



**VILNIAUS UNIVERSITETAS**  
**MATEMATIKOS IR INFORMATIKOS FAKULTETAS**  
**BIOINFORMATIKOS BAKALAURO STUDIJŲ PROGRAMA**

**Pavadinimas lietuviškai**

**Pavadinimas angliškai**

**Kursinis projektas**

Autorė: Danielė Stasiūnaitė  
VU el. p.: daniela.stasiunaite@gmc.vu.lt

Darbo vadovė: J. m. d. Kotryna Kvederavičiūtė

**Vilnius**  
**2023**

# Turinys

<b>1 Įvadas</b>	<b>4</b>
<b>2 Pasirinktų mėginių charakteristika</b>	<b>5</b>
<b>3 Sintenijos blokai ir jų nustatymo metodai</b>	<b>6</b>
3.1 Sintenijos bloko apibūdinimas . . . . .	6
3.2 Identifikavimo metodai . . . . .	7
3.2.1 Giminingų organizmų genomų lyginimas . . . . .	7
3.2.2 Mažai giminingų arba negiminingų organizmų genomų lyginimas . . . . .	8
<b>4 Tyrimo metodai</b>	<b>10</b>
4.1 Sintenijos blokų nustatymas ir vizualizavimas . . . . .	10
4.2 Sintenijos blokų pozicijų gavimas . . . . .	10
4.3 Sintenijos blokų persidengimas su pikais . . . . .	10
4.4 Pikų anotavimas . . . . .	10
4.5 Persidengiančių genų nustatymas . . . . .	10
<b>5 Naudotų įrankių metodų aprašymas</b>	<b>12</b>
5.1 Cintenų įrankio metodas . . . . .	12
<b>6 Rezultatai ir jų aptarimas</b>	<b>14</b>
6.1 Sintenijos blokų nustatymas ir vizualizavimas . . . . .	14
6.2 Sintenijos blokų pasiskirstymas chromosomose . . . . .	15
<b>7 Išvados</b>	<b>19</b>
<b>8 Priedas</b>	<b>23</b>

# Santrauka

**Raktiniai žodžiai:** homologija, sintenija, lyginamoji genomika, transkripcijos faktorius, Tbx5, *Mus musculus*, *Danio rerio*, R.

# Summary

**Keywords:** homology, sinteny, comparative genomics, transcription factor, Tbx5, *Mus musculus*, *Danio rerio*, R.

# 1 Įvadas

## Darbo temos aktualumas

## Darbo tikslas

Nustatyti *Mus musculus* ir *Danio rerio* organizmų homologiją, naudojantis skirtingais poveikiais veikty *Mus musculus* širdies ląstelių ChIP sekoskaitos duomenimis.

## Uždaviniai

- Išanalizuoti metodus, taikomus sintenijos tarp skirtingų organizmų nustatymui;
- Įvertinti sintenijos blokams priklausančių sekų fragmentų pasiskirstymą chromosomose;
- Nustatyti, kokia mėginių pikų dalis priklauso sintenijos blokams;
- Anotavus pikus nustatyti pikus atitinkančius genus *Danio rerio* genome;
- Patikrinti, ar *Danio rerio* genome nustatytų genų promotoriuose galimas Tbx5 transkripcijos faktoriaus prisijungimas.

## 2 Pasirinktų mėginių charakteristika

Analizė atlikta, naudojantis Kursinio darbo metu analizuotais duomenimis. Duomenys atsisiųsti iš GTRD (Gene Transcription Regulation Database)[1] duomenų bazės, saugančios informaciją apie transkripcijos sekas bei atviro chromatinio regionus.

Analizei naudoti 7 skirtingi naminių pelių (lot. *Mus musculus*) širdžių ląstelių mėginiai, tirti 3 nepriklausomų eksperimentų metu širdžių ląsteles veikiant skirtingais poveikiais. Šių mėginių charakteristikos pateiktos pirmoje lentelėje.

1 lentelė. Mėginių charakteristikos

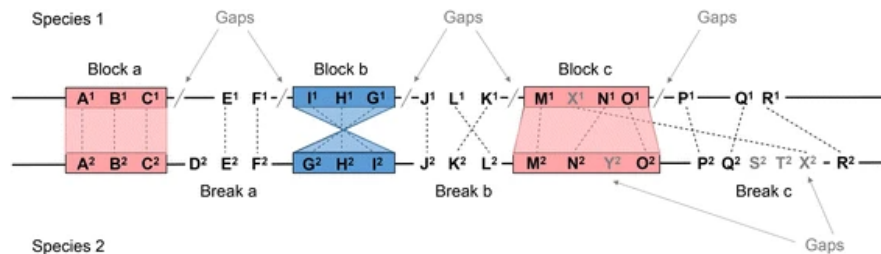
Žymėjimas grafikuose	Ląstelių tipas	Kamienas	Poveikis	Antikūnai	GTRD ID
<i>crdc_mscl_r1_2</i>	HL - 1 (širdies raumens)	C57BL/6J	TRE promotorius (2 d.)	-	EXP030898
<i>heart_r1_1</i>	Širdies prieširdžių	C57BL/6	-	Tbx5 (sc-17866)	EXP058852
<i>emb_fibr_r1_1</i>	MEF (embrionų fibroblastai)	C57BL/6	AGHMT (2 d.)	anti-Tbx5 (sc-17866)	EXP058843
<i>emb_fibr_r2_2</i>	MEF (embrionų fibroblastai)	C57BL/6	GHMT (2 d.)	Tbx5 (sc-17866)	EXP058847
<i>emb_fibr_r3_3</i>	MEF (embrionų fibroblastai)	C57BL/6	GMT (2 d.)	Tbx5 (sc-17866)	EXP058850
<i>emb_fibr_r4_4</i>	MEF (embrionų fibroblastai)	C57BL/6	Tbx5 (2 d.)	Tbx5 (sc-17866)	EXP058856
<i>crdc_fibr_r1_3</i>	Pelių naujagimių širdies fibroblastų, ekspresuojančių didelį kiekį T antigeno, linija	CD1	sb431542, xav939	anti-TBX5 (sc-17866x)	EXP062056

- **TRE**: tetraciklino atsako elementas (angl. *Tetracycline Response Element*). Tai yra 7 DNR sekos fragmentai, sudaryti iš 19 nukleotidų ir atskirti trumpesniais sekų fragmentais.
- **sb431542**: stipriai veikianti, selektyvi cheminė medžiaga; augimo faktoriaus inhibitorius.
- **xav939**: stipriai veikianti cheminė medžiaga; tankirazės, slopinančios specialaus baltymo, stabdančio telomerazės veiklą, junginasi prie telomerinių DNR sekų, inhibitorius.
- **AGHMT**: Transkripcijos faktorių, A - Akt1 kinazė, G - GATA4, H - HAND2, M - MEF2C, T - Tbx5, kompleksas.
- **GHMT, GMT**: sumažintas transkripcijos faktorių kompleksas.
- **sc-17866x/sc-17866**: žmonių, pelių ir žiurkių Tbx5 antigeną atpažįstantys antikūnai (išskirti iš oškų).

### 3 Sintenijos blokai ir jų nustatymo metodai

#### 3.1 Sintenijos bloko apibūdinimas

**Sintenijos blokas** - tarp skirtingų genomų identifikuojami chromosomų regionai, kuriems būdingi bendri homologiniai genai, išsidėstę nebūtinai vienoda tvarka. Sintenijos blokai neretai gali būti vadinami kolineariais blokais. Pastarieji blokai nuo sintenijos blokų skiriasi tik tuo, kad kolineariams blokams būdingas identiškas genų išsidėstymas. Žemiau pateikiamame pirmame paveiksle (1 pav.) vaizduojami galimi sintenijos blokų tipai:



1 pav. Sintenijos blokų tipai

Paveiksle vaizduojamuose blokuose yra po tris ortologinius genus (inkarus):

1. **A blokas:** blokas tarp vienoda tvarka išsidėsčiusių ortologinių genų;
2. **B blokas:** blokas tarp priešinga tvarka išsidėsčiusių ortologinių genų;
3. **C blokas:** blokas tarp ortologinių genų, kuriuos skiria neortologiniai genai.

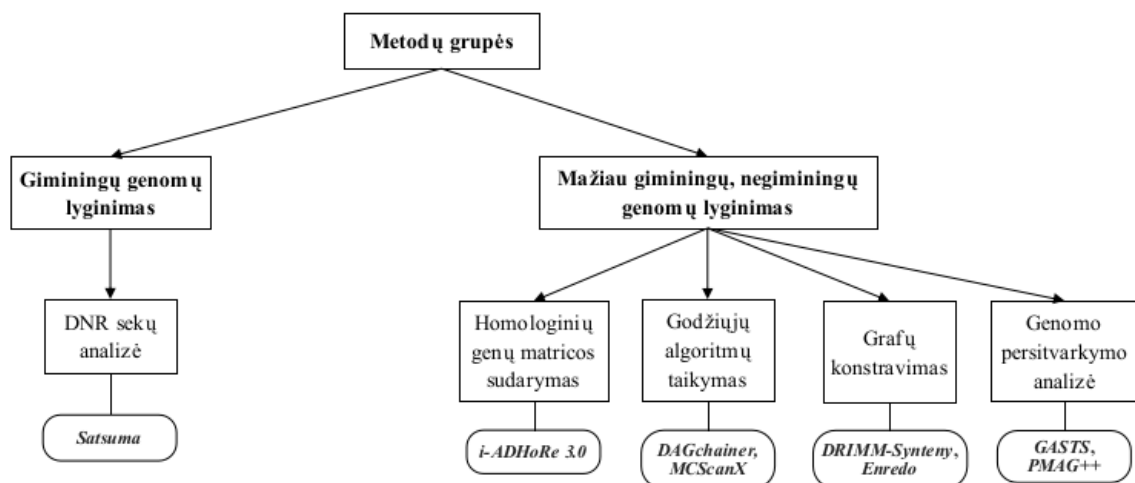
*Synteny* blokų taikymas leidžia tirti skirtingų organizmų sąsajas ir genomų evoliuciją, vykstant genomų persitvarkymo procesams. Atliekant šių blokų analizę galima rekonstruoti filogenetinius ir filogenominius medžius[2], siekiant išsiaiškinti bendrą tiriamų organizmų kilmę. Taip pat konservatyvių genų blokų nustatymas tarp skirtingų organizmų genomų gali indikuoti funkcinės genų produktų (baltymų) sąsajas[3].

## 3.2 Identifikavimo metodai

Internete galima rasti ne vieną įrankį, leidžiantį vizualizuoti gerai anotuotų genomų sinteniją DNR lygmenyje (pavyzdžiui, Ensembl SyntenyView[4], NCBI MapViewer[5], Genomicus[6]), tačiau nepaisant to, kad taikant šiuos įrankius konservatyvius regionus galima vizualizuoti ne vienu būdu, įrankiai pateikia iš anksto suskaičiuotus rezultatus. Naudojant šiuos įrankius tyrėjai negali pasirinkti parametrų, pavyzdžiui, leidžiančių keisti sintenijos blokų ilgį ar minimalų genų skaičių bloke.

Dėl šių įrankių trūkumų yra sukurta daug įvairių programinių įrankių, leidžiančių atlikti pastarųjų blokų paiešką savarankiškai, tiriant nemodelinių organizmų genomų evoliuciją bei keičiant algoritmų parametrus. Šiuos įrankius galima pasirinkti pagal tai, koks yra organizmų giminingumas, koks geno tyrimas yra atliekamas ir kokią hipotezę siekiama patvirtinti arba paneigti.

Antrame paveiksle (2 pav.) pateiktoje schemeje vaizduojami sintenijos nustatymo metodai išskaidyti pagal lyginamų genomų giminingumą:



2 pav. Sintenijos blokų identifikavimo metodai ir populiariausių įrankių pavyzdžiai

### 3.2.1 Giminingų organizmų genomų lyginimas

Lyginant genomus, priklausančius itin giminingiems organizmams, *synteny* blokai gali būti identifikuojami taikant įrankius, paremtus DNR sekų išlyginimais, nes genomų sekų atitikimo laipsnis yra didelis (didelė dalis nukleotidų sutampa).

Konservatyvių chromosomų regionų nustatymui naudojami DNR sekų išlyginimu paremti metodai yra intuityvūs, tačiau reikalaujantys didelių kompiuterio resursų (jeigu tarpusavyje lyginami eukariotinių organizmų genomai). Nepaisant to, yra sukurta įrankių, kurie efektyviai sprendžia ne tik kompiuterio resursų naudojimo problemą, bet ir pateikia patikimus rezultatus, leidžiančius daryti išvadas apie atliktą tyrimą.



### 3.2.2 Mažai giminingų arba negiminingų organizmų genomų lyginimas

Tiriant tos pačios biologinės klasifikacijos taksonominio rango - klasės - organizmus, kurie kilę iš bendro protėvio ir turi panašią sandarą, metodai, kurie remiasi DNR sekų išlyginimu, negali būti taikomi, nes lyginamų organizmų DNR sekos yra pakitusios (evoliucijos eigoje įvyko daugiau įvairių mutacijų), todėl sintenijos blokų paieška turi būti atliekama aminorūgščių lygyje. Nepaisant evoliucijos eigoje vykstančių DNR sekų pokyčių, viena aminorūgštis gali būti koduojama keliais skirtingais kodonais. Dėl šios priežasties metodo taikymas tinka evoliucijos eigoje mutavusių mažai giminingų organizmų genomų sekų lyginimui ir homologinių regionų paieškai.

Norint palyginti skirtingoms taksonominėms klasėms priklausančių organizmų genomus taikomi metodai, analizuojantys baltymus koduojančias aminorūgštis, bei metodai, konstruojantys profilius, grafus ir naudojančius įvairius statistinius modelius.

Taikant šiuos metodus yra svarbu, kad analizuojamų organizmų genomai būtų kuo kokybiškesni: jeigu genome trūksta sekų, gali nebūti tam tikrų genų anotacijų, todėl homologinių sąsajų taip pat nebūtų galima tyrinėti - kai kurie sintenijos blokai būtų neidentifikuoti.

#### Homologinių genų matricos sudarymo metodas

Taikant šį metodą sintenijos blokai identifikuojami, naudojant kaimyninių genų porų klasterizavimą. Genomų genų atitikimai arba neatitikimai fiksuojami matricoje. Jeigu genai homologiniai, matricoje įrašomas '1', jei nehomologiniai - įrašomas '0'. Remiantis šiuo metodu sintenijos blokai nustatomi, ieškant, kurioje matricos diagonalės vietoje yra didžiausia vienetų sankaupa.

**Metodo privalumas:** taikant šį metodą nustatytiems sintenijos blokams galima paprastai pritaikyti statistinę validaciją ir aptikti *false positive* reikšmes.

**Metodo trūkumai:** taikant šį metodą yra svarbu tiksliai žinoti, į kokią biologinę klausimą norima atsakyti, nes nuo to priklauso, koks paieškai reikalingas parametras arba jų konfigūracijos bus pasirinktos, jog būtų gauti optimalūs rezultatai. Pavyzdžiui, tarpo tarp dviejų genų dydis, kuris gali egzistuoti identifikuotame sintenijos bloke. Kuo šio parametro vertė mažesnė, tuo daugiau nedidelių ir sunkiai analizuojamų sintenijos blokų bus nustatyta, tačiau, jei tarpo ilgio parametro reikšmė didelė, nustatytas blokas bus didelis - analizės atlikimas taps paprastesniu. Kita vertus, tokiam bloke gali atsirasti tokių sekų fragmentų, kurie iš tiesų neturėtų priklausyti blokui (daugiau *false positive* reikšmių), nes nėra aptinkami kitame genome.

#### Godžiųjų algoritmų taikymo metodas

Realizuojant šį metodą yra konstruojamos kolinearių genų porų grandinės. Metodo įgyvendinimui atliekama daug daugiau skaičiavimų nei aprašytas homologinių genų matricos sudarymo metodas. Įrankiuose, kurie remiasi šiuo metodu, yra naudojami *BLASTP* rezultatai. Jais remiantis

galima taikyti dinaminį programavimą ir sukonstruoti kolinearių genų porų grandines, turinčias maksimalias grandinės vertes (angl. *chainscore*). Taikant šį metodą sintenijos blokus sudaro kolineariose pozicijose esančių unikalių homologinių sekų (sintenijos inkarų) genai bei tarp jų išsidėstę nehomologiniai genai, kurie evoliucijos eigoje buvo paveikti mutacijų.

**Metodo privalumai:** Taikant šį metodą galima lyginti daug skirtingų genomų ir analizuoti genų pokyčius chromosomose. Taip pat genomų kolinearių genų paieškos algoritmai veikia itin sparčiai ir pateikia išsamius rezultatus.

**Metodo trūkumas:** Taikant metodą reikia iš anksto žinoti, ko reikia ieškoti, bei, kokios yra tiriamų genomų bei sintenijos blokų charakteristikos.

## Grafų konstravimo metodas

Kai yra žinomos baltymus koduojančių genų pozicijos ir sekų išlyginimo tikimybė, galima konstruoti grafus, surenkančius visas įmanomas homologinių genų poras tarp lyginamų genomų.

Iš pradžių yra atliekamas lokalus genomų išlyginimas, po kurio gauti išlyginimai naudojami grafo konstravimui bei genomų vidinių struktūrų paieškai. Priklausomai nuo to, kokį grafą bus pasirinkta konstruoti, gali būti gauti labai skirtingi identifikuotų sintenijos blokų rezultatai. Taikant šį metodą galima konstruoti keturių skirtingų tipų grafus: išlyginimų, *de-Bruijn*, *Enredo*, *Cactus*.

Taikant skirtingus grafų tipus tiriamos vidinės genomo struktūros, jog būtų identifikuoti sintenijos blokai. Šiomis struktūromis gali būti: iš eilės einantys ir tarp dviejų genomų sutampančios genų sekų blokai, tarp kurių nėra tarpų ir kuriems būdinga vienoda kryptis abiejuose genomuose; mikroblokai - blokai tarp kurių gali būti tarpų; trumpi ciklai, kurie įrodo, kad genomas pakito, vykstant persitvarkymo procesams.

## 4 Tyrimo metodai

Homologijos tarp *Mus musculus* ir *Danio rerio* organizmų tyrimo analizė atlikta su R programavimo kalba[13] (4.2.2 versija)

Tarpiniai analizės rezultatai pateikti, naudojant komandinės eilutės įrankį Scickick[14] (0.2.0 versija). Naudojant šį įrankį buvo apjungti keli Rmd failai ir gauta HTML formato ataskaita (6 skyriaus priedas).

### 4.1 Sintenijos blokų nustatymas ir vizualizavimas

Sintenijos blokai tarp *Mus musculus* ir *Danio rerio* organizmų nustatyti ir vizualizuoti pasinaudojus internetinės *Synteny Portal*[15] aplikacijos *SynCircos* įrankiu.

### 4.2 Sintenijos blokų pozicijų gavimas

Šiame analizės etape sintenijos blokai buvo gauti pasinaudojus *Cinteny* įrankiu, kurio pateiktoje išvestyje buvo pateiktos pirmojo genomo *Mus musculus* genomines pozicijas ir antrojo genomo *Danio rerio* genomines pozicijas, kuriose abiejų genomų sekų fragmentai sutapo - gauti sintenijos blokai.

### 4.3 Sintenijos blokų persidengimas su pikais

Sintenijos blokų failas buvo apdorotas R programoje, kurioje apskaičiuota kiekvieno mėginio pikų bei su sintenijos blokais persidengiančių pikų dalis, panaudojus *rtracklayer*[23] bibliotekos funkciją *subsetByOverlaps()*. Stulpelinės diagramos sukurtos su *ggplot2*[17] bibliotekos *geom\_bar()* funkcija.

### 4.4 Pikų anotavimas

Analizuotų mėginių pikai anotuoti, pasinaudojus R programavimo kalbos bibliotekos *ChIP-seeker*[18, 19] funkcija *annotatePeak()*. Dveji iš šiai funkcijai perduodamų parametrų buvo gauti įdiegus *TxDb.Mmusculus.UCSC.mm10.knownGene*[20] ir *org.Mm.eg.db*[21] bibliotekas.

### 4.5 Persidengiančių genų nustatymas

*Danio rerio* organizmo genų rinkinys gautas, pasinaudojus iš Ensembl[22] duomenų bazės atsisiųsta *Danio rerio* organizmo GTF formato anotacija. GTF failas importuotas į R programą kaip objektas, pritaikius *rtracklayer*[23] bibliotekos funkciją *import()*.

Zebražuvės genų identifikatoriai ir genų pavadinimai gauti, atliekant objekto kovertavimą į *data.frame* struktūrą bei atliekant informacijos išgavimą iš šios duomenų struktūros.

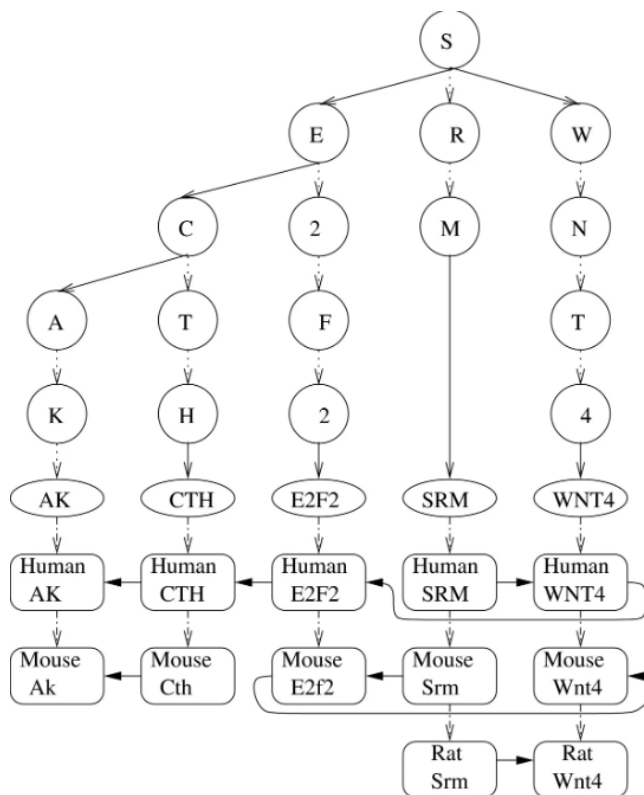
Tarp naminės pelės anotuotų pikų ir zebražuvės genų rinkinių persidengiantys genai nustatyti su standartine R funkcija *intersect()*.

## 5 Naudotų įrankių metodų aprašymas

Šioje analizės dalyje buvo siekiama patikrinti skirtingų *syntenų* blokų identifikavimo įrankių veikimą, ieškant homologinių genomų sričių tarp *Mus musculus* ir *Danio rerio* organizmų. Todėl šioje dalyje buvo išbandyti 3 skirtingi įrankiai: *Cinteny*.

### 5.1 Cinteny įrankio metodas

Siekiant efektyviai reprezentuoti linijinę genomų markerių tvarką šiame įrankyje implemenuota išplėsta trejetainio paieškos medžio struktūra (angl. *ternary search tree* (TST)). Ši struktūra išplėsta, pavaizdavus „perėjimus“ tarp medžio lapų. Šie „perėjimai“ atitinka linijinius geno markerių „perėjimus“. Šie markeriai gali būti apibūdinti kaip ortologiniai (homologiniai) genai.



4 pav. TST medžio struktūra su perėjimais

Pateiktame paveiksle vaizduojama TST medžio struktūra su „perėjimais“ keliems genams iš žmogaus, pelės ir žiurkės genomų. Paveiksle viršūnės (apskritimai) atitinka genų simbolius, medžio lapai (elipsės) atitinka homologinę grupę. Individualūs genai (meta viršūnės), priklausantys kiekvienai homologinei grupei, atvaizduoti po medžio lapais. Linijiniai perėjimai sukonstruoti, sujungus meta viršūnes ta tvarka, kuria genai nustatyti chromosomose.

Konservatyvių genų blokai, išlaikantys savo eiliškumą, yra agreguojami į didesnius blokus,

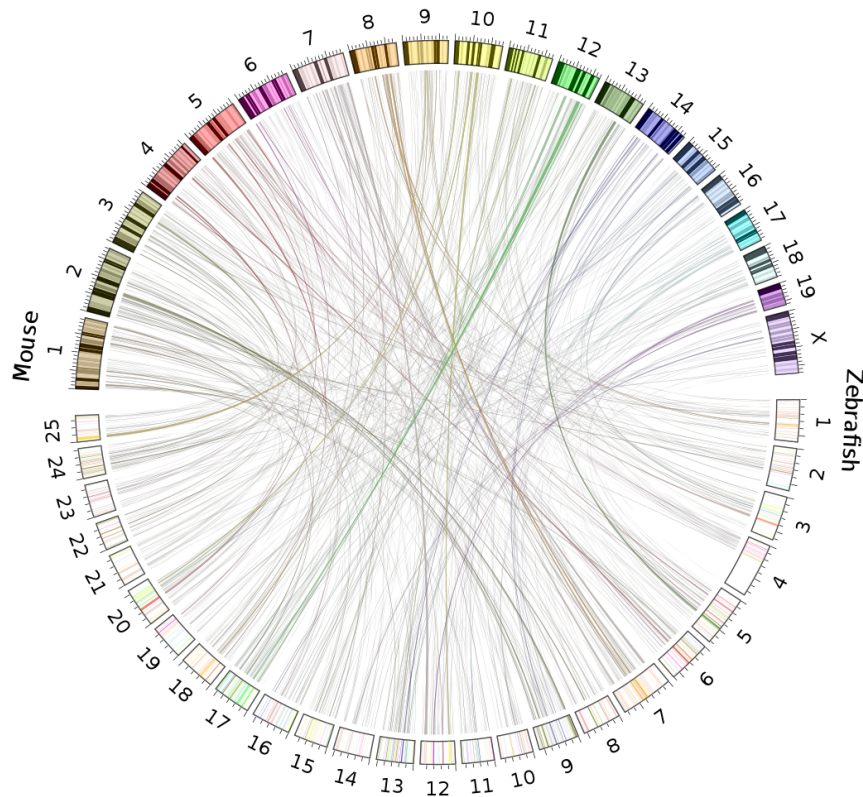
neįtraukiant blokų, kurie gauti atlikus sekų mikropertvarkymus. Keičiant įrankio parametrus (minimalų sintenijos blokų ilgį, maksimalų tarpą tarp agreguotų blokų ir minimalų genų (markerių) skaičių bloke) gali būti gauti skirtingi sintenijos blokai.

Šios hibridinės struktūros taikymas leidžia atlikti genų paiešką per logaritminį laiką, įvardinti visus galimus ortologus bei pereiti visus genų perėjimus iš 5' galo link 3' chromosomos galo arba atvirkščiai. Struktūrą realizuojantis algoritmo kompleksiskumas yra  $N \log_{10} N$ , kur  $N$  skaičius nurodo genome nustatytų genų skaičių, todėl šį metodą galima taikyti ir dideliems (didelių bazių porų skaičių turintiems) genomams.

## 6 Rezultatai ir jų aptarimas

### 6.1 Sintenijos blokų nustatymas ir vizualizavimas

Sintenijos blokai tarp *Mus musculus* ir *Danio rerio* organizmų genomų vizualizuoti *circo* diagramoje (3 pav.).



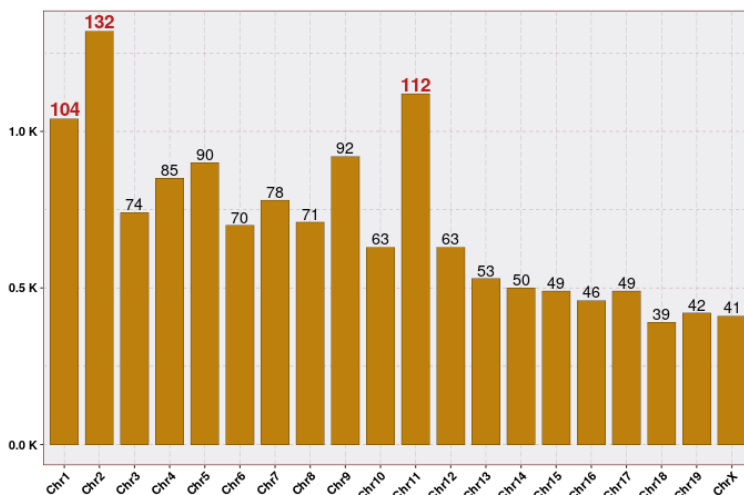
3 pav. Sintenijos blokų *circo* diagrama

Remiantis sukurta diagrama, galima pastebėti, kad tarp viso naminės pelės chromosomų rinkinio nebuvo nei vienos chromosomos, kurios fragmentai nebūtų nustatyti zebražuvės genome. Nepaisant to, naminės pelės fragmentų pozicijos akivaizdžiai nesutapo su zebražuvės genome nustatytomis tų pačių arba evoliucijos eigoje mutavusių fragmentų pozicijomis (naminės pelės fragmentai buvo nustatyti kitose zebražuvės chromosomose).

Zebražuvės genome nustatytas ne vienas homologiškai su naminės pelės genomu susijęs fragmentas, tačiau ketvirtoje zebražuvės chromosomoje (remiantis NCBI[24] duomenimis tai yra didžiausia *Danio rerio* chromosoma, kurią sudaro 78,093,715 bazių poros) sintenijos blokų sandauba pastebima tik vienoje chromosomos pusėje (pavienių blokų nenustatyta).

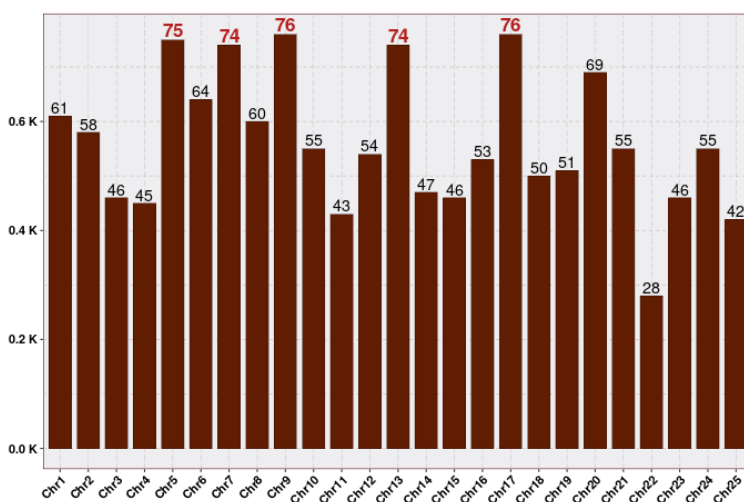
## 6.2 Sintenijos blokų pasiskirstymas chromosomose

Pateiktoje diagramoje vaizduojama, kiek kiekvienos naminės pelės chromosomos fragmentų buvo nustatyta skirtingose zebražuvės chromosomose bei pozicijose, neatsižvelgus į sekų fragmentų ilgį.



4 pav. Fragmentų skaičius *Mus musculus* chromosomose

Pateiktoje stulpelinėje diagramoje išskirtos trys chromosomos (*Chr1*, *Chr2* ir *Chr11*). Šių naminės pelės chromosomų fragmentai pasitaikydavo zebražuvės genome dažniausiai. Atsižvelgus į šių chromosomų dydžius, kur vienuolikta chromosoma yra mažesnė nei kitos išskirtos chromosomos (ją sudaro 122 Mbp[25]), buvo nustatyta sąlyginai didelė dalis fragmentų, turinčių homologijų su zebražuvės genomo sekų fragmentais.



5 pav. Fragmentų skaičius *Danio rerio* chromosomose

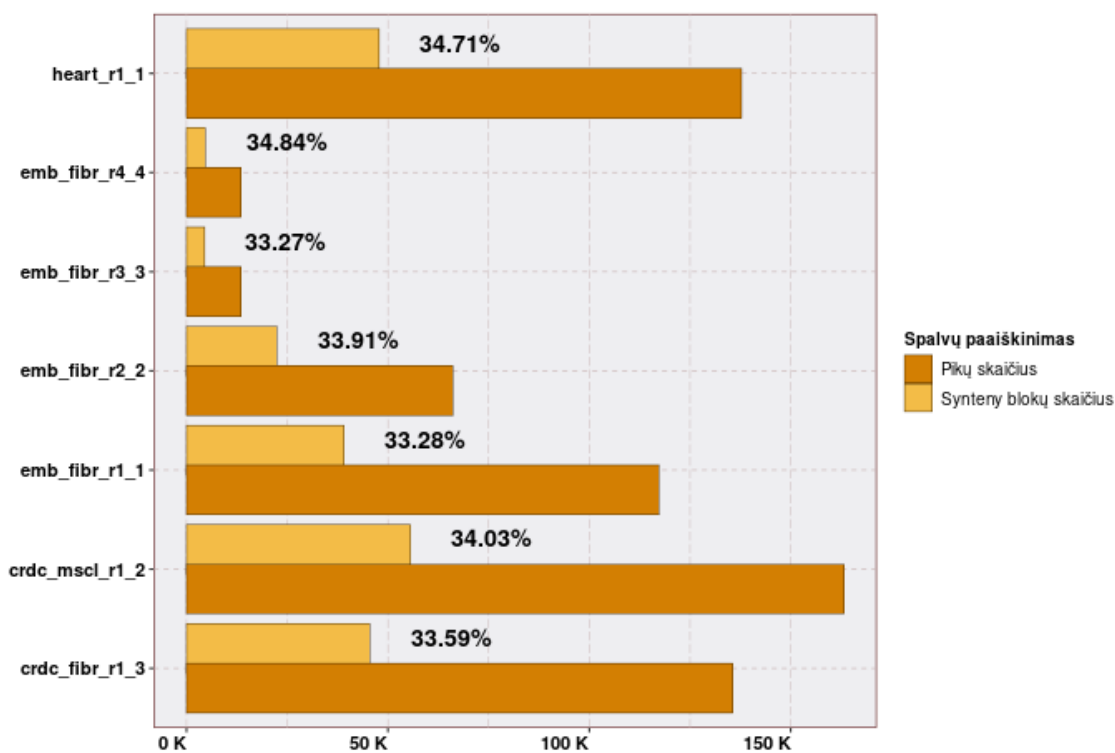
Tuo tarpu patikrinus, kiek yra *Danio rerio* chromosomų fragmentų, kurie sudaro sinteni-



jos blokus, gauti rezultatai vizualizuoti penktame paveiksle (5 pav.). Galima pastebėti, kad sintenijos blokams priklausančių chromosomų fragmentų skaičius buvo itin nevienodas, palyginus su *Mus musculus* rezultatais. Chromosomose, kuriose esantys fragmentai buvo dažniausiai homologiški naminės pelės genomui (*Chr9* ir *Chr7*), buvo nustatyta po 76 sintenijos blokui priklausančius fragmentus. Mažiausias sintenijos blokui priklausančių fragmentų skaičius nustatytas 22 chromosomoje (28 fragmentai).

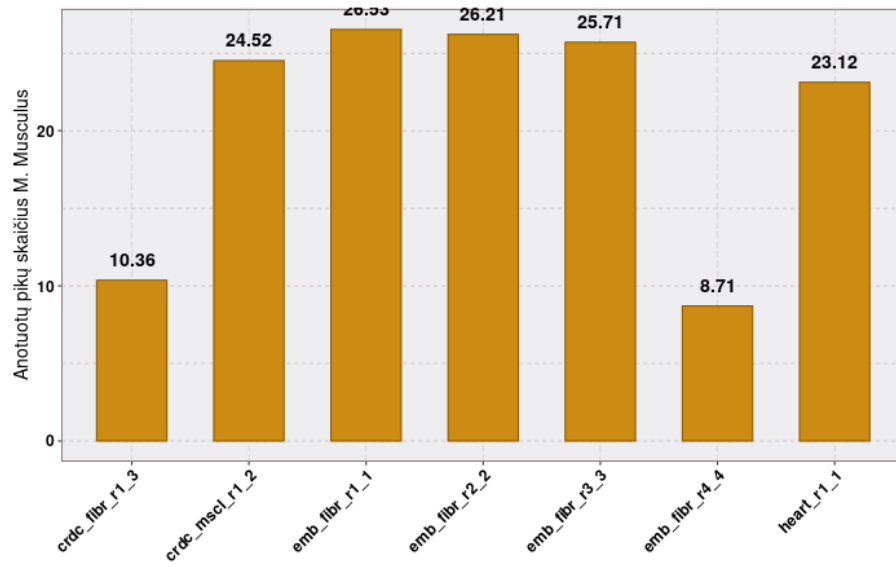
## Cintenų įrankio rezultatas

Įrankio pateikta sintenijos blokų informacija buvo panaudota, siekiant išsiaiškinti, kokia pikų dalis turi persidengimų su *Cintenų* programos nustatytais sintenijos blokais. Iš viso buvo nustatyti 1403 blokai. Gauti rezultatai pateikti 5 pav.

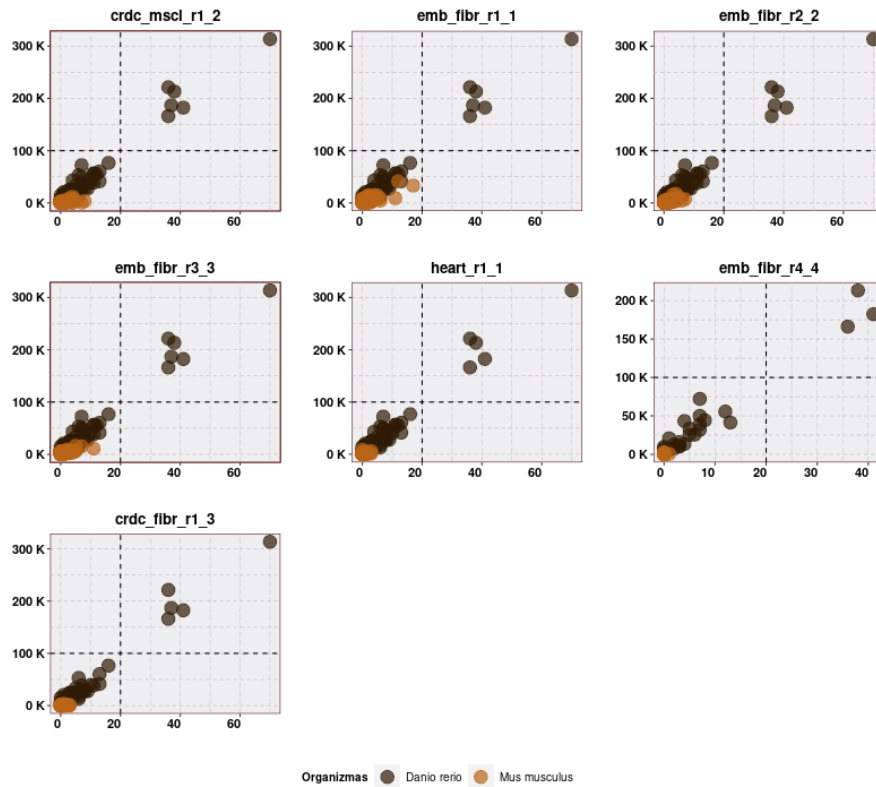


5 pav. Sintenijos blokų procentinė dalis

Remiantis gautu rezultatu galima pastebėti, kad didžiausias nustatytų sintenijos blokų skaičius (11.04%), palyginus su bendru pikų skaičiumi, būdingas *emb\_fibr\_r3\_3* mėginiui, kuriame naminių pelių embrionų fibroblastų ląstelės buvo GMT (GATA4, MEF2C ir Tbx5) kardiogeniniais transkripcijos faktoriais. Mažiausias sintenijos blokų skaičius nustatytas *heart\_r1\_1* mėginyje, kuriame naminės pelės prieširdžių ląstelės buvo du kartus po 24 valandas laikytos mišinyje su retrovirusais, tačiau netaikant papildomų poveikių.



Galbūt čia reikia vis dėlto pateikti lentelę, kurioje nurodyta, koks procentas yra Exon, Intron, Intergenic ir t.t.?



Čia pateikti lentelę, kurioje nurodyti, kiek atitikimų surasta. T.y. kiek sintenų blokų buvo nustatyta tarp pikų (kurie blokai iš gautų su Cinteny įrankiu buvo tinkami ir priklausė pikams.)

2 lentelė.

Mėginys	Blokų skaičius	Procentinė dalis
<i>crdc_mscl_r1_2</i>	39	2.78%
<i>emb_fibr_r1_1</i>	41	2.92%
<i>emb_fibr_r2_2</i>	40	2.85%
<i>emb_fibr_r3_3</i>	40	2.85%
<i>heart_r1_1</i>	39	2.78%
<i>emb_fibr_r4_4</i>	18	1.28%
<i>crdc_fibr_r1_3</i>	25	1.78%

Apibendrinus gautus rezultatus nustatyta sintenijos blokų dalis pikų atžvilgiu buvo maža (neviršijo 20%), todėl sintenijos blokų paieška nėra tinkamiausias būdas, siekiant įvertinti *Mus musculus* ir *Danio rerio* organizmų homologiją, nes taikomas sintenijos blokų paieškos metodas, tikėtina, nesuranda visų galimų homologiškų *Mus musculus* ir *Danio rerio* organizmų genomų fragmentų.

## 7 Išvados

# Literatūra

- [1] GTRD: an integrated view of transcription regulation. Kolmykov S, Yevshin I, Kulyashov M, Sharipov R, Kondrakhin Y, Makeev VJ, Kulakovskiy IV, Kel A, Kolpakov F *Nucleic Acids Res.* 2021 Jan 8;49(D1):D104-D111.
- [2] Nadeau JH, Taylor BA: Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc Nat Acad Sci* 1984, 81: 814–818. 10.1073/pnas.81.3.814.
- [3] Frazer KA, Elnitski L, Church DM, Dubchak I, Hardison RC: Cross-species comparisons: a review of methods and available resources. *Genome Res* 2003, 13: 1–12. 10.1101/gr.222003.
- [4] Clamp M, Andrews D, Barker D, Bevan P, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Hubbard T, Kasprzyk A, Keefe D, Lehvaslaiho H, Iyer V, Melsopp C, Mongin E, Pettett R, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Birney E: Ensembl 2002: accommodating comparative genomics. *Nucleic Acids Res* 2003, 31: 38–42. 10.1093/nar/gkg083.
- [5] Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Ostell J, Miller V, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E: Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2007, (35 Database):D5-D12. 10.1093/nar/gkl1031.
- [6] Matthieu Muffato, Alexandra Louis, Charles-Edouard Poisnel, Hugues Roest Crollius, Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes, *Bioinformatics*, Volume 26, Issue 8, 15 April 2010, Pages 1119–1121, <https://doi.org/10.1093/bioinformatics/btq079>.
- [7] Sinha, A.U., Meller, J. Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. *BMC Bioinformatics* 8, 82 (2007). <https://doi.org/10.1186/1471-2105-8-82>
- [8] Grabherr, M. G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., Di Palma, F., & Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics*, 26(9), 1145-51.
- [9]

- [10] Proost, S.; Fostier, J.; De Witte, D.; Dhoedt, B.; Demeester, P.; Van de Peer, Y.; Vandepoele, K. i-ADHoRe 3.0—Fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res.* 2012, 40, e11.
- [11] Liu, D., Hunt, M. & Tsai, I.J. Inferring synteny between genome assemblies: a systematic evaluation. *BMC Bioinformatics* 19, 26 (2018). <https://doi.org/10.1186/s12859-018-2026-4>
- [12] Lallemand, T.; Leduc, M.; Landès, C.; Rizzon, C.; Lerat, E. An Overview of Duplicated Gene Detection Methods: Why the Duplication Mechanism Has to Be Accounted for in Their Choice. *Genes* 2020, 11, 1046. <https://doi.org/10.3390/genes11091046>
- [13] R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [14] Scikick. Utility for executing collections of computational notebooks. URL [https://petronislab.camh.ca/pub/scikick/stable/docs/report/out\\_html/introduction.html](https://petronislab.camh.ca/pub/scikick/stable/docs/report/out_html/introduction.html).
- [15] Lee J, Hong WY, Cho M, Sim M, Lee D, Ko Y, Kim J. Synteny Portal: a web-based application portal for synteny block analysis. *Nucleic Acids Res.* 2016 Jul 8;44(W1):W35-40. doi: 10.1093/nar/gkw310. Epub 2016 May 6. PMID: 27154270; PMCID: PMC4987893.
- [16] M. Lawrence, R. Gentleman, V. Carey: "rtracklayer: an R package for interfacing with genome browsers". *Bioinformatics* 25:1841-1842.
- [17] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- [18] Wang Q, Li M, Wu T, Zhan L, Li L, Chen M, Xie W, Xie Z, Hu E, Xu S, Yu G (2022). "Exploring epigenomic datasets by ChIPseeker." *Current Protocols*, 2(10), e585. doi: 10.1002/cpz1.585.
- [19] Yu G, Wang L, He Q (2015). "ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization." *Bioinformatics*, 31(14), 2382-2383. doi: 10.1093/bioinformatics/btv145.
- [20] Team BC, Maintainer BP (2019). *TxDb.Mmusculus.UCSC.mm10.knownGene*: Annotation package for TxDb object(s). R package version 3.10.0.
- [21] Carlson M (2022). *org.Mm.eg.db*: Genome wide annotation for Mouse. R package version 3.15.0.
- [22] Magali Ruffier, Andreas Kähäri, Monika Komorowska, Stephen Keenan, Matthew Laird, Ian Longden, Glenn Proctor, Steve Searle, Daniel Staines, Kieron Taylor, Alessandro Vullo,

Andrew Yates, Daniel Zerbino, Paul Flicek Ensembl core software resources: storage and programmatic access for DNA sequence and genome annotation Database 2017, bax020 doi: 10.1093/database/bax020

- [23] M. Lawrence, R. Gentleman, V. Carey: "rtracklayer: an R package for interfacing with genome browsers". *Bioinformatics* 25:1841-1842.
- [24] National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2023 Jan 22]. Available from: <https://www.ncbi.nlm.nih.gov/>
- [25] Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE, the Mouse Genome Database Group. 2019. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* 2019 Jan. 8;47 (D1): D801–D806.

## 8 Priedas

Priedų sąrašė pateikiamos tarpinių rezultatų puslapio, sugeneruoto su Scikick, bei Git repozitorijos, kurioje saugomi analizei naudoti duomenų failai, parašyti skriptai bei pagrindinė R programa, nuorodos.

- **Kursinio projekto analizės Git repozitorija:**

[https://github.com/dansta0804/Tbx5\\_analysis\\_II.git](https://github.com/dansta0804/Tbx5_analysis_II.git)

- **Kursinio darbo Scikick ataskaita:**

[https://karklas.mif.vu.lt/~dast6577/KursinisDarbas/v2.0/peaks\\_MM.html](https://karklas.mif.vu.lt/~dast6577/KursinisDarbas/v2.0/peaks_MM.html)

- **Kursinio darbo analizės Git repozitorija:**

[https://github.com/dansta0804/Tbx5\\_analysis.git](https://github.com/dansta0804/Tbx5_analysis.git)