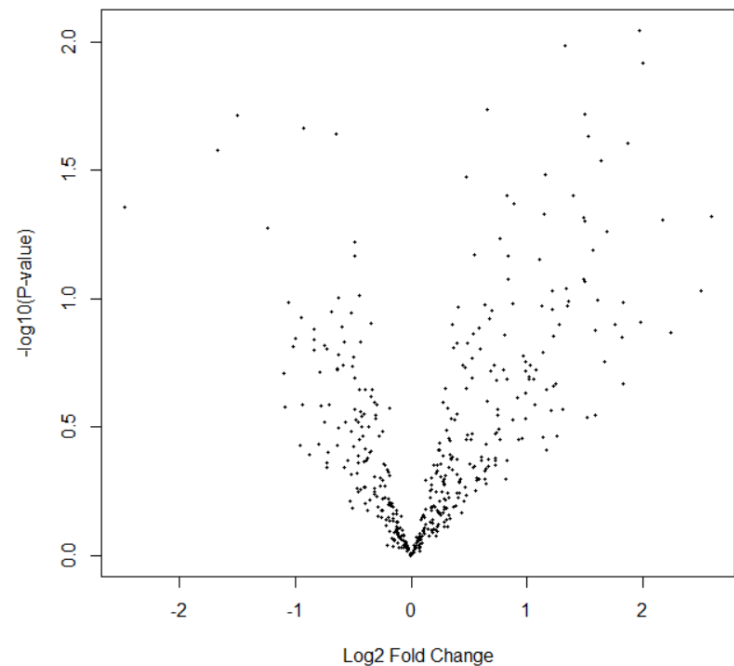
Jeśli chodzi o MDS plot to nie można tu z pewnością zakładać jakiś większych różnic między próbkami, mimo, że widać obecność niektórych genów miRNA odbiegających od średnich, co mogłoby sugerować ekspresję różnicową, jednak nie wiadomo czy są to różnice istotne statystycznie, ponieważ wiadomo, że próbek M2 jest znacznie więcej niż M1. Poniżej wyniki:

```
> summary(results)
      M1 M2vsM1
Down    1      0
NotSig 125    450
Up     324      0
```

Nie wykryto tutaj żadnych genów różnicowych między M1 i M2. Zarówno przy progu pvalue 0.05 jak i 0.01. Możemy wykluczyć różnicę w ekspresji badanych miRNA między podtypami białaczki.



Podobnie jak pokazywał poprzedni wykres, są geny z różnymi poziomami ekspresji jednak ich pvalue jest zbyt wysokie.



- Porównanie typu białaczki M1 i próbek kontrolnych:

```
#porównanie typu białaczki M1 i kontroli
y1=DGEList(counts=counts1, genes=rownames(counts1))

subtypes <- Homo.eSet$subtype
m1 = which(subtypes!='M2')
y1 <- y1[, m1]
subtypes <- subtypes[subtypes!='M2']

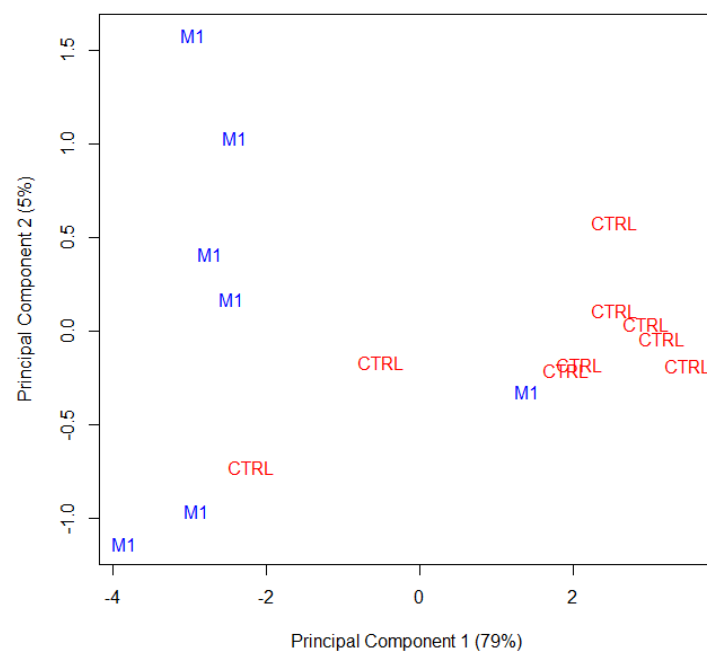
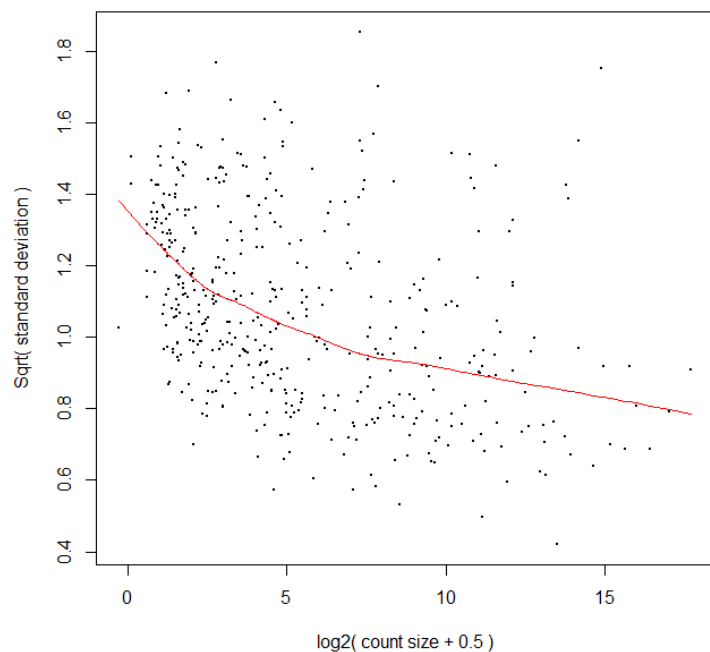
Group <- factor(subtypes, levels=c("CTRL","M1"))
design <- model.matrix(~Group)
colnames(design) <- c("CTRL","M1vsCTRL")
design

y1 <- calcNormFactors(y1)
v1 <- voom(y1,design,plot=TRUE)
plotMDS(v1,top=50,labels=subtypes, col=ifelse(subtypes=="M1", "blue","red"), gene.selection= "common")

### wyniki ###
fit1 = lmFit(v1,design)
fit1 <- eBayes(fit1)
top = topTable(fit1,coef="M1vsCTRL",n=Inf)
top1 = top[top$adj.P.val<=0.01,]
top1

results <- decideTests(fit1, p.value=0.01)
summary(results)
volcanoplot(fit1, coef=2)
```

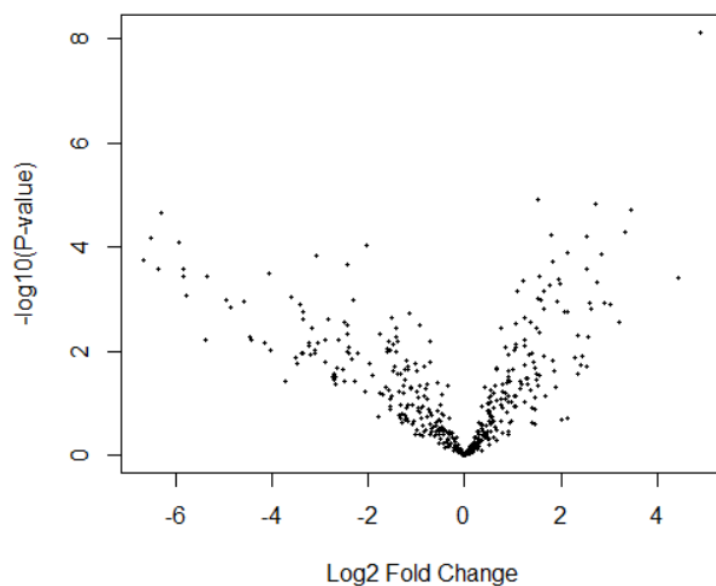
voom: Mean-variance trend



```
> summary(results)
      CTRL M1vsCTRL
Down    0      12
NotSig  79     418
Up     371     20
```

W tym porównaniu otrzymujemy wynik, że mamy 32 geny miRNA ulegające różnicowej ekspresji w M1 względem kontroli.

	genes	logFC	AveExpr	t	P.Value	adj.P.Val	B
hsa-miR-196b-5p	hsa-miR-196b-5p	4.896815	7.9023146	10.269606	7.306614e-09	3.287976e-06	10.52331340
hsa-miR-92a-3p	hsa-miR-92a-3p	1.524932	15.1055589	6.014619	1.195255e-05	1.951079e-03	3.33449826
hsa-miR-181d	hsa-miR-181d	2.712177	7.2683709	5.901935	1.501594e-05	1.951079e-03	3.24298715
hsa-miR-199b-5p	hsa-miR-199b-5p	3.458409	8.4763287	5.785913	1.902486e-05	1.951079e-03	2.99283053
hsa-miR-127-3p	hsa-miR-127-3p	-6.280807	6.8466181	-5.722245	2.167866e-05	1.951079e-03	2.91766527
hsa-miR-196b-3p	hsa-miR-196b-3p	3.340914	0.9824889	5.298661	5.233792e-05	3.380532e-03	2.06042077
hsa-miR-181c-3p	hsa-miR-181c-3p	1.802645	7.6781951	5.264304	5.626852e-05	3.380532e-03	1.94674998
hsa-miR-92a-1-5p	hsa-miR-92a-1-5p	2.535210	3.5059326	5.209770	6.313957e-05	3.380532e-03	1.91464027
hsa-miR-654-3p	hsa-miR-654-3p	-6.504534	4.3284390	-5.177458	6.761065e-05	3.380532e-03	1.85699476
hsa-miR-381-3p	hsa-miR-381-3p	-5.917337	4.3339324	-5.093373	8.082912e-05	3.637310e-03	1.69027804
hsa-miR-141-3p	hsa-miR-141-3p	-2.040109	7.0709912	-5.020501	9.441492e-05	3.862429e-03	1.46349482
hsa-miR-181b-3p	hsa-miR-181b-3p	2.852971	4.1001470	4.846590	1.370867e-04	4.745309e-03	1.16682861
hsa-miR-181b-5p	hsa-miR-181b-5p	2.138889	12.1370795	4.876198	1.286274e-04	4.745309e-03	1.04155548



- Porównanie typu białaczki M2 i próbek kontrolnych

```
#porównanie typu białaczki M2 i kontroli
y1=DGEList(counts=Counts1, genes=rownames(Counts1))

subtypes <- Homo_eSet$subtype
m2 = which(subtypes!='M1')
y1 <- y1[, m2]
subtypes <- subtypes[subtypes!='M1']

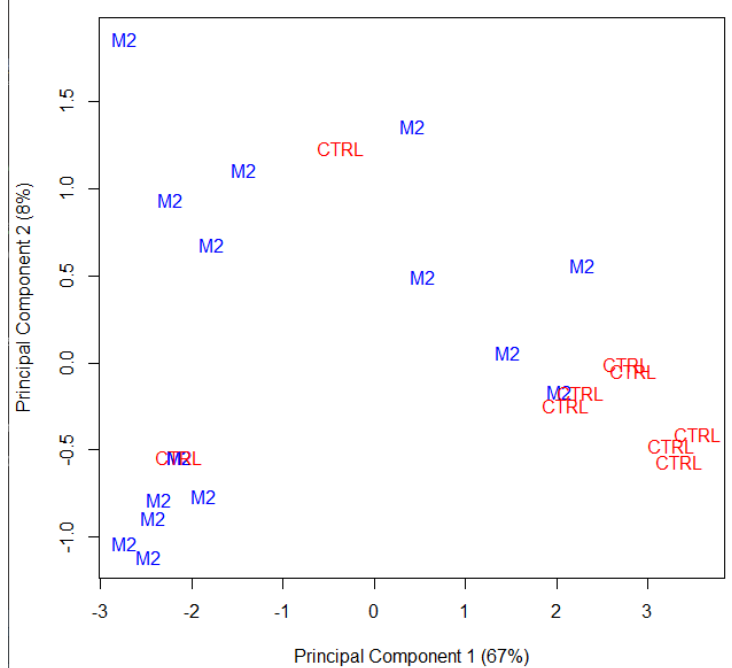
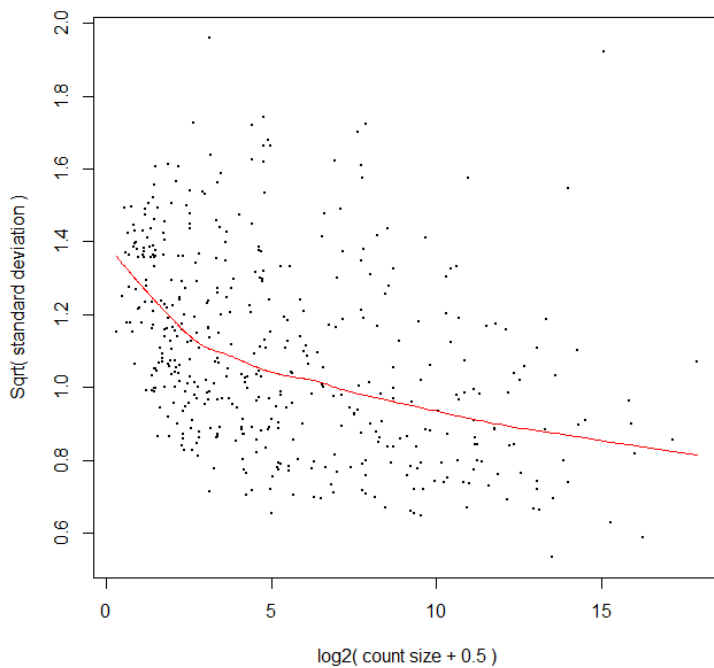
Group <- factor(subtypes, levels=c("CTRL","M2"))
design <- model.matrix(~Group)
colnames(design) <- c("CTRL","M2vsCTRL")
design

y1 <- calcNormFactors(y1)
v1 <- voom(y1,design,plot=TRUE)
plotMDS(v1,top=50,labels=subtypes, col=ifelse(subtypes=="M2", "blue","red"), gene.selection= "common")

### wyniki ###
fit1 = lmFit(v1,design)
fit1 <- eBayes(fit1)
top = topTable(fit1,coef="M2vsCTRL",n=Inf)
top1 = top[top$adj.P.Val<=0.01,]
top1

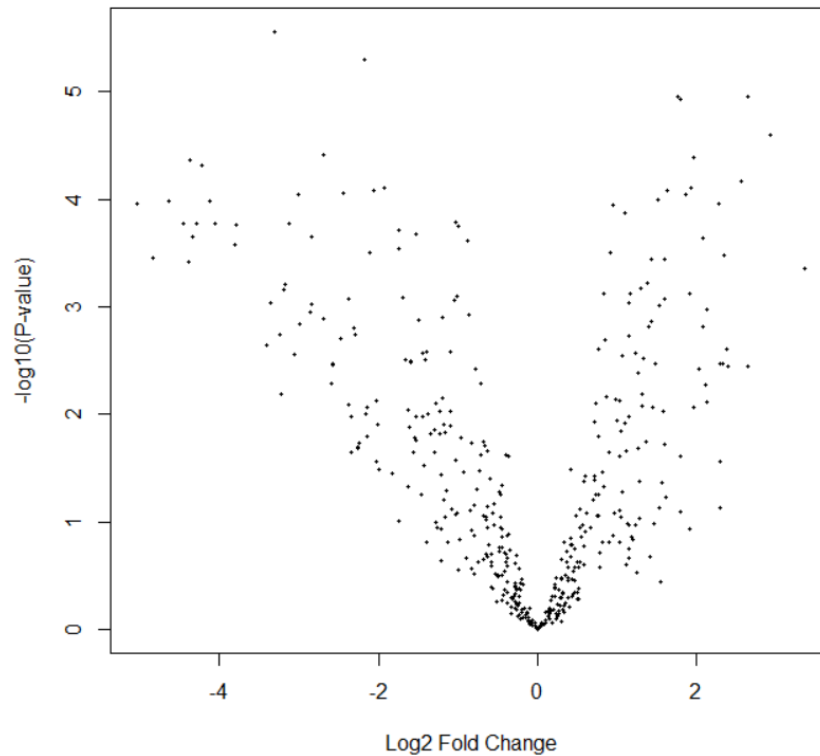
results <- decideTests(fit1, p.value=0.01)
summary(results)
volcanoplot(fit1, coef="M2vsCTRL")
```

voom: Mean-variance trend



```
> summary(results)
      CTRL M2vsCTRL
Down    0      44
NotSig  82     375
Up     368      31
```

Otrzymano 75 genów ulegających ekspresji różnicowej.



- Porównanie próbek kobiet i mężczyzn (chorych)

```
#porównanie F i M
y1=DGEList(counts=Counts1, genes=rownames(Counts1))

sexs <- Homo_eSet$sex
mf = which(sexs!='CTRL')
y1 <- y1[, mf]
sexs <- sexs[sexs!='CTRL']

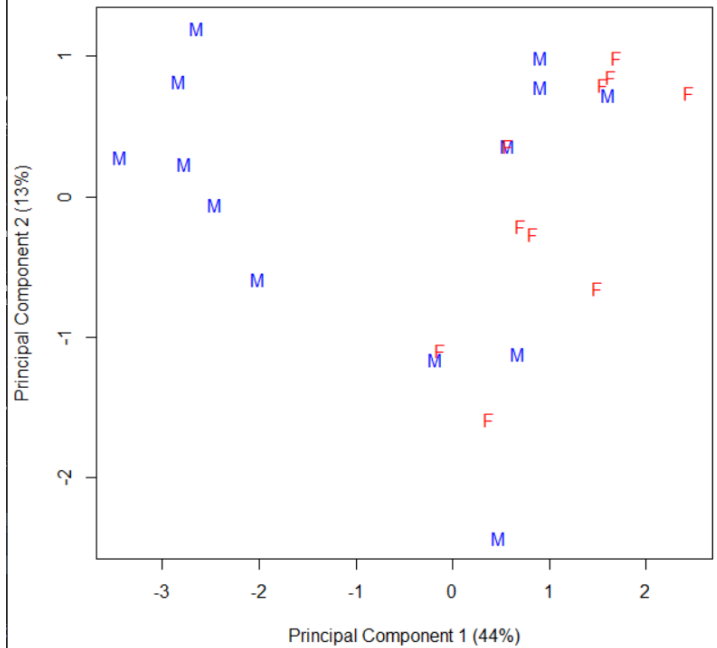
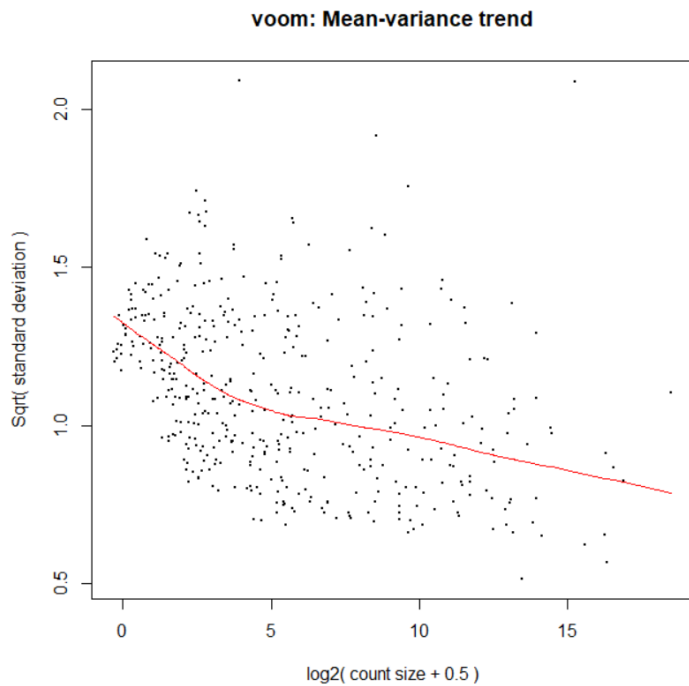
Group <- factor(sexs, levels=c("F","M"))
design <- model.matrix(~sexs)
colnames(design) <- c("F","MVSF")
design

y1 <- calcNormFactors(y1)
v1 <- voom(y1,design,plot=TRUE)
plotMDS(v1,top=50,labels=sexs, col=ifelse(sexs=="M", "blue","red"), gene.selection= "common")

### wyniki ###
fit1 = lmFit(v1,design)
fit1 <- eBayes(fit1)
top = topTable(fit1,coef=2,n=Inf)
top1 = top[top$adj.P.Val<=0.01,]
top1

results <- decideTests(fit1)
summary(results)
volcanoplot(fit1, coef=2)
```

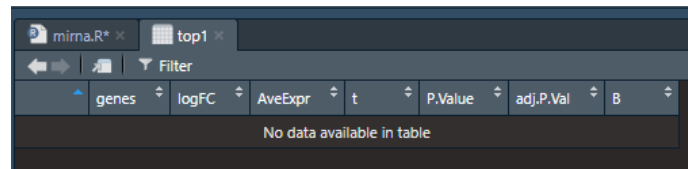




Po wykresie MDS widzimy, że można się spodziewać jakiś różnic między próbkami męskimi a damskimi. Jednak część próbek męskich mimo wszystko jest wymieszana z damskimi.

```
> summary(results)
      F MvsF
Down   23   0
NotSig  84 450
Up    343   0
```

Widać jednak, że nie ma istotnych różnic między próbkami męskimi ani damskimi. Zarówno przy p-value 0.01 i 0.05 otrzymujemy pustą tabelę:



- Porównanie próbek męskich i kontrolnych

```
#porównanie M i kontrola
y1=DGELIST(counts=Counts1, genes=rownames(Counts1))

sexs <- Homo_eSet$sex
mf = which(sexs!='F')
y1 <- y1[, mf]
sexs <- sexs[sexs!='F']

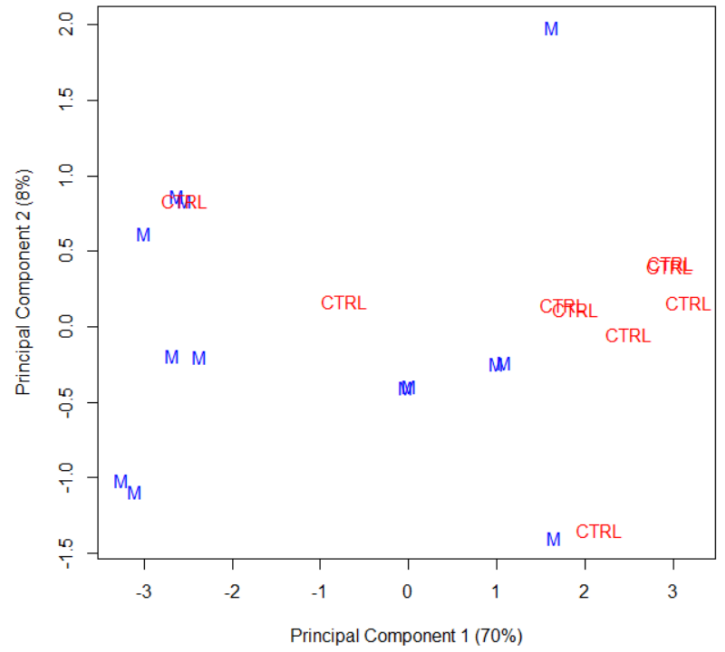
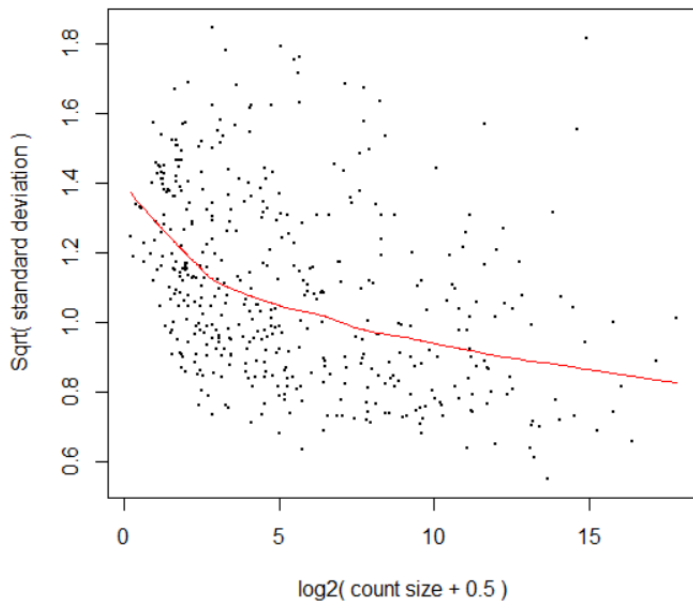
Group <- factor(sexs, levels=c("CTRL","M"))
design <- model.matrix(~Group)
colnames(design) <- c("CTRL","MvsCTRL")
design

y1 <- calcNormFactors(y1)
v1 <- voom(y1,design,plot=TRUE)
plotMDS(v1,top=50,labels=sexs, col=ifelse(sexs=="M", "blue","red"), gene.selection= "common")

### wyniki ###
fit1 = lmFit(v1,design)
fit1 <- eBayes(fit1)
top = topTable(fit1,coef=2,n=Inf)
top1 = top[top$adj.P.Val<=0.01,]
top1

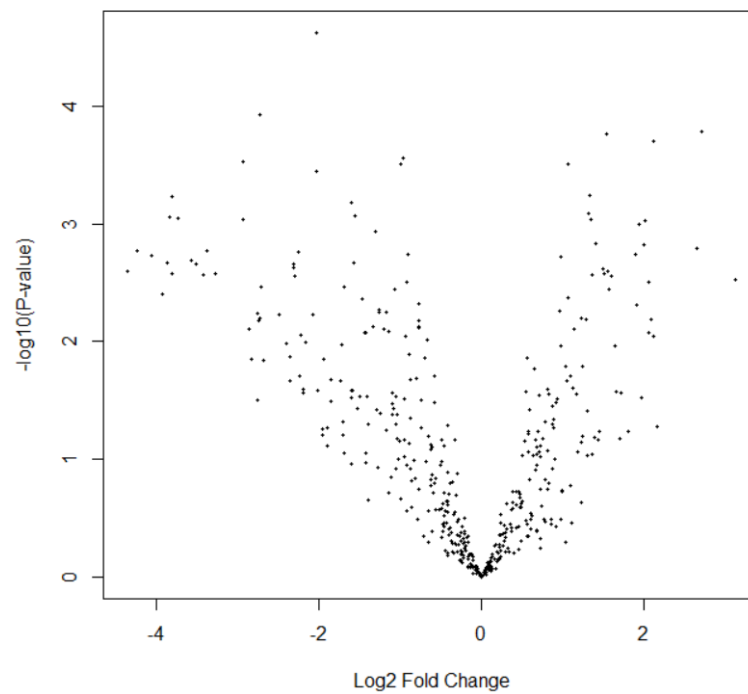
results <- decideTests(fit1)
summary(results)
volcanoplot(fit1, coef=2)
```

voom: Mean-variance trend



```
> summary(results)
      CTRL MvsCTRL
Down    0      55
NotSig  44     364
Up     406      31
```

Przy pvalue 0.01 nie otrzymano żadnych miRNA, natomiast już przy p-value 0.05 otrzymano ich aż 86.



- Porównanie próbek kobiecych i kontrolnych

```
#porównanie F i kontrola
y1=DGEList(counts=Counts1, genes=rownames(Counts1))

sexs <- Homo_eSet$sex
mf = which(sexs!='M')
y1 <- y1[, mf]
sexs <- sexs[sexs!='M']

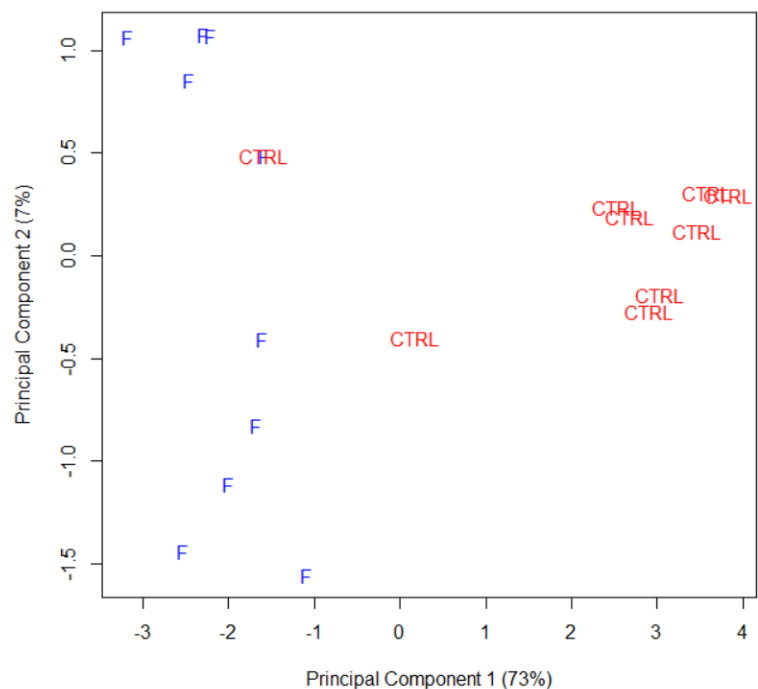
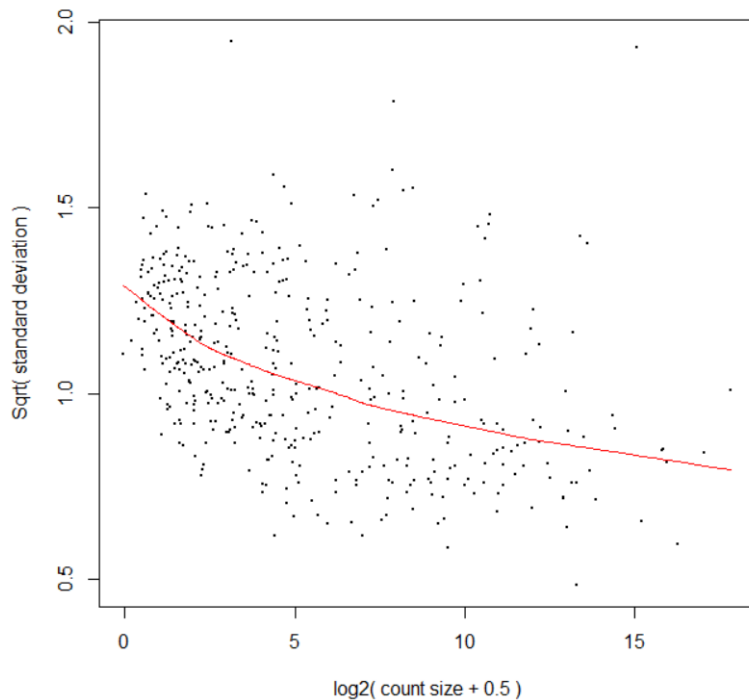
Group <- factor(sexs, levels=c("CTRL","F"))
design <- model.matrix(~Group)
colnames(design) <- c("CTRL","FvsCTRL")
design

y1 <- calcNormFactors(y1)
v1 <- voom(y1,design,plot=TRUE)
plotMDS(v1,top=50,labels=sexs, col=ifelse(sexs=="M", "blue","red"), gene.selection= "common")

### wyniki ###
fit1 = lmFit(v1,design)
fit1 <- eBayes(fit1)
top = topTable(fit1,coef=2,n=Inf)
top1 = top[top$adj.P.Val<=0.01,]
top1

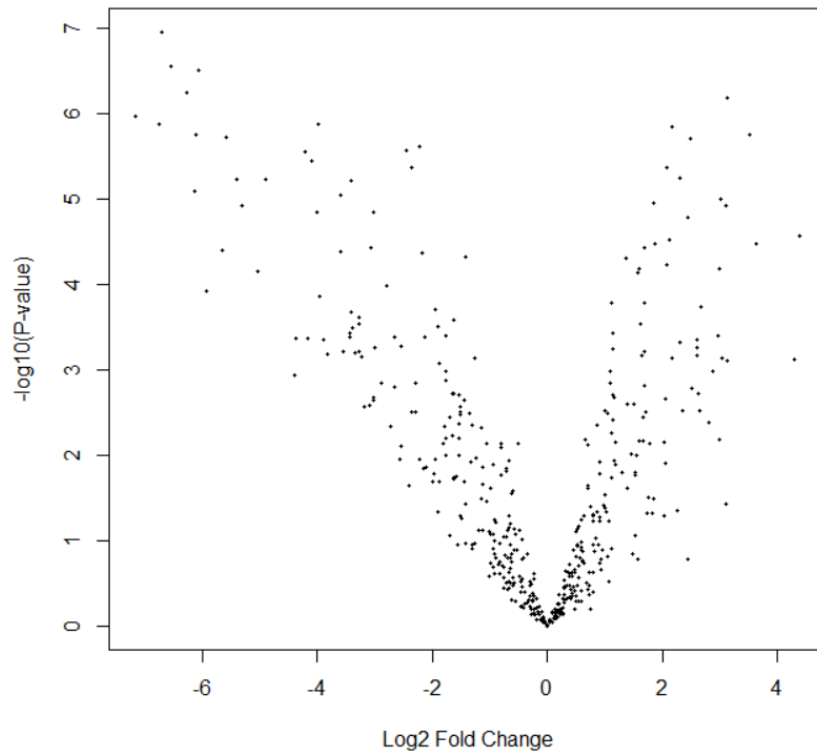
results <- decideTests(fit1, p.value=0.01)
summary(results)
volcanoplot(fit1, coef=2)
```

voom: Mean-variance trend



Tutaj już można się spodziewać większych różnic niż w przypadku mężczyzn, jednak porównanie kobiety-mężczyźni pokazało brak różnic. I faktycznie patrząc na poniższe wyniki otrzymujemy aż 117 genów ulegających ekspresji różnicowej przy pvalue 0.01. Co trochę kłóci się z poprzednimi porównaniami dot. płci. Jednak przy porównaniu M-F widać, że część próbek faktycznie jest mocno różna od siebie, a część męskich próbek pokrywa się z damskimi i być może to wygładza te różnice przy znajdowaniu genów różnicowych.

```
> summary(results)
      CTRL FvsCTRL
Down    0      68
NotSig  84     333
Up     366     49
```



- Porównanie w zależności od wieku:

```
### wiek wszyscy ###
y1=DGEList(counts=Counts1, genes=rownames(Counts1))

ages <- Homo_eSet$age1
ag = which(ages!='CTRL')
y1 <- y1[, ag]
ages <- ages[ages!='CTRL']

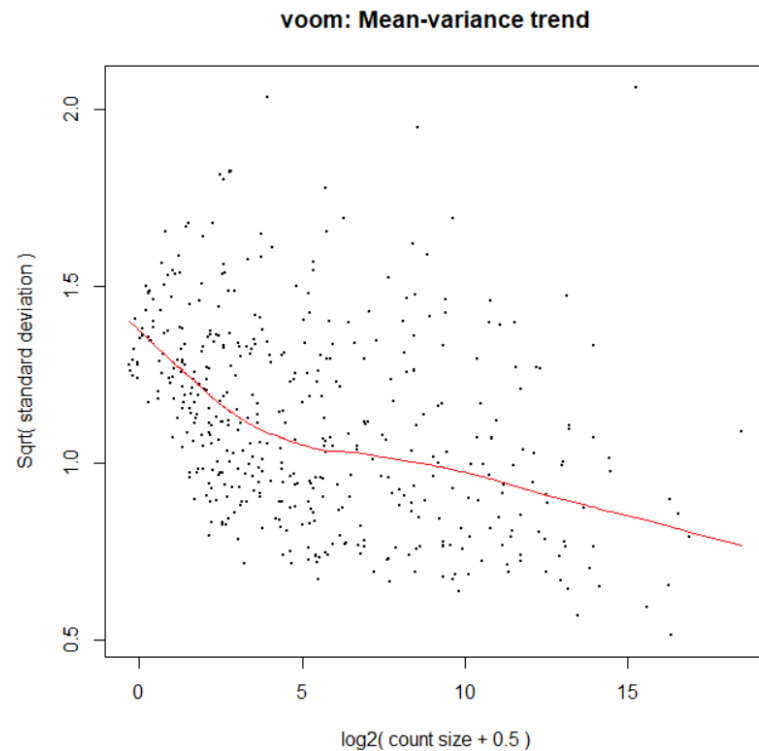
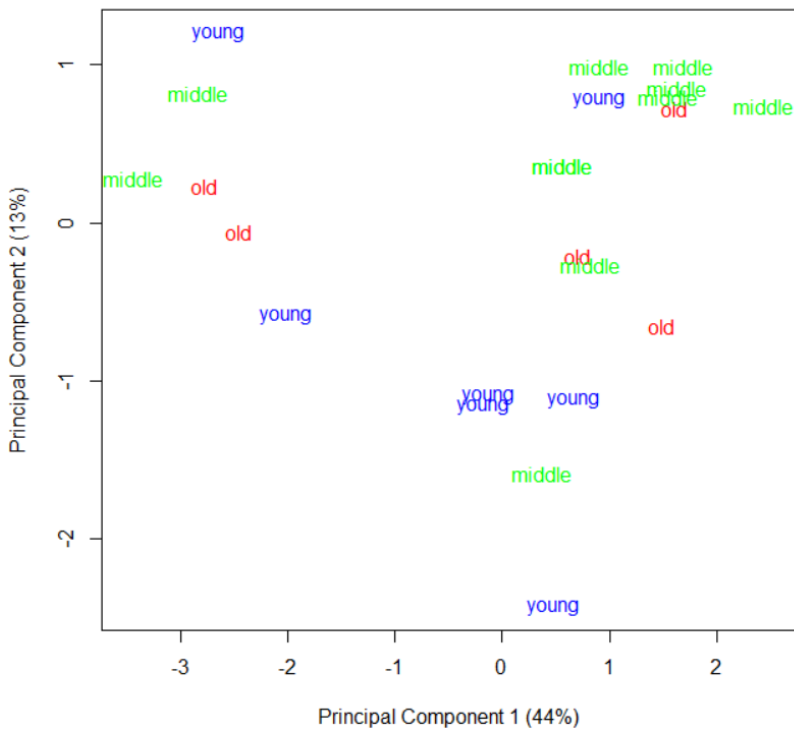
Group <- factor(ages, levels=c("young","middle","old"))
design <- model.matrix(~0 + Group)
colnames(design) <- c("young","middle", "old")
design

colors <- ifelse(ages == "young", "blue", ifelse(ages == "middle", "green", "red"))

y1 <- calcNormFactors(y1)
v1 <- voom(y1,design,plot=TRUE)
plotMDS(v1,top=50,labels=ages, col=colors, gene.selection= "common")

fit1 = lmFit(v1,design)
contrast.matrix <- makeContrasts(young-middle, young-old, middle-old,levels=design)
fit2 <- contrasts.fit(fit1, contrast.matrix)
fit2 <- eBayes(fit2)
```

Najpierw sporządzono design i macierz kontrastu w taki sposób, aby parami można było porównać ze sobą wszystkie grupy wiekowe, a ostatecznie zbadać różnice porównując wszystkie 3 grupy jednocześnie.



```
#porównanie young-middle
top = topTable(fit2, coef=1, n=Inf)
top1 = top[top$adj.P.Val<=0.05,]
top1

#porównanie old-middle
top = topTable(fit2, coef=2, n=Inf)
top1 = top[top$adj.P.Val<=0.05,]
top1

#porównanie middle-old
top = topTable(fit2, coef=3, n=Inf)
top1 = top[top$adj.P.Val<=0.05,]
top1
```

```
> top1
[1] genes logFC AveExpr t P.Value adj.P.Val B
<0 wierszy> (lub 'row.names' o zerowej długości)
> #porównanie young-middle
> top = topTable(fit2, coef=1, n=Inf)
> top1 = top[top$adj.P.Val<=0.05,]
> top1
[1] genes logFC AveExpr t P.Value adj.P.Val B
<0 wierszy> (lub 'row.names' o zerowej długości)
> #porównanie old-middle
> top = topTable(fit2, coef=2, n=Inf)
> top1 = top[top$adj.P.Val<=0.05,]
> top1
[1] genes logFC AveExpr t P.Value adj.P.Val B
<0 wierszy> (lub 'row.names' o zerowej długości)
> #porównanie middle-old
> top = topTable(fit2, coef=3, n=Inf)
> top1 = top[top$adj.P.Val<=0.05,]
> top1
```

```
> summary(results)
              young - middle young - old middle - old
Down              0              0              0
NotSig            450            450            450
Up                0              0              0
```

Nie wykryto żadnych różnic w porównaniach parami między grupami wiekowymi zarówno przy p-value 0.05 jak i 0.01.

Następnie porównano wszystkie grupy jednocześnie:

```
> top = topTable(fit2, number=Inf)
> top1 = top[top$adj.P.Val<=0.05,]
> top1
[1] genes          young...middle young...old   middle...old   AveExpr        F          P.Value
[8] adj.P.Val
<0 wierszy> (lub 'row.names' o zerowej długości)
```

Również nie otrzymano istotnych statystycznie wyników.

- Porównanie young vs kontrola

```
#wiek i kontrola
y1=DGEList(counts=Counts1, genes=rownames(counts1))

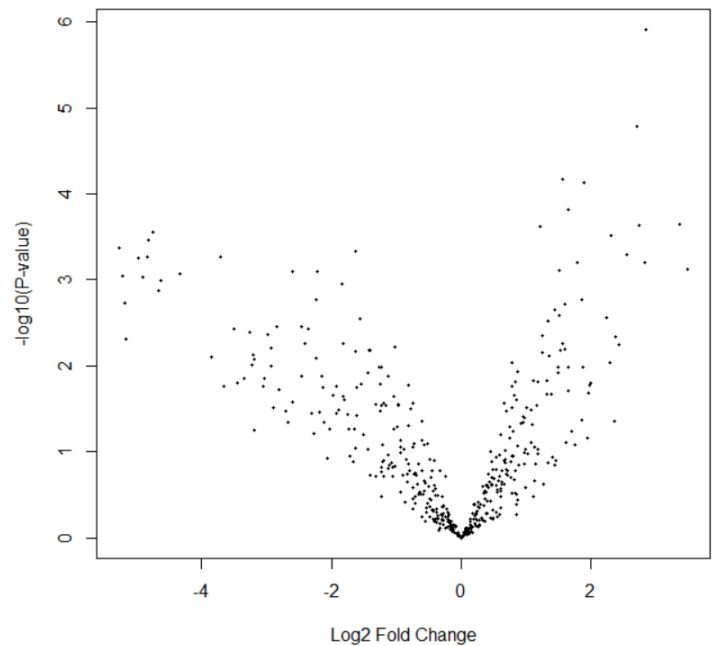
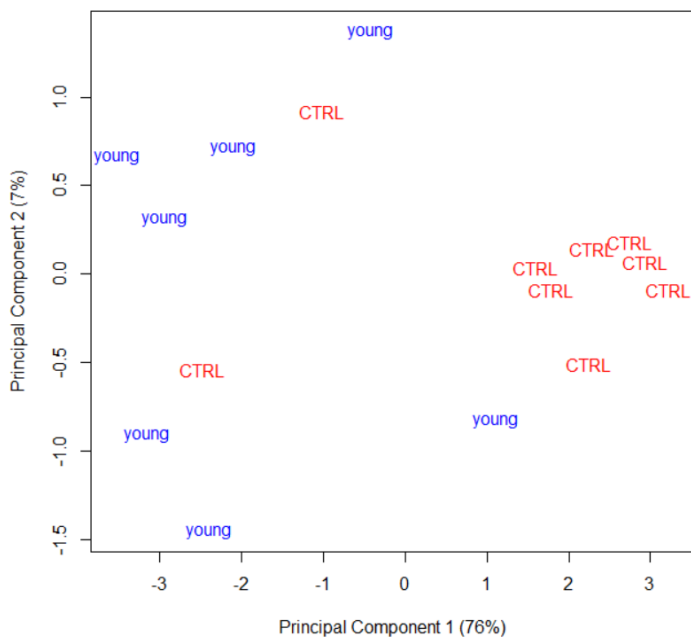
ages <- Homo_eSet$age1
mf = which(ages!='middle' & ages!='old')
y1 <- y1[, mf]
ages <- ages[mf]

Group <- factor(ages, levels=c("CTRL","young"))
design <- model.matrix(~Group)
colnames(design) <- c("CTRL","youngvsCTRL")
design

y1 <- calcNormFactors(y1)
v1 <- voom(y1,design,plot=TRUE)
plotMDS(v1,top=50,labels=ages, col=ifelse(ages=="young", "blue","red"), gene.selection= "common")

### wyniki ###
fit1 = lmFit(v1,design)
fit1 <- eBayes(fit1)
top = topTable(fit1,coef=2,n=Inf)
top1 = top[top$adj.P.Val<=0.01,]
top1

results <- decideTests(fit1, p.value=0.01)
summary(results)
volcanoPlot(fit1, coef=2)
```



```
> top1
      genes      logFC      AveExpr      t      P.Value      adj.P.val      B
hsa-miR-181d hsa-miR-181d 2.848396  7.318754  7.133446  1.235288e-06  0.0005558798  5.595600
hsa-miR-181b-5p hsa-miR-181b-5p 2.710471  12.385616  5.825382  1.642613e-05  0.0036958787  3.156959
hsa-miR-425-5p hsa-miR-425-5p 1.565643  10.500226  5.159695  6.667923e-05  0.0083742348  1.823059
hsa-miR-181c-3p hsa-miR-181c-3p 1.898799  7.715355  5.108386  7.443764e-05  0.0083742348  1.748690
> results <- decideTests(fit1, p.value=0.01)
> summary(results)
      CTRL youngvsCTRL
Down      0           0
NotSig    78         446
Up       372          4
```

Powyżej przedstawię 4 miRNA, które ulegają podwyższonej ekspresji u osobników młodych względem kontroli. (próg pvalue – 0.01).

- Porównanie middle vs kontrola

```
#wiek i kontrola
y1=DGELIST(counts=Counts1, genes=rownames(Counts1))

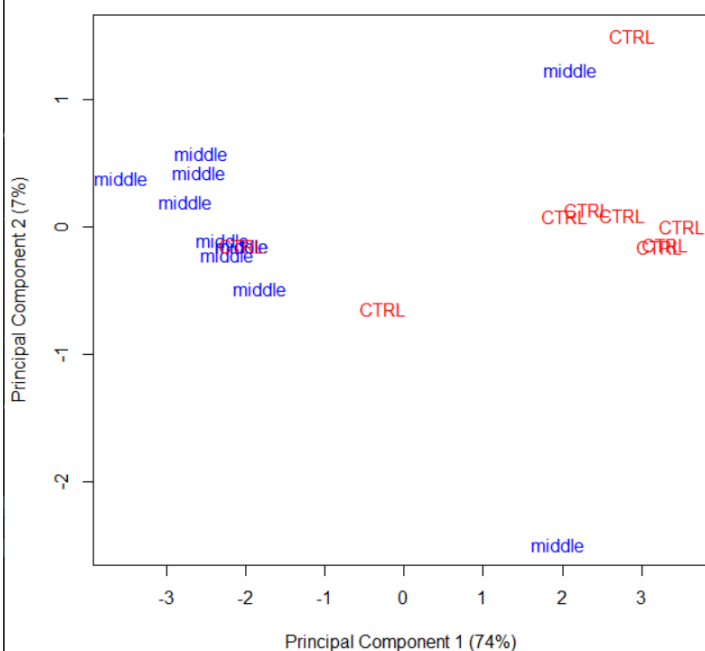
ages <- Homo_eSet$age1
mf = which(ages!='young' & ages!='old')
y1 <- y1[, mf]
ages <- ages[mf]

Group <- factor(ages, levels=c("CTRL","middle"))
design <- model.matrix(~Group)
colnames(design) <- c("CTRL","middlevsCTRL")
design

y1 <- calcNormFactors(y1)
v1 <- voom(y1,design,plot=TRUE)
plotMDS(v1,top=50,labels=ages, col=ifelse(ages=="middle", "blue","red"), gene.selection= "common")

### wyniki ###
fit1 = lmFit(v1,design)
fit1 <- eBayes(fit1)
top = topTable(fit1,coef=2,n=Inf)
top1 = top[top$adj.P.Val<=0.01,]
top1

results <- decideTests(fit1, p.value=0.01)
summary(results)
volcanoPlot(fit1, coef=2)
```



```
> summary(results)
      CTRL middlevsCTRL
Down      0           37
NotSig    85          376
Up       365          37
```

Mamy 74 miRNA ulegające różnicowej ekspresji przy pvalue 0.01. Dużo więcej niż w przypadku porównania young-CTRL, mimo że porównanie young-middle nie pokazało żadnych różnic.

- Porównanie old vs kontrola

```
#wiek i kontrola
y1=DGEList(counts=Counts1, genes=rownames(Counts1))

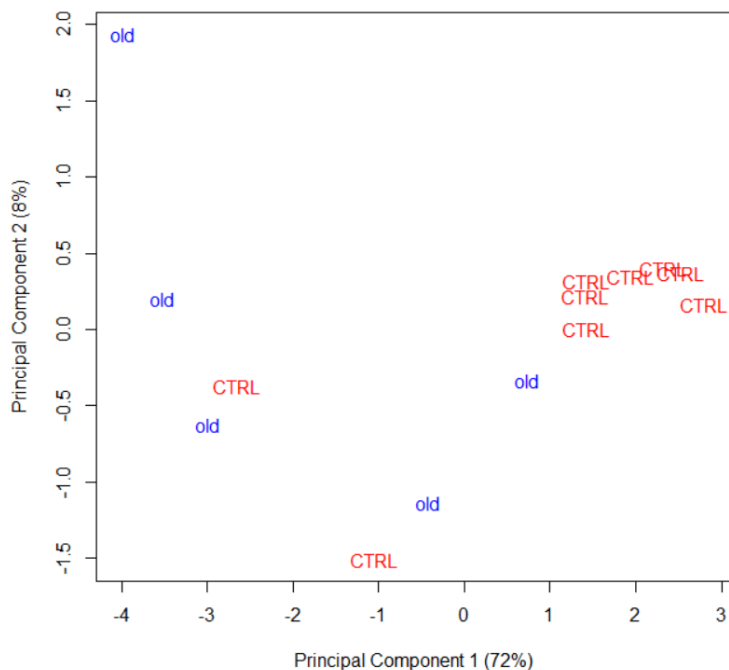
ages <- Homo_eSet$age1
mf = which(ages!='young' & ages!='middle')
y1 <- y1[, mf]
ages <- ages[mf]

Group <- factor(ages, levels=c("CTRL","old"))
design <- model.matrix(~Group)
colnames(design) <- c("CTRL","oldvsCTRL")
design

y1 <- calcNormFactors(y1)
v1 <- voom(y1,design,plot=TRUE)
plotMDS(v1,top=50,labels=ages, col=ifelse(ages=="old", "blue","red"), gene.selection= "common")

### wyniki ###
fit1 = lmFit(v1,design)
fit1 <- eBayes(fit1)
top = topTable(fit1,coef=2,n=Inf)
top1 = top[top$adj.P.Val<=0.01,]
top1

results <- decideTests(fit1, p.value=0.01)
summary(results)
volcanoplot(fit1, coef=2)
```



```
> fit1 <- lmFit(v1,design)
> fit1 <- eBayes(fit1)
> top = topTable(fit1,coef=2,n=Inf)
> top1 = top[top$adj.P.Val<=0.01,]
> top1
[1] genes logFC AveExpr t P.value adj.P.val B
<0 wierszy> (lub 'row.names' o zerowej długości)
> results <- decideTests(fit1, p.value=0.01)
> summary(results)
      CTRL oldvsCTRL
Down    0          0
NotSig  68        450
Up      382         0
> volcanoplot(fit1, coef=2)
> top = topTable(fit1,coef=2,n=Inf)
> top1 = top[top$adj.P.Val<=0.05,]
> top1
      genes logFC AveExpr t P.value adj.P.val B
hsa-miR-19b-3p hsa-miR-19b-3p 2.011566 10.87661 5.158582 0.0001038768 0.04674457 1.487281
>
```

```
> summary(results)
      CTRL oldvsCTRL
Down    1          0
NotSig  38        449
Up      411         1
```

W tym porównaniu nie otrzymano żadnych miRNA przy progu pvalue 0.01. Po zwiększeniu do 0.05 otrzymano jeden miRNA przedstawiony powyżej. Ulegał podwyższonej ekspresji względem kontroli.