

# Reikšmingai skirtingos raiškos genų paieška tarp plaučių vėžio tipų

*Mantas Bernotas*

Šio darbo tikslas - rasti reikšmingai skirtingos raiškos genus tarp dviejų plaučių vėžio tipų. Naudojami duomenys yra “cel” formato failai, su vėžinių ląstelių pavyzdžių genų raiškos profiliais. Prieš analizuojant duomenis, jie yra apdorojami. Tada patikrinama mikrogardelių kokybė. Galiausiai duomenys analizuojami ir reikšmingai skirtingos raiškos, tarp dviejų plaučių vėžio tipų, genai vizualizuojami vulkano tipo grafikuose. Taip pat sugeneruojami “csv” failai (išsaugomi aplanke “All”), kuriuose pateikiami reikšmingai skirtingos raiškos genai su jiems apskaičiuotais statistiniais parametrais.

Importuojamos reikalingos Bioconductor bibliotekos

```
library(affy)
library(simpleaffy)
library(RColorBrewer)
library(affyPLM)
library(limma)
library(hgu133plus2.db)
library(annotate)
require(ggplot2)
```

Nuskaitomi CEL failai is “celfiles” aplanko, pagal “annotation.txt” faile nurodytą tvarką.

##		Name	FileName	Factors
## 1	GSM1065772.CEL	GSM1065772.CEL	ACs1	
## 2	GSM1065774.CEL	GSM1065774.CEL	ACs1	
## 3	GSM1065776.CEL	GSM1065776.CEL	ACs1	
## 4	GSM1065778.CEL	GSM1065778.CEL	ACs1	
## 5	GSM1065783.CEL	GSM1065783.CEL	ACs1	
## 6	GSM1065784.CEL	GSM1065784.CEL	ACs1	
## 7	GSM1065788.CEL	GSM1065788.CEL	ACs1	
## 8	GSM1065789.CEL	GSM1065789.CEL	ACs1	
## 9	GSM1065793.CEL	GSM1065793.CEL	ACs1	
## 10	GSM1065797.CEL	GSM1065797.CEL	ACs1	
## 11	GSM1065765.CEL	GSM1065765.CEL	ACs2	
## 12	GSM1065767.CEL	GSM1065767.CEL	ACs2	
## 13	GSM1065769.CEL	GSM1065769.CEL	ACs2	
## 14	GSM1065770.CEL	GSM1065770.CEL	ACs2	
## 15	GSM1065777.CEL	GSM1065777.CEL	ACs2	
## 16	GSM1065780.CEL	GSM1065780.CEL	ACs2	
## 17	GSM1065781.CEL	GSM1065781.CEL	ACs2	
## 18	GSM1065785.CEL	GSM1065785.CEL	ACs2	
## 19	GSM1065791.CEL	GSM1065791.CEL	ACs2	
## 20	GSM1065798.CEL	GSM1065798.CEL	ACs2	
## 21	GSM1065739.CEL	GSM1065739.CEL	SCCs1	
## 22	GSM1065740.CEL	GSM1065740.CEL	SCCs1	
## 23	GSM1065743.CEL	GSM1065743.CEL	SCCs1	
## 24	GSM1065756.CEL	GSM1065756.CEL	SCCs1	
## 25	GSM1065758.CEL	GSM1065758.CEL	SCCs1	
## 26	GSM1065766.CEL	GSM1065766.CEL	SCCs1	

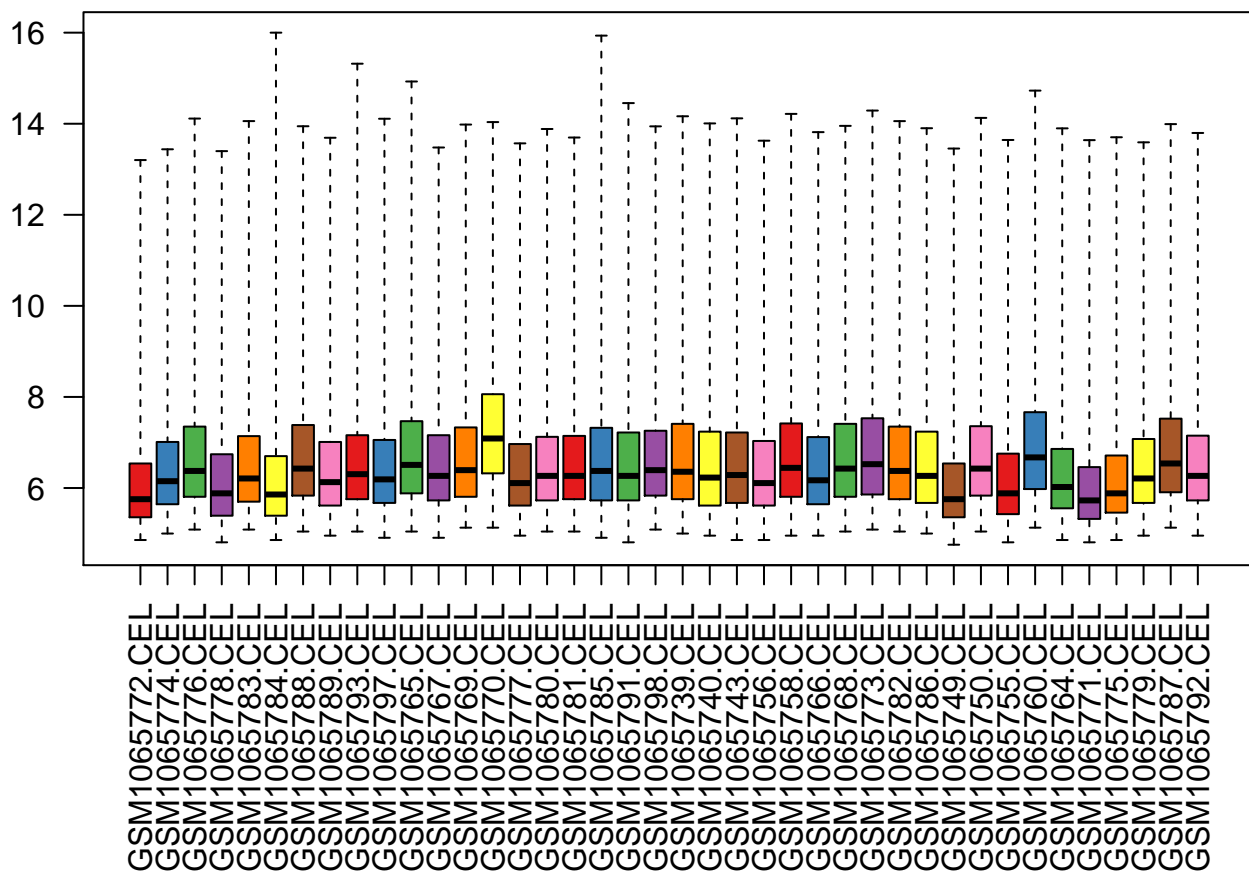
```

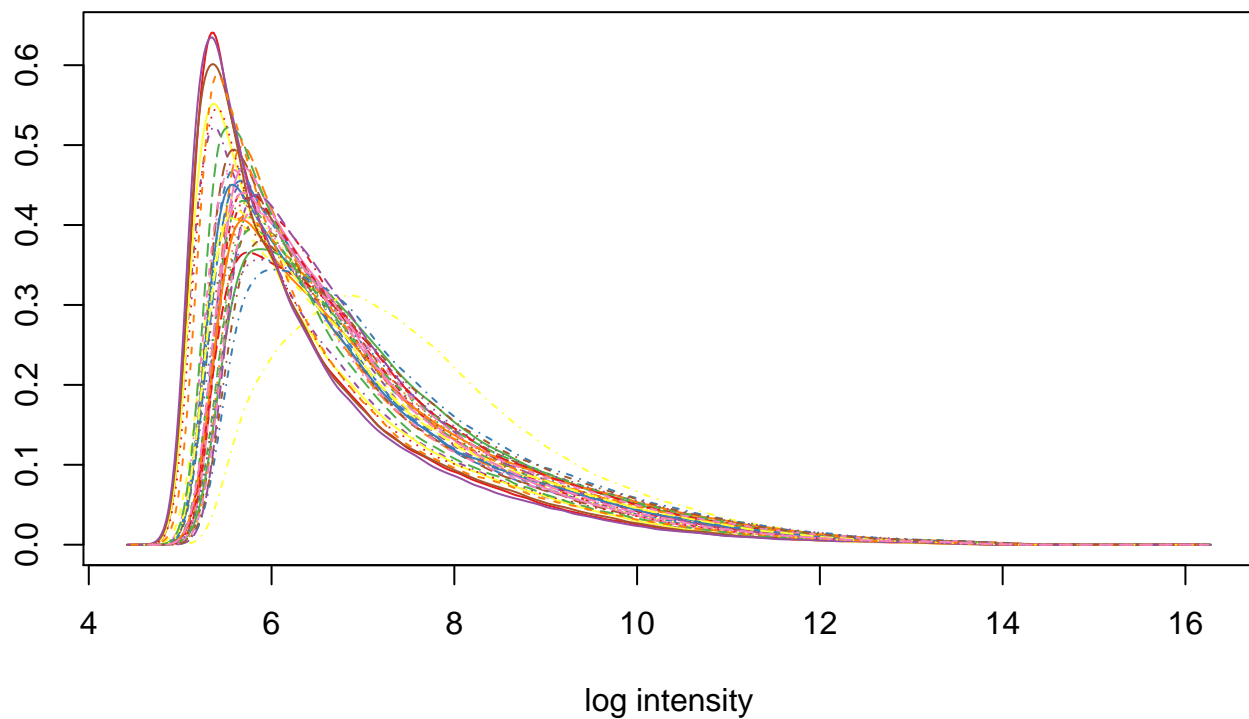
## 27 GSM1065768.CEL GSM1065768.CEL SCCs1
## 28 GSM1065773.CEL GSM1065773.CEL SCCs1
## 29 GSM1065782.CEL GSM1065782.CEL SCCs1
## 30 GSM1065786.CEL GSM1065786.CEL SCCs1
## 31 GSM1065749.CEL GSM1065749.CEL SCCs2
## 32 GSM1065750.CEL GSM1065750.CEL SCCs2
## 33 GSM1065755.CEL GSM1065755.CEL SCCs2
## 34 GSM1065760.CEL GSM1065760.CEL SCCs2
## 35 GSM1065764.CEL GSM1065764.CEL SCCs2
## 36 GSM1065771.CEL GSM1065771.CEL SCCs2
## 37 GSM1065775.CEL GSM1065775.CEL SCCs2
## 38 GSM1065779.CEL GSM1065779.CEL SCCs2
## 39 GSM1065787.CEL GSM1065787.CEL SCCs2
## 40 GSM1065792.CEL GSM1065792.CEL SCCs2

```

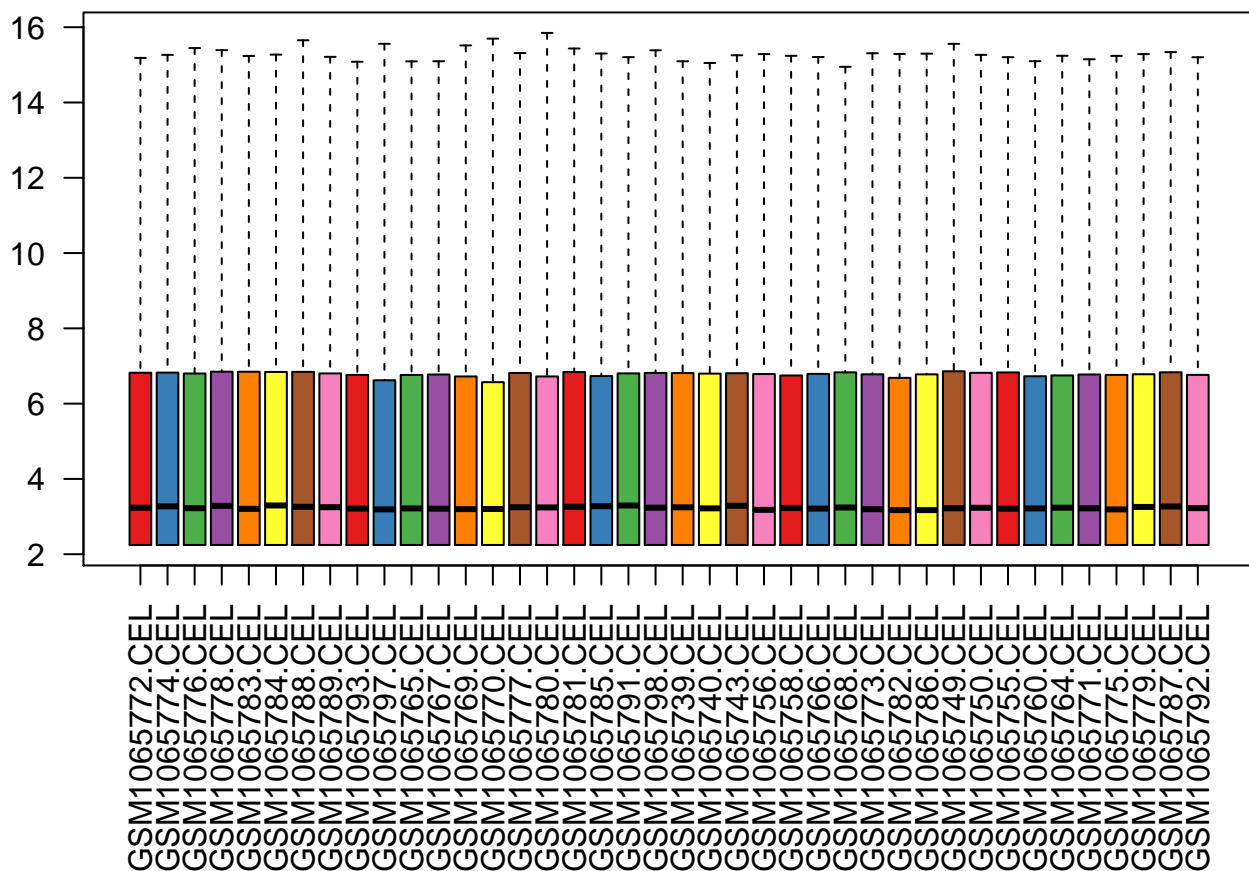
Apdorojami duomenys

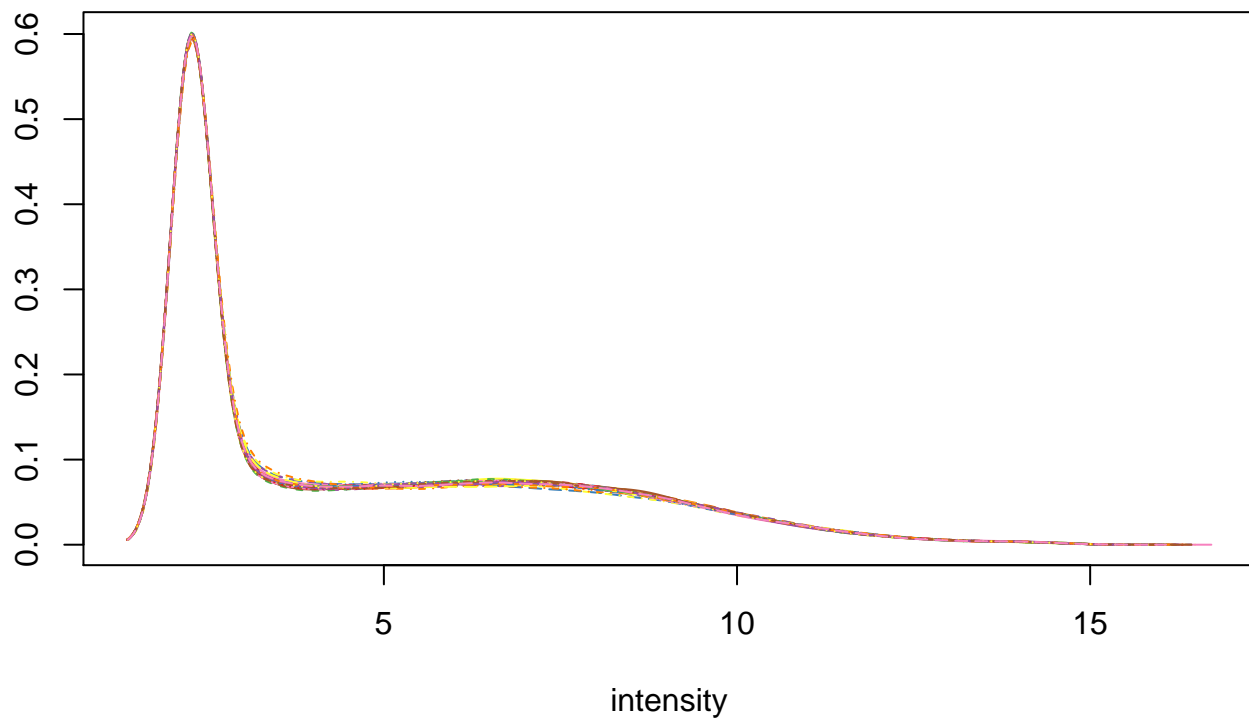
Pateikiami duomenys prieš normalizaciją “Boxplot” ir logaritminio intensyvumo tankio grafikais.





Toliau pateikiami duomenys po normalizacijos “Boxplot” ir logaritminio intensyvumo tankio grafikais.

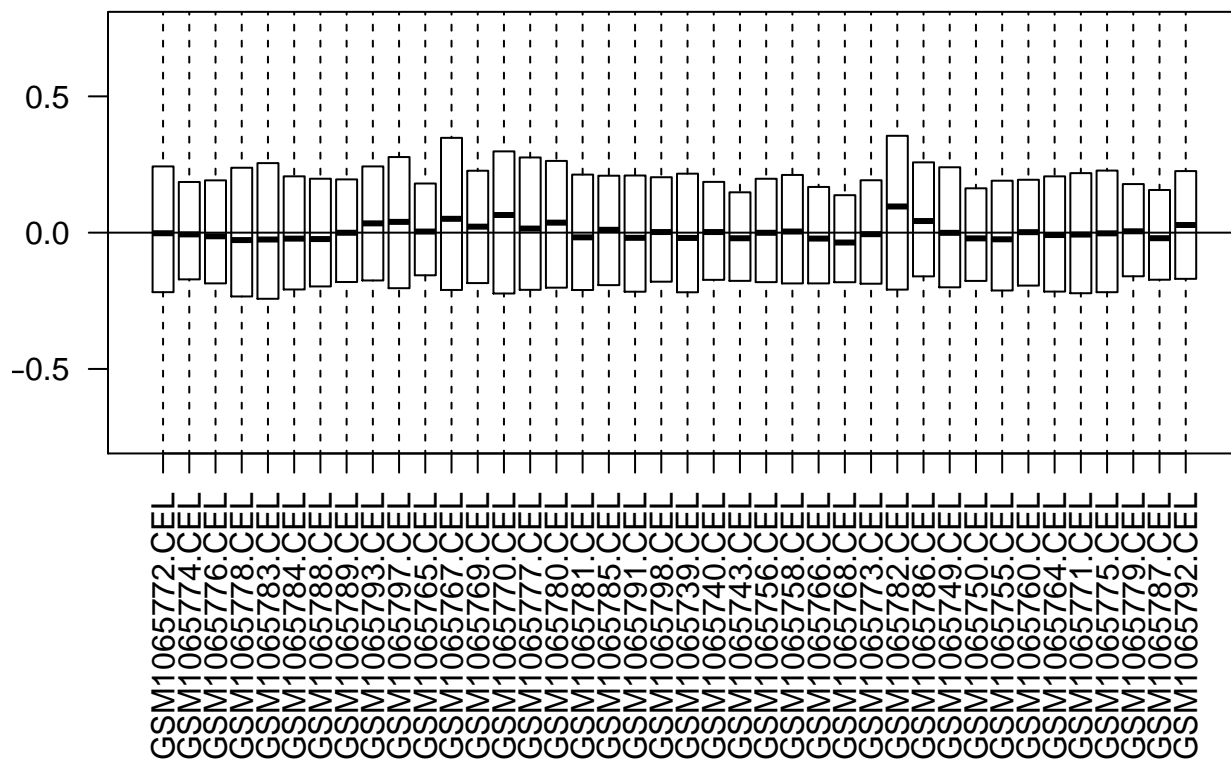




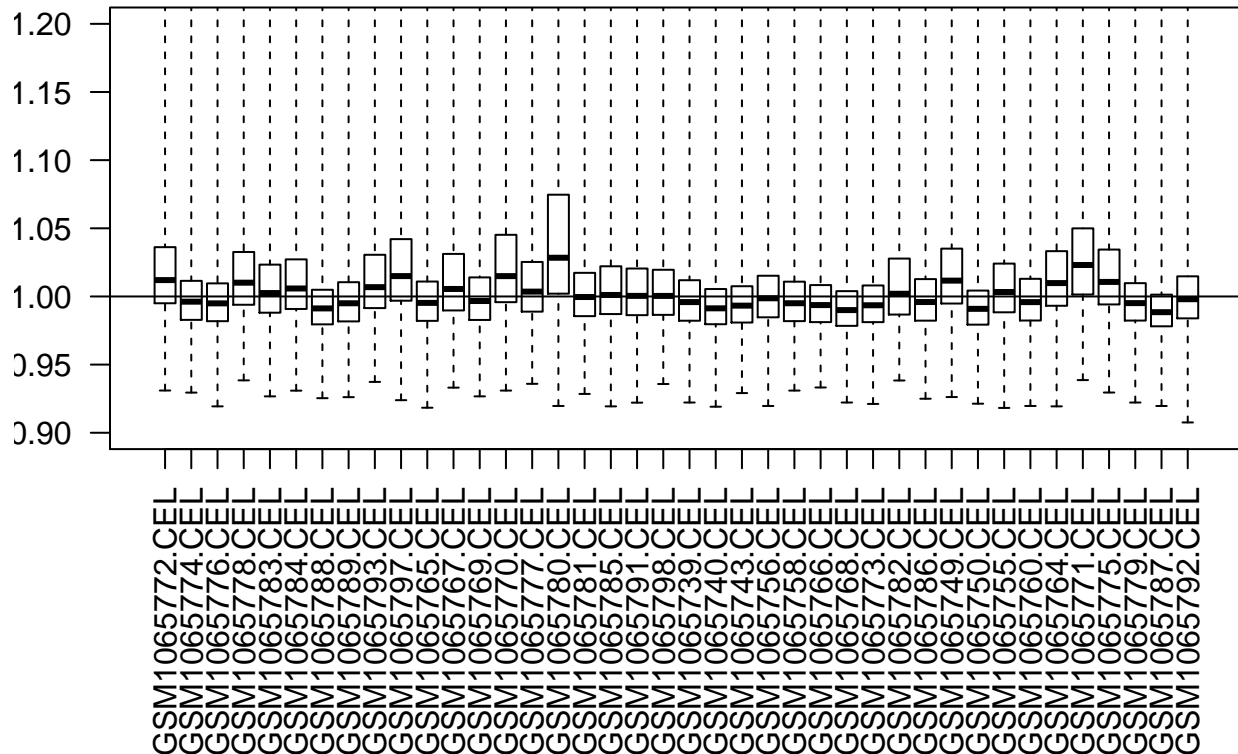
Nustatinėjama mikrogardelių kokybė

Mikrogardelių kokybės įvertinimui pavaizduojami visų analizuojamų mikrogardelių RLE ir NUSE grafikai.

## RLE



## NUSE

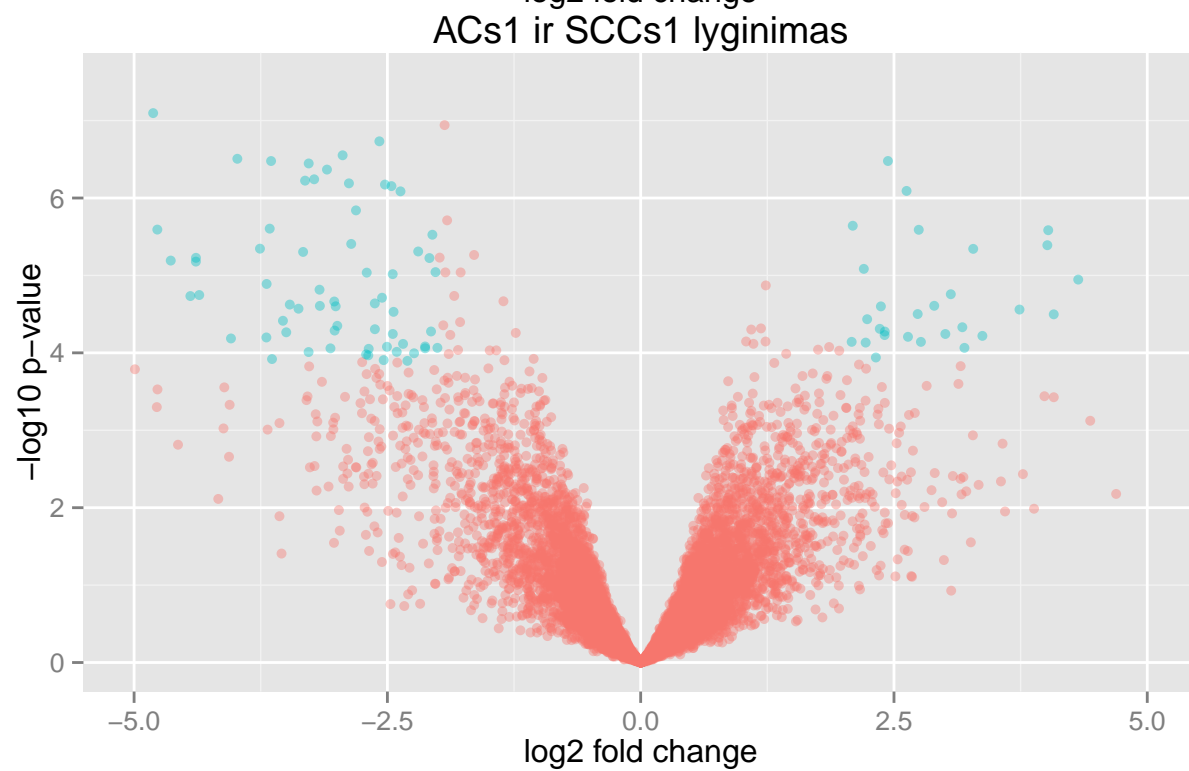
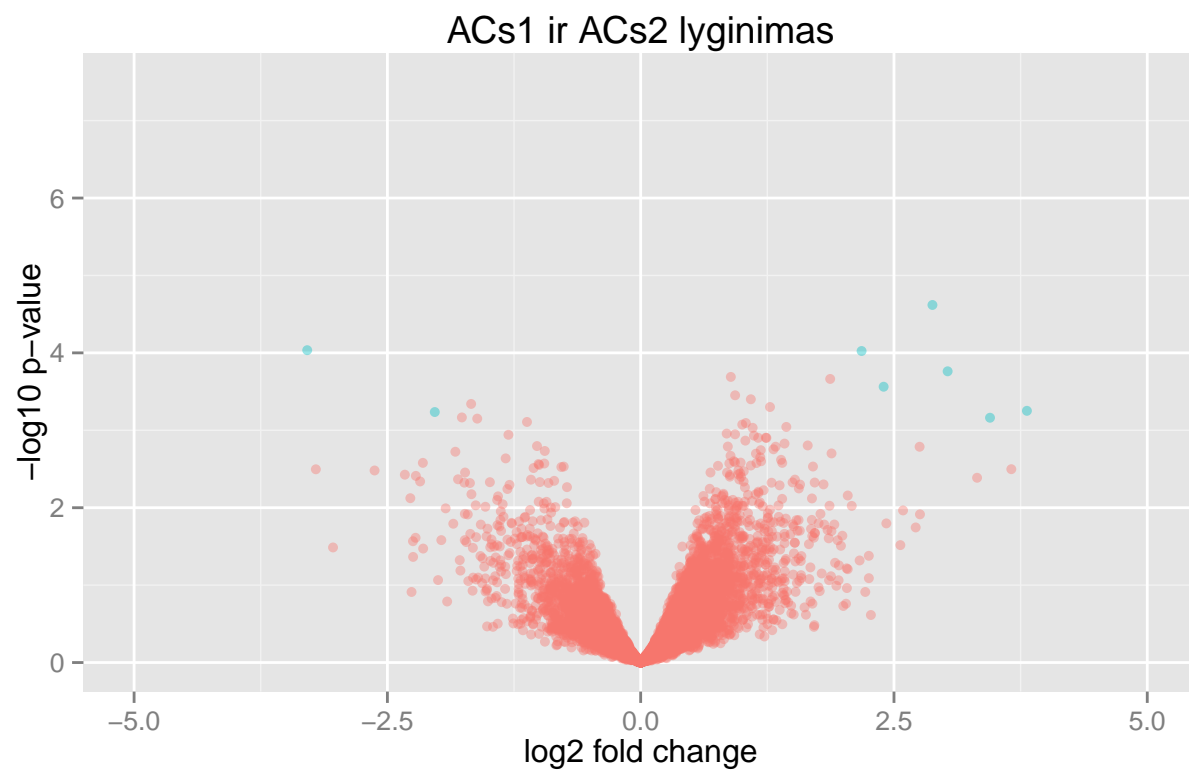


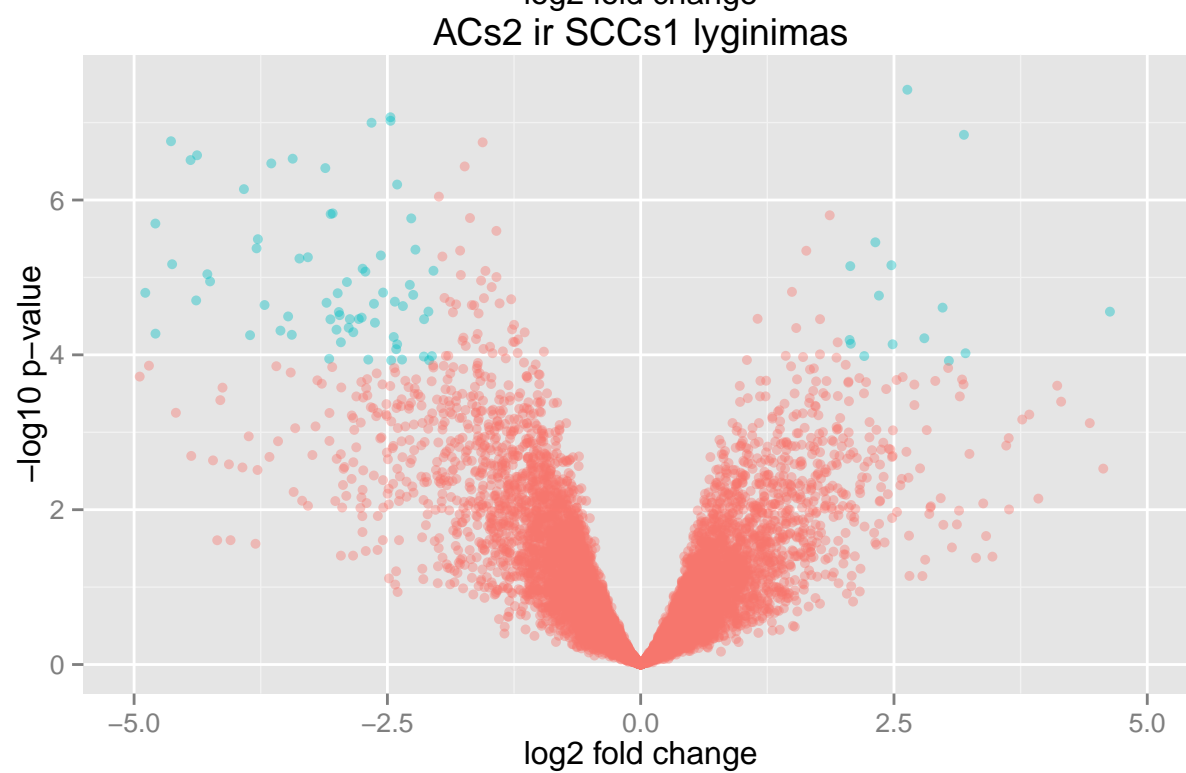
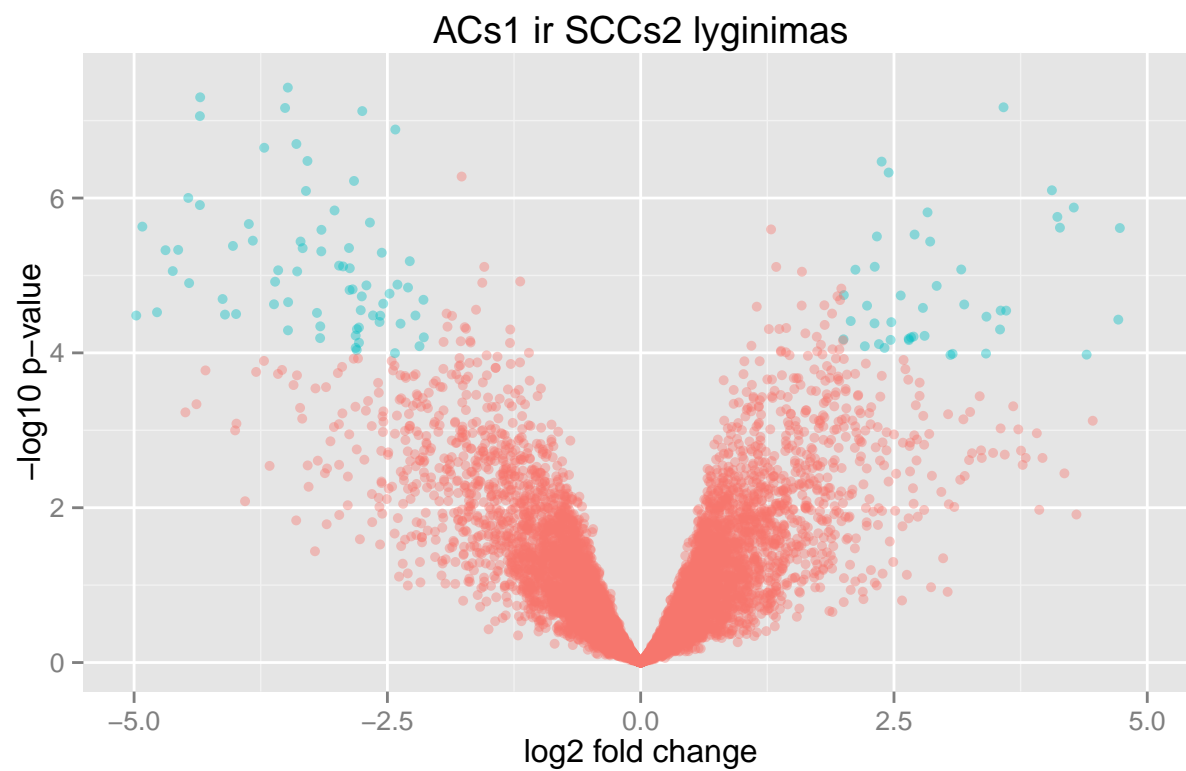
Iš RLE grafiko matome, kad „GSM1065782.CEL“ stulpelio centras ganėtinai nutolęs nuo 0, ir galbūt jį reikėtų pakeisti kitu failu. Tačiau šis nuokrypis nėra toks reikšmingas, kad įtakotus rezultatus, todėl paliekame tą patį „cel“ failą.

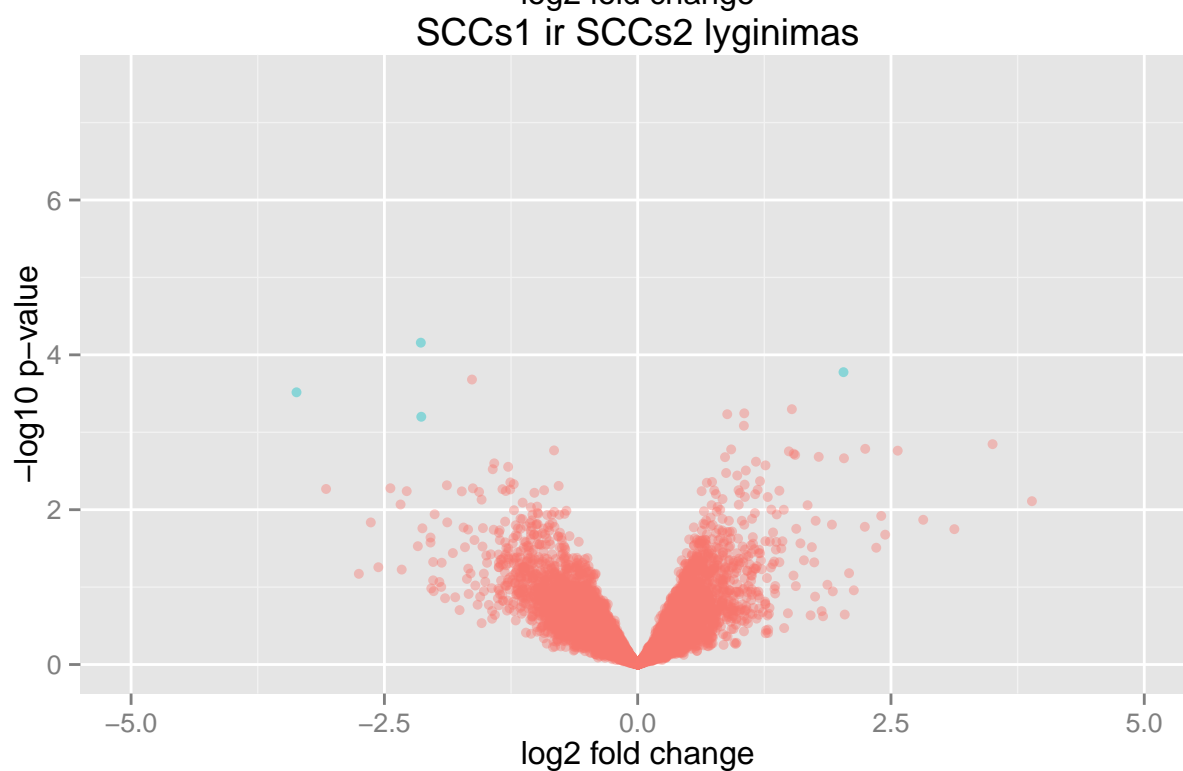
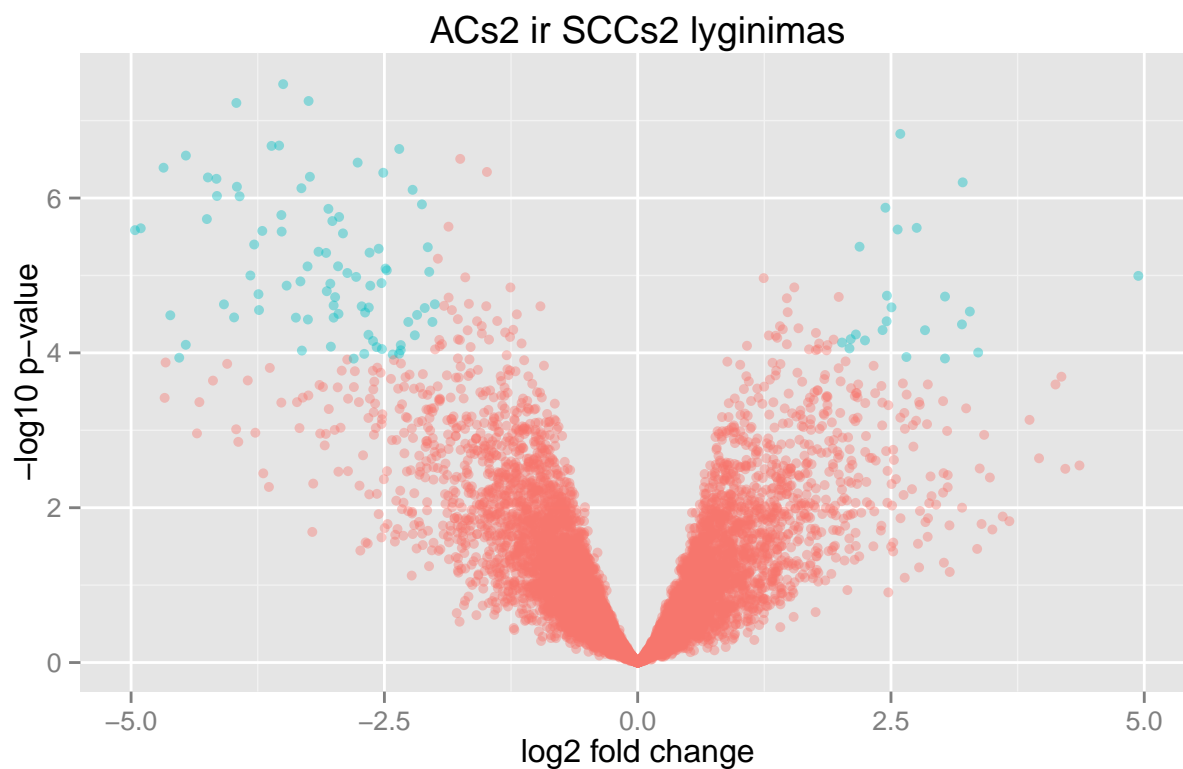
Iš NUSE grafiko galima matyti, kad nėra „cel“ failo, kurio stulpelio centras viršytų 1.1 ribą. Taigi galima teigti, kad pasirinktos mikrogardelės yra geros kokybės ir tolimesni rezultatai, nebus iškreipti dėl kokybės problemų.

Reikšmingai skirtingos raiškos genai pateikiami vulkano grafikais

Vulkano grafikas yra taškinis grafikas, kuris dažnai naudojamas mikrogardelių duomenų rinkinių analizei, svarbių genų peržiūrai. Grafiką sudaro ant x ašies vaizduojama geno logaritminė pokyčio reikšmingumo slenksčio reikšmė ir ant y ašies neigiamas dešimtainis P-reikšmės logaritmas. Vaizduojami biologiškai ir statistiškai reikšmingi genai. Biologinė įtaka pokyčiui stebima ant x ašies, y ašis rodo statistinį pokyčio patikimumą. Iš tokių grafikų galima atrinkti svarbiausius atvejus tolimesniems tyrimams. Toliau pateikiami visų lyginimų vulkano grafikai. Vulkano tipo grafike išskiriami genai, kurių logaritminės pokyčio reikšmingumo slenksčio reikšmės modulis yra didesnis už 2 ir kurių P-reikšmė yra mažesnė už 0.05 ir genų, kurių logaritminės pokyčio reikšmingumo slenksčio reikšmės modulis yra didesnis už 2, skaičiaus, santykį.







Iš vulkano grafikų taip pat galime pastebėti, kad daug daugiau reikšmingai skirtingos raiškos genų yra tarp AC ir SCC tipų lyginimų, nei tarp AC ir AC arba SCC ir SCC tipų lyginimų. Tai rodo, kad lyginant tos pačios plaučių vėžio rūšies tačiau skirtingų vėžio stadijų pavyzdžius yra randama gerokai mažiau reikšmingai skirtingos raiškos genų, negu lyginant pavyzdžius tarp skirtingų plaučių vėžio rūšių.