

Su GBM asocijuojamų miRNA paieška

Dovydas Kičiatovas

2018-04-12

Turinys

1. Duomenys
2. Įrankiai
3. GBM klasifikavimas
4. Terapijų grupės
5. Statistiniai testai
6. Papildomi darbai
7. Galutinis miRNA sąrašas

1. Duomenys

- miRNA ekspresijos duomenys – iš **TCGA Legacy** duomenų bazės - <https://portal.gdc.cancer.gov/legacy-archive/>
- miRNA skaičius – 534
- Pacientų skaičius – 495
- Taip pat 10-ies kontrolinių pacientų duomenys.
- Duomenys nestandartizuoti - ekspresijos reikšmės intervale (5:15)

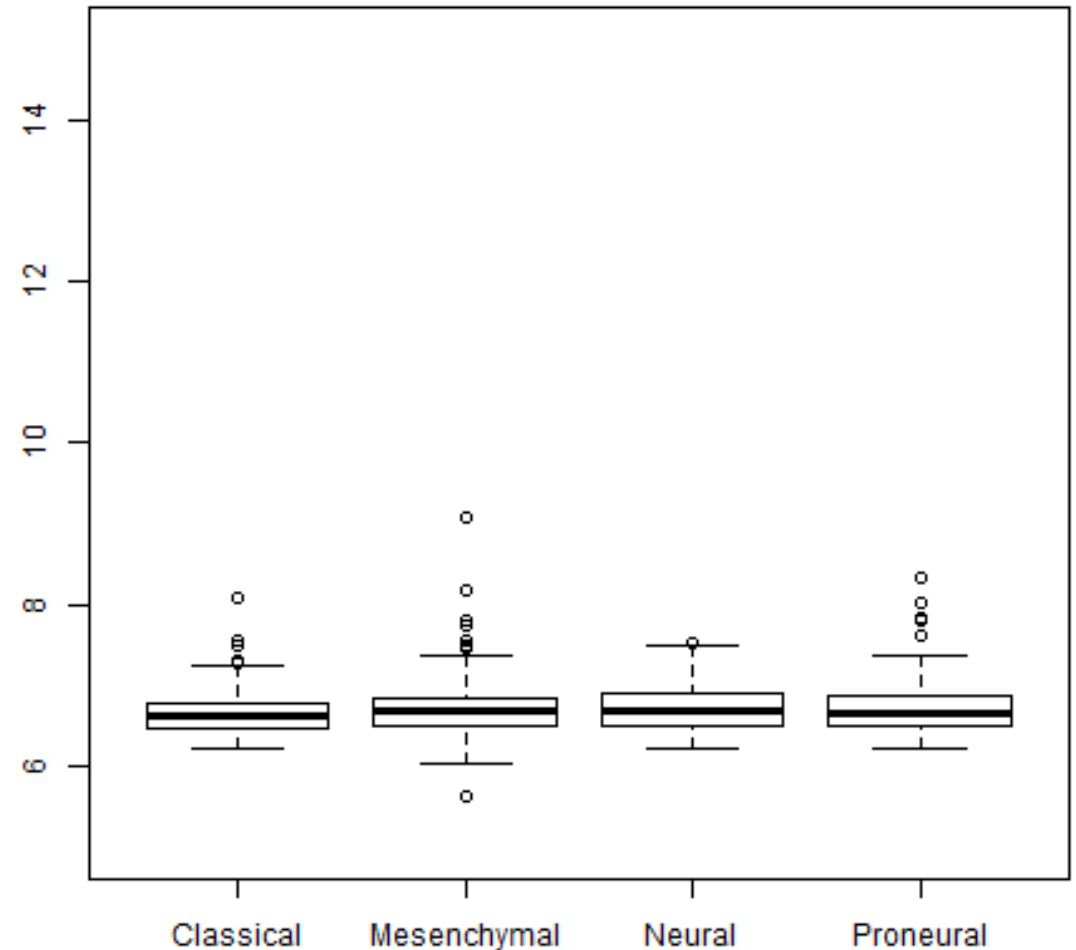
2. Įrankiai

- Didžioji dalis darbo atlikta naudojant statistinių skaičiavimų paketą **R** - <https://www.r-project.org/>
- **R** daugiausiai naudotas kartu su vienu pagrindinių šiuolaikinio bioinformatiko įrankių – R papildomu paketu **Bioconductor** – <https://bioconductor.org/>
- Šis paketas skirtas įvairiausių biologinių duomenų paruošimui, analizei bei atvaizdavimui.



3. GBM klasifikavimas

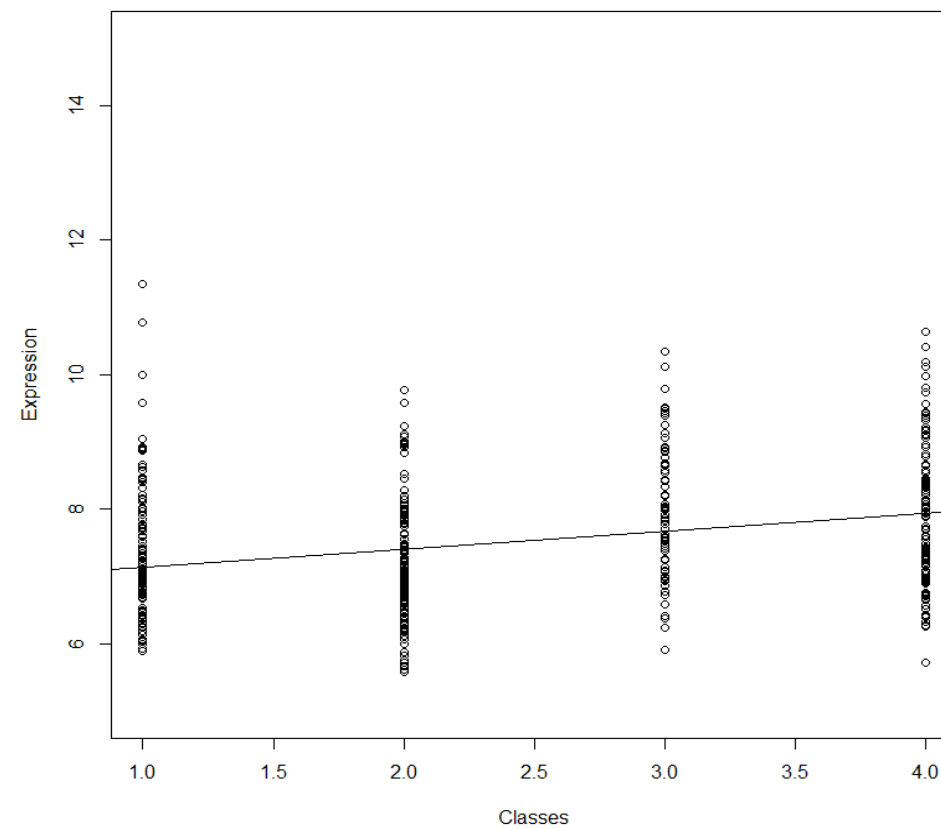
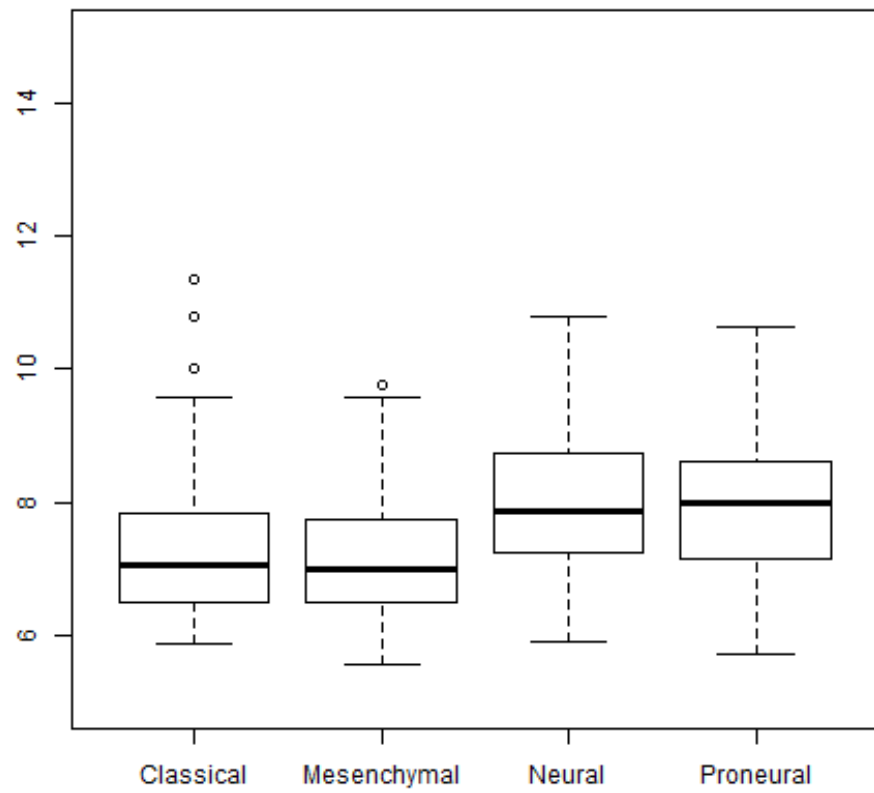
- Pirma užduotis – sumažinti miRNA skaičių tolimesniems tyrimams.
- Didelė dalis miRNA iš turimų duomenų neturi įtakos GBM klasifikavimui (ekspresijos lygiai per visus 4 subtipus vienodi)
- Dešinėje – hsa-miR-362 ekspresijos boxplot



3. GBM klasifikavimas

- miRNA skaičių sumažinau pasinaudodamas tiesine regresija.
- Į 4 subtipus sugrupuotoms ekspresijoms pritaikomas tiesinis modelis, kuriam galima taikyti ANOVA (dispersijos analizę), kuri leis įvertinti, ar stebima prasminga tiesinė regresija.
- Taip pat reikia nepamiršti grupių perstatų.

3. GBM klasifikavimas

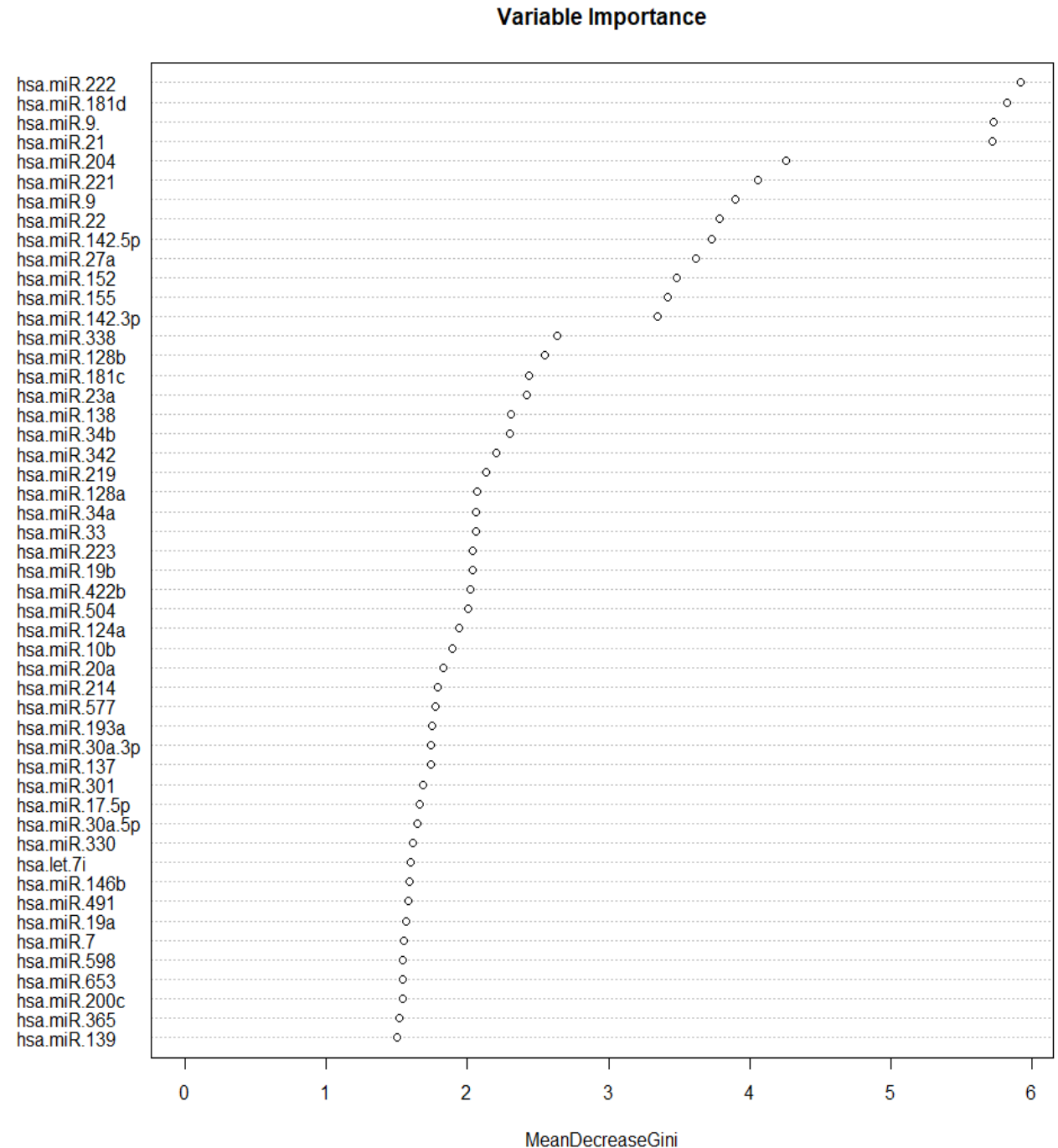


3. GBM klasifikavimas

- Po tiesinės regresijos analizės gautas 150 miRNA sąrašas, kuriame esančiose miRNA tarp GBM subtipų stebima tiesinė regresija ($Q = 0,95$).
- Pacientai klasifikuojami naudojant **RandomForest**, tikslumas ~72 proc. (100 bandymų vidurkis).
- Modelio treniravimo ir testavimo proporcijos – 80 proc. pacientų apmokymui ir 20 proc. testavimui.
- Sąrašą dar galima mažinti – galbūt koks nors mažiau nei 150 miRNA derinys klasifikuos taip pat tiksliai (gal net tiksliau)?

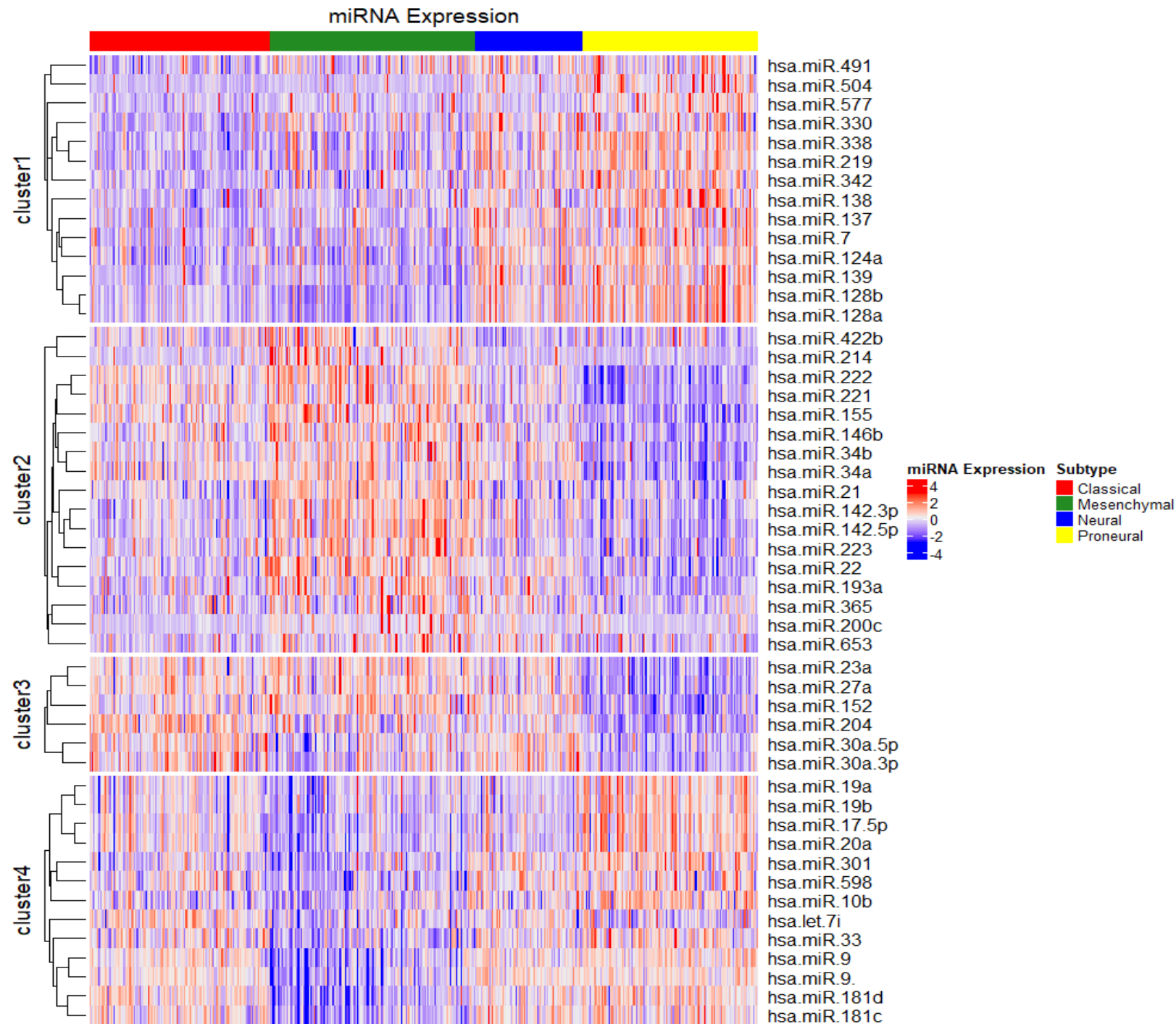
Variable Importance

- Vykdoma kintamųjų (miRNA) perstata ir skaičiuojamas naujo medžio Gini indeksas: svarbūs kintamieji indeksą stipriai sumažins, mažiau svarbūs – indeksas pakis nedaug.
- Grafike – Gini indekso sumažėjimo vidurkis per medžius, pakeitus (pašalinus) atitinkamą miRNA.



Heatmap

- Šių 50-ies miRNA klasifikatoriaus tikslumas ~74 proc. (taip pat 100 bandymų).
- Dešinėje – šio sąrašo intensyvumo žemėlapis (eilučių klasteriai pagal k-means algoritmą, $k = 4$)



4. Terapijų grupės

- Pacientus nutarta skirstyti į 3 terapijų grupes:
 1. Radioterapija (apjungtos standartinės radiacijos, nestandartinės radiacijos ir nenurodytos radiacijos grupės), pacientų sk. = 129
 2. Radioterapija kartu su temozolomido (TMZ) terapija, pacientų sk. = 72
 3. TMZ terapija, pacientų sk. = 191
- GBM klasifikavimas, remiantis ankstesniu 50-ies miRNA sąrašu, tarp grupių skiriasi:
 - Radioterapija – 72 proc.
 - Radioterapija/TMZ – 76 proc.
 - TMZ – 75 proc.

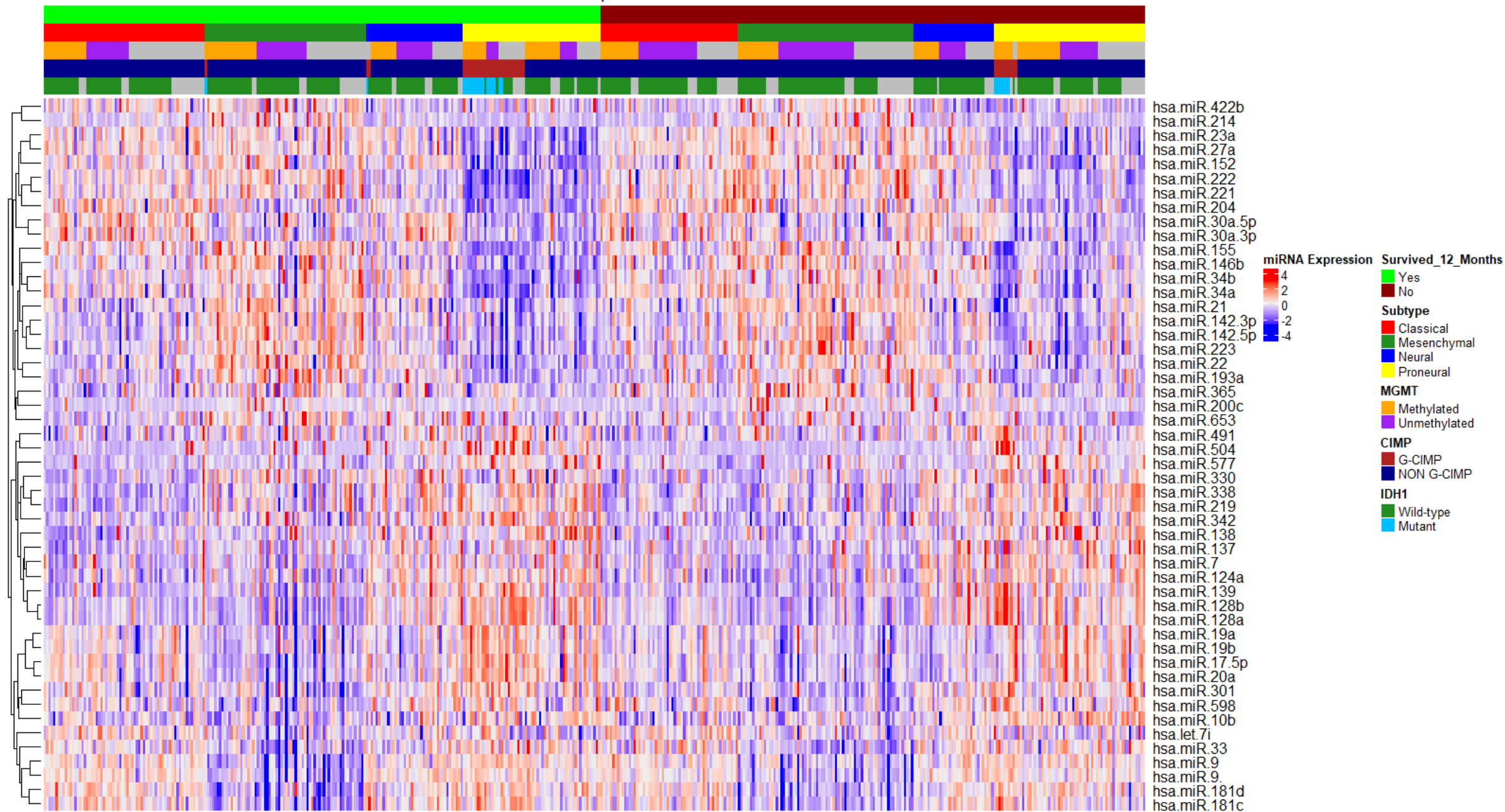
4. Terapijų grupės

- Terapijų grupės papildomai skaidomos į 2 išgyvenamumo grupes:
 1. Išgyvenusieji daugiau (arba lygiai) 12 mėnesių;
 2. Neišgyvenusieji 12 mėnesių.
- Grupių dydžiai:
 - Visi pacientai: **235** ir **230** (51 ir 49 proc.)
 - Radioterapija: **35** ir **94** (27 ir 73 proc.)
 - Radioterapija/TMZ: **51** ir **21** (71 ir 29 proc.)
 - TMZ: **113** ir **78** (59 ir 41 proc.)

4. Terapijų grupės

- Išgyvenamumo grupės klasifikuojamos **RandomForest** (kaip ir GBM subtipai):
 - Bendras tikslumas (be terapijos grupių): 57 proc.
 - Radioterapija: 72 proc.
 - Radioterapija/TMZ: 72 proc.
 - TMZ: 61 proc.

miRNA Expression



5. Statistiniai testai

- Reikia nustatyti, kurių miRNA ekspresijos profilis tarp išgyvenamumo grupių skiriasi.
- Panaudojau Mann-Whitney U testą ($Q = 0,95$) kiekvienai miRNA kiekvienam GBM subtipo, tikrinantį, ar reikšmių pasiskirstymas tarp dviejų populiacijų skiriasi;
- Pagal šį testą, šių miRNA pasiskirstymas išgyvenamumo grupėse (neatsižvelgiant į terapiją) turi prasmingus skirtumus:
hsa-miR-106b, hsa-miR-130b, hsa-miR-148a, hsa-miR-17-3p, hsa-miR-182, hsa-miR-183, hsa-miR-204, hsa-miR-221, hsa-miR-222, hsa-miR-29c, hsa-miR-30c, hsa-miR-34a, hsa-miR-34b, hsa-miR-9, hsa-miR-92, hsa-miR-93

5. Statistiniai testai

- Terapijos grupėse gauti tokie rezultatai:
 - Radioterapijos grupėje:
hsa-miR-335, hsa-miR-138, hsa-miR-99a, hsa-miR-148a
 - Radioterapijos/TMZ grupėje:
hsa-miR-10b, hsa-miR-142-3p, hsa-miR-142-5p, hsa-miR-22, hsa-miR-23b, hsa-miR-92b
 - TMZ grupėje:
hsa-miR-124a, hsa-miR-182, hsa-miR-183, hsa-miR-18a, hsa-miR-222, hsa-miR-223, hsa-miR-93
- Šios miRNA savo atitinkamose grupėse išlaiko klasifikavimo tikslumą.
- Įtraukus pacientus su pasikartojančia GBM (*recurrent*) rezultatai pasikeičia.

6. Papildomi darbai

- Be šių metodų, taikyti kiti:
 - PCA analizė
 - MiRaGE – miRNA kontrolės inferencija iš genų ekspresijos
 - *k-means* algoritmas (kartu su intensyvumo žemėlapiais)

7. Galutinis miRNA sąrašas

- hsa-miR-106b
- hsa-miR-10b
- hsa-miR-124a
- hsa-miR-130b
- hsa-miR-138
- hsa-miR-142-3p
- hsa-miR-142-5p
- hsa-miR-148a
- hsa-miR-17-3p
- hsa-miR-182
- hsa-miR-183
- hsa-miR-18a
- hsa-miR-204
- hsa-miR-22
- hsa-miR-221
- hsa-miR-222
- hsa-miR-223
- hsa-miR-23b
- hsa-miR-29c
- hsa-miR-30c
- hsa-miR-335
- hsa-miR-34a
- hsa-miR-34b
- hsa-miR-9
- hsa-miR-92
- hsa-miR-92b
- hsa-miR-93
- hsa-miR-99a