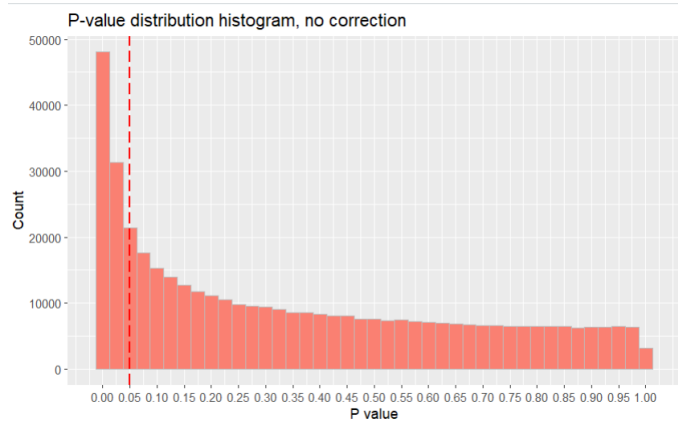


3 užduotis

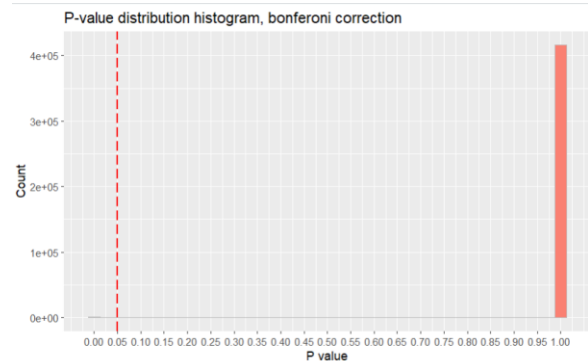
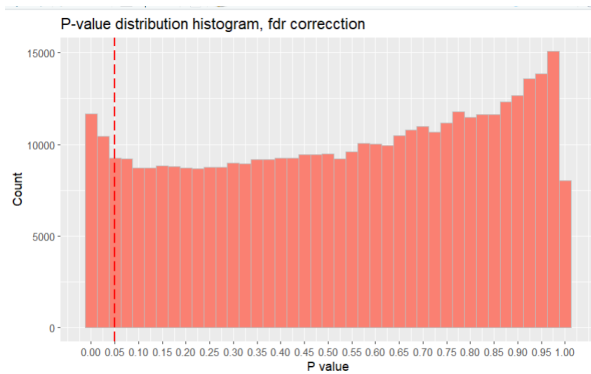
Histogramos

Mes tyrinėjome rūkymo poveikį. Todėl padalinome žmones į rūkančius/nerūkančius. Mes neskyrėme žmones pagal amžių, rasę, lytį ir pan.



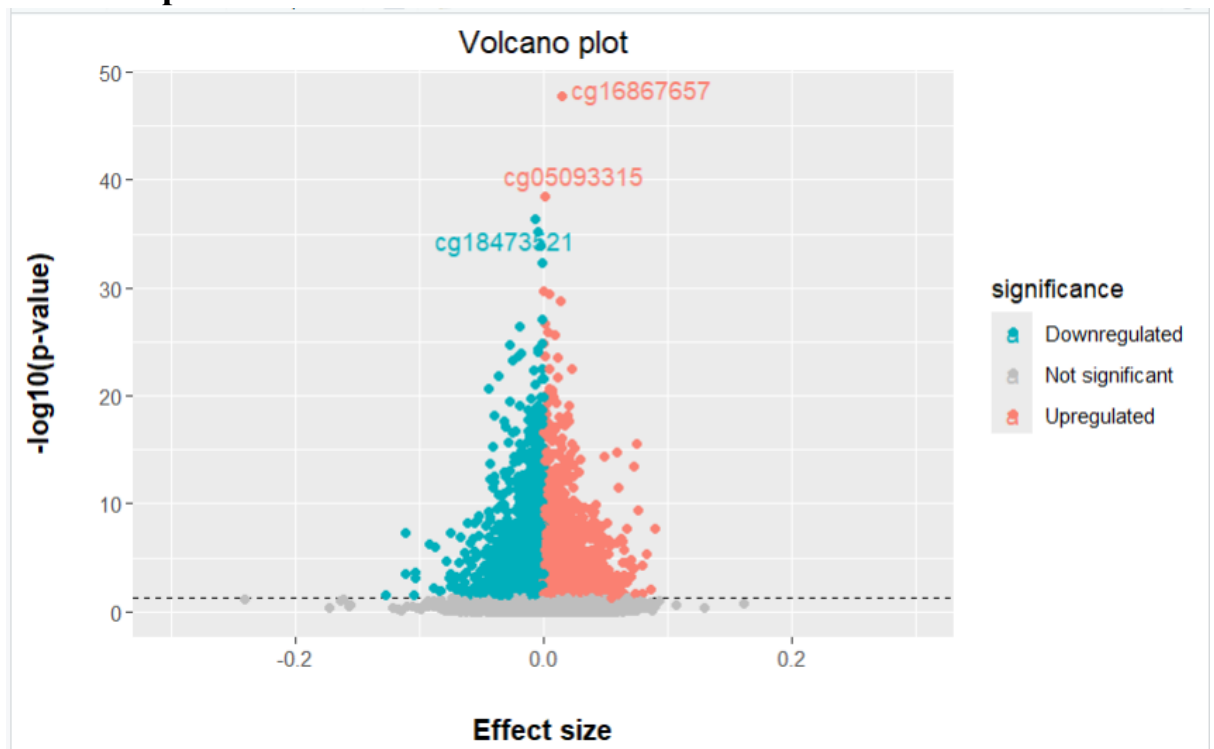
P-value < 0.05 pozicijų kiekis:

- Be korekcijos: 90692
- Po fdr korekcijos: 26752
- Po Bonferroni korekcijos: 1579



Taigi, sudarėme 3 histogramas: be korekcijos, po fdr korekcijos ir po bonferoni korekcijos. Matome, kad histograma be korekcijos atrodo, gana neblogai. Turime daug pozicijų, kur $p\text{-value} < 0.05$. Tuomet, korekcija buvo pritaikyta nelabai sėkmingai - matome, kad „gerų“ pozicijų kiekis netgi sumažėjo. Todėl, tolimesnei analizei naudojame duomenis be korekcijos.

Volcano plot



Matome, kad visi taškai tarsi koncentruojasi vienoje vietoje, dėl to, kad efekto dydžio diapazonas yra ganėtinai mažas. Tačiau, galime matyti ir keletą „emissions“.

Pavyzdžiui, cg16867657, kurio p-value yra labai mažas, todėl taškas yra uždėtas labai aukštai.

cg16867657:

ELOVL2 – genas, koduojantis baltymą, kuris yra atsakingas už lipidų elongaciją. Tačiau, šio laboratorinio darbo kontekste įdomi ne baltymo funkcija. Tyrimai rodo, kad ELOVL2 geno metilinimas yra vienas iš geriausių epigenetinio amžiaus prediktorių!

cg05093315:

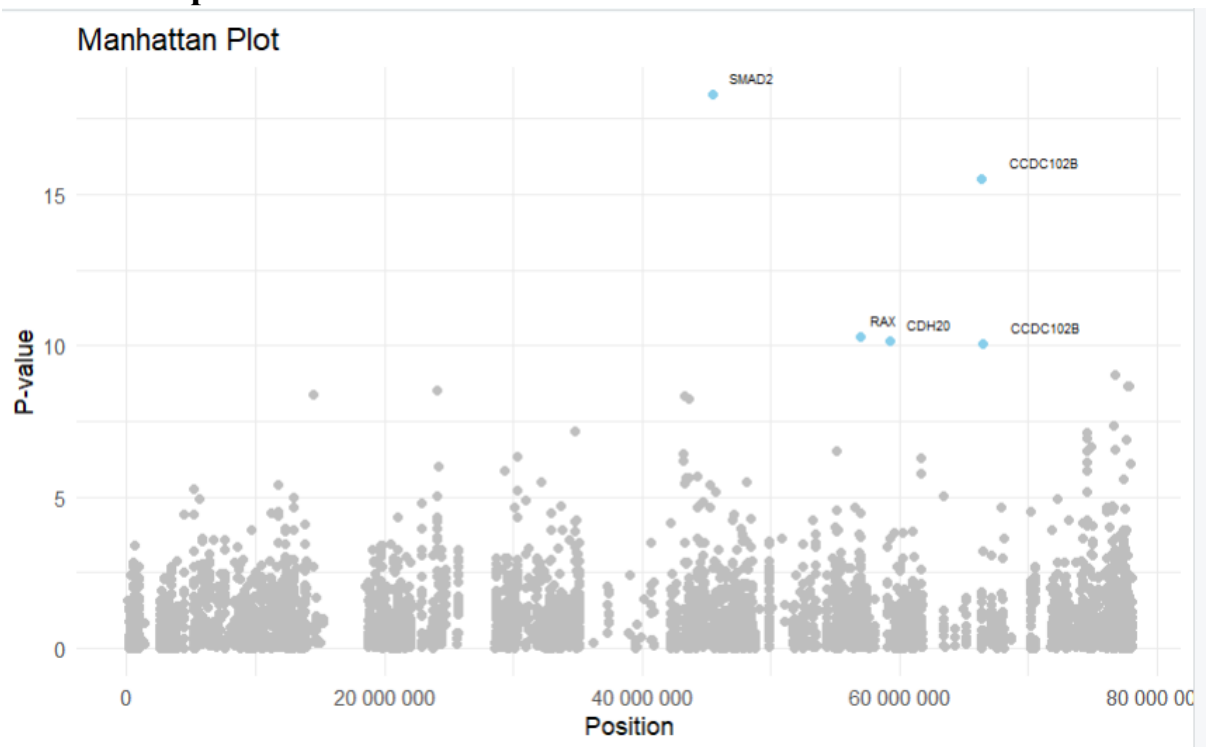
SAAL1 – genas, koduojantis baltymą, kuris yra atsakingas už atsaką į uždegimą. Tikriausiai, kažkokiu būdu tai galima būtų susieti su rūkimu, bet uždegimą gali sukelti labai daug dalykų, todėl kažkokias konkrečias išvadas padaryti būtų sunku.

Interesting-fact: šio baltymo disfunkcija gali būti susieta su ekshibicionizmu.

cg18473521:

HOX4 – genas, koduojantis transkripcijos faktorių. Geno disfunkcija yra susieta su imunodeficitu. Kadangi turėjome duomenų būtent iš imuninės sistemos ląstelių, šios pozicijos mažas p-value yra ganėtinai įdomus reiškinys.

Manhattan plot

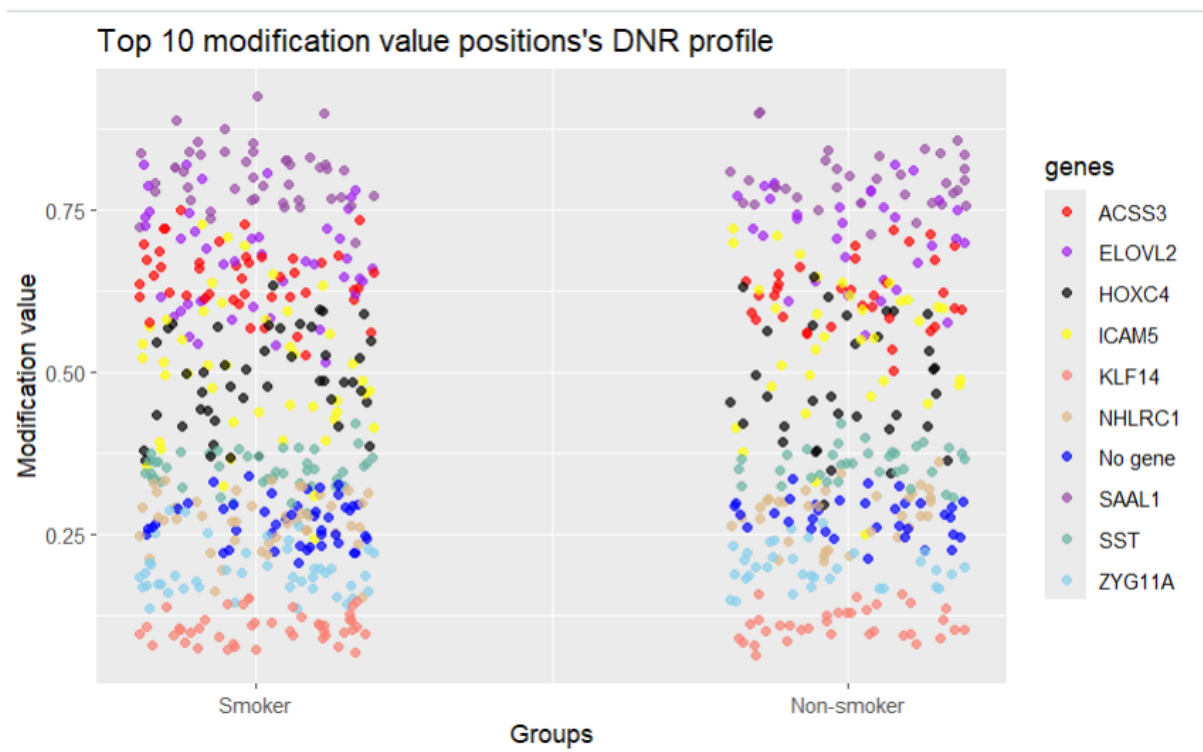


Savo *manhattan plot* parinkome 18 chromosomą, dėl to, kad procentaliai patikimų pozicijų kiekis buvo didžiausias būtent 18 chromosomoje (~22%)

Matome keletą genų su labai mažu p-value. Tačiau, kažkokių įdomių dalykų su šiais genais atrasti nesugebėjome, todėl, manau, nėra tikslo juos išsamiai nagrinėti.

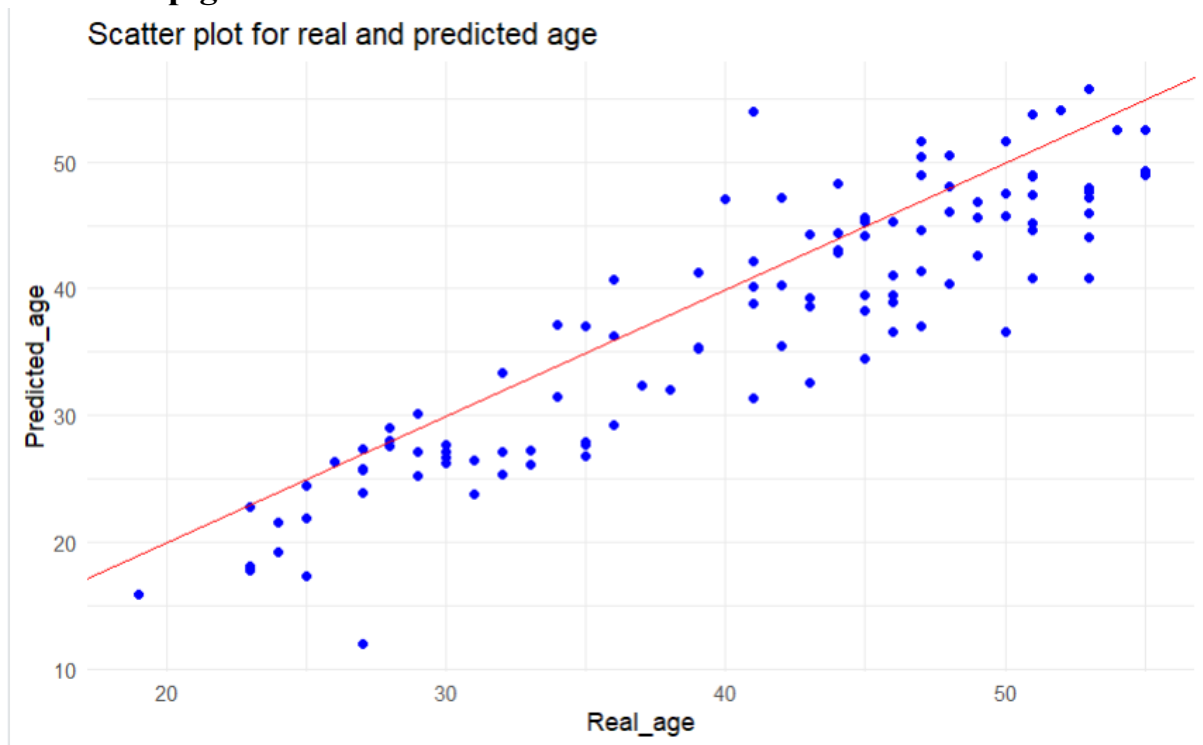
.

DNR-profiliai:



Geresniam atvaizdavimui parinkome tik b-limfocitų duomenis. Tačiau, kaip matome, kažkokio labai didelio skirtumo tarp grupių nėra. Modifikacijos lygiai yra maždaug tokie pat, tik matosi, kad visgi „Smoker” grupėje ELOVL2 geno modifikacijos lygis turi šiek tiek platesnę pasiskirstymą, tačiau tai gali būti tiesiog atsitiktinis dalykas. Kažkokių išvadų remiantis vien šiuo faktu sudaryti nebūtų tikslinga.

Horvath epigenetiniai laikrodžiai



Pasinaudodami duomenimis iš straipsnio, sugebėjome apskaičiuoti epigenetinį amžių. Iš *scatter plot'o*, subjektyviai, galime teigti, kad rezultatas yra labai geras. Žinoma, nėra 100% atitikimo, tačiau galime matyti keletą išskirtinių ribinių atvejų. .

Donor	True age	Predicted age	Age difference
F055	41	~54	-13
F190	27	12	15

Tačiau, daryti išvadas tik iš ribinių atvejų būtų nelabai tikslinga. Todėl, svarbu įvertinti skirtumų vidurkį. Vidurkis išėjo ~3, taigi net geriau nei buvo tikėtasi.

Nustatykite donorų amžių naudojantis savo pasirinktais epigenetiniais laikrodžiais

Parinkome “methylCIPHER” paketą. Tam, kad palyginti mūsų įgyvendintą epigenetinio amžiaus skaičiavimo funkciją, parinkome 2 „methylCIPHER“ paketo funkcijas: „calcHorvath1“ ir „calcHorvath2“. Tam, kad palygintume mūsų įvykdytą realizaciją, vėl pažvelgsime į ribinius atvejus bei amžių vidurkius.

Funkcija	Min	Max	Mean
Mūsų	-12,98	15,037	2,98
calcHorvath1	-12,99	15,03	2,97
calcHorvath2	-12,99	15,03	2,97

Matome, kad skirtumo beveik nėra. Mūsų funkcijos tikslumas yra labai geras, žinoma, kaip teigėme, yra tam tikrų ribinių atvejų, kur mūsų modelis negali tinkamai atspėti žmogaus amžiaus. Kita vertus, to tiksliai negali padaryti ir kiti modeliai, kuriuos ištyrėme, taigi, vis dėlto, galime teigti, kad šis modelis yra ganėtinai veiksmingas.

Papildoma

Iš “Age prediction in living: Forensic epigenetic age estimation based on blood samples” sužinojame, kad yra tam tikri epigenetinio amžiaus predikatoriai. Tai būtų įdomų patikrinti: Ar galime nustatyti amžių, jeigu naudosisime tik prediktorių pozicijas? Tam, kad patikrinti mūsų hipotezę, mes paėmėme prediktorių pavadinimus, kurie buvo paminėti straipsnyje „*ELOVL2, FHL2, EDARADD, ASPA, PDE4C, PENK, C1orf132, TRIM59* and *KLF14*“. Toliau mes parinkome pozicijas, kurios atitinka šitas genų aibes ir su „calcHorvath1“ įrankiu pabandėm gauti rezultatą. Taigi, štai ką mes gavome: „atspėtų“ amžių diapozonas yra tarp 30 ir 35,5. Vidurkis yra apie 32. Taigi, galime teigti, kad nors predikatoriai gali turėti didelę įtaką amžiaus nustatymo tyrimuose, tačiau pavieniai jie „nieko nedaro“, kitaip tariant, mes negalime nustatyti žmogaus amžiaus, jeigu naudojame tiek mažai požymių, nors tie požymiai ir yra reikšmingi.

Šaltiniai:

Examination of DNA methylation status of the *ELOVL2* marker may be useful for human age prediction in forensic science

<https://www.sciencedirect.com/science/article/abs/pii/S187249731400218X>

Age prediction in living: Forensic epigenetic age estimation based on blood samples

<https://www.sciencedirect.com/science/article/abs/pii/S1344622320300973>

ELOVL2 Gene - ELOVL Fatty Acid Elongase 2

<https://www.genecards.org/cgi-bin/carddisp.pl?gene=ELOVL2>

SAAL1 Gene - Serum Amyloid A Like 1

<https://www.genecards.org/cgi-bin/carddisp.pl?gene=SAAL1>