

Programa DNR metilinimo duomenų tyrimams

Aušrinė Zeleckytė, Ernestas Stulgis, Gabrielė Vilutytė
Bioinformatika, III kursas
Taikomoji kompiuterija
2024m. gegužė

Turinys

- Programos paskirtis
- Duomenys
- Programos naudojimas
- Programos įgyvendinimas
- Rezultatai

Programos paskirtis

- Programa skirta DNR metilinimo duomenų apdorojimui ir analizei.
- Programa geba:
 - skaityti metilinimo duomenis;
 - atlikti duomenų kokybės kontrolę sukurdamą grafiką, vaizduojantį metilinimo vidurkius skirtingose genomo vietose;
 - atlikti statistinį T-test pasirinktam vienam ląstelės tipui ir dviem amžių grupėms, atvaizduoja gautų p-reikščių ir vidurkių skirtumų histogramas bei vulkano grafiką;
 - atvaizduoti grafikus “Heatmap” pavidalu, kurie rodo koreliaciją tarp duomenų prieš ir po hierarchinio klasterizavimo;
 - atlikti duomenų normalumo testą pasirinktam ląstelės tipui.

Duomenys

Duomenys, kuriuos naudojame savo programoje yra paskelbti šiose publikacijose:

Wang, X., Campbell, M.R., Cho, H.-Y., Pittman, G.S., Martos, S.N. and Bell, D.A. (2023). Epigenomic profiling of isolated blood cell types reveals highly specific B cell smoking signatures and links to disease risk. 15(1).

online: <https://doi.org/10.1186/s13148-023-01507-8>.

Programos naudojimas

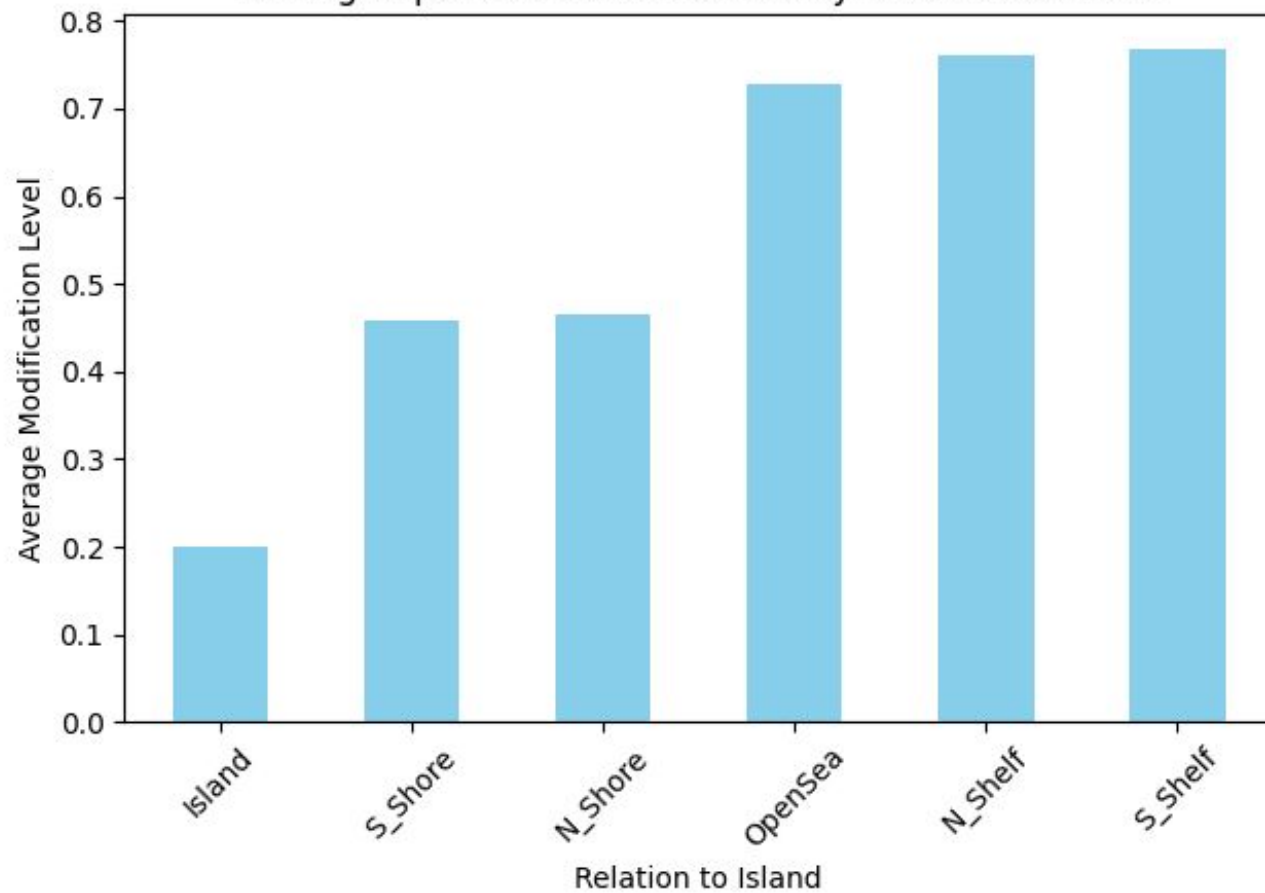
- `git clone <github_repo>*`
 - `github_repo` - https://github.com/gabrielevil/Taikomoji_kompiuterija
- `python main.py <dna_methylation_matrix> <genome_map> <samplekey>`
 - `dna_methylation_matrix` - matrica su CpG vietų metilinimo lygiais skirtinguose mėginiuose
 - `genome_map` - data frame su informacija apie kiekvieną CpG vietą
 - `sample_key` - data frame su informaciją apie kiekvieną mėginį

Rezultatai

Duomenų kokybės kontrolė

- Programos modulis control.py sugeneruoja grafiką pagal pateiktus <dna_methylation_matrix> ir <genome_map> duomenis

Average CpG Modification Levels by Relation to Island

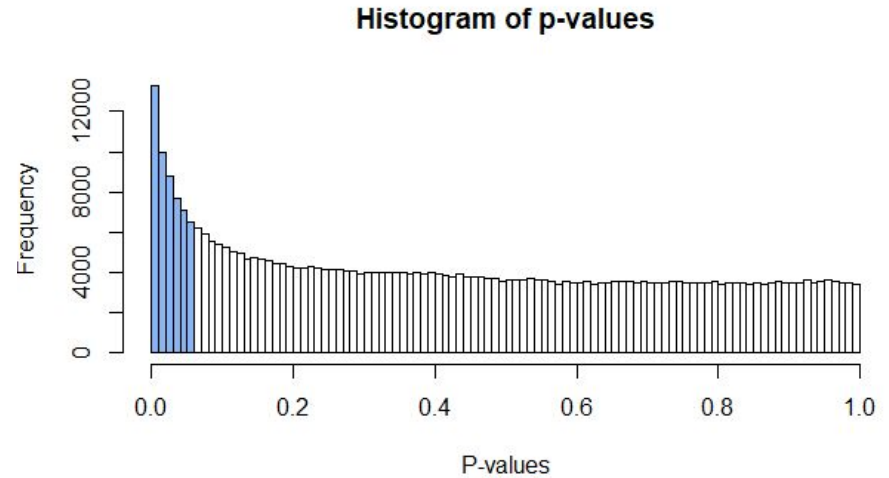
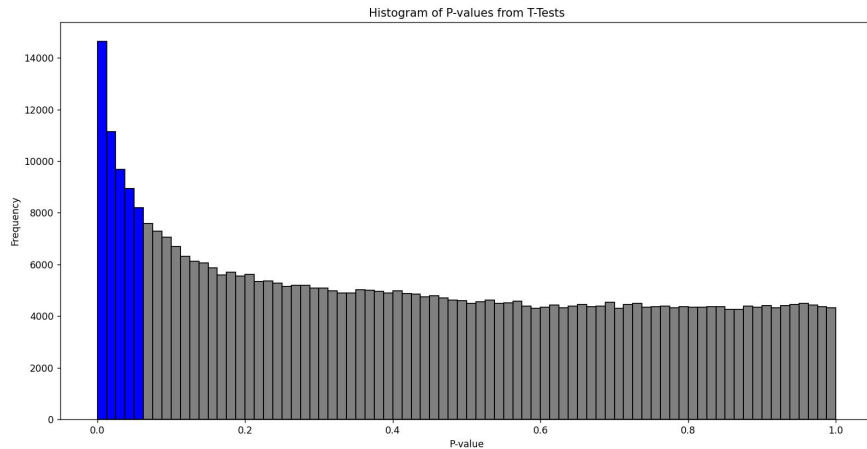


Student T-testas

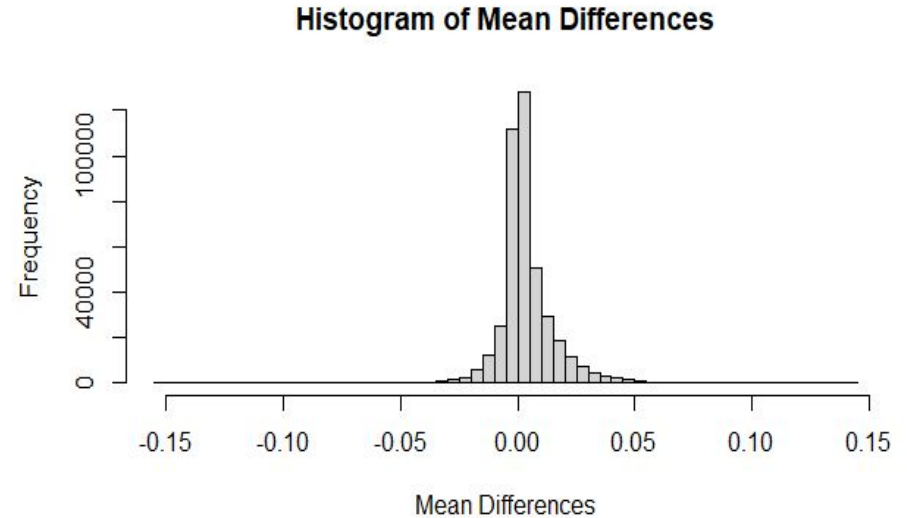
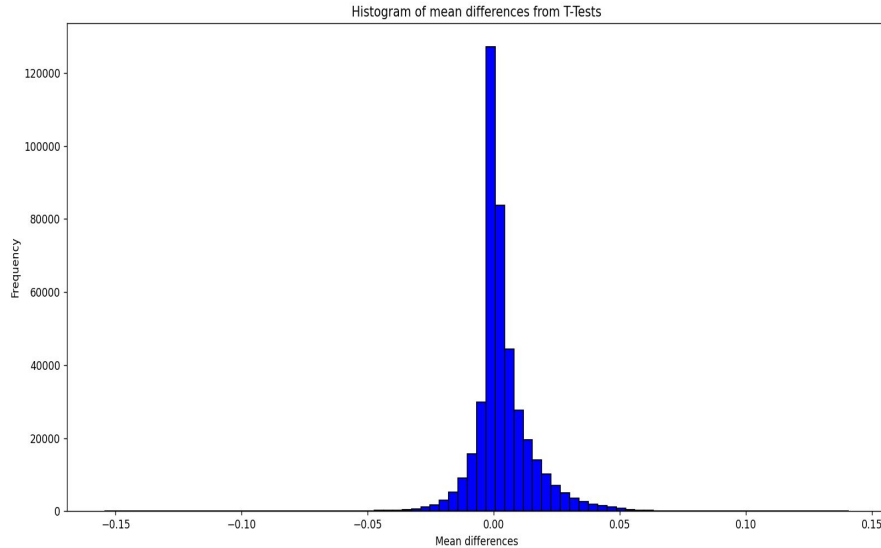
Programos modelis ttest.py sugeneruoja grafikus pagal pateiktas <dna_methylation_matrix>, <sample_key>, ląstelės tipą ir amžiaus grupių atskyrimo ribą.

Šiuo atveju grafikai generuoti pasirinkus B ląsteles ir amžiaus ribą 43 metai.

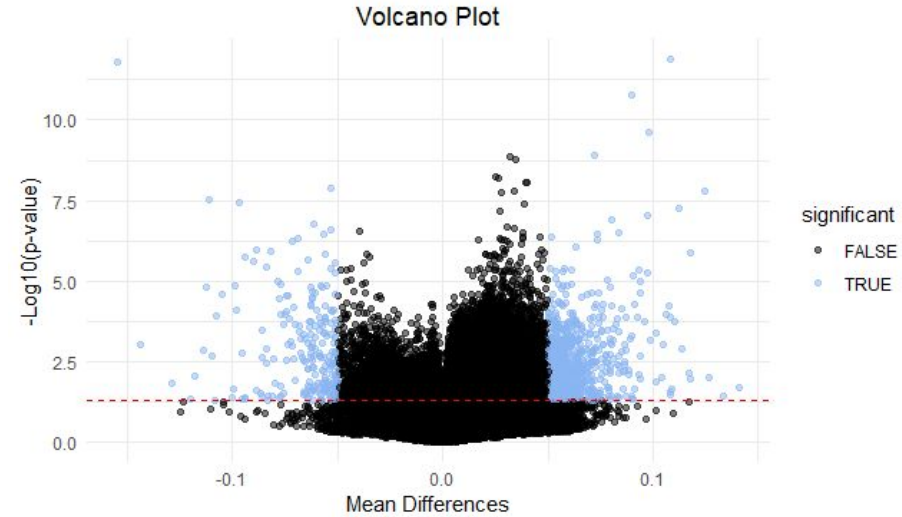
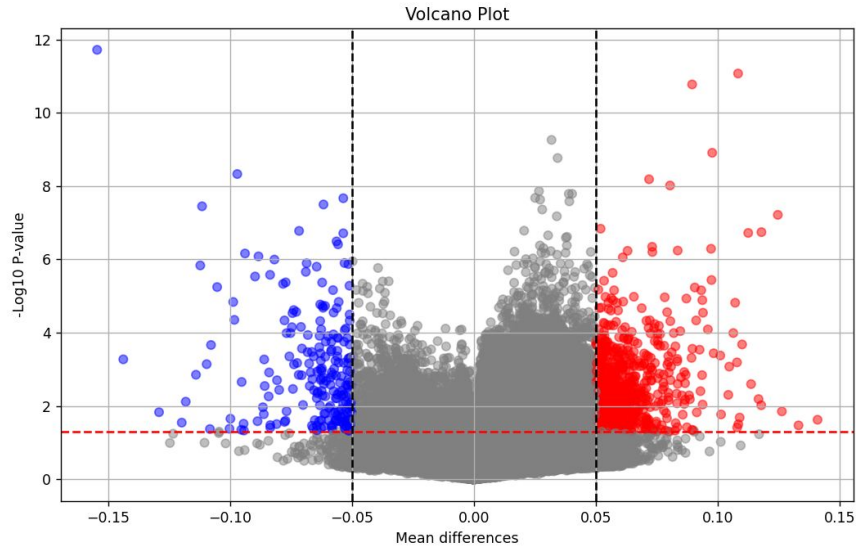
P-reikšmių histogramos python vs R



Vidurkių skirtumų histogramos python vs R



Vulkano grafikas python vs R

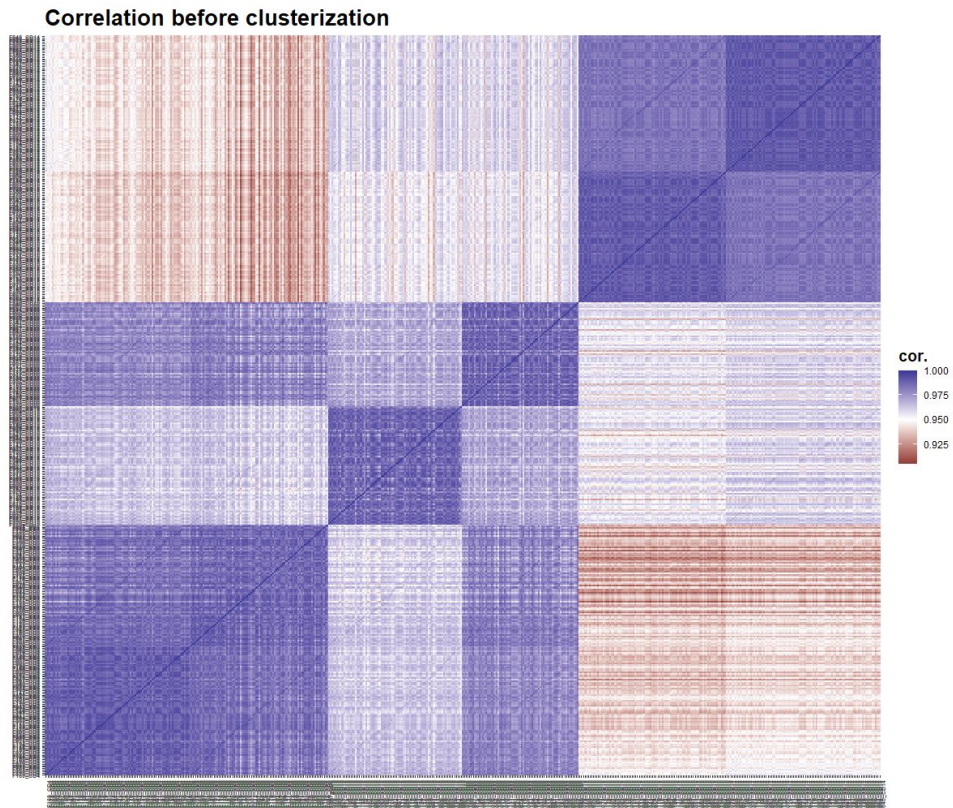


Koreliacija ir hierarchinis klasterizavimas

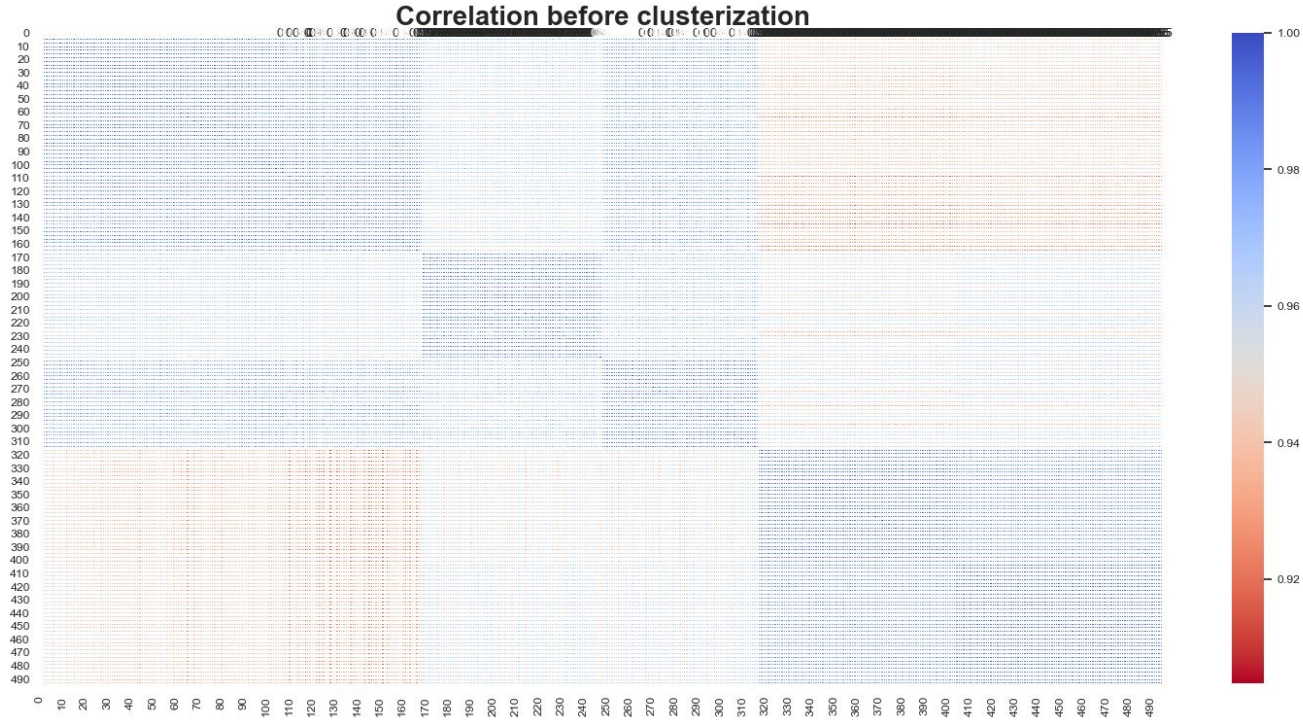
Programos modelis corrPlot.py sugeneruoja grafikus pagal pateiktą `<dna_methylation_matrix>`.

Gaunami grafikai yra koreliacijos “Heatmap”, prieš ir po duomenų hierarchinio klasterizavimo.

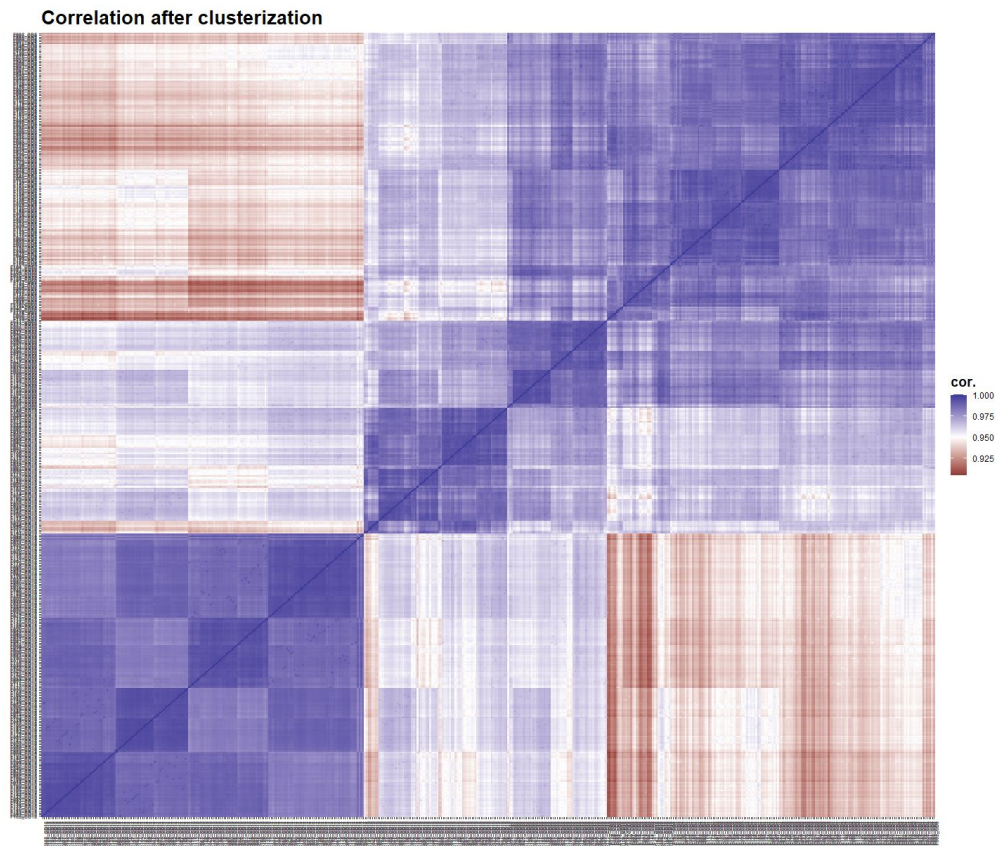
Koreliacija prieš klasterizavimą: R



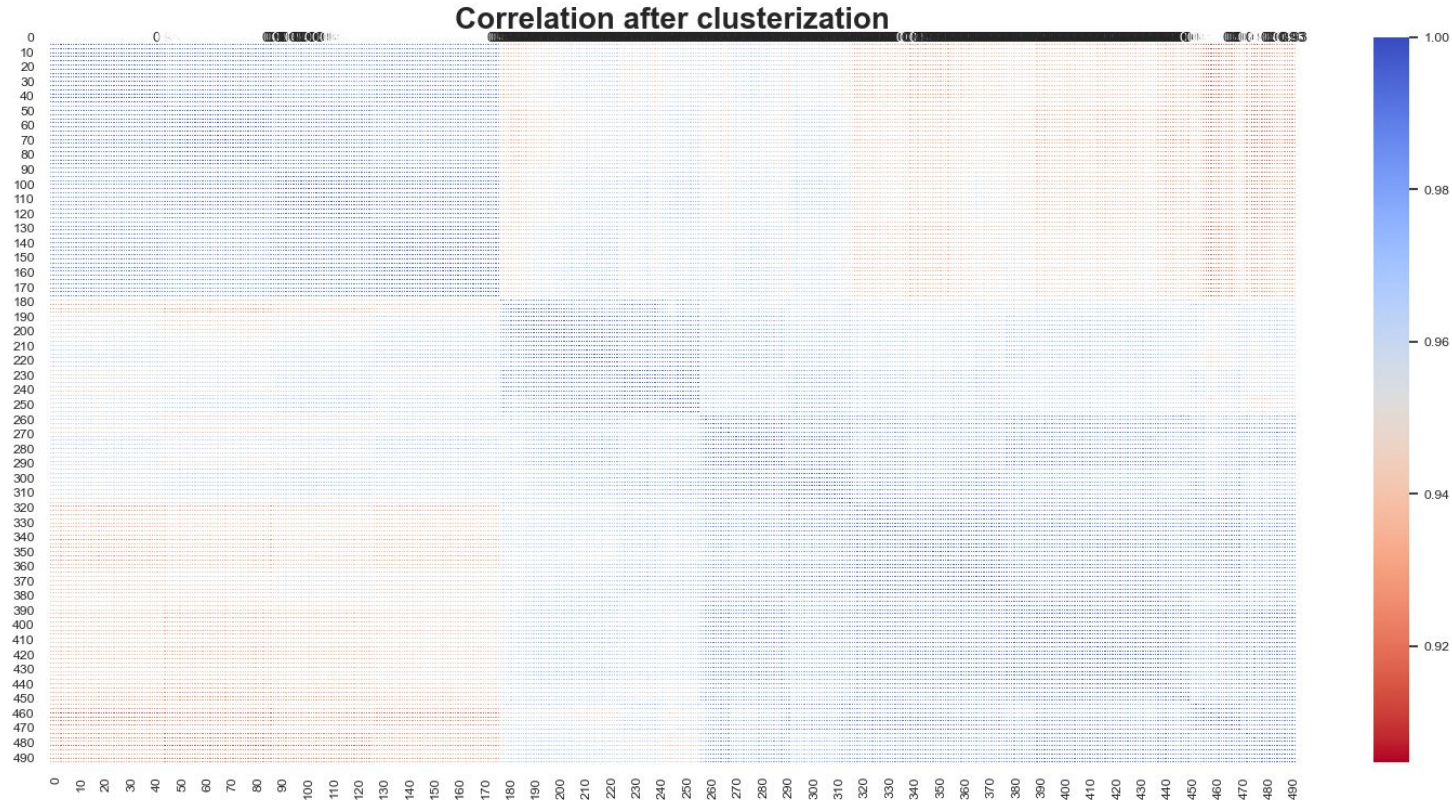
Koreliacija prieš klasterizavimą: Python



Koreliacija po klasterizavimo: R



Koreliacija po klasterizavimo: Python

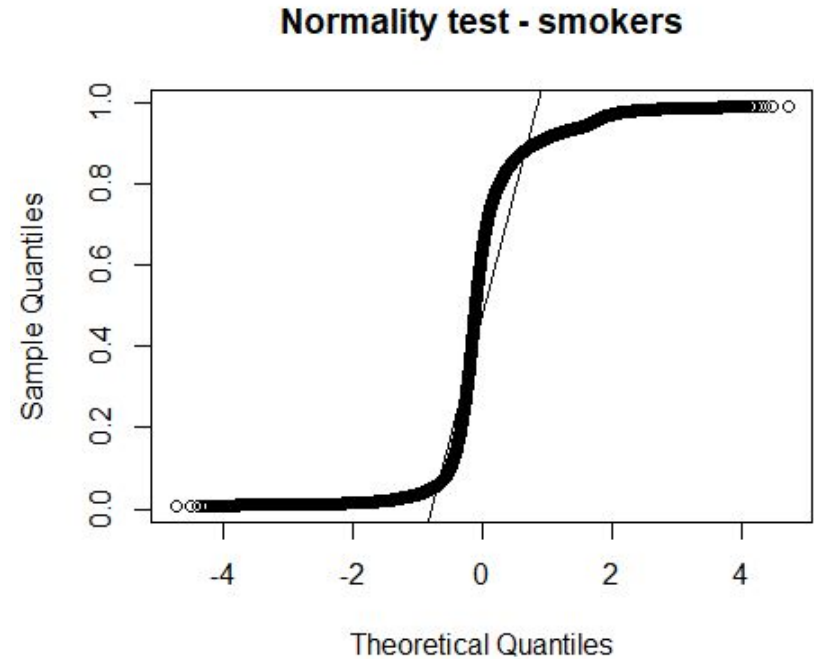
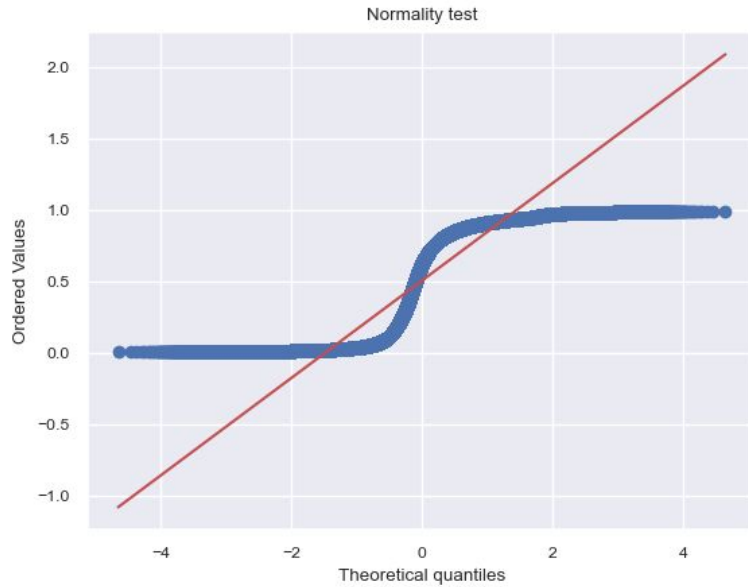


Normality testas

Programos modelis analysis.py sugeneruoja grafiką pagal pateiktą <dna_methylation_matrix>, <sample_key> ir ląstelės tipą.

Galimi ląstelių tipai: bcell, mono, neu, nk, tcd4 ir tcd8

Grafikas su tcd4: Python vs R



GitHub repository

Projekts gali būti pasiekiamas per github:

https://github.com/gabrielevil/Taikomoji_kompiuterija