

<https://doi.org/10.15388/vu.thesis.478>

<https://orcid.org/0000-0003-1726-5825>

VILNIAUS UNIVERSITETAS

Gabrielė Žukauskaitė

Poveikio genomo variantų ir teigiamos gamtinės atrankos veikiamų genomo sričių vertinimas lietuvių populiacijoje

DAKTARO DISERTACIJA

Medicinos ir sveikatos mokslai,
Medicina (M 001)

VILNIUS 2023

Disertacija rengta 2018–2023 metais Vilniaus universiteto Medicinos fakulteto Biomedicinos mokslų instituto Žmogaus ir medicininės genetikos katedroje (VU MF BMI ŽMGK).

Mokslinius tyrimus rėmė Lietuvos mokslo taryba (Nr. S-MIP-20-35; projektas „Apsauginiai DNR mechanizmai – plataus masto Lietuvos Černobylio katastrofos likviduotojų viso genomo tyrimas“, akronimas – ADAPT). Tyrimui naudoti du bendrosios lietuvių populiacijos grupės duomenų rinkiniai, gauti įgyvendinus šiuos projektus:

1) VU MF BMI ŽMGK 2011–2015 m. vykdytą projektą „Lietuvos populiacijos genetinė įvairovė ir sandaros kitimai, susiję su evoliucija ir dažniausiai paplitusiomis ligomis“ (akronimas – LITGEN) (projekto nr. VP1-3.1-ŠMM-07-K-01-013; bioetikos leidimo nr. 158200-05-329-79);

2) VU MF BMI ŽMGK 2020–2022 m. vykdytą projektą „Senovės ir nauji aleliai lietuvių genome: mutacijos, gamtinė atranka ir adaptacija“ (akronimas – ANELGEMIA) (Nr. S-MIP-20-34; bioetikos leidimo nr. 2020/6-1243-724).

Mokslinis vadovas – prof. habil. dr. Vaidutis Kučinskas (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina, M 001).

Mokslinė konsultantė – doc. dr. Laima Ambrozaitytė (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina, M 001).

Gynimo taryba:

Pirmininkas – prof. dr. Vytautas Kasiulevičius (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina, M 001).

Nariai:

prof. habil. dr. Rūta Dubakienė (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina, M 001),

prof. dr. Rimantas Jankauskas (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina, M 001),

prof. dr. Janis Klovins (Latvian Biomedical Research and Study Centre, medicinos ir sveikatos mokslai, medicina, M 001),

prof. habil. dr. Limas Kupčinskas (Lietuvos sveikatos mokslų universitetas, medicinos ir sveikatos mokslai, medicina, M 001).

Disertacija ginama viešame Gynimo tarybos posėdyje 2023 m. birželio mėn. 15 d. 13 val. Vilniaus universiteto ligoninės Santaros klinikų V-oje (325B) auditorijoje. Adresas: Santariškių g. 2, B korpusas, LT-08410 Vilnius, Lietuva, el. paštas gabriele.zukauskaite@mf.vu.lt

<https://doi.org/10.15388/vu.thesis.478>

<https://orcid.org/0000-0001-2345-6789>

VILNIUS UNIVERSITY

Gabrielė Žukauskaitė

Evaluation of Effect Genome Variants and Loci of Positive Selection in the Lithuanian Population

DOCTORAL DISSERTATION

Biomedical Sciences,
Medicine (M 001)

VILNIUS 2023

The dissertation was prepared between 2018 and 2023 at Vilnius University, Faculty of Medicine, Institute of Biomedical Sciences, Department of Human and Medical Genetics (VU MF BMI DHMG). The research was supported by the Research Council of Lithuania (project “Adaptive genetic mechanisms – a comprehensive study of whole genome variation in the group of Lithuanian Chernobyl catastrophe liquidators”, acronym – ADAPT, Project No. S-MIP-20-35).

Two data sets of the general Lithuanian population group were used for the study, which were obtained from:

- 1) VU MF BMI DHMG 2011–2015 project “Genetic diversity and structural changes of the Lithuanian population related to evolution and the most common diseases” (acronym – LITGEN) (Project No. VP1-3.1-ŠMM-07-K-01-013, bioethical approval No. 158200-05-329-79);
- 2) VU MF BMI DHMG 2020–2022 project “Ancient and new alleles in the Lithuanian genome: mutations, natural selection and adaptation” (acronym – ANELGEMIA) (Project No. S-MIP-20-34, bioethical approval No. 2020/6-1243-724).

Academic supervisor –

Prof. Habil. Dr. Vaidutis Kučinskas (Vilnius University, Medical and Health Sciences, Medicine, M 001).

Academic consultant – Assoc. Prof. Dr. Laima Ambrozaitytė (Vilnius University, Medical and Health Sciences, Medicine, M 001).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:

Chairman – Prof. Dr. Vytautas Kasiulevičius (Vilnius University, Medical and Health Sciences, Medicine, M 001).

Members:

Prof. Habil. Dr. Rūta Dubakienė (Vilnius University, Medical and Health Sciences, Medicine, M 001),

Prof. Dr. Rimantas Jankauskas (Vilnius University, Medical and Health Sciences, Medicine, M 001),

Prof. Dr. Janis Klovins (Latvian Biomedical Research and Study Centre, Medical and Health Sciences, Medicine, M 001),

Prof. Habil. Dr. Limas Kupčinskas (Lithuanian University of Health Sciences, Medical and Health Sciences, Medicine, M 001).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 13 (hour) on 15th of June 2023 in meeting room V (325B) of the Vilnius University Hospital Santaros Clinics. Address: Santariškių str. 2, LT-08410 Vilnius, Lithuania. e-mail: gabriele.zukauskaite@mf.vu.lt

SANTRUMPOS (ABBREVIATIONS)

ADAPT – Lietuvos Černobylio katastrofos likviduotojų tiriamųjų grupė, dalyvavusi ADAPT projekte (angl. *data of Lithuanian clean-up workers of Chernobyl nuclear disaster from the ADAPT project*),

ANELGEMIA – bendrosios lietuvių populiacijos tiriamieji, dalyvavę ANELGEMIA projekte (angl. *data of the general Lithuanian population from the ANELGEMIA project*),

CEU – Šiaurės ar Vakarų Europos kilmės Jutos gyventojų populiacija (angl. *the population of Utah of Northern or Western European ancestry*),

Chr. – chromosoma (angl. *chromosome*),

DNR – deoksiribonukleorūgštis (angl. *DNA – deoxyribonucleic acid*),

EUR – bendroji Europos populiacija (angl. *general European population*),

FIN – suomių iš Suomijos populiacija (angl. *Finnish population from Finland*),

GS – galimybių santykis (angl. *OR – odds ratio*),

GWAS – plataus masto genomo asociacijos tyrimai (angl. *genome wide association studies*),

HETERO – heterozigotinių genotipų skaičius (angl. *count of heterozygous genotypes*),

HOMO – homozigotinių genotipų skaičius (angl. *count of homozygous genotypes*),

JS – jonizuojančioji spinduliuotė (angl. *IR – ionizing radiation*),

Kb – kilobazė (1 000 bazių porų) (angl. *kilobase (1,000 base pairs)*),

LEP – lyginamoji Europos populiacija (angl. *EP – the population to which the comparison is made*),

LČKL – Lietuvos Černobylio katastrofos likviduotojai (angl. *LCWC – Lithuanian clean-up workers of Chernobyl nuclear disaster*),

LITGEN – bendrosios lietuvių populiacijos tiriamieji, dalyvavę LITGEN projekte (angl. *data of the general Lithuanian population from the LITGEN project*),

LTU – bendroji lietuvių populiacija (angl. *general Lithuanian population*),

MAF – retojo alelio dažnis (angl. *minor allele frequency*),

Mb – megabazė (1 000 000 bazių porų) (angl. *megabase (1,000,000 base pairs)*),

PI – pasikliautinis intervalas (angl. *CI – confidence interval*),

Tir. – tiriamieji (angl. *Ind. – individuals*),

VNP – vieno nukleotido polimorfizmas (angl. *SNP – single nucleotide polymorphism*),

VU MF BMI ŽMGK – Vilniaus universiteto Medicinos fakulteto Biomedicinos mokslų instituto Žmogaus ir medicininės genetikos katedra (angl. *VU MF BMI DHMG – Vilnius University, Faculty of Medicine, Institute of Biomedical Sciences, Department of Human and Medical Genetics*).

TURINYS

ĮVADAS	10
1. LITERATŪROS APŽVALGA	14
1.1. Adaptacijos ir gamtinės atrankos ryšys.....	14
1.2. Poveikio genomo variantai ir jų savybės	18
1.3. Gamtinės atrankos nustatymo metodai	20
1.4. Teigiamos gamtinės atrankos tyrimai pasaulyje ir Lietuvoje	24
1.5. Išorinių aplinkos veiksnių įtaka genomui ir adaptacijai – Černobylio katastrofos pavyzdys	26
2. TIRIAMIEJI IR TYRIMO METODAI	33
2.1. Tyrimo eiga	33
2.2. Tiriamieji	34
2.3. Lietuvos Černobylio katastrofos likviduotojų anketinės apklausos duomenų rinkimas ir analizės strategija	36
2.4. Genotipavimas, naudojant VNP žymenų lustus	36
2.4.1. Genotipavimo duomenų analizė.....	37
2.5. Viso egzomo ir genomo sekoskaita	43
2.5.1. Bendrosios lietuvių populiacijos (LITGEN) viso egzomo sekoskaita	43
2.5.2. Bendrosios lietuvių populiacijos (ANELGEMIA) ir Lietuvos Černobylio katastrofos likviduotojų viso genomo sekoskaita	43
3. TYRIMO REZULTATAI IR JŲ APTARIMAS	47
3.1. Lietuvos Černobylio katastrofos likviduotojų anketinės apklausos ir ligų istorijų duomenų aprašomoji statistinė analizė	47
3.2. Bendrosios lietuvių populiacijos (LITGEN) poveikio genomo variantų analizė	49
3.3. Lietuvos Černobylio katastrofos likviduotojų poveikio genomo variantų asociacijų analizė	54
3.3.1. Lietuvos Černobylio katastrofos likviduotojų ir bendrosios lietuvių populiacijos (LITGEN) tiriamųjų grupių poveikio genomo variantų asociacijų analizė	54
3.3.2. Lietuvos Černobylio katastrofos likviduotojų ir bendrosios lietuvių populiacijos (LITGEN) vyrų, suskirstytų pagal amžiaus grupes, poveikio genomo variantų asociacijų analizė	58

3.3.3. Lietuvos Černobylio katastrofos likviduotojų ir bendrosios lietuvių populiacijos (LITGEN) 50–78 m. vyrų visų genotipavimo duomenų asociacijų analizė	60
3.4. Lietuvos Černobylio katastrofos likviduotojų ir bendrosios lietuvių populiacijos (LITGEN) 50–78 m. vyrų kryptinga, su ligomis siejamų genų analizė	60
3.4.1. Psichikos būklių asociacijų analizė	61
3.4.2. Burnos ertmės ligų asociacijų analizė	66
3.5. Bendrosios lietuvių populiacijos ir Lietuvos Černobylio katastrofos likviduotojų teigiamos gamtinės atrankos veikiamų genomo sričių analizė	69
3.5.1. Teigiamos gamtinės atrankos įtakos įvertinimas nustatytiems poveikio genomo variantams bendrojoje lietuvių populiacijoje	69
3.5.2. Teigiamos gamtinės atrankos įtakos įvertinimas nustatytiems poveikio genomo variantams Lietuvos Černobylio katastrofos likviduotojų grupėje	71
3.5.3. Teigiamos gamtinės atrankos veikiamų genomo sričių nustatymas ir reikšmės adaptacijai įvertinimas Lietuvos Černobylio katastrofos likviduotojų grupėje	74
IŠVADOS	81
LITERATŪROS SĄRAŠAS (REFERENCES)	83
MOKSLINIO DARBO APROBACIJA	100
PRIEDAI	103
SUMMARY	148
PADEKA	183
CURRICULUM VITAE	184

ĮVADAS

Tiriamoji problema ir jos aktualumas

Išgyvenamumas ir prisitaikymas prie besikeičiančių aplinkos sąlygų yra svarbūs populiacijų adaptacijos komponentai. Jie tiesiogiai priklauso nuo veikiančių gamtinės atrankos jėgų.

Gamtinė atranka – pagrindinis evoliucijos veiksnys, turintis galią keisti populiacijos genetinę architektūrą. Molekuliniu lygmeniu veikiant populiacijų fenotipus, adaptacines savybes, struktūrą, buvo formuojamos praeities ir bus formuojamos ateities populiacijos. Specifinės genomo sritys ir jų variacija, susijusi su išgyvenamumu, tarp populiacijų skiriasi, galimai, dėl vykstančių skirtingų mikroevoliucinių procesų, o tai lemia skirtingas populiacijų savybes. Siekiant išsiaiškinti populiacijos unikalumą lemiančius genomo veiksnius, svarbu nustatyti ir tirti genomo variantus (naudingus ir patogeninius) ir Lietuvos populiacijoje. Taip pat svarbu identifikuoti populiacijai specifines gamtinės atrankos veikiamas genomo sritis, kurių nustatymas naudingas populiacijos adaptyvumą determinuojančioms genomo sritims įvardinti. Nustačius populiacijos išskirtinumą lemiančius genomo veiksnius, gauti rezultatai galėtų būti panaudoti populiacijos specifinei diagnostikai ir terapiniams taikiniams identifikuoti asmeninės medicinos eroje.

Aplinkos sąlygų kaita – svarbus gamtinės atrankos komponentas. Aplinkai keičiantis ir vykstant kataklizminiams reiškiniams, tokiems kaip ekologinės katastrofos, apsaugantys genomo variantai gali tapti žalingi, todėl vyksta mikroevoliuciniai procesai, galintys lemti prisitaikymą prie populiacijos genetinės architektūros virsmo. Lietuva unikali tuo, kad yra patyrusi Černobylio¹ atominės elektrinės katastrofos padarinius ir dalyvavusi juos likviduojant. Tarp likusių gyvų Lietuvos Černobylio katastrofos likviduotojų yra ne tik sergančių, bet ir santykinai sveikai senstančių asmenų. Jie vertingi kaip tyrimo objektas, nes išgyveno ekstremaliomis sąlygomis ir adaptavosi prie visą gyvenimą besitęsiančių jonizuojančiosios spinduliuotės padarinių ir Černobylio katastrofos metu patirtų išgyvenimų. Potencialiai jų išgyvenamumą ir prisitaikymą gali lemti unikali genomo variacija, atsiradusi prieš katastrofą. Šios ir kitų unikalių populiacijų genomo variacijos ir teigiamos gamtinės atrankos veikiamų genomo sričių tyrimai gali padėti

¹ Lietuvių kalboje taip pat vartotina šio vietovardžio forma *Čornobyliš* (sulietuvinta forma iš ukrainiečių (ukrainų) kalbos). Šiame darbe pasirinkta vartoti plačiau paplitusią, lietuvių kalboje nusistovėjusią sulietuvintą vietovardžio formą, transkribuotą iš rusų kalbos, – *Černobyliš*.

identifikuoti išgyvenamumui ir adaptacijai reikšmingas geno sritis ir geriau suvokti vykstančius mikroevoliucinius procesus.

Šiame tyrime atliktas geno variantų, kurie gali būti tiek naudingi, tiek žalingi, t.y. *poveikio* geno variantų, vertinimas bendrosios lietuvių populiacijos ir Lietuvos Černobylio katastrofos likviduotojų tiriamųjų grupėse. Siekiama nustatyti geno variantus, galimai turinčius įtakos minėtų grupių adaptacinėms savybėms (sergamumui, atsakui į patirtą stresą ir į jonizuojančiosios spinduliuotės poveikį). Bendrojoje lietuvių populiacijos tiriamųjų grupėje asociacijų analizės metu nustatytų poveikio geno variantų galimai protekcinėi funkcijai vertinti ne tik pasirinkta naudoti *in silico* variantų funkcinio vertinimo metodus, bet ir siekta atlikti teigiamos gamtinės atrankos veikiamų geno sričių analizę. Tai nauja tyrimo strategija, sujungianti asociacijos ir gamtinės atrankos tyrimus. Tyrimo metu pirmą kartą nustatytos teigiamos gamtinės atrankos veikiamos geno sritys, unikali Lietuvos Černobylio katastrofos likviduotojų tiriamųjų grupė, palyginti su bendrąja lietuvių populiacijos grupe. Šios sritys galimai lemia likviduotojų išgyvenamumą ir adaptaciją prie ekstremalių sąlygų bei visą gyvenimą besitęsiančių Černobylio katastrofos padarinių.

Tyrimo naujumas

Atliekant tyrimą, remiantis bendrosios lietuvių populiacijos plataus masto VNP žymenų genotipavimo ir egzomo duomenimis ir naudojant Lietuvos Černobylio katastrofos likviduotojų tiriamųjų grupės plataus masto VNP žymenų genotipavimo duomenis, abiejose tiriamųjų grupėse tikslingai analizuoti poveikio geno variantai, galimai turintys įtakos populiacijos adaptacinėms savybėms. Tyrimo metu pritaikyta inovatyvi tyrimo strategija, variantų funkcinei reikšmei vertinti naudojant ne tik *in silico* analizės įrankius, bet ir teigiamos gamtinės atrankos analizės metodą.

Tyrimo metu, siekiant nustatyti žmogaus adaptacijai ir išgyvenamumui svarbias geno sritis, veikiant ekstremalioms aplinkos sąlygoms, Lietuvos Černobylio katastrofos likviduotojų ir bendrosios lietuvių populiacijos viso geno sekoskaitos duomenys pirmą kartą panaudoti teigiamos gamtinės atrankos veikiamoms geno sritims išaiškinti. Teigiamos gamtinės atrankos poveikiui vertinti naudotas metodas santykinai naujas, apimantis keletą svarbių teigiamos gamtinės atrankos žymių ir analizuojantis jas vienu metu, taip kompleksiskai įvertinant teigiamos gamtinės atrankos poveikį. Tyrimo metu pirmą kartą ištirti Lietuvos Černobylio katastrofos likviduotojų genomai molekulinio lygmeniu. Aukštas jonizuojančiosios spinduliuotės dozės gavusių asmenų genomai didelės skiriamosios gebos lygmeniu iki šiol mažai

analizuoti. Naujas, gilesnis genomo analizės lygmuo ir identifikuotos genomo sritys, potencialiai lemiančios geresnę individų adaptaciją, leidžia geriau suvokti vykstančius mikroevoliucijos procesus.

Tyrimo tęstinumas

Tyrimo metu sukaupti DNR mėginių ir anketinės apklausos duomenų rinkiniai bei gauti plataus masto vieno nukleotido polimorfizmų (VNP) žymenų genotipavimo ir viso genomo sekoskaitos duomenys leidžia juos panaudoti pakartotinai, suteikia galimybę toliau plėtoti mokslinius tyrimus. Lietuvos Černobylio katastrofos likviduotojams nustačius kandidatinius protekcinius genomo variantus ir kandidatines teigiamos gamtinės atrankos veikiamas genomo sritis bei įvertinus jų reikšmę *in silico* įrankiais, galima tikslinti jų reikšmę funkciniais metodais.

Disertacijos tikslas

Remiantis bendrosios lietuvių populiacijos ir Lietuvos Černobylio katastrofos likviduotojų tiriamųjų grupių plataus masto VNP žymenų genotipavimo ir viso genomo sekoskaitos duomenimis, nustatyti ir įvertinti kandidatinių poveikio genomo variantų ir teigiamos gamtinės atrankos veikiamų genomo sričių reikšmę adaptacijai.

Disertacijos uždaviniai

1. Sudarius kandidatinių poveikio genomo variantų katalogą, nustatyti ir palyginti alelių ir genotipų dažnius bendrosios lietuvių populiacijos ir Lietuvos Černobylio katastrofos likviduotojų tiriamųjų grupėse. Remiantis plataus masto VNP žymenų genotipavimo ir egzomo sekoskaitos duomenimis, identifikuoti poveikio genomo variantus, svarbius abiejų grupių tiriamųjų adaptacijai.
2. Remiantis viso genomo sekoskaitos duomenimis, nustatyti Lietuvos Černobylio katastrofos likviduotojų tiriamųjų grupei unikalias, palyginti su bendrąja lietuvių populiacija, teigiamos gamtinės atrankos veikiamas kandidatines genomo sritis.
3. Įvertinti teigiamos gamtinės atrankos intensyvumą nustatytiems kandidatiniams poveikio genomo variantams bendrosios lietuvių populiacijos ir Lietuvos Černobylio katastrofos likviduotojų tiriamųjų grupėse.

4. Taikant bioinformatikos įrankius ir remiantis moksline literatūra, įvertinti atrinktų poveikio geno variantų ir teigiamos gamtinės atrankos veikiančių sričių funkcijas ir nustatyti jų reikšmę fenotipui.

Ginamieji teiginiai

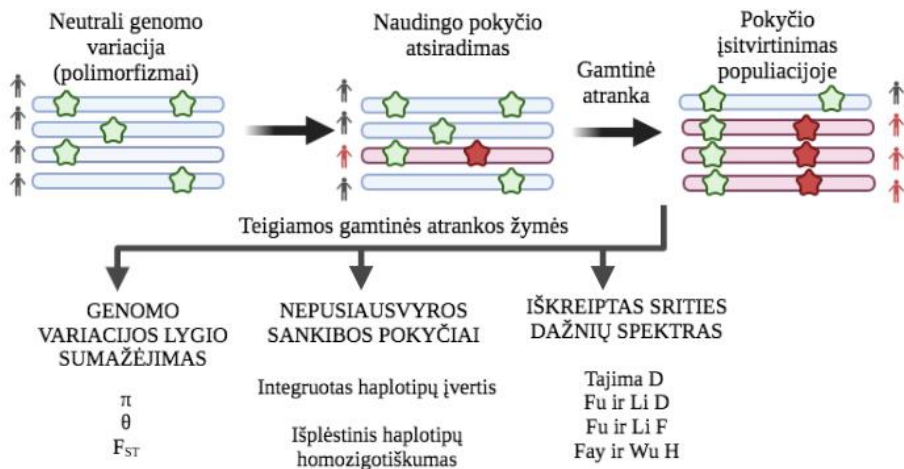
1. Bendrosios lietuvių populiacijos ir Lietuvos Černobylio katastrofos tiriamųjų grupių genuose poveikio geno variantai formuoja tiriamosioms grupėms specifines fenotipo ir adaptacines savybes.
2. Poveikio geno variantų funkcinis reikšmingumas gali būti įvertintas naudojant ne tik tradicinius *in silico* analizės įrankius, bet ir teigiamos gamtinės atrankos poveikį vertinančius metodus.
3. Dėl patirto ekstremalaus aplinkos poveikio spaudimo Lietuvos Černobylio katastrofos likviduotojų genuose yra unikalių teigiamos gamtinės atrankos veikiančių sričių, kurios galimai lemia likviduotojų išgyvenamumą ir adaptacines savybes.

1. LITERATŪROS APŽVALGA

1.1. Adaptacijos ir gamtinės atrankos ryšys

Išgyvenamumas ir išlikimas neatsiejami nuo prisitaikymo prie besikeičiančių aplinkos sąlygų – adaptacijos. Adaptacija glaudžiai susijusi su gamtine atranka, kuri yra vienas iš pagrindinių evoliucijos veiksnių, galinčių keisti genomo struktūrą.

Populiacijų genomai gali adaptuotis prie besikeičiančios aplinkos trimis būdais. Populiacijai naudingas alelis gali kilti įvykus *de novo* pokyčiui, t. y. iš polimorfizmo, esančio genome jau kurį laiką (angl. *standing variation*), arba atsirasti priemaišos (angl. *admixture*) su kitomis populiacijomis būdu. Suprasti alelio kilmę svarbu, siekiant išsiaiškinti, kaip adaptacija veikia ir kodėl skirtingų gamtinės atrankos poveikių metu genome paliekama skirtingų žymių (angl. *sweep signatures*). Nuo to taip pat priklauso ir tolesnių tyrimo metodų parinkimas šių genomo sričių analizei (1.1 pav.).



1.1 pav. Teigiamos gamtinės atrankos poveikio ir žymių schema. Prieš veikiant teigiamai atrankai, genome susiformuoja tam tikras genomo variacijos profilis. Atsiradus naudingam pokyčiui (raudona žvaigždutė) ir veikiant teigiamai gamtinei atrankai, šis pokytis perduodamas populiacijoje, kol įsitvirtina arba jo dažnis tampa netolimas fiksicijai (dešinė paveikslėlio dalis). Specifinę gamtinės atrankos žymę sukuria ir genomo variantai, esantys nepusiausvyroje sankiboje su naudingumu aleliu, todėl jų dažnis taip pat padidėja (žalios žvaigždutės greta raudonos). Išskiriamos trys pagrindinės teigiamos gamtinės atrankos žymės. Joms nustatyti naudojami skirtingi statistiniai metodai (F_{ST} , integruotas haplotipų įvertis, Tajima D ir kt.).

Teigiamos gamtinės atrankos poveikis *de novo* pokyčiui pagrįstas faktu, kad naujai susidaręs alelis yra naudingas populiacijai atsiradimo momentu,

nes selekcinis spaudimas (angl. *selective pressure*) atsiranda dar prieš pokytį. Šio naujai atsiradusio pokyčio dažnis populiacijoje sparčiai didėja. Vis dėlto minėtina, kad naudingas pokytis turi atsirasti tinkamu metu tinkamoje populiacijoje, kitaip jis gali būti prarandamas genetinio dreifo metu [1].

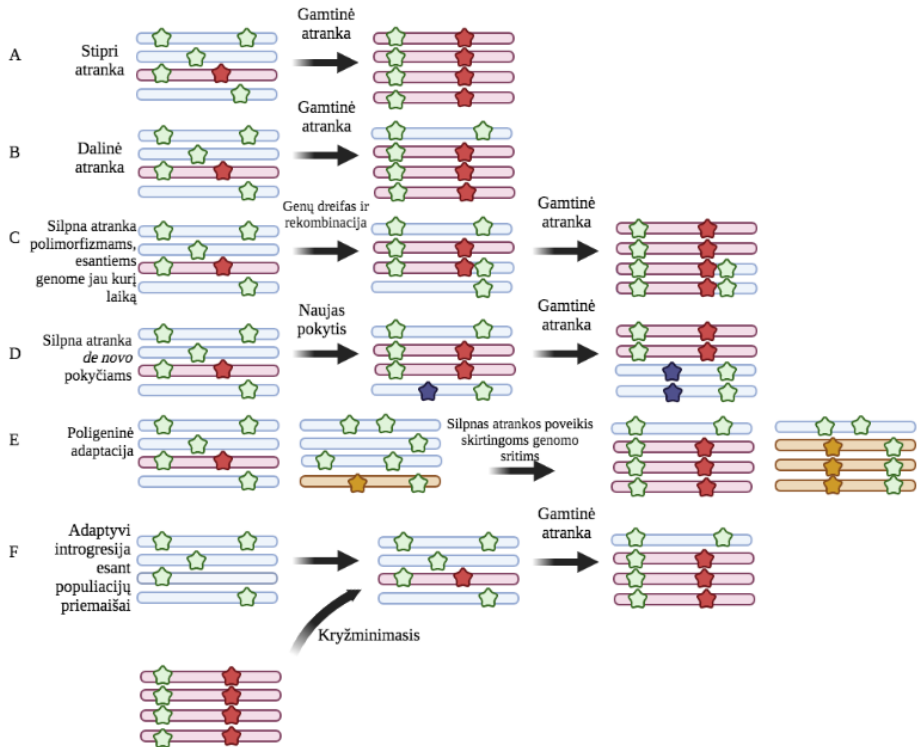
Teigiamos gamtinės atrankos poveikis polimorfizmui, esančiam genome jau kurį laiką, įvyksta, kai jau egzistuojanti genomo variacija tampa naudinga veikiant išorės veiksniams, dažniausiai aplinkos veiksniams kritiškai pasikeičiant [2]. Šiuo atveju teigiamos gamtinės atrankos veikiamas alelis yra atsiradęs anksčiau negu selekcinis spaudimas. Atranka teigiamai veikia alelius, kurie yra silpnai naudingi arba žalingi (dėl staiga pasikeitusio selekcinio spaudimo) arba yra palaikomi balansuojančios gamtinės atrankos. Pastarieji aleliai turi didelį potencialą greitai prisitaikyti prie besikeičiančios aplinkos, tiesiogiai veikia fenotipą ir yra pakankamai didelio dažnio, dar prieš pradėdant veikti selekciniam spaudimui [3].

Genetinės priemaišos su kitomis populiacijomis būdu populiacijoje taip pat gali atsirasti naudingų genomo variantų. Šis procesas reguliuojamas genų tėkmės (angl. *gene flow*) [4]. Klasikiniai šių procesų išdavos pavyzdžiai yra laktozės toleravimas ir pjautuvinė anemija. Aleliai, lemiantys geresnį laktozės metabolizmą, plito iš ankstyvųjų Europos ganytojų populiacijų, jie nustatomi ir šių laikų Europos populiacijose [5]. Afrikinės kilmės *Duffy* kraujo grupės aleliai, veikiami stiprios teigiamos atrankos, lėmė atsparumą maliarijai šių laikų malagasių (Madagaskaras) [6] ir kitose Afrikos populiacijose.

Teigiama gamtinė atranka, veikdama genome, palieka specifinių žymių, kurios skiriasi priklausomai nuo fenotipo genetinio mechanizmo, alelio kilmės ir atrankos stiprumo bei greitumo. Vykstant monogeninei adaptacijai, nustatomas stiprus teigiamos gamtinės atrankos poveikis vienoje genomo srityje. Šios adaptacijos atveju teigiama gamtinė atranka skirstoma į stipriąją (angl. *hard sweep*) ir silpnąją (angl. *soft sweep*). Stiprioji teigiama gamtinė atranka nustatoma atrankai veikiant *de novo* pokyčius (1.2 pav. A ir B). Silpnoji nustatoma veikiant *de novo* pokyčius arba polimorfizmus, egzistuojančius genome jau kurį laiką ir tampančius naudingais veikiant išorės jėgai (1.2 pav. C ir D) [7].

Alelio, veikiamo gamtinės atrankos, dažnis priklauso nuo atrankos stiprumo ir kilmės, todėl neįprastai didelis alelių dažnių svyravimas laikomas viena iš universalios stipriosios gamtinės atrankos paliekamų žymių [1]. Kartu su teigiamos gamtinės atrankos veikiamu aleliu nepusiausvyros sankibos būdu susijusių alelių dažniai linkę sparčiai didėti, o greta jų buvusi genomo variacija linkusi mažėti [8]. Tai antroji teigiamos gamtinės atrankos žymė, mokslinėje literatūroje ji įvardijama kaip *iškreiptas srities dažnių*

spektras (angl. *site frequency spectrum*; matas, apibūdinantis retųjų alelių dažnių pasiskirstymą).



1.2 pav. Teigiamos gamtinės atrankos žymių tipai. Paveiksle vaizduojamas DNR sekos fragmentas populiacijoje. Žvaigždutėmis pažymėti DNR sekos variantai ir jų pasiskirstymas bei išvirtinimas populiacijos viduje, veikiant gamtinei atrankai (sukurta naudojantis *BioRender.com* įrankiu).

Teigiamai atrankai veikiant *de novo* pokyčius, greta pokyčio esantys geno regionai yra nepusiausvyroje sankiboje su gamtinės atrankos veikiamu aleliu, todėl geno srčiai būdingas didesnis haplotipo homozigotiškumas populiacijoje (1.2 pav. A), ypač ilguose geno regionuose [1]. Pastarąjį geriausiai iliustruoja laktozės toleravimo pavyzdys. *LCT* genas, koduojantis laktozę, patenka į haplotipų bloką, kuriame Europos ir Afrikos populiacijose yra didelis haplotipo homozigotiškumas. Tokiame haplotipe nustatyti funkcionalūs geno pokyčiai, susiję su laktozės metabolizmu. Dėl šių pokyčių ir haplotipo homozigotiškumo populiacijoje galimas laktozės metabolizmas, o tai buvo naudinga praeities populiacijoms, kurios vertėsi gyvulininkyste ir kurioms pieno produktai buvo svarbus maistinių medžiagų ir vandens šaltinis (ypač populiacijoms, kurios gyveno sausringose vietovėse) [9].

Teigiamai atrankai veikiant polimorfizmus, esančius genome jau kurį laiką, šios aptartos teigiamos gamtinės atrankos žymės nustatomos silpnesnės ir variabilesnės (1.2 pav. C ir D), priklausomai nuo veikiamo alelio dažnio ir laiko, nuo kada ši alelį pradėjo veikti gamtinės atrankos jėgos. Taigi mokslinėje literatūroje nagrinėjamas diferencijavimo tarp stiprios ir silpnos atrankos sudėtingumas [2].

Čia svarbu paminėti ir poligeninę adaptaciją (1.2 pav. E). Ji vyksta, kai keli ar keliasdešimt alelių populiacijoje veikiami gamtinės atrankos. Tai lemia suderintus alelių dažnių pokyčius ir, savo ruožtu, fenotipo pokyčių populiacijoje. Poligeninė adaptacija yra dažnas procesas, nes dauguma žmogaus požymių yra poligeniniai, todėl ši adaptacijos rūšis paaiškina tokius požymius, kaip žmonių mitybos įpročiai ar atsparumas patogenams [10–12]. Poligeninės adaptacijos žymės įprastai yra labai variabilios, silpnos, rodančios daugybinius nedidelio dažnio poslinkius populiacijoje, kurie gali vykti skirtingu metu. Dėl šios priežasties poligeninę adaptaciją sunku atskirti nuo neutralaus dreifo [13]. Teigiamos gamtinės atrankos poveikis nustatomas ir vykstant adaptyviai introgresijai, kai naudingi aleliai atsiranda priemaišos būdu, vykstant tarppopuliaciniam kryžminimuisi (1.2 pav. F). Naudingasis alelis paplinta populiacijoje ir artėja link fiksacijos [14].

Adaptacija prisideda prie genetinės ir fenotipinės žmonių diferenciacijos tarp populiacijų grupių, įskaitant ir su sveikata susijusius požymius, kurie lemia tarp populiacijų egzistuojančius sveikatos skirtumus. Specifinės genomo sritys ir variantai, susiję su išgyvenamumu, skiriasi tarp populiacijų galimai dėl vykstančių skirtingų mikroevoliucinių procesų. Besikeičiančioje aplinkoje genomo variantai, kurie atliko protekcinę genomo funkciją, gali tapti žalingi, todėl mikroevoliuciniai procesai turi galią keisti populiacijos genetinę architektūrą [15]. Genomo nesugebėjimas prisitaikyti gali lemti ir daugiaveiksnių požymių, lemiančių didesnę populiacijos mirtingumą, tokių kaip hipertenzija [16], kraujo krešėjimo pokyčiai [17], hiperlipidemija [18], pasireiškimą, todėl gamtinės atrankos veikiamų genomo sričių tyrimai yra svarbūs specifinių (geografinių ir etniniu požiūriu) populiacijų diagnostikai ir gydymui [19] tiksliosios (personalizuotos) medicinos eroje. Be to, šių laikų žmogus gyvena nevienalypėje aplinkoje ir yra spaudžiamas skirtingų gamtinės atrankos jėgų, lemiančių skirtingas genomo adaptacines savybes prie aplinkos sąlygų. Tai apima prisitaikymą prie klimato sąlygų ir ultravioletinės spinduliuotės poveikio [1], ligų [20], mitybos įpročių [21], geografinių aplinkos ypatybių (pavyzdžiui, aukščio virš jūros lygio) [22]. Skirtinga adaptacija prie šių veiksnių formuoja svarbius genetinius ir fenotipinius skirtumus tarp populiacijų. Dėl šių priežasčių svarbu nustatyti ir tirti genomo variantus (naudingus ir žalingus) ir Lietuvos populiacijoje, siekiant nustatyti

populiacijos unikalumą lemiančius veiksnius. Galiausiai, populiacijai specifinių gamtinės atrankos veikiamų geno sričių identifikavimas ir funkcinio vaidmens nustatymas naudingas populiacijai fenotipiškai svarbioms geno sritims išaiškinti ir prisideda prie biologinių adaptacinių procesų sampratos.

1.2. Poveikio geno variantai ir jų savybės

Siekiant išsiaiškinti ir suprasti mechanizmus, lemiančius daugiaveiksnių požymių pasireiškimą, svarbu ne tik atlikti variantų funkcinę analizę, bet ir nagrinėti gamtinės atrankos ir adaptacijos procesus per geno variantų dažnių svyravimo prizmę. Gilinantis į šiuos procesus, reikia suvokti, kad kai kurie geno variantai negali būti tiesiogiai skirstomi į žalingus arba apsaugančius (protekcinus). Variantas, kuris gali būti žalingas ir lemiantis vieno daugiaveiksnių požymio atsiradimą, gali būti protekcinis kito požymio atžvilgiu, todėl, šiame tyrime, daugiafunkciai geno variantai įvardyti *poveikio* geno variantais. Šių geno variantų dažnis, poveikis koduojamo baltymo struktūrai ir funkcijai bei varianto evoliucinė kilmė yra esminiai aspektai, kuriuos reikėtų nagrinėti siekiant geriau suprasti daugiaveiksnių požymių, ligų mechanizmus ir populiacijoje vykstančius mikroevoliucinius procesus.

Visų pirma, poveikio geno variantai įprastai yra reti, tačiau lemiantys požymius, kurie padidina asmens pranašumą populiacijoje ir geresnes adaptacines savybes. Šie variantai populiacijoje linkę tapti vis dažnesni ir taip įsitvirtinti. Dauguma minėtų variantų nustatomi ne tik koduojančiose, bet ir nekoduojančiose ar biologiškai nefunkcionaliose geno srityse. Taip išvengiama gryninančios gamtinės atrankos įtakos ir išlaikomi aukšti poveikio geno variantų dažniai skirtingose populiacijose [23]. Tai ypač aktualu žalingų geno variantų atveju ir paaiškina, kodėl kai kurių ligų dažniai populiacijose yra aukšti, nors šiose populiacijose yra poveikio geno variantų, saugančių nuo minėtų ligų. Pavyzdžiui, jei asmuo turi poveikio geno variantą, apsaugantį nuo nutukimo, šis asmuo turi mažesnę tikimybę susirgti ir didesnę tikimybę perduoti savo požymį palikuonims, veikiant teigiamai gamtinei atrankai. Naudingo poveikio geno varianto alelio dažnis padidėja, o su variantu susijusioje geno srityje neutrali variacija sumažėja, taip sukurdamą gamtinės atrankos žymę (angl. *selective sweep*). Remiantis šia logika, tam tikrų daugiaveiksnių ligų dažnis laikui bėgant turėtų mažėti, tačiau, būtent dėl poveikio geno variantų aukšto dažnio palaikymo mažiau reikšmingose ir rečiau analizuojamose geno srityse, ligų dažniai išlieka pastovūs ir aukšti. Be to, aukštas ligų dažnis galėtų būti paaiškintas ir

eksponentiniu žmonių populiacijos didėjimu, kurio rezultatas – itin retų ir sunkiai nustatomų poveikio geno variantų atsiradimas ir kaupimasis [24]. Tokiems variantams analizuoti dažniausiai atliekami plataus masto geno asociacijos tyrimai (angl. *genome wide association studies*, GWAS). Ši tyrimo strategija leidžia nustatyti sąsają tarp geno varianto ir fenotipo, tačiau vienas iš metodo ribotumų yra retų (kai retojo alelio dažnis 0,5–1 proc.) ir itin retų (kai retojo alelio dažnis <0,5 proc.) geno variantų nustatymas ir interpretacija. Būtent šie nepakankamai nagrinėti geno variantai galėtų paaiškinti dalį trūkstamo paveldimumo daugiaveiksnėms ligoms ir požymiams [25]. Taip pat tokie reti geno variantai GWAS metu gali būti klaidingai charakterizuojami kaip žalingi ar didinantys riziką [26]. Taigi poveikio geno variantų funkcinei reikšmei pagrįsti svarbus kompleksinis tyrimo metodų naudojimas, papildant asociacijų analizės rezultatus evoliucinių ir funkcinių tyrimų (*in silico*) rezultatais, kaip tai buvo atlikta šiame tyrime.

Geno variantų dažnis nėra vienintelis kriterijus, kuriuo remiantis galima identifikuoti poveikio geno variantus. Aiškinantis jų poveikį, svarbu, kad variantai koduojančiose geno dalyse turėtų poveikį koduojamai aminorūgščiai ar kitaip veiktų koduojamo baltymo struktūrą. Dauguma poveikio geno variantų, atliekančių protekcinę funkciją, veikia baltymą per funkcijos praradimo ar geno išjungimo (angl. *knockout*) mechanizmus, taip sukeldami tam tikrus fenotipinius požymius. Tačiau taip dauguma funkciškai reikšmingų protekcinų geno variantų gali būti prarandami veikiant gryninančiai gamtinei atrankai [27]. Visa tai gali paaiškinti ir šių geno variantų retumą populiacijose. Kita vertus, naujuose moksliniuose tyrimuose kalbama ir apie koduojamos aminorūgšties nekeičiančių (angl. *synonymous*) geno variantų įtaką geno koduojamo baltymo kiekiui – nustatyta, kad egzistuoja veiklieji (angl. *optimal*) kodonai, kurie, nors ir nekeičia koduojamos aminorūgšties, leidžia ląstelei greičiau vykdyti transkripcijos ir translacijos procesus, taip padidindami koduojamo baltymo produkcijos pajėgumus [28]. Tai įrodo, kad ir koduojamos aminorūgšties nekeičiantys geno variantai, kurie įprastai laikomi nefunkcionaliais, netiesioginiais veikimo būdais gali lemti tam tikrų požymių pasireiškimą populiacijoje, todėl turėtų būti atidžiau tiriami.

Poveikio geno variantui nustatyti labai svarbi alelio kilmė. Kiekvienas toks variantas susideda iš dviejų alelių, iš kurių vienas yra laikomas ligą ar požymį lemiančiu, o kitas – protekcinu, t. y. daroma klaidinga prielaida, kad ligos ar požymio nelemiantis alelis yra protekcinis [19]. Tam, kad būtų galima daryti tokią išvadą apie alelio protekcinę savybę, svarbu atsigręžti į

evoliucinę šio alelio kilmę. Retasis (angl. *minor*) alelis ne visais atvejais yra išvestinis (angl. *derived, mutant*). Kilmė gali būti nustatoma lyginant dominančius alelius tarp skirtingų rūšių, pavyzdžiui, žmonių ir primatų. Jei retasis alelis yra protekcinis, jis turėtų būti išvestinis, nes protėvinės kilmės (angl. *ancestral*) aleliai yra paveldėti iš kitų rūšių evoliucijoje ir yra konservatyvūs funkcinio požiūriu [19]. Taigi, jei nustatytas retasis alelis nustatomas ir sutampa tarp skirtingų rūšių, tikėtina, kad jis yra neutralus. Jei retasis alelis būdingas tik tiriamai rūšiai, bet nėra tapatus su kitomis evoliucine prasme artimomis rūšimis, tikėtina, kad jis turi tam tikrą rūšiai specifinę funkcinę reikšmę.

Poveikio genomo variantų nustatymas yra kompliktuotas ir nevienareikšmis, tačiau jų nagrinėjimas yra reikšmingas, norint geriau suprasti mikroevoliucinius procesus. Tolesnis šių poveikio genomo variantų tarpusavio sąsajų aiškinimasis padėtų apibrėžti jų funkcinę reikšmę skirtingose populiacijose [29, 30].

1.3. Gamtinės atrankos nustatymo metodai

Adaptacijos proceso nustatymas gali būti kompliktuotas teigiamą gamtinę atranką painiojant su kitais evoliuciniais procesais. Pavyzdžiui, alelių dažnis gali padidėti ne dėl teigiamos gamtinės atrankos poveikio, bet dėl atsitiktinio genetinio dreifo (nulemta demografinių įvykių) ir kartais taip greitai, kad veikiamas evoliucinės jėgos žymių profilis genome būna labai panašus į teigiamą gamtinę atranką. Todėl kartais sunku atskirti demografinius procesus nuo gamtinės atrankos poveikio [2]. Be demografinių procesų, siekiant nustatyti teigiamos atrankos poveikį, reikia atsižvelgti ir į gryninančią atranką. Pavyzdžiui, foninė (angl. *background*) atranka (ši atranka veikia genomo sirtis, nesiejamas su žinomais požymiais) gali lemti padidėjusį atsitiktinės atrankos poveikį ir lemti populiacijų diferenciaciją, o šie procesai nustatomi ir adaptacijos atveju [31]. Taigi skirtumų tarp šių procesų paieška ir tinkamas procesų įvardijimas vis dar yra vienas iš populiacinės genetikos iššūkių.

Pagrindinė teigiamos atrankos nustatymo strategija apima tris etapus. Pirma, tiriama daug pasirinktos populiacijos ar tiriamųjų grupės genomo sričių vienu metu, pavyzdžiui, viso genomo VNP arba tam tikro ilgio genomo fragmentai (angl. *genomic windows*). Antra, pasirenkamas apibendrinamasis (angl. *summary*) statistinis metodas, leidžiantis nustatyti tam tikras teigiamos atrankos žymes kiekvienoje srityje. Galiausiai, gauti statistiniai įverčiai lyginami su įverčiais, darant prielaidą, kad didžioji genomo sritis yra neutrali, arba su empiriniu genomo variacijos pasiskirstymu. Pastarasis lyginimas

atliekamas siekiant nustatyti gamtinės atrankos žymių statistinių parametru išskirtis tiriamoje populiacijoje [32].

Egzistuoja daugybė metodų, leidžiančių įvertinti gamtinės atrankos poveikį ir jos paliekamas žymes genome. Šie metodai varijuoja nuo paprastesnių, apibendrinamųjų statistinių iki inovatyvesnių, pavyzdžiui, naudojami Bajeso ar mašininio mokymosi metodai. Nors esama ir sudėtingų analizės metodų, didžioji dalis statistinių įrankių sukurti klasikinėms gamtinės atrankos žymėms nustatyti (1.1 pav.), kai pokytis atsiranda *de novo*, jo dažnis greitai padidėja ir (beveik) pasiekia fiksaciją. Tokie metodai paremti neutralumo vertinimu, kuris remiasi alelių dažnių spektru (pavyzdžiui, Tajimos D testas [33], Fay ir Wu H testas [34], Fu ir Li D bei F testai [35]). Taip pat gali būti naudojama kita strategija, vertinanti neįprastą nepusiausvyros sankibos profilį genome. Tokios strategijos metodų pavyzdžiai – išplėstinis haplotipų homozigotiškumas (angl. *extended haplotype homozygosity*) [36] ir integruotas haplotipų įvertis (angl. *integrated haplotype score*) [37]. Trečioji analizės strategija paremta populiacijos struktūra, kai lyginamas alelių dažnių skirtumas tarp populiacijų, pavyzdžiui F_{ST} [38, 39] ar sričiai specifinis sekos ilgio metodas (angl. *locus-specific branch length*) [40]. Visų šių metodų pagrindinis trūkumas – statistškai vertinama tik viena pasirinkta gamtinės atrankos žymė, nors svarbus kompleksinis šių žymių vertinimas. Vis dėlto egzistuoja ir šių metodų kombinacijų, pavyzdžiui, tarppopuliacinis išplėstinis haplotipų homozigotiškumo metodas [41] ar daugybinių žymių testas (angl. *composite of multiple signals test*) [42], tačiau jie reikalauja didesnių kompiuterinių resursų. Kitas metodų trūkumas susijęs su duomenų statistiniu haplotipų sudarymu iš genomo duomenų (angl. *phasing*) ir jų panaudojimu. Nepusiausvyra sankiba grįstiems metodams reikalingas haplotipų įvertinimas ir rekombinacijos žemėlapis. Šiuose analizės etapuose atsiradusios klaidos gali lemti netikslius gamtinės atrankos nustatymo rezultatus. Be to, genomo heterogeniškumas, vykstant gryninančiai atrankai [43], mutacijų dažnis, genomo struktūrinė variacija ir rekombinacijos dažnis taip pat gali turėti įtakos klaidingam kandidatinių gamtinės atrankos veikiančių genomo sričių įvertinimui [44]. Šie ir kiti metodai susisteminti ir apžvelgti 1.1 lentelėje ir 1.1 paveiksle.

1.1 lentelė. Apibendrinti tradiciniai teigiamos gamtinės atrankos poveikio vertinimo metodai

Įvertis / testas	Statistinio modelio aprašas	Literatūros šaltinis
Nukleotidų įvairovės įverčiai		
π	Nukleotidų įvairovės išraiška – vidutinis nukleotidų skirtumų skaičius genomo pozicijoje.	Jukes ir Cantor (1969) [45]; Nei ir Gojobori (1986) [46]
θ	Tikėtinų polimorfinių genomo pozicijų proporcija.	Watterson (1975) [47]; Tajima (1983, 1996) [48, 49]
Atrankos nustatymo metodai, paremti alelių dažnių pokyčiais ir variabilumo lygiu		
Tajima D	Polimorfizmų skaičius, esant vidutiniam poriniam skirtumui (angl. <i>mean pairwise difference</i>), tarp sekų. Skaičiuoja skirtumą tarp π ir θ . Veikiant gamtinei atrankai, šios dvi reikšmės yra lygios viena kitai, todėl Tajima D lygi nuliui. Jei reikšmė nėra artima nuliui, tai rodo nukrypimą nuo alelių dažnių pasiskirstymo. Teigiamos reikšmės reiškia padidėjusį vidutinio dažnio alelių skaičių, kuris gali būti butelio kaklelio efekto ar subalansuotos atrankos priežastis. Neigiamos reikšmės reiškia mažo dažnio alelių perteklių, kurį galėjo lemti populiacijos ekspansija arba teigiama gamtinė atranka.	Tajima (1989) [33]
Fu ir Li D	Išvestinių genomo variantų skaičius nustatytas mėginyje tik kartą, palyginti su bendroju išvestinių genomo variantų skaičiumi. Metodas panašus į Tajima D, tačiau leidžia genome atskirti senus ir naujus pokyčius, atsižvelgiant į jų atsiradimo vietą genealogijos šakose.	Fu ir Li (1993) [35]
Fu ir Li F	Išvestinių genomo variantų skaičius, nustatytas mėginyje tik kartą, esant vidutiniam poriniam skirtumui tarp sekų.	Fu ir Li (1993) [35]
Fay ir Wu H	Išvestinių, aukšto ir žemo dažnių genomo variantų skaičiaus palyginimas su vidutinio dažnio variantų skaičiumi. Išskirtinai didelis aukštą dažnį turinčių alelių skaičius yra teigiamos gamtinės atrankos indikatorius.	Fay ir Wu (2000) [34]
Atrankos nustatymo metodai, paremti nepusiausvyros sankibos nustatymu		

Įvertis / testas	Statistinio modelio aprašas	Literatūros šaltinis
Integruotas haplotipų įvertis, iHS	<p>Paremtas alelių dažnių skaičiavimu regionuose, esančiuose didelėje nepusiausvyroje sankiboje. Tinkamas gamtinės atrankos žymėms nustatyti, kai haplotipai yra vidutinio (50–80 %) dažnio, tačiau prastai nustato aukšto dažnio sritis ar sritis, pasiekusias fiksaciją. Smarkiai teigiamos ir smarkiai neigiamos iHS reikšmės reiškia, kad ilguose haplotipuose atitinkamai nustatyti arba protėvių kilmės (angl. <i>ancestral</i>), arba išvestinės kilmės (angl. <i>derived</i>) aleliai.</p> <p>Paremtas ilgų, didelio dažnio haplotipų populiacijoje analize, norint nustatyti netolimos gamtinės atrankos žymes. Apibrėžiamas kaip tikimybė nustatyti dvi atsitiktinai pasirinktas chromosomas, kuriose dominantis VNP ar haplotipas yra identiški pagal kilmę. Turi daugiau galios nustatyti aukšto (>80 %) dažnio haplotipus, veikiamus gamtinės atrankos.</p>	Voight et al (2006) [37]
Išplėstinis haplotipų homozigotiškumas, EHH	Atrankos nustatymo metodai, paremti populiacijos diferenciacija	
F_{ST}	<p>Matuoja vidutinę genų tėkmę, paremtą alelių dažniu. Interpretuojamas kaip heterozigotiškumo sumažėjimas populiacijose, palyginti su teorine reikšme, veikiant atsitiktiniam kryžminimuisi. Esant atrankos neutralumui, F_{ST} veikiamas genų dreifo ir migracijos.</p>	Hudson et al (1992) [38]

Naujos kartos sekoskaitos metodas leido detaliau analizuoti skirtingų geografinių vietovių populiacijų, prisitaikiusių prie skirtingų aplinkos sąlygų, genomus. Atsižvelgiant į gautus gausius pirminius duomenis (palyginti su VNP žymenų genotipavimo metodo ar egzomo sekoskaitos metu gaunamais duomenimis), pastebėta, kad egzistuojantys gamtinės atrankos statistiniai metodai reikalauja didelių kompiuterinių (atminties ir (ar) skaičiavimų pajėgumo) resursų. Tradiciniai gamtinės atrankos nustatymo metodai gali būti naudojami ir viso genomo duomenims analizuoti, tačiau dauguma metodų naudoja „slenkančio lango“ (angl. *sliding-window*) principą, analizuodami genomą fiksuoto dydžio fragmentais, todėl negalima nustatyti bendrosios gamtinės atrankos veikimo masto [50].

Minėtai problemai spręsti gali būti taikytini pažangesni įrankiai, naudojantys sudėtinio tikėtinumo santykio (angl. *composite likelihood ratio*) metodą, pavyzdžiui, *SweepFinder* [8] ar *SweeD* [51], tačiau šių metodų trūkumas – ilgas atliekamų skaičiavimų laikas. 2018 m. sukurtas naujas įrankis *RAiSD* (padidinto tikslumo gamtinės atrankos žymių nustatymo įrankis, angl. *Raised Accuracy in Sweep Detection*) naudoja μ reikšmę, leidžiančią įvertinti teigiamos gamtinės atrankos poveikio stiprumą [52]. Ši reikšmė vertina genomo regionus, remdamasi VNP vektorių skaičiavimu, sujungdama pagrindines tris, anksčiau aptartas, gamtinės atrankos žymes: alelių dažnių pokyčius, iškreiptą srities dažnių spektrą ir specifinį nepusiausvyros sankibos profilį. Be to, kad šis metodas sujungia ir analizuoja kelias gamtinės atrankos žymes, kiti šio metodo privalumai yra santykinai trumpas, palyginti su kitais metodais, analizės laikas ir didesnis tikslumas (didesnis 1,66 karto, palyginti su minėtais *SweepFinder* ar *SweeD* įrankiais) bei specifiskumas (didesnis 4,27 karto, palyginti su minėtais *SweepFinder* ar *SweeD* įrankiais) dėl VNP vektorių skaičiavimo metodikos. Minėtina ir tai, kad šis metodas, palyginti su kitais naudojamais gamtinės atrankos poveikio nustatymo metodais, reikalauja santykinai nedidelių kompiuterinių resursų (įprastai kelių megabaitų), net ir esant dideliems duomenų kiekiams [52]. Dėl minėtų priežasčių pastarasis metodas pasirinktas šiame tyrime gamtinės atrankos poveikiui vertinti. Tai pirmas kartas, kai minėtas metodas taikomas bendrosios lietuvių populiacijos ir Lietuvos Černobylio katastrofos likviduotojų teigiamos gamtinės atrankos veikiams genomo sritims analizuoti.

1.4. Teigiamos gamtinės atrankos tyrimai pasaulyje ir Lietuvoje

Vis lengviau prieinami plataus masto genomo duomenys lemia įvairių GWAS tyrimų gausą. Išaugo ir minėtų tyrimų skaičius skirtingų populiacijų adaptacijos tema. Šie tyrimai leidžia daryti įžvalgas apie naujus kandidatinius genus ir jų tarpusavio sąveikas, dalyvaujančias adaptacijos prie skirtingų aplinkos sąlygų, mitybos ir ligų procesuose.

Remiantis archeologinių radinių ir genomo tyrimų įrodymais, galima teigti, kad anatominiu požiūriu šių laikų žmogus kilo iš Afrikos prieš kelis šimtus tūkstančių metų ir greitai išplito pasaulyje, prisitaikydamas prie vis kitų aplinkos sąlygų [53]. Šios migracijos pagrindinė pasekmė buvo skirtingo gamtinės atrankos spaudimo poveikis besiformuojančioms subpopuliacijoms. Dėl šio spaudimo susiformavo tokie svarbūs žmonių požymiai, kaip termoreguliacija šalto klimato zonose, prisitaikymas prie hipoksijos aukštumose ar žmonių, gyvenančių mažai saulėtose vietovėse, šviesi oda [54].

Minėti požymiai pagerino žmonių prisitaikymą prie aplinkos sąlygų, t. y. lėmė geresnę jų adaptaciją. Gamtinės atrankos veikiamų geno sričių identifikavimas ne tik paaiškina, kaip adaptacija formuoja geno variaciją skirtingose populiacijose, bet ir suteikia svarbių žinių apie šių geno sričių įtaką fenotipui ir biologinę funkciją.

Be jau minėtų žmogaus evoliucijai ir adaptacijai reikšmingų savybių, nustatyta ir kitų požymių, kurie, veikiant gamtinei atrankai, lėmė geresnį žmonių prisitaikymą ir išgyvenamumą. Pavyzdžiui, mokslinėse publikacijose minima, kad genai, turintys mažesnę nesinoniminių pokyčių dažnį (tai rodo gryninančios atrankos poveikį), tiesiogiai koreliuoja su mendelinių ligų pasireiškimu dažniu [55]. Taip pat nustatyta, kad dauguma genų, veikiamų teigiamos gamtinės atrankos, dalyvauja procesuose, susijusiuose su vėžiniais molekuliniais mechanizmais [56]. Be to, nustatyti genai, veikiami teigiamos gamtinės atrankos ir galimai turintys įtakos daugiaveiksnių ligų, pavyzdžiui, širdies ir kraujagyslių ligų (*MMP3*, *CYP3A*, *AGT* genai) [57–59], II tipo diabeto (*CAPN10* genas) [60], astmos (*IL13*, *IL4*, *IL1A* genai) [61–63], etiopatogenezei. Tai tik keli pavyzdžiai, sietini su ligomis, – teigiamos gamtinės atrankos veikiamų geno sričių sąrašas nuolat pildomas daugėjant statistinių analizės metodų ir gamtinės atrankos tyrimų. Skirtingais statistiniais analizės metodais nustatytų teigiamos gamtinės atrankos veikiamų geno sričių skaičiaus analizė parodė, kad yra nustatyta daugiau kaip 2 000 genų, veikiamų teigiamos gamtinės atrankos ir siejamų su daugybe fenotipinių požymių [64].

Kiekviena populiacija išsiskiria jai būdingomis geno charakteristikomis, galinčiomis lemti skirtingą polinkį į ligas ar skirtingas adaptacines savybes. Lietuvių populiacija yra genetiškai artima kaimyninėms Europos populiacijoms, tokioms kaip slavai ar finougrai [65]. Y chromosomos haplogrupių analizė įvardija genetinę lietuvių populiacijos panašumą į estus ir latvius [65]. Išvardytos populiacijos veikiamos panašių aplinkos veiksnių ir gamtinės atrankos jėgų spaudimo, todėl jų geno charakteristikos yra artimos. Vis dėlto svarbu išsiaiškinti ir tik lietuvių populiacijai būdingus geno bruožus. Jau esama tyrimų, analizuojančių lietuvių etnolingvistinių grupių genetinę struktūrą [66] ar nagrinėjančių teigiamos gamtinės atrankos veikiamų geno sričių poveikį [67], tačiau tokių tyrimų vis dar nedaug. Svarbu tęsti minėtus tyrimus, siekiant nustatyti populiacijos unikalumą lemiančius geno veiksnius.

Remiantis lietuvių populiacijos tyrimais, galima teigti, kad tarp didžiausių teigiamos gamtinės atrankos poveikį patiriančių geno sričių yra genai, susiję su mitybos įpročiais (*PNLIP*, *PPARD*), odos pigmentacija (*SLC24A5*, *TYRP1*, *PPARD*) ir imuniniu atsaku (*BRD2*, *HLA-DOA*, *IL26*,

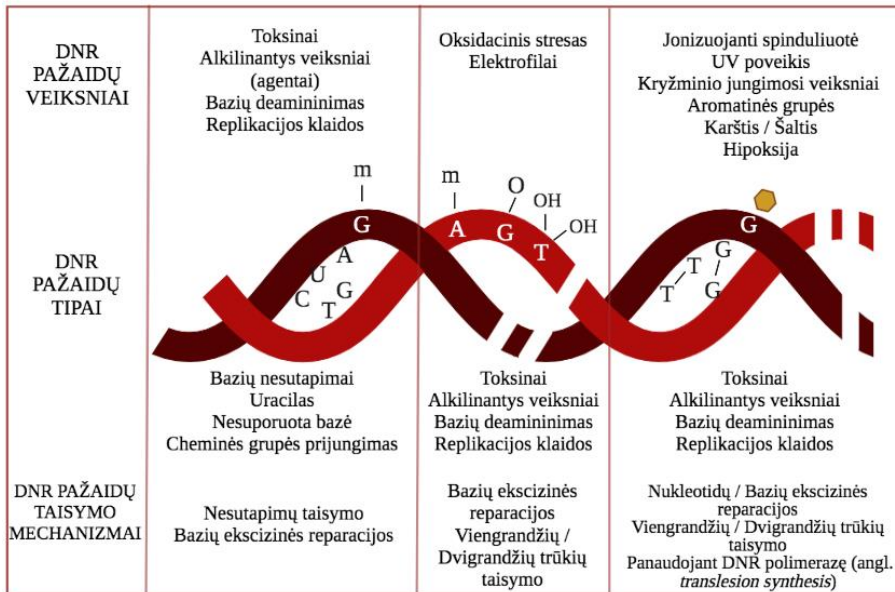
IL22) [68]. Iš viso nustatytos 36 genomo sritys, galimai veikiamos netolimos teigiamos gamtinės atrankos [67]. Rezultatai rodo, kad teigiama gamtinė atranka veikia ir formuoja tam tikrus lietuvių populiacijos gyvensenos įpročius ir bruožus, susijusius su adaptacija (pavyzdžiui, odos pigmentacija). Iki šiol atliktuose teigiamos gamtinės atrankos lietuvių populiacijoje tyrimuose naudoti tradiciniai (F_{ST} , išplėstinio haplotipų homozigotiškumo, Tajimos D) statistiniai metodai. Siekiant praturtinti esamus rezultatus, šiame tyrime teigiamai gamtinei atrankai nustatyti taikytas naujas statistinis metodas – *RAiSD* [52] įrankis. Tyrimo metu analizuotas teigiamos gamtinės atrankos poveikis statistiškai reikšmingiems poveikio genomo variantams, nustatytiems bendrojoje lietuvių populiacijoje taikant asociacijų analizę. Tai nauja strategija, sujungianti asociacijos ir teigiamos gamtinės atrankos poveikio nustatymo metodus ir leidžianti nustatyti galimą poveikio genomo variantų naudingumą fenotipui. Genomo įvairovės apibūdinimas nedidelėse populiacijose, tokiose kaip lietuvių, yra svarbus, norint suprasti populiacijų skirtumus, o tai gali būti naudinga norint suprasti biologinius adaptacijos, išgyvenamumo, daugiaveiksnių požymių ir ligų mechanizmus.

1.5. Išorinių aplinkos veiksnių įtaka genomui ir adaptacijai – Černobylio katastrofos pavyzdys

Išoriniai aplinkos veiksniai yra vieni iš pagrindinių gamtinės atrankos spaudimo (angl. *selective pressure*) jėgų. Anksčiau aptarti termoreguliacijos, prisitaikymo prie hipoksijos aukštose vietovėse, odos pigmentacijos pavyzdžiai rodo, kaip aplinkos veiksniai gali daryti tiesioginę įtaką populiacijų genomams ir fizinėms savybėms. Keičiantis aplinkai ir vykstant kataklizminiams reiškiniams, tokiems kaip ekologinės katastrofos, kažkada buvę protekciniai variantai gali tapti žalingi, todėl vyksta mikroevoliuciniai procesai, galintys lemti prisitaikymą prie populiacijos genetinės architektūros virsmo [15]. Kataklizminių įvykiu metu veikiantis gamtinės atrankos spaudimas gali santykinai per trumpą laiką pakeisti populiacijos savybes, nes ji turi greitai adaptuotis prie kilusios grėsmės. Genomo nesugebėjimas prisitaikyti gali sukelti įvairių pasekmių, pavyzdžiui, ligų, lemiančių didelį mirtingumą, paplitimą [16–18].

Žmonija nuolat patiria DNR pažeidimus sukeliančių veiksnių poveikį, tai turi įtakos sveikatai. Egzistuoja pažeidimų taisymo mechanizmai, kurie tiesiogiai ištaiso atsiradusią pažeidimą arba lemia ląstelės toleranciją šiai pažeidimui, taip padidindami ląstelės išgyvenamumo tikimybę. Sutrikusi šių mechanizmų veikla sutrikdo ląstelės homeostazę ir sukelia genomo nestabilumą, todėl gali lemti, pavyzdžiui, vėžines ligas [69] ar priešlaikinius senėjimo procesus [70].

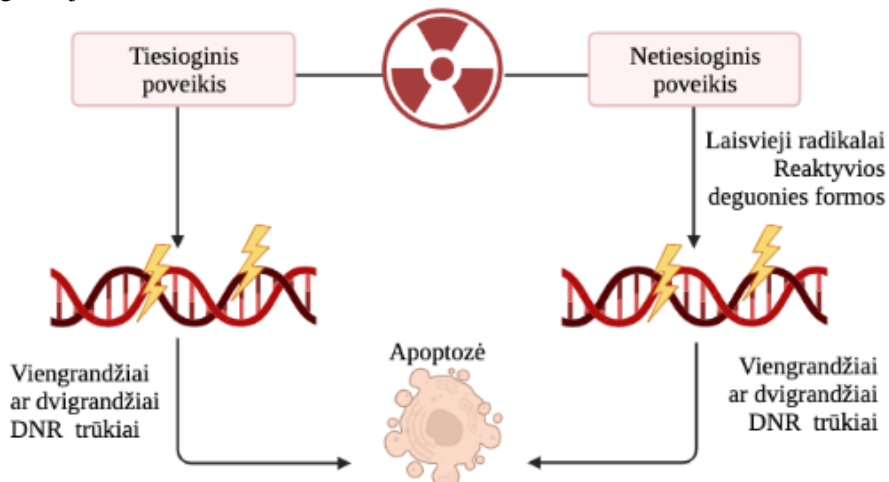
Egzistuoja DNR pažeidimų taisymo, pažeidimų tolerancijos, ląstelės ciklo patikros punktų ir ląstelių žūties mechanizmai, kurie funkcionuoja kartu, kad ištaisytų ar sušvelnintų DNR pažeidimų poveikį. Mokslinėje literatūroje aprašomi penki pagrindiniai DNR pažeidimų taisymo mechanizmai, aktyvūs skirtingose ląstelės ciklo stadijose: bazių ekscizinė reparacija (angl. *base excision repair*), nukleotidų ekscizinė reparacija (angl. *nucleotide excision repair*), nesutapimų taisymo (angl. *mismatch repair*), homologinė rekombinacija (angl. *homologous recombination*) ir nehomologinis galų sujungimas (angl. *nonhomologous end joining*) [71]. Šie mechanizmai yra kertiniai, palaikant genomo stabilumą ląstelėje, šalinant DNR pažeidimus. To nepadarius, ląstelėse pradeda kauptis mutacijos ar gali būti stabdomi metaboliniai ląstelės procesai, tokie kaip replikacija ar transkripcija. Visa tai sukelia ląstelės senėjimą ar žūtį (1.3 pav.).



1.3 pav. DNR pažeidimus sukeliančių veiksnių, DNR pažeidimų ir jų taisymo mechanizmų schema. Skirtingi DNR pažeidimus sukeliantys veiksniai lemia skirtingas DNR pažeidimus, kurias taiso skirtingi DNR pažeidimų taisymo mechanizmai (parengta pagal N. Chatterjee ir G. C. Walkerį (2017) [71]).

Vienas iš veiksnių, sukeliančių DNR pažeidimus, yra jonizuojančioji spinduliuotė (JS). Jonizuojančioji spinduliuotė – tai aukštą energetinį potencialą turinti spinduliuotė, kuri geba atpalaiduoti elektronus iš atomų ir molekulių taip sugeneruodama jonus, kad jie gali nutraukti kovalentinius ryšius. Jonizuojančioji spinduliuotė tiesiogiai veikia DNR struktūrą, sukeldama viengrandžius ar dvigrandžius DNR trūkius. Šią genetinės

medžiagos pažeidimą nelengva ištaisyti, todėl ji gali lemti genetinės medžiagos praradimą ląstelės dalijimosi metu ir ląstelės žūtį. Netinkamai ištaisyta pažeidima gali lemti mutacijas ir vėžinius procesus [72]. JS taip pat veikia ir netiesiogiai, susidarant reaktyvioms deguonies formoms (angl. *reactive oxygen species*), galinčioms DNR molekulės struktūroje pakeisti ar iš jos pašalinti heterociklinę bazę ir sukelti viengrandžius DNR trūkius. Taip galiausiai sukeliama apoptozė (1.4 pav.) [73, 74]. Taškiniai pokyčiai (pavyzdžiui, VNP) dažniausiai sukeliama veikiant nedidelėms JS dozėms. Vidutinės dozės sukelia skaitymo rėmelio poslinkio pokyčius (pavyzdžiui, vieno nukleotido tarpai ar iškritos), o didelės JS dozės gali lemti daugybinių pokyčių atsiradimą, dėl kurių gali būti prarandami ištisi genomo fragmentai ir genai juose [75].



1.4 pav. Jonizuojančiosios spinduliuotės poveikis DNR. Jonizuojančioji spinduliuotė veikia tiesiogiai ir netiesiogiai, sukeldama DNR struktūros pokyčius ir apoptozę.

Visame pasaulyje, taip pat ir Lietuvoje, dominuoja tyrimai, dideles JS dozes gavusių asmenų genomus tiriant citogenetiniais tyrimo metodais. Kiti tyrimai analizuoja nedideles JS dozes (pavyzdžiui, medicininių intervencijų metu) gavusių asmenų genomus DNR lygmeniu. Iki šiol mažai tirti aukštas JS dozes gavusių asmenų genomai, esant didelei skiriamajai gebai. Unikalių populiacijų, patyrusių aukštas JS dozes kataklizminių įvykių metu, pavyzdžiui, išgyvenusių branduolines katastrofas (minėtina Hirošima, Nagasakis, Černobylis), tyrimai molekulinio genomo analizės lygmeniu, vykdant genomo variantų analizę ir nustatant alelių dažnius, jų ryšius ir susijusių genomo sričių tinklus, leistų nustatyti naujus ir padėtų geriau suvokti jau žinomus mikroevoliucijos (adaptacijos) procesus ir ligų mechanizmus. Be

to, nagrinėjant ekologines katastrofas patyrusias populiacijas, atsiranda galimybė atlikti plataus masto ekologinius ir evoliucinius tyrimus, kurių nebūtų įmanoma atlikti laboratorijoje simuliuojant mažesnes imtis. Pavyzdžiui, Černobylio atominės elektrinės katastrofa leidžia atlikti išsamią JS poveikio individams ir populiacijoms analizę ekosistemos mastu.

Atominius sprogimus išgyvenusių populiacijų duomenų rinkiniai yra pirmieji ir vis dar vieni svarbiausių, leidžiančių įvertinti JS poveikį žmogaus sveikatai. Antrojo pasaulinio karo metu (1945 m.) Japonija Hirošimos ir Nagasakio regionuose patyrė atominio ginklo poveikį. Sprogimai sukėlė smūgines bangas, išskyrė šiluminę energiją ir jonizuojančiąją spinduliuotę. Viso to pasekmė – fiziniai sužalojimai, ūmūs radiaciniai apsinuodijimai ir įvairios lėtinės ligos, dėl kurių per pirmus metus po katastrofos žuvo apie 216 tūkst. žmonių [76]. Po katastrofos įsikūręs Radiacijos poveikių tyrimų fondas (angl. *Radiation Effects Research Foundation*) iki dabar yra viena didžiausių įstaigų, vykdančių mokslinius tyrimus JS poveikio žmonijai tematika. Nustatyta, kad žmonės, išgyvenę atominius sprogimus, turi didesnę polinkį sirgti tam tikrų organų (ypač kraujodaros, skydliaukės, krūties) vėžiu [77]. Aukštas JS dozes patyręs asmenys dažniau sirgo ir akių, skydliaukės, širdies, psichikos ligomis [78]. Tyrimų metu nustatyta ir tai, kad JS neigiamai veikia imuninę sistemą [79]. Vėžiniai procesai tiesiogiai susiję su DNR pažaidomis ir jų reparacijos mechanizmų disfunkcija, o minėtos daugiaveiksnės būklės taip pat turi stiprų genetinį komponentą, todėl nėra abejonių, kad JS veikia atominį sprogimą išgyvenusių populiacijų genomus. Kita vertus, atliekant genetinius šių populiacijų tyrimus, daugiausia tiriami aukštas JS dozes patyrusių asmenų palikuonys. Kol kas nėra nustatytas JS poveikis atominę katastrofą išgyvenusių asmenų palikuonių genomams [80]. Palikuonių grupės analizė parodė, kad *de novo* VNP ir didelių struktūrinių persitvarkymų dažniai nėra padidėję [81, 82], o tai leidžia daryti prielaidą, kad JS sukelti pokyčiai genome nėra perduodami ateities kartoms. Vis dėlto palikuonių epigenomas ar transkriptomas mažai nagrinėti, todėl tai galėtų būti tolesnių tyrimų planas.

Kitas žmonijai radiacine prasme svarbus įvykis – 1986 m. Ukrainoje, Černobylyje, sugedus atominės elektrinės reaktoriui, įvykusi branduolinė katastrofa. Apie 600 tūkst. žmonių 1986–1990 m. laikotarpiu buvo įdarbinti katastrofos padariniams likviduoti, todėl patyrė didelį ir ilgalaikį (ne smūginį, kaip Hirošimos ar Nagasakio atveju) JS poveikį [76]. Tarp šių asmenų buvo ir lietuvių. Pagrindinis Hirošimos ir Nagasakio bei Černobylio katastrofų skirtumas – apšvitos tipas. Hirošimos ir Nagasakio atvejais atominis sproginas lėmė smarkų gama ir neutroninės spinduliuotės poveikį. Černobylyje ūmią apšvitą keitė daugiausia nedidelių dozių gama spindulių

ilgalaikė vidinė apšvita (per augalinės ir gyvulinės kilmės maisto produktus, vandenį, orą). Šie skirtumai paaiškina ir skirtingų vėžinių ligų dažnumą minėtose populiacijose. Hirošimos ir Nagasakio atvejais padaugėjo vaikų leukemijos ir krūties vėžio atvejų, o dėl Černobylio išaugo skydliaukės vėžinių ligų skaičius [83]. Briansko (Rusija) teritorijoje, kur, įvykus Černobylio katastrofai, taip pat nustatytas didelis radioaktyvus užterštumas, statistiškai reikšmingai daugiau fiksuota naujagimių įgimtų anomalijų (pavyzdžiui, polidaktilija) [84].

Smarkiausiai Černobylio katastrofa paveikė Ukrainą, Baltarusiją ir Rusiją. Šių šalių teritorijose padaugėjo radioaktyvių, apie 30 m. skilimo pusperiodį turinčių cezio (^{137}Cs , ^{134}Cs), stroncio (^{90}Sr) radionuklidų ir kitų, trumpesnio skilimo pusperiodžio izotopų, pavyzdžiui, jodo (^{131}I) [85]. Tai, kad šių radionuklidų skilimo pusperiodis skaičiuojamas dešimtmečiais, reiškia, kad jų poveikis bus jaučiamas ilgą laiką, keliose kartose. Patekusių per vandenį ir maistą, minėtų radionuklidų poveikis tapo ilgalaikis. Ypač kenksmingas stroncis, kuris ilgam laikui įsiterpia į kaulines struktūras [86].

Černobylio katastrofos likviduotojų ir šios katastrofos padarinius patyrusių asmenų genomo tyrimai daugiausia apima citogenetinius tyrimo metodus. Įvykus Černobylio katastrofai, nustatytas padidėjęs nestabilių ir stabilių chromosominių ir chromatidinių persitvarkymų dažnis, padidėjęs žiedinių, dicentrinių chromosomų skaičius, o tai vienas iš pagrindinių spinduliuotės sukeltos mutagenzės indikatorių chromosomos lygmeniu [87, 88].

Daug genomo tyrimų atlikta vaikams, išgyvenusiems Černobylio katastrofą arba gimusiems iškart po jos. Nustatyta, kad vaikų, gyvenusių didelėse JS užterštumo teritorijose, chromosominių pažaidų skaičius yra 53 proc. didesnis negu pažaidų skaičiaus vidurkis neužterštose teritorijose [89]. Vaikų, paveiktų JS *in utero*, tyrimai taip pat atskleidė padidėjusį chromosominių pažaidų dažnį. Dažniausi nustatomi pokyčiai – translokacijos, inversijos ir delecijos [90]. Padidėjęs chromosominių pažaidų skaičius nustatytas tiek žemesnes, tiek ir aukštesnes JS dozes patyrusiems vaikams [91]. Minėtini ir baltarusių vaikų, patyrusių Černobylio katastrofos padarinius, tyrimai, atskleidę, jog net po 10 m. limfocituose yra padidėjęs DNR trūkių ir translokacijų kiekis [92]. Padidėjęs chromosominių pokyčių kiekis nustatytas ne tik didesnes JS dozes patyrusių asmenų grupėje. Černobylio ekosistemų tyrimai rodo ir JS poveikį skirtingoms gyvūnų ir augalų rūšims [93]. Visa tai atskleidžia JS poveikį visos gyvosios gamtos ilgalaikiam genomo nestabilumui.

Atkreiptinas dėmesys, kad molekulinį genomo tyrimų minėta tema nėra daug, o ir atliktų tyrimų išvados nevienareikšmės. Minėtinas 130 vaikų,

gimusių aukšto JS užterštumo zonose 1987–2002 m. laikotarpiu, viso genomo sekoskaitos tyrimas. Gauti rezultatai patvirtino Hirošimoje ir Nagasakyje atliktų tyrimų rezultatus, padidėjęs *de novo* pokyčių dažnis nenustatytas [94]. Vis dėlto kito tyrimo duomenys parodė, kad Černobylio katastrofos likviduotojų palikuonims mutacijų dažnis statistiškai reikšmingai padidėjęs [95]. Šis nevienareikšmiškumas pagrindžia unikalių populiacijų tolesnių tyrimų molekulinio lygmeniu svarbą.

Lietuva taip pat nukentėjo nuo Černobylio katastrofos sukeltų padarinių. Radioaktyvieji debesys praslinko virš Lietuvos per pirmąsias paras po avarijos, palikdami žymias radioaktyviųjų medžiagų iškritas pietvakarinėje ir vakarinėje šalies dalyje. Labiausiai nukentėjo Varėnos, Alytaus, Neringos ir Klaipėdos rajonai. Didžiausia tarša buvo radioaktyviuoju jodu, pasklido ir dideli ilgaamžio radionuklido cezio kiekiai, kurio pėdsakų grunte, miško grybuose ir uogose nustatoma iki šiol. Radioaktyvaus cezio ir stroncio skilimo pusperiodis yra apie 30 metų, todėl šių radionuklidų pėdsakų vis dar galima aptikti vietinės gamybos maisto produktuose, žuvyse [96]. Tai lemia papildomą gyventojų apšvitą.

Lietuva, kaip tyrimo objektas, unikali tuo, kad yra patyrusi Černobylio atominės elektrinės katastrofos padarinius ir dalyvavusi juos likviduojant. Didelė dalis katastrofos likviduotojų serga įvairiomis ligomis, potrauminio streso sutrikimu ar depresija [97]. Be to, aukštos JS dozės didina vėžinių ligų, skydliaukės, akių ligų riziką [98]. Ištyrus Estijos Černobylio katastrofos likviduotojus, nustatytas padidėjęs šios grupės asmenų mirtingumo dažnis, palyginti su bendrąja estų populiacija. Padidėjusio mirtingumo dažnio priežastys – burnos ar ryklės vėžys, vėžinės ligos, sietinos su gausesniu alkoholio vartojimu, savižudybės [99]. Tarp likusių gyvų Lietuvos Černobylio katastrofos likviduotojų (LČKL) yra ne tik sunkiai sergančių ar kenčiančių nuo priklausomybių, bet ir santykinai sveikai senstančių asmenų. Jie vertingi kaip tyrimo objektas, nes išgyveno ekstremaliomis sąlygomis, adaptavosi prie visą gyvenimą besitęsiančių JS ir kitų katastrofos sukeltų padarinių. Potencialiai šias savybes gali lemti unikali genomo variacija. Populiacijose dažnų genomo variantų pasiskirstymas ir patogeninių variantų išitvirtinimas genome, vykstant gamtinei atrankai, yra svarbūs išgyvenamumui ir ilgaamžiškumui.

Lietuvoje atlikti LČKL tyrimai leido nustatyti šiai tiriamųjų grupei būdingus statistiškai reikšmingus genomo nestabilumo pokyčius. Išvadose akcentuojama tiriamųjų grupės tolesnio tyrimo svarba [100, 101]. Pagrindinės chromosominės aberacijos didelės JS dozes patyrusių LČKL grupėje, kaip ir kitų likviduotojų genuose, – dicentrinės, žiedinės ar fragmentuotos chromosomos. Patyrus didelės JS dozes, chromosominių aberacijų kiekis

ilgainiui mažėja [102], tačiau LČKL ląstelėse, nepriklausomai nuo patirtos JS dozės stiprumo, nustatytas aukštesnis negu įprasta citogenetinių persitvarkymų skaičius, net ir praėjus 8 m. po Černobylio atominės elektrinės katastrofos [100, 101]. Šios išvados sutampa su anksčiau aptartų tyrimų [87–92] rezultatais ir juos papildo. Ilgalaikis JS poveikis yra iššūkis genomo vientisumui ir ląstelių gyvybingumui palaikyti, todėl reikalingi tyrimai ne tik chromosominiu, bet ir DNR lygmeniu. Dėl šios priežasties šiame tyrime buvo siekta charakterizuoti unikalią LČKL genomo variaciją, nustatyti poveikio genomo variantus ir identifikuoti šiai grupei unikalias, teigiamos gamtinės atrankos veikiamas genomo sritis, turinčias įtakos šios tiriamųjų grupės adaptacijai prie ekstremalių išorės veiksnių.

2. TIRIAMIEJI IR TYRIMO METODAI

2.1. Tyrimo eiga

Tyrimas vykdytas nuo 2018 m. spalio 1 d. iki 2022 m. rugsėjo 30 d. Vilniaus universiteto Medicinos fakulteto Biomedicinos mokslų instituto Žmogaus ir medicininės genetikos katedroje (VU MF BMI ŽMGK). Įgyvendintas projektas „Apsauginiai DNR mechanizmai – plataus masto Lietuvos Černobylio katastrofos likviduotojų viso genomo tyrimas“ (akronimas – ADAPT) (Nr. S-MIP-20-35). Tyrimui atlikti išduotas Vilniaus regioninio biomedicininį tyrimų etikos komiteto leidimas (Nr. 2019/4-1119-612, DNAPROTECT).

Tyrimui naudoti du bendrosios lietuvių populiacijos grupės duomenų rinkiniai, gauti įgyvendinus šiuos projektus:

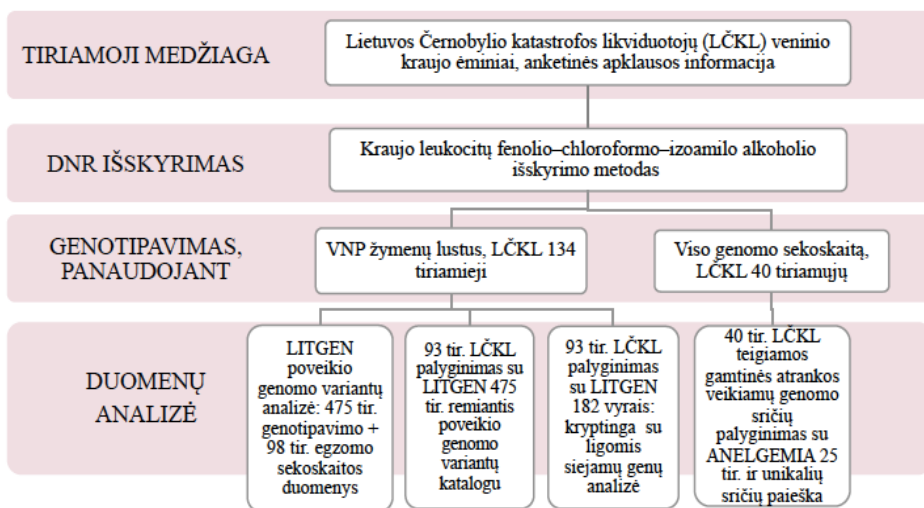
- 1) VU MF BMI ŽMGK 2011–2015 m. vykdytą projektą „Lietuvos populiacijos genetinė įvairovė ir sandaros kitimai, susiję su evoliucija ir dažniausiai paplitusiomis ligomis“ (akronimas – LITGEN) (projekto nr. VP1-3.1-ŠMM-07-K-01-013, bioetikos leidimo nr. 158200-05-329-79);
- 2) VU MF BMI ŽMGK 2020–2022 m. vykdytą projektą „Senovės ir nauji aleliai lietuvių genome: mutacijos, gamtinė atranka ir adaptacija“ (akronimas – ANELGEMIA) (projekto nr. S-MIP-20-34, bioetikos leidimo nr. 2020/6-1243-724).

Visi tiriamieji pasirašė sutikimus dalyvauti biomedicininuose tyrimuose.

Disertacijos tyrimą sudaro trys dalys:

- 1) kandidatinių poveikio genomo variantų paieška ir analizė bendrojoje lietuvių populiacijos (LITGEN) grupėje, remiantis plataus masto VNP žymenų genotipavimo ir egzomo sekoskaitos duomenimis;
- 2) kandidatinių poveikio genomo variantų paieška ir analizė Lietuvos Černobylio katastrofos likviduotojų (LČKL) grupėje ir gautų rezultatų lyginimas su bendrosios lietuvių populiacijos (LITGEN) grupės rezultatais, remiantis plataus masto VNP žymenų genotipavimo duomenimis;
- 3) teigiamos atrankos veikiamų genomo sričių analizė LČKL grupėje ir gautų rezultatų lyginimas su bendrosios lietuvių populiacijos grupės rezultatais (ANELGEMIA), remiantis viso genomo sekoskaitos duomenimis.

Tyrimo metu naudotas *Homo sapiens* GRCh37 (hg19) referentinis genomus. Tyrimo strategija ir darbo eiga pateikta 2.1 paveiksle. Autorės atliktų darbų indėlis pateiktas I priede.

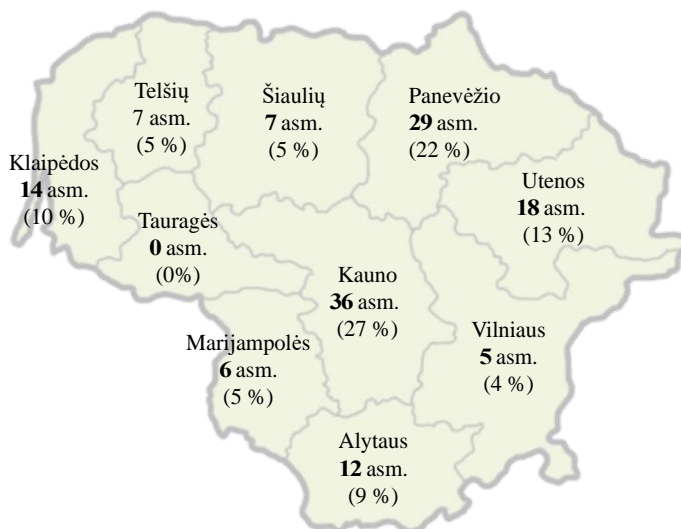


2.1 pav. Tyrimo strategija ir darbo eiga (LČKL – Lietuvos Černobylio katastrofos likviduotojai, LITGEN – bendrosios lietuvių populiacijos tiriamieji, dalyvavę LITGEN projekte, ANELGEMIA – bendrosios lietuvių populiacijos tiriamieji, dalyvavę ANELGEMIA projekte, tir. – tiriamieji)

2.2. Tiriamieji

Tyrimo metu surinkti 134 LČKL veninio kraujo ėminiai ir šių tiriamųjų anketinės apklausos duomenys. Tiriamųjų grupę sudarė 128 vyrai ir 6 moterys, dalyvavę Černobylio atominės elektrinės katastrofos padarinių likvidavimo darbuose. Tiriamieji yra Lietuvos Respublikos gyventojai, iš jų: 116 – lietuvių, 13 – rusų, 2 – ukrainiečių, 2 – baltarusių, 1 – kazachų tautybės. Daugiausia tiriamųjų gyvena Kauno apskrityje ($n = 36$). Visų LČKL grupės tiriamųjų pasiskirstymas pagal gyvenamąją vietą Lietuvos apskričių atžvilgiu pateikiamas 2.2 paveiksle.

Naudojant skirtingus genomo žymenis, tyrimo duomenys buvo gaunami netolygiai, tyrimas tęsėsi visą doktorantūros ir disertacijos rengimo laikotarpį (tiriamųjų skaičius buvo papildytas jau pradėjus duomenų analizės etapą). Siekiant kuo labiau suvienodinti šio tyrimo metu lyginamas populiacijas (pašalinant išskirtis, pavyzdžiui, pagal lytį), disertacijoje pasirinkta į tolesnę analizę įtraukti 93 LČKL vyrus (iš 134 tirtų asmenų).



2.2 pav. Lietuvos Černobylio atominės elektrinės katastrofos padarinių likviduotojų grupės visų tiriamųjų pasiskirstymas pagal Lietuvos apskritis. Daugiausia tiriamųjų – iš Kauno ir Panevėžio apskričių.

Tiriamieji užpildė dalyvio apklausos anketą (II priedas (žr. 2.3 skyrių). Tyrimo metu LČKL vidutinis amžius siekė 64 metus (amžius varijavo nuo 50 iki 78 m.).

Kaip jau minėta, tyrimui taip pat naudoti du bendrosios lietuvių populiacijos grupės duomenų rinkiniai:

1. LITGEN projekto tiriamųjų grupę sudarė 475 asmenys (239 moterys ir 236 vyrai). Vidutinis tiriamųjų amžius – 53 metai. Tiriamųjų veninio kraujo mėginiai surinkti 2011–2013 m. LITGEN projekto metu iš atsitiktinai atrinktų asmenų, priklausančių skirtingoms Lietuvos etnolingvistinėms grupėms, kurių tėvai ir seneliai gyveno toje pačioje etnolingvistinės grupės teritorijoje. Genominę DNR iš veninio kraujo ėminių išskyrė, plataus masto VNP žymenų genotipavimą ir egzomo sekoskaitą atliko VU MF BMI ŽMGK LITGEN projekto vykdytojai. Disertacijoje pristatomam tyrimui panaudoti projekto metu gauti genotipavimo ir egzomo sekoskaitos duomenys.

2. ANELGEMIA projekto tiriamųjų grupę sudarė 25 vyrai. Vidutinis tiriamųjų amžius – 36 metai (amžius varijavo nuo 29 iki 49 m.). Tiriamųjų veninio kraujo mėginiai surinkti 2020–2022 m. ANELGEMIA projekto metu iš lietuvių tautybės asmenų, kurių tėvai taip pat yra Lietuvoje gyvenantys, lietuvių tautybės asmenys, o seneliai gyvenę Lietuvos teritorijoje. Genominę DNR iš veninio kraujo ėminių išskyrė VU MF BMI ŽMGK ANELGEMIA projekto vykdytojai. Disertacijoje pristatomam tyrimui panaudoti projekto metu gauti viso genomo sekoskaitos duomenys.

2.3. Lietuvos Černobylio katastrofos likviduotojų anketinės apklausos duomenų rinkimas ir analizės strategija

Visi tiriamieji užpildė dalyvio apklausos anketas. Šios anketos leido surinkti informaciją apie tiriamųjų vaidmenį likviduojant Černobylio katastrofą, buvusias darbo ir gyvenimo sąlygas, gautą jonizuojančiosios spinduliuotės dozę, gyvensenos įpročius, psichologinę tiriamųjų būklę (II priedas). Psichologinė tiriamųjų anketinės apklausos dalis sudaryta remiantis HAD skalės psichologiniu klausimynu. Šis klausimynas naudojamas nerimo sutrikimą turintiems ir (ar) depresija sergantiems asmenims nustatyti. Klausimą sudaro keturiolika klausimų – po septynis klausimus, skirtus atitinkamai nerimo sutrikimui ir depresijai įvertinti. Kai klausimyno suminė depresijos ar nerimo sutrikimo klausimų reikšmė pasiekė 11 ar daugiau, laikyta, kad tiriamasis galimai serga depresija ir (ar) turi nerimo sutrikimą. Remiantis atliktais moksliniais tyrimais ir statistinio patikimumo įvertinimu, HAD klausimynas yra patikimas įrankis depresijai ir nerimo sutrikimui įvertinti [103].

Anketinės apklausos metu buvo renkamos ir tiriamųjų ligos istorijos. Tai leido nustatyti tiriamųjų klinikinę būklę. Šį darbą atliko VU MF BMI ŽMGK dirbanti gydytoja-genetikė ir (ar) medicinos gydytoja. Surinkti duomenys buvo apdoroti, aprašomoji statistinė analizė atlikta *Microsoft Office Excel* programiniu įrankiu. Statistiniam reikšmingumui įvertinti naudotas *Rstudio* v3.5.2 programinės įrangos *Rcmdr* paketas. Anketinių ir klinikinių duomenų analizė leido LČKL grupėje išskirti dažniausias ligų grupes. Tolesnei analizei pasirinktos burnos ertmės ir psichikos ligų grupės, dėl šių ligų sąsajų su LČKL patirtais išgyvenimais ir jonizuojančiosios spinduliuotės poveikiu. Be to, tiriamųjų grupėje fiksuotas didelis šių ligų dažnis. Remiantis mokslinė literatūra ir duomenų bazėmis, sudarytas genų, siejamų su burnos ertmės ligomis, sąrašas, kuris panaudotas kryptingai LČKL genotipavimo duomenų asociacijų analizei atlikti (2.4.2.3 punktas). Psichikos ligų asociacijų analizė, atrinkus asmenis, galimai turinčius psichikos sutrikimų, atlikta remiantis anketinės apklausos duomenimis ir ligų istorijomis bei atlikus asociacijų analizę trimis skirtingomis kryptimis, išskirstant LČKL grupės tiriamuosius į pogrupius (detaliau žr. 2.4.2.3 skyriuje).

2.4. Genotipavimas, naudojant VNP žymenų lustus

Lietuvos Černobylio katastrofos likviduotojų genominė DNR išskirta iš kraujo leukocitų fenolio–chloroformo–izoamilo alkoholio metodu. Išskirtos DNR koncentracija ir švarumas nustatyti *NanoDrop ND-1000* spektrofotometru. Tyrimui naudoti DNR išskyrimo bei koncentracijos ir

švarumo matavimo protokolai detaliai aprašyti dr. Ingridos Domarkienės daktaro disertacijos I priede [104].

Lietuvos Černobylio katastrofos likviduotojų platus masto genotipavimas pagal VNP žymenis atliktas VU MF BMI ŽMGK, naudojant *Infinium OmniExpress–24v1.3* genominius lustus (714 238 VNP) ir *Illumina HiScanSQ* sistemą. Genotipavimo procedūra atlikta pagal dr. Ingridos Domarkienės daktaro disertacijos II priede pateikiamą protokolą [104].

Genotipavimas bendrosios lietuvių populiacijos (LITGEN) tiriamiesiems atliktas VU MF BMI ŽMGK darbuotojų, naudojant *Illumina HumanOmniExpress–12 v1.1* (719 666 VNP), *Illumina HumanOmniExpress–24 v.1.1–v.1.2* (713 599 VNP) genominius lustus.

2.4.1. Genotipavimo duomenų analizė

2.4.1.1. Poveikio genomo variantų katalogas

Siekiant atlikti kryptingą poveikio genomo variantų analizę, iš *ClinVar* [105], *OMIM* [106] duomenų bazių ir mokslinių publikacijų sudarytas poveikio genomo variantų katalogas. Šį katalogą sudarė 144 genomo variantai (III priedas). Pagrindiniai variantų įtraukimo iš duomenų bazių į katalogą kriterijai:

1) klinikinio reikšmingumo statusas duomenų bazėse (protekcinis ar neaiškios klinikinės reikšmės),

2) varianto pateikimo duomenų bazei skaičius (angl. *Count of submission*), jis siekė >1.

Variantų įtraukimo iš mokslinių publikacijų kriterijai:

1) varianto poveikis geno funkcijai (variantas turi veikti geno funkciją; daugiausia atrinkta *funkcijos praradimo* (angl. *Loss-of-function*) variantų),

2) varianto dažnis (įtraukti reti ir dažni variantai). Variantų dažnis – svarbus kriterijus genomo variantams, siejamiems su adaptacija, identifikuoti. Minėti variantai populiacijoje dažniausiai yra reti, tačiau dėl teigiamos įtakos fenotipui jų gausėja ir jie vis labiau įsitvirtina populiacijoje. Dauguma minėtų variantų dažnesni genuose, neatliekančiuose biologiškai svarbių funkcijų. Taip išvengiama gryninančios gamtinės atrankos poveikio ir išlaikomas didelis dažnis skirtingose populiacijose [107]. Dėl šių priežasčių į genomo variantų katalogą įtraukti tiek reti, tiek dažni genomo variantai.

Į katalogą įtraukti genomo variantai siejami su skirtingomis paveldimomis ligomis ir būklėmis (III priedas). Sudarytas poveikio genomo variantų katalogas buvo naudojamas tolesnėje, kryptingoje asociacijų

analizėje, siekiant iširti poveikio genomo variantus, atrinktus iš plataus masto VNP žymenų genotipavimo duomenų.

2.4.1.2. Genotipavimo duomenų pirminė analizė (paruošimas)

Genotipavimo duomenims paruošti naudota *GenomeStudio 2.0* programinė įranga [108]. Pagrindiniai pirminės analizės etapai:

1. Naujas analizės projekto sukūrimas – pradinių genotipavimo duomenų ir papildomų failų (angl. *Sample sheet* failas .csv formatu, angl. *SNP manifest* failas .bpm formatu) įkėlimas.

2. Genotipavimo duomenų VNP grupavimas (klasterizavimas). Tai pirminis duomenų kokybės kontrolės etapas, kurio metu apskaičiuojami mėginių ir VNP žymenų kokybės parametrai.

3. Prastos kokybės VNP žymenų pašalinimas. Pagrindiniai parametrai, nurodantys VNP žymenų kokybę: identifikavimo dažnis (angl. *Call frequency*), jo reikšmė pasirinkta $>0,97$, *GenTrain* įverčio reikšmė, kuo artimesnė 1, *ClusterSep* įvertis yra $>0,4$ [109]. VNP žymenys, kurių *ClusterSep* įvertis mažesnis negu 0,27, pašalinti. Žymenys, kurių *ClusterSep* įvertis buvo tarp 0,27 ir 0,4, pergrupuoti ir įtraukti į tolesnę analizę.

4. Prastos kokybės mėginių pašalinimas. Mėginiai vertinti pagal genotipų kokybės santykį (angl. *Call Rate*). Pasirinktas >97 % įvertis [110]. Dėl prasto genotipų kokybės santykio, atliekant tyrimą, iš LČKL 93 tiriamųjų grupės pašalinti 2 mėginiai, iš bendrosios lietuvių populiacijos (LITGEN) 475 tiriamųjų grupės pašalinta 10 mėginių.

5. Įvesties failų, reikalingų tolesnei analizei, sukūrimas .ped ir .map formatais, naudojantis *GenomeStudio* programos papildomu įrankiu *PLINK Input Report Plug-in v2.1.4*.

2.4.1.3. Genotipavimo duomenų antrinė analizė

Antrinei analizei naudota *PLINK v1.90* programinė įranga. Pirmiausia, atlikta bendrosios lietuvių populiacijos (LITGEN) grupės (465 tiriamieji) antrinė duomenų kokybės kontrolė, poveikio genomo variantų atranka iš genotipavimo duomenų ir asociacijų su kitomis Europos populiacijomis analizė.

1. Bendrosios lietuvių populiacijos (LITGEN) duomenų iš skirtingų genotipavimo lustų sujungimas. Tai atlikta *PLINK v1.90* programos *--merge* komanda, nurodant norimų sujungti failų sąrašą .txt formatu.

2. Chromosomų 0, 23, 24, 25 pašalinimas, naudojant komandinę eilutę *plink --file --chr 1-22 --recode --out*. Atliekant disertacijoje pristatomą tyrimą, dėl lyties chromosomų analizės problematiškumo (jis kyla dėl homologinių regionų tarp X ir Y chromosomų) pasirinkta analizuoti tik autosomas. 0 chromosoma įvesties (.map) faile žymi VNP žymenis, nepriskirtus nė vienai chromosomai, 23 – X chromosomą, 24 – Y chromosomą, 25 – žymenis, priskirtus X ir Y chromosomoms (homologiniuose regionuose).

3. Kokybės kontrolė pagal įverčius: *--mind 0,1* (individų, kuriems trūksta >10 proc. genotipo, pašalinimas); *--geno 0,01* (VNP, kuriems trūksta >1 proc. genotipo, pašalinimas); *--hwe 0,001* (VNP, kurie neatitinka Hardžio-Vainbergo pusiausvyros dėsnio, kai p reikšmė <10⁻³, pašalinimas). Atlikus kokybės kontrolę, toliau į analizę įtraukti 421 bendrosios lietuvių populiacijos (LITGEN) tiriamieji.

4. Poveikio genomo variantų filtravimas iš genotipavimo duomenų, panaudojant poveikio genomo variantų katalogą .txt failo formatu ir *--extract* komandą. Iš 144 poveikio genomo variantų, esančių kataloge, tolesnė analizė buvo atlikta 46 VNP.

Kitame etape atlikta ir LČKL grupės tiriamųjų genotipavimo duomenų analogiška analizė. Kokybės kontrolės parametrus po pirminės analizės etapo atitiko 91 tiriamasis.

2.4.1.4. Genotipavimo duomenų statistinės analizės strategija

Genotipavimo duomenų statistinė analizė suskirstyta į du pagrindinius etapus:

1) bendrosios lietuvių populiacijos (LITGEN) tiriamųjų grupės asociacijų analizė, prijungiant 98 tiriamųjų egzomo sekoskaitos duomenis (žr. 2.5.1) ir lyginant juos su tam tikrų Europos populiacijų duomenimis,

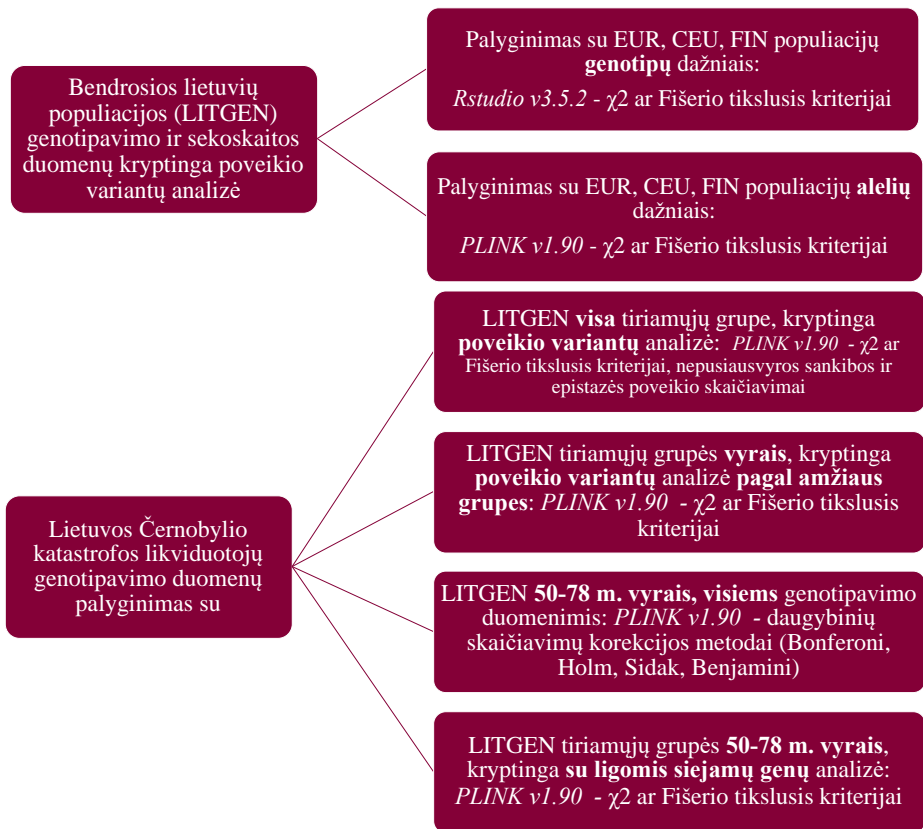
2) LČKL grupės tiriamųjų duomenų asociacijų analizė, jų lyginimas su bendrosios lietuvių populiacijos (LITGEN) tiriamųjų grupės duomenimis.

Detali statistinės analizės schema pateikta 2.3 paveiksle.

Pirmojo statistinės analizės etapo metu apskaičiuoti poveikio alelių ir genotipų dažniai. Jie lyginti su bendrosios Europos (EUR) ir atskirų Europos populiacijų (Šiaurės ar Vakarų Europos kilmės Jutos gyventojų (angl. *Utah Residents with Northern and Western European Ancestry*, CEU) ir suomių iš Suomijos (angl. *Finnish in Finland*, FIN) alelių ir genotipų dažniais, naudojantis 1 000 genomų projekto duomenimis [111], pasiekiamais *NCBI dbSNP* duomenų bazėje [112]. Šios atskiros Europos populiacijos pasirinktos remiantis moksliniu tyrimu, kurio metu nustatyta, kad Lietuvos populiacija turi genomo evoliucinių tapatumų su minėtomis populiacijomis, o kandidatinių, teigiamos atrankos veikiamų genomo sričių analizė šiuos tapatumus patvirtino [68]. Bendrosios Europos populiacijos duomenų rinkinį

sudarė duomenys iš Šiaurės ar Vakarų Europos kilmės Jutos gyventojų, suomių iš Suomijos, Anglijos ir Škotijos britų, Ispanijos iberų ir Italijos toskaniečių subpopuliacijų. Skaičiavimai atlikti naudojant *PLINK v1.90* ir *Rstudio v3.5.2* programinės įrangos *Rcmdr* paketą ($\alpha = 0,05$).

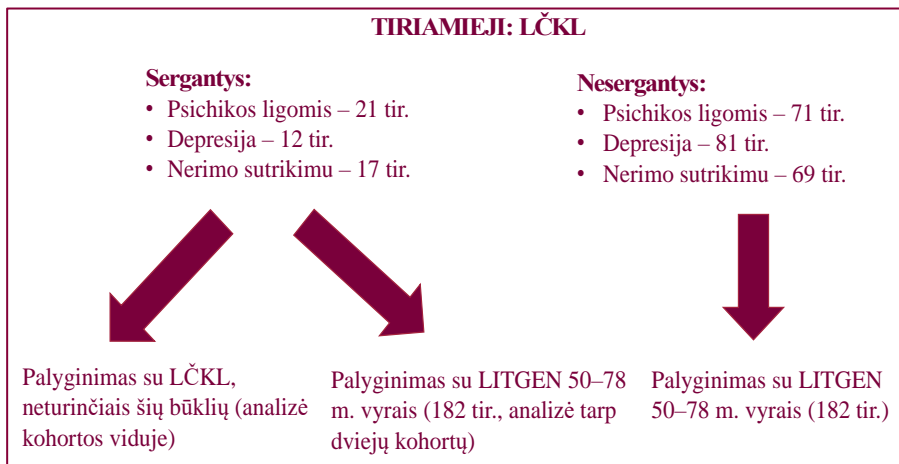
Antrojo statistinės analizės etapo metu LČKL duomenų analizė atlikta keturiomis kryptimis (2.3 pav.). Prieš atliekant LČKL ir bendrosios lietuvių populiacijos (LITGEN) tiriamųjų duomenų palyginimą, atskiri šių tiriamųjų duomenų rinkiniai po genotipavimo duomenų antrinės analizės etapo buvo sujungti naudojantis *PLINK v1.90* programos *--merge* komanda. Šiame etape χ^2 (*--assoc* komanda) ir Fišerio tikslaus (*--fisher* komanda) kriterijų ($\alpha = 0,05$), nepusiausvyros sankibos (*--ld* komanda, $\alpha = 0,01$) ir epistazės (*--epistasis* komanda, $\alpha = 0,01$) skaičiavimai atlikti naudojantis *PLINK v1.90* programine įranga.



2.3 pav. Genotipavimo duomenų statistinės analizės schema. Analizę sudarė dvi pagrindinės dalys pagal tiriamųjų grupes, t. y. atlikta bendrosios lietuvių populiacijos ir Lietuvos Černobylio katastrofos likviduotojų genotipavimo duomenų analizė. Statistiniai analizės įrankiai pasirinkti pagal išsikelto hipotezę – *Rstudio v.3.5.2* įrankis naudotas genotipų dažniams palyginti, o *PLINK v.1.90* įrankiu atliktas alelių dažnių palyginimas ir kiti skaičiavimai.

Atliekant LČKL ir bendrosios lietuvių populiacijos (LITGEN) kryptingą, su ligomis siejamų geno variantų asociacijų analizę, pasirinkta analizuoti LČKL, kurie galimai turi burnos ertmės ligų ar psichikos sutrikimų. Šios grupės atrinktos naudojantis anketinės apklausos ir ligų istorijų duomenimis – LČKL grupėje fiksuotas didelis minėtų būklių dažnis. Burnos ertmės ligų ir VNP žymenų asociacijų analizė atlikta sudarius su burnos ertmės ligomis siejamų VNP ir jų genų sąrašą iš duomenų bazių bei mokslinės literatūros. Šis sąrašas *.txt* failo formatu panaudotas atlikti kryptingai 59-iuose genuose esančių 909-ių VNP (IV priedas) asociacijų analizei. Tai įgyvendinta panaudojus *--extract* komandą. Asociacijų analizė atlikta kohortos viduje, lyginant LČKL, neturinčius burnos ertmės ligų, su šių ligų turinčiais LČKL. Minėta analize siekta identifikuoti genus, kurie potencialiai turi sąsajų su prasta LČKL burnos sveikata.

Tiriamųjų psichinė sveikata vertinta dvejopai: klinikinės apžiūros metu renkant tiriamųjų ligų istorijos duomenis (taip surinkta informacija apie LČKL psichikos ligas) ir taikant HAD skalės psichologinį klausimyną (surinkta informacija išskirtinai apie depresiją ir nerimo sutrikimą). Gauti duomenys panaudoti klinikinio fenotipo asociacijai su genetinė variacija nustatyti. Kontrolinę grupę sudarė 182 vyrai iš bendrosios lietuvių populiacijos (LITGEN). Psichikos ligų asociacijų analizė LČKL grupėje atlikta trimis kryptimis (2.4 pav.): sergančių depresija, nerimo sutrikimu ir apskritai turinčių psichikos sutrikimų asmenų duomenis lyginant su: 1) LČKL, neturinčių šių būklių, duomenimis, 2) bendrosios lietuvių populiacijos (LITGEN) tiriamųjų grupės duomenimis ir 3) nesergančių minėtomis būklėmis LČKL duomenis lyginant su bendrosios lietuvių populiacijos asmenų duomenimis, siekiant nustatyti geno variantus, galimai turinčius protekcinį efektą. Analizė atlikta visiems genotipavimo luste esantiems VNP žymenims, siekiant nustatyti naujus, su psichikos ligomis siejamus genus ir jų tarpusavio sąveikos mechanizmus.



2.4 pav. Lietuvos Černobylio katastrofos likviduotojų (LČKL) psichikos ligų asociacijų analizės schema. Psichikos ligos analizuotos trimis asociacijų analizės pjūviais – atlikta sergančiųjų analizė kohortos viduje (su neturinčiais psichikos sutrikimų LČKL), sergančiųjų analizė tarp dviejų kohortų (su kontroline LITGEN populiacija) bei sveikų LČKL palyginimas su kontroline LITGEN populiacija, siekiant nustatyti genomo variantus, galimai saugančius LČKL nuo psichikos ligų. Kai kurie tiriamieji galėjo būti priskirti kelioms skirtingoms grupėms.

2.4.1.5. Nustatytų statistškai reikšmingų genomo variantų funkcinio įvertinimo *in silico* įrankiais strategija

Siekiant įvertinti nustatytų statistškai reikšmingų poveikio genomo variantų įtaką genomui ir fenotipui, variantai analizuoti pasitelkiant *in silico* analizės įrankius ir mokslines duomenų bazes: *Varsome* [113], *Uniprot* [114], *Ensembl* [115], *ClinVar* [105], *OMIM* [106]. Funkcinių molekulinėjų sąsajų tarp genų ieškota ir funkciniai genų tinklai generuoti *GeneMANIA* [116], *STRING* [117], *Humanbase* [118], *G:Profiler* [119] *in silico* analizės įrankiais.

Ensembl duomenų bazė taip pat naudota bendrosios lietuvių populiacijos (LITGEN) analizės metu nustatytiems statistškai reikšmingiems poveikio aleliams palyginti su skirtingomis primatų rūšimis (*Gorilla gorilla*, *Pongo abelii*, *Theropithecus gelada* ir *Chlorocebus sabaeus*), siekiant nustatyti, ar poveikio alelis yra išvestinės kilmės (ir potencialiai lemia protekcinę funkciją). Siekiant pagrįsti ar paneigti nustatytų poveikio genomo variantų galimą protekcinį vaidmenį, be naudotų *in silico* funkcinio reikšmingumo įvertinimo įrankių, hipotezei patiprinti atlikta teigiamos gamtinės atrankos analizė *RAiSD* programine įranga [52] (žr. 2.5.2.1 poskyrį).

2.5. Viso egzomo ir genomo sekoskaita

2.5.1. Bendrosios lietuvių populiacijos (LITGEN) viso egzomo sekoskaita

Bendrosios lietuvių populiacijos (LITGEN) tiriamųjų grupės viso egzomo sekoskaita atlikta 98 lietuvių kilmės asmenims, panaudojant *5500 series SOLiD™* sistemą [66]. Sekoskaita atlikta VU MF BMI ŽMGK darbuotojų. Šiam tyrimui naudoti LITGEN projekto egzomo sekoskaitos duomenys, juos sujungiant ir atliekant asociacijų analizę kartu su LITGEN projekto metu gautais genotipavimo duomenimis.

2.5.2. Bendrosios lietuvių populiacijos (ANELGEMIA) ir Lietuvos Černobylio katastrofos likviduotojų viso genomo sekoskaita

Bendrosios lietuvių populiacijos (ANELGEMIA) grupės viso genomo sekoskaita atlikta 25, o LČKL grupės – 40 tiriamųjų. LČKL grupės atrankos kriterijai viso genomo sekoskaitai:

1) patirtos jonizuojančiosios spinduliuotės dozės dydis (jis turėjo būti dokumentuotas ir kuo didesnis),

2) asmuo dirbo ir (ar) gyveno 30 km zonoje aplink Černobylių, turėjo tai įrodančius dokumentus,

3) darbo pobūdis (įtraukti tiriamieji, dirbę pavojingiausius ir arčiausiai Černobylio atominės elektrinės vykdytus darbus).

Viso genomo sekoskaita ir pirminė pradinių duomenų failų (*.fastq*) kokybės kontrolė atlikta Genomikos ir transkriptomikos centre Vokietijoje (angl. *Center for Genomics and Transcriptomics (CeGaT)*), naudojant *Illumina NovaSeq 6000* sekoskaitos sistemą (pagal sudarytą VU sutartį).

2.5.2.1. Bendrosios lietuvių populiacijos (ANELGEMIA) ir Lietuvos Černobylio katastrofos likviduotojų kandidatinių teigiamos gamtinės atrankos veikiamų genomo sričių analizės strategija

Teigiama gamtinė atranka lemia adaptacinius procesus populiacijoje, o šios atrankos veikiamų genomo sričių išaiškinimas padėtų atsakyti į klausimą, kurios sritys yra naudingos asmens išgyvenamumui. Siekiant pagrįsti ar paneigti nustatytų poveikio variantų naudingumą ir įtaką tiriamosioms populiacijoms, taip pat įvardyti kandidatines teigiamos gamtinės atrankos veikiamas genomo sritis, būdingas išskirtinai LČKL, atlikta teigiamos gamtinės atrankos veikiamų genomo sričių analizė, naudojant laisvai prieinamą *RAiSD v2.9* (angl. *Raised Accuracy in Sweep Detection*)

programinę įrangą [52]. Ši programinė įranga bendrosios lietuvių populiacijos ir LČKL analizei naudota pirmą kartą.

Pagrindinis minėtos programinės įrangos privalumas – trijų svarbiausių žymių (angl. *sweep*), padedančių identifikuoti teigiamos gamtinės atrankos veikiamas geno sritis, sujungimas į vieną parametrą (μ). Pagrindinės teigiamos gamtinės atrankos žymės: 1) polimorfizmų skaičiaus sumažėjimas geno srityje, 2) geno srities dažnių spektro sumažėjimas, 3) nepusiausvyros sankibos lygio specifinis profilis (angl. *pattern*) geno srityje. Kiti panašūs įrankiai sukurti kiekvienai iš šių žymių atskirai nustatyti. Be minėtojo, kiti *RAiSD* privalumai yra santykinai nedidelių kompiuterinių resursų naudojimas analizei ir didelių duomenų kiekių apdorojimas per santykinai trumpą laiką. Programinė įranga leidžia nustatyti teigiamos gamtinės atrankos poveikį pasitelkiant geno variaciją, egzistavusią genuose dar prieš selekcinį spaudimą (LČKL atveju – įvykusią katastrofą).

Analizei naudoti viso geno sekoskaitos duomenys, gauti *.vcf.gz* failo formatu. Pirminiame etape taip pat reikalingas *bcftools v1.8* [120] programinis įrankis, naudojamas skirtingų tiriamųjų *.vcf* failams sujungti į vienos tiriamosios grupės (LČKL arba ANELGEMIA) failą. Tam atlikti panaudota ši komandinė eilutė (pavyzdžiuose pateikiama bendrosios lietuvių populiacijos (ANELGEMIA) analizė):

```
bcftools merge -l ANELGEMIAmerge.txt -O v -o mergedANELGEMIA.vcf,  
kur .txt faile nurodytos norimų sujungti .vcf.gz failų direktorijos.
```

Tas pats veiksmas pakartotas su LČKL duomenimis.

Dėl per didelio LČKL ir ANELGEMIA failų dydžio toliau atlikta atskirai kiekvienos autosomos analizė. Iš pirmame etape sugeneruotų jungtinių failų sugeneruoti atskiri failai kiekvienai chromosomai (iš viso – 22 failai vienai tiriamųjų grupei). Tai atlikta panaudojus komandinę eilutę:

```
grep -w chr1 mergedANELGEMIA.vcf > ANELGEMIAchr1.vcf.
```

Gautam pirmos chromosomos failui turi būti sutvarkyta failo antraštinė (pirmoji) eilutė, todėl panaudotos komandinės eilutės:

```
head -1 ANELGEMIAchr1.vcf – parodo failo pirmą eilutę, ji nukopijuojama ir perkeliama į antrą komandą tarp  $\langle \rangle$  ženklų;
```

```
sed -i '1s/##contig=<ID=chr1,length=249250621>/' ANELGEMIAchr1.vcf.
```

Paruošiamas *header.vcf* failas, kuriame išsaugota teisinga *.vcf* failo antraštinė eilutė. Šis failas sujungiamas su pirmos chromosomos failu:

```
cat ANELGEMIAchr1.vcf >> header.vcf – ši komandinė eilutė perkelia informaciją iš ANELGEMIAchr1.vcf į header.vcf failą, pastarasis failas pervadinamas ANELGEMIAchr1H.vcf.
```

Galiausiai naudojama *RAiSD* programinė įranga, panaudojant komandinę eilutę:

RAiSD -n ANELGEMIAchr1 -I ANELGEMIAchr1H.vcf -X centromeros.txt -R -P.

Šioje komandinėje eilutėje *-X* argumentas iš analizės pašalina pasirinktas genomines sritis, kurios surašytos *.txt* faile. Disertacijoje pristatomo tyrimo metu iš analizės pašalinti tirtų autosomų centromeriniai regionai, nes šiose srityse dirbtinai padidinama μ reikšmė dėl galimai nedidelės geno variacijos ir techninių μ skaičiavimo ypatybių. Dėl šios priežasties tolesnėje analizėje centromerai gretimi regionai, turintys dirbtinai padidintą μ reikšmę, neinterpretuojami. Dėl programinės įrangos klaidos (angl. *bug*), 4 ir 16 chromosomose centromeriniai regionai nebuvo pašalinti, todėl buvo analizuoti rankiniu būdu, naudojant sugeneruotus *RAiSD_Report* failus.

Nurodyta komandinė eilutė sugeneruoja tris failus *RAiSD_Info.ANELGEMIAchr1*, *RAiSD_Report.ANELGEMIAchr1* ir *RAiSD_Plot.ANELGEMIAchr1.pdf*. *Report* faile saugoma informacija apie teigiamos atrankos veikiamų genominių sričių genomines pozicijas, trijų pagrindinių teigiamos gamtinės atrankos žymių įverčius ir jungtinio μ parametro reikšmes. *Plot* faile grafiškai vaizduojamas trijų pagrindinių teigiamos gamtinės atrankos žymių ir μ įverčių (*y* ašis) pasiskirstymas tiriamoje chromosomoje genomines pozicijos atžvilgiu (*x* ašis). Visi šie etapai, pateikti pavyzdyje su bendrosios lietuvių populiacijos (ANELGEMIA) duomenimis pirmai chromosomai, pakartojami visoms chromosomoms ir LČKL duomenims. Iš viso sugeneruoti 44 failų rinkiniai (po 22 chromosomas ANELGEMIA tiriamiesiems ir LČKL). Laikoma, kad teigiamos gamtinės atrankos poveikis nustatytas, kai $\mu > 0$. Kuo didesnė μ reikšmė, tuo stipresnis teigiamos atrankos poveikis geno srityje. Siekiant nustatyti reikšmingiausias teigiamos gamtinės atrankos veikiamas geno sritis, daryta prielaida, kad didžioji dalis geno yra neutrali, t. y. nėra veikiamas teigiamos gamtinės atrankos, todėl laikyta, kad reikšmingos geno sritys yra turinčios 5 % didžiausių μ įverčių – šios sritys pasirinktos kaip kandidatiniai teigiamos atrankos veikiami geno regionai. Ši 5 % didžiausių įverčių reikšmė atitinka p reikšmę ir yra naudojama kaip reikšmingumo slenkstinė vertė [52].

Siekiant palyginti LČKL ir bendrosios lietuvių populiacijos (ANELGEMIA) kandidatines teigiamos gamtinės atrankos veikiamas geno sritis bei surasti išskirtinai LČKL būdingas sritis, galimai turinčias įtakos LČKL adaptacijai, panaudota komandinė eilutė:

RAiSD -n commonoutlierschr1 -CO RAiSD_Report.LCKLchr1 1 7 RAiSD_Report.ANELGEMIAchr1 1 7 -COT 0.05 -COD 5000.

Reikšmingiausioms sritims nustatyti pasirinkta naudoti 5 % didžiausių įvertį turinčias μ reikšmes (*COT* argumentas; $\alpha = 0,05$). Ši komandinė eilutė skirta bendriems teigiamos gamtinės atrankos veikiamiems regionams rasti.

Generuojami keturi failai kiekvienai analizuojamai chromosomai. Iš šių failų analizei svarbiausias *RAiSD_CommonOutlierPlot.pdf*. Vaizduojami du grafikai, kuriuose pateikiamas μ reikšmių palyginimas chromosomoje tarp dviejų tiriamųjų grupių (3.9 pav.). Raudona spalva pažymėtos chromosomos sritys, kuriose teigiama gamtinė atranka *veikia abiejose* tiriamųjų grupėse. Lietuvos Černobylio katastrofos likviduotojams *unikalios* teigiamos gamtinės atrankos veikiamos chromosomos sritys vaizduojamos *pilka* spalva. Tikslios srities genominei pozicijai nustatyti naudojamas *RAiSD_Report* failas. Šiame tyrime tolesnei genominių sričių analizei pasirinkta įtraukti unikalios LČKL grupei teigiamos gamtinės atrankos veikiamas geno sritys, turinčias aukščiausią μ reikšmę.

RAiSD_Report failas taip pat panaudotas siekiant įvertinti teigiamos gamtinės atrankos įtaką iš LČKL ir bendrosios lietuvių populiacijos (LITGEN) genotipavimo (VNP žymenų lustais) duomenų nustatytiems poveikio geno variantams. Pagal turimą varianto genomine pozicija minėtame faile surandama teigiamos gamtinės atrankos sritys, į kurią patenka analizuojamas variantas ir įvertinama μ reikšmė chromosomos kontekste bei μ reikšmės reikšmingumas slenkstinės vertės atžvilgiu.

Genominių sričių analizei naudota UCSC geno naršyklė [121], genų funkcinė reikšmė įvertinti naudotos *Ensembl* [115], *OMIM* [109], *GeneMANIA* [116], *PopHumanScan* [122], *PANTHER* [123], *Gene Ontology Browser* [124], *Human DNA Repair Genes* [125] duomenų bazės ir įrankiai.

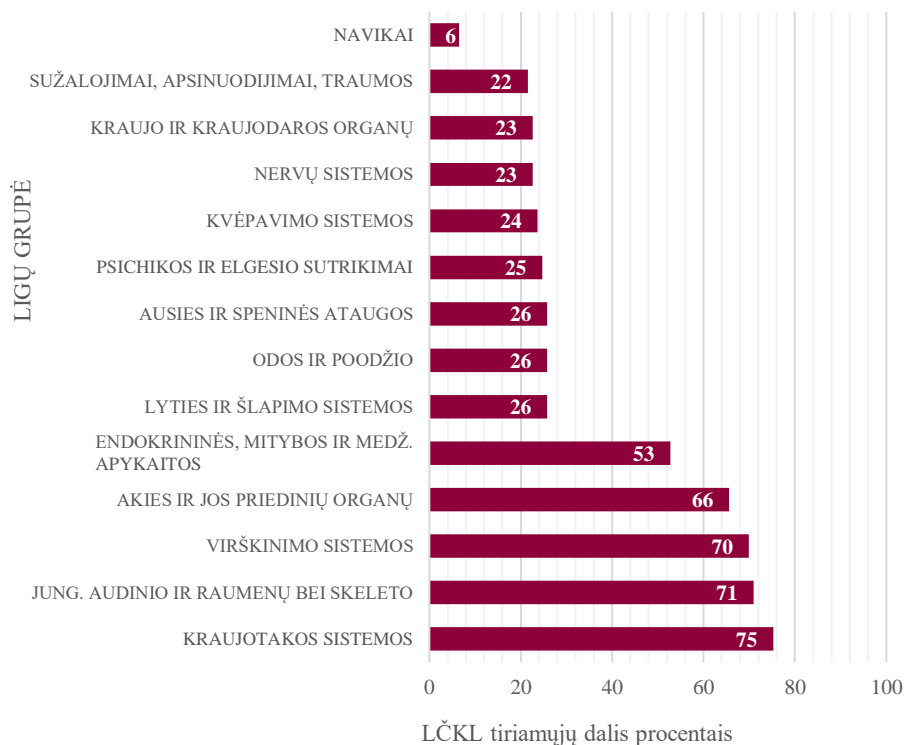
3. TYRIMO REZULTATAI IR JŲ APTARIMAS

3.1. Lietuvos Černobylio katastrofos likviduotojų anketinės apklausos ir ligų istorijų duomenų aprašomoji statistinė analizė

Tiriamąją imtį sudarė 93 LČKL – 50–78 m. vyrai. Lietuvos Černobylio katastrofos likviduotojai buvo išsiųsti dirbti katastrofos likvidavimo darbų į aukščiausio radioaktyvumo zoną (apie 30 km nuo branduolinio reaktoriaus). Šioje zonoje jie atliko įvairias užduotis, turėjusias užkirsti kelią radioaktyviam užterštumui plisti. Labiausiai žalingos veiklos buvo granito gabalų rinkimas rankomis, atominės elektrinės teritorijos tvarkymas, įrangos, buvusios elektrinės viduje arba arti jos, tvarkymas, dezinfekavimas. Kitos dažniau vykdytos veiklos buvo transporto priemonių į Černobylio atominę elektrinę ir iš jos vairavimas ar kelių tvarkymas.

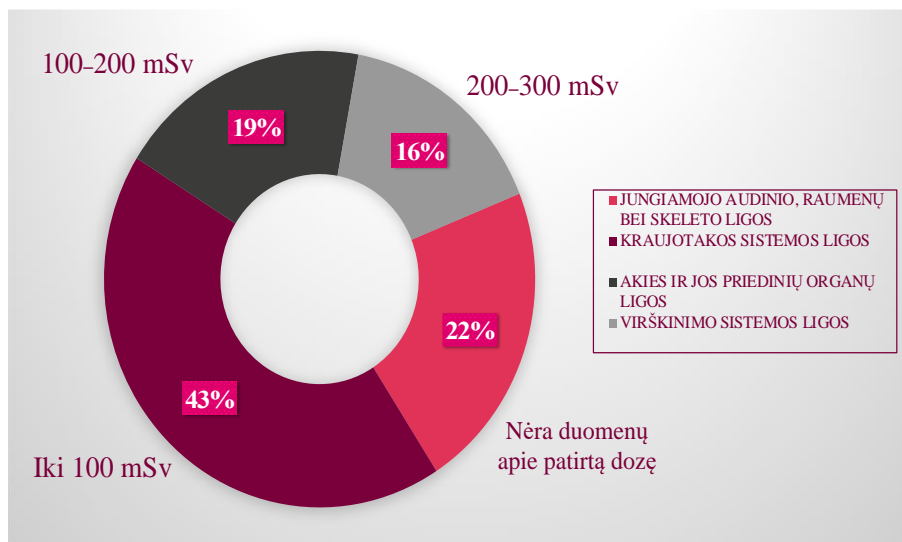
Remiantis oficialiais duomenimis [126], likviduotojų 1986–1989 m. patirtos jonizuojančiosios spinduliuotės (JS) dozės siekė iki 100 mSv. Šie duomenys patvirtina gautus rezultatus, kad net 43 proc. LČKL patyrė dozes iki 100 mSv (3.2 pav). Minėtina, kad duomenys apie patirtas JS dozes buvo surinkti anketinės apklausos metu ir ne visada buvo patvirtinti dokumentais (pavyzdžiui, dozė įrašyta kariniame biliete), todėl galimi netikslumai. Dozės, aukštesnės negu 200 mSv, nustatytos 16 proc. LČKL.

Remiantis klinikinės apžiūros duomenimis (II priedas, 66–83 klausimai), matyti, kad daugiausia LČKL serga kraujotakos sistemos, jungiamojo audinio, raumenų ir skeleto sistemos bei virškinimo sistemos ligomis. Šis dažniausių ligų grupių pasiskirstymas (3.1 pav.) beveik sutampa su pasiskirstymu bendrojoje lietuvių populiacijoje (2019 m. Lietuvos Higienos instituto duomenimis) [127]. Dauguma minėtų ligų susijusios su tiriamųjų amžiumi. Įdomu tai, kad tik keliems LČKL diagnozuoti navikiniai susirgimai, nors yra žinoma, kad JS veikia kancerogeniškai. Be to, navikiniai susirgimai – viena dažniausių mirties priežasčių bendrojoje lietuvių populiacijoje tarp 45 m. ir vyresnių asmenų [127]. Ši priešprieša galėjo atsirasti dėl unikalios, adaptyvios LČKL genomų struktūros arba dėl LČKL grupės struktūros (tikėtina, kad dauguma tiriamųjų, kuriems diagnozuota navikinių susirgimų, jau yra mirę).



3.1 pav. Dažniausių ligų grupių pasiskirstymas LČKL grupėje. X ašyje vaizduojama LČKL dalis procentais tam tikroje ligų grupėje. Kai kurie tiriamieji priklauso kelioms ligų grupėms. Y ašyje vaizduojamos ligų grupės.

Toliau charakterizuoti LČKL siekta atlikus ligų grupių ir patirtos JS asociacijų analizę. Didžioji dalis LČKL (40 iš 93; 43 %) patyrė JS dozes iki 100 mSv. Šioje grupėje dominavo kraujotakos sistemos ligos (3.2 pav.). Jonizuojančiosios spinduliuotės dozę tarp 100 ir 200 mSv patyrė 19 proc. (18 iš 93) LČKL, šioje grupėje dažniausios buvo akies ir jos priedinių organų ligos. Virškinimo sistemos ligos dominavo 200–300 mSv JS dozes patyrusių LČKL grupėje (16 %; 15 iš 93). Asociacija tarp skirtingų ligų grupių ir patirtos JS dozės nebuvo nustatyta.



3.2 pav. LČKL patirtos jonizuojančiosios spinduliuotės dozių pasiskirstymas. Procentai nurodo tiriamųjų dalį, priklausomai nuo patirtos skirtingos jonizuojančiosios spinduliuotės dozės. Diagramoje nurodytos dažniausios ligų grupės skirtingose patirtų jonizuojančiosios spinduliuotės dozių grupėse.

3.2. Bendrosios lietuvių populiacijos (LITGEN) poveikio genomo variantų analizė

Pirminiame tyrimo analizės etape siekta išaiškinti bendrajai lietuvių populiacijai būdingus poveikio genomo variantus, genotipavimo ir egzomo sekoskaitos duomenų analizei naudojant sudarytą poveikio genomo variantų katalogą. Iš 144 genomo variantų, esančių kataloge, atlikus Hardžio–Veinbergo pusiausvyros skaičiavimus, į tolesnę analizę įtraukta 70 poveikio genomo variantų. 39 variantai nustatyti ir genotipavimo, ir egzomo sekoskaitos duomenyse, 7 variantai nustatyti išskirtinai genotipavimo, 24 – egzomo sekoskaitos duomenyse. Bendrojoje lietuvių populiacijos grupėje (LITGEN) keturių variantų alelių dažniai statistiškai reikšmingai skyrėsi, palyginti su Šiaurės ar Vakarų Europos kilmės Jutos gyventojų (angl. *Utah Residents with Northern and Western European Ancestry*, CEU) ir suomių iš Suomijos (angl. *Finnish in Finland*, FIN) bei bendrosios Europos populiacijos (EUR) duomenimis (3.1 ir 3.2 lentelės). Dviejų variantų (*PPARG* ir *ADH1C* genuose) alelių dažnių lyginimo su bendrąja Europos populiacija rezultatas buvo artimas statistiškai reikšmingam. Apskaičiuota p reikšmė gali kisti priklausomai nuo imties dydžio, t. y. padidinus imties dydį, rezultatas gali tapti statistiškai reikšmingas arba nereikšmingas, todėl į tolesnę analizę pasirinkta įtraukti ir galimai statistiškai reikšmingus (artimus statistiniam

reikšmingumui) rezultatus. Atlikus genotipų dažnių tarp minėtų tiriamųjų grupių palyginimą, statistiškai reikšmingų skirtumų nenustatyta.

3.1 lentelė. Poveikio variantai ir tiriamųjų skaičius. Tiriamųjų skaičius LITGEN grupėje skiriasi priklausomai nuo naudotų duomenų rinkinių analizei (genotipavimo, egzomo sekoskaitos ar abiejų rinkinių).

Genomo varianto identifikacinis numeris	Genas	Susijusi būklė	LITGEN	CEU	FIN	EUR
rs1801282	PPARG	II tipo diabetas	168	99	99	504
rs13266634	SLC30A8	II tipo diabetas	464			
rs11556924	ZC3HC1	Koronarinė širdies liga	465			
rs2274223	PLCE1	Stemplės vėžys	463			
rs7498665	SH2B1	Nutukimas	98			
rs698	ADH1C	Priklausomybė nuo alkoholio	98			

LITGEN – bendroji lietuvių populiacija, CEU – Šiaurės ar Vakarų Europos kilmės Jutos gyventojai, FIN – suomių iš Suomijos populiacija, EUR – bendroji Europos populiacija.

3.2 lentelė. Poveikio genomo variantų alelių dažnių palyginimas tarp bendrosios lietuvių populiacijos (LITGEN) ir Europos populiacijų (CEU, FIN, EUR)

Genomo varianto identifikacinis numeris	Genas	Pokytis *	LEP	HOMO ALT		HETERO		HOMO REF		MAF (LTG)	MAF	χ^2	p
				LTG	LEP	LTG	LEP	LTG	LEP				
rs1801282	PPARG	NM_0013546682: c.34C>G	EUR	4	5	46	111	118	387	0,161	0,120	3,632	0,05
rs13266634	SLC30A8	NM_0011728152: c.826C>T	FIN	50	18	195	43	219	38	0,318	0,399	4,839	0,03
			CEU		6		36		57		0,242	4,387	0,04
rs11556924	ZC3HC1	NM_0012821901: c.1025G>A	FIN	77	5	225	40	163	54	0,408	0,253	16,642	<0,01
rs2274223	PLCE1	NM_0011659792: c.4856A>G	EUR	65	59	223	222	175	222	0,381	0,338	3,919	0,04
rs7498665	SH2B1	NM_0011458121: c.1450A>G	EUR	19	52	36	227	43	224	0,378	0,329	7,015	0,03
rs698	ADH1C	NM_000669.5: c.1048A>G	EUR	25	93	48	221	25	189	0,5	0,405	5,878	0,05

* Visi pokyčiai yra iš referentinio alelio į poveikio alelį, visi poveikio aleliai yra retieji (angl. *minor*).

LEP – lyginamoji Europos populiacija (CEU ar FIN, ar EUR), kurioje pasiekta statistinio reikšmingumo, CEU – Šiaurės ar Vakarų Europos kilmės Jutos gyventojų populiacija, EUR – bendroji Europos populiacija, FIN – suomių populiacija, LTG – bendroji lietuvių populiacija (LITGEN), MAF – poveikio alelio dažnis skirtingose Europos populiacijų grupėse, MAF LTG – poveikio alelio dažnis bendrojoje lietuvių populiacijos grupėje (LITGEN), HOMO ALT – homozigotinių poveikio alelių genotipų skaičius, HETERO – heterozigotinių genotipų skaičius, HOMO REF – homozigotinių referentinių alelių genotipų skaičius.

Remiantis mokslinė literatūra, galima teigti, kad koduojamo baltymo aminorūgštis keičiantys poveikio variantai galimai saugo nuo alkoholio priklausomybės (*ADH1C*, rs698), II tipo diabeto (*PPARG*, rs1801282; *SLC30A8*, rs13266634), koronarinės širdies ligos (*ZC3HC1*, rs11556924), nutukimo (*SH2B1*, rs7498665) ir stemplės vėžio (*PLCE1*, rs2274223). Taikant laisvai prieinamus *in silico* analizės įrankius (*Varsome*, *UniProt*, *Ensembl*), nustatyta, kad penki iš šešių poveikio geno variantų įvardijami kaip nepatogeniniai arba riziką lemiantys (angl. *risk-determining*). Šie penki variantai (*PPARG*: rs1801282, *SLC30A8*: rs13266634, *ZC3HC1*: rs11556924, *PLCE1*: rs2274223, *SH2B1*: rs7498665) įtraukti į poveikio geno variantų katalogą, sudarytą iš mokslinių publikacijų [19, 27]. Publikacijose minėti poveikio geno variantai įvardyti kaip galimai protekciniai ir veikiami teigiamos gamtinės atrankos.

Remiantis duomenų bazėmis *Ensembl*, *ClinVar* ir *OMIM*, šeštasis variantas (*ADH1C*, rs698) klasifikuojamas kaip protekcinis ir lemiantis etanolio metabolizmo pokyčius. Vis dėlto mokslinių tyrimų duomenys šiuo atžvilgiu kontroversiški. Nustatyta, kad pokytis susijęs su lėtesniu etanolio metabolizmu, o tai gali lemti pailgėjusį alkoholio vartojimo laiką ir didesnių kiekių suvartojimą, siekiant norimo efekto. Taigi žmonėms, turintiems šį alelį, kyla didesnė rizika ilgiau ir pertekliškai vartoti alkoholį [128, 129]. Nustatyta, kad dažni VNP gali būti atsakingi už 30 proc. požymio (priklausomybė alkoholiui) variabilumo [130]. Nepaisant to, VNP, susijusių su padidėjusiu polinkiu į perteklinį alkoholio vartojimą, tyrimai atskleidė nedidelį pavienių variantų reikšmingumą fenotipui ir šių variantų sąsają su įvairiomis psichikos būklėmis [131]. Mokslinių išvadų nevienareikšmiškumas tik parodo molekulinį procesų, susijusių su padidėjusiu alkoholio vartojimu, žinių stoką, o tai galimai padėtų išspręsti dideli, aiškiai charakterizuoti sergančiųjų ir kontrolinių grupių duomenų rinkiniai. Analizės metu, remiantis duomenų bazėmis, taip pat nustatytas kalbamojo varianto protekcinis poveikis, nors mokslinėje literatūroje nurodoma kitaip. Todėl, vertinant duomenų bazėse pateikiamą informaciją, svarbu analizę papildyti naujausių mokslinių publikacijų paieška, siekiant kuo tiksliau išsiaiškinti funkcinę geno varianto reikšmę.

Dauguma analizuotų poveikio geno variantų yra dažni (retojo alelio dažnis >1 %). Galima manyti, kad, jei retasis alelis vis tik atlieka protekcinę funkciją ir individui suteikia evoliucinio pranašumo, jis, veikiamas teigiamos gamtinės atrankos, turėtų tapti vienu iš dažniausių alelių populiacijoje [19]. Didelis dažnis galėtų paaiškinti ir nustatytą poveikio geno variantų funkcinės reikšmės įvertinimą duomenų bazėse – dauguma šių variantų identifikuojami kaip nepatogeniniai.

Analizuojant poveikio genomo variantus, be funkcinės reikšmės, taip pat svarbu atkreipti dėmesį į poveikio alelio kilmę. Siekiant pagrįsti hipotezę, kad nesininiminius VNP yra naudingas asmens adaptacijai, tikrinta, kuris iš alelių yra išvestinis (angl. *derived*), o kuris – protėvių kilmės (angl. *ancestral*). Retasis alelis nebūtinai yra išvestinis, todėl, norint nustatyti alelio kilmę, genotipai gali būti lyginami su primatų rūšimis. Toks lyginimas tiriamojame populiacijoje leido įvardyti statistiškai reikšmingų poveikio genomo variantų sąrašo variantus (variantai *PLCE1*, *ADH1C* ir *SH2B1* genuose), kurie žmogaus genome vis dėlto yra protėvių kilmės. Genomo variantų adaptacinis (protekcinis) vaidmuo gali būti svarstomas, kai tiriamas alelis yra išvestinis, todėl variantai *PLCE1*, *ADH1C* ir *SH2B1* genuose nebuvo priskirti prie protekcinių ir lemiančių geresnes lietuvių adaptacines savybes, taigi nebuvo įtraukti į tolesnę analizę.

Galiausiai, *PPARG*, *SLC30A8* ir *ZC3HC1* genuose nustatyti išvestinės kilmės aleliai buvo laikomi reikšmingais bendrosios lietuvių populiacijos (LITGEN) tiriamųjų grupės adaptacijai. Šie genai gali turėti įtakos sumažėjusiam tiriamųjų polinkiui sirgti II tipo diabetu (variantai *PPARG* ir *SLC30A8* genuose) ir koronarine širdies liga (*ZC3HC1* geno variantas). Didelės apimties plataus masto genomo asociacijos tyrimai (angl. *genome wide association studies*, GWAS) leido identifikuoti nemažai genomo variantų, susijusių su II tipo diabetu [132], ir tik kelis variantus, siejamus su mažesniu polinkiu į šią ligą. Neseniai atliktoje metaanalizėje [133] nustatytas rs1801282 variantas *PPARG* gene, sietinas su mažesne rizika sirgti II tipo diabetu (minėtina, kad tas pats variantas nustatytas ir disertacijoje pristatomo tyrimo metu). Vieno nukleotido polimorfizmai *PPARG* (branduolio receptoriuje) atlieka reikšmingą vaidmenį kontroliuojant lipidų ir gliukozės metabolizmą. Protekcinė poveikio genomo varianto funkcija statistiškai dažniau pastebima europiečių (18 %), rytų Azijos (20 %) ir pietryčių Azijos (18 %) populiacijose [133]. Minėtoje metaanalizėje nurodoma, kad šiaurės europiečiai, esantys šio varianto nešiotojai, turi mažesnę riziką sirgti II tipo diabetu, palyginti su centrinės ar pietų Europos populiacijomis. Tai papildoma ir disertacijoje pristatomo tyrimo rezultatais.

Kitas variantas (rs13266634), nustatytas atliekant tyrimą, taip pat siejamas su II tipo diabetu. Šis variantas yra *SLC30A8* gene. Genas koduoja baltymą, kuris dalyvauja insulino išskyrimo procese, todėl yra tiesiogiai susijęs su diabeto biologiniu mechanizmu. J. Flannicko ir bendraautorių [134] bei D. Brunke-Reese ir kt. [135] tyrimais nustatyta, kad minėtas poveikio variantas, kartu su kitais retesniais funkcijos praradimo variantais, yra susijęs su mažesne tikimybe sirgti II tipo diabetu.

II tipo diabetas glaudžiai susijęs su koronarine širdies liga [136]. Atliekant tyrimą, bendrosios lietuvių populiacijos tiriamųjų grupėje nustatytas poveikio genomo variantas rs11556924 *ZC3HCl* gene, kuris mokslinėse publikacijose klasifikuojamas kaip, tikėtina, lemiantis koronarinę širdies ligą. *ZC3HCl* yra biologiniais molekuliniiais mechanizmais susijęs su *KLHDC10* genu, dalyvaujančiu oksidacinio streso sukeltos ląstelių žūties ir uždegiminio atsako procesuose. Žinoma, kad šie procesai svarbūs aterosklerozės ir, atitinkamai, koronarinės širdies ligos etiopatogenezėje [136]. Teigiamos gamtinės atrankos analizė, apskaičiuojant integruotą haplotipų įvertį, parodė, kad šis ir kiti poveikio genomo variantai *ZC3HCl* gene galimai yra veikiami teigiamos gamtinės atrankos ir potencialiai naudingi tiriamųjų genomams [19, 37], jie gali būti protekciniai.

Analizuojant populiacinius duomenis, svarbu apžvelgti ir aplinkos veiksnių įtaką tiriamiems fenotipiniams požymiams (šiuo atveju – ligoms). Lietuvos sveikatos informacijos centro duomenimis, kuriuos pateikė Higienos institutas, širdies ir kraujagyslių sistemos ligos 2019 m. Lietuvoje sudarė didžiausią mirčių dalį (54,6 %) [137]. Tais pačiais metais 4,8 proc. populiacijos sirgo II tipo diabetu, o 6,9 proc. – koronarine širdies liga. Palyginimui, širdies ir kraujagyslių sistemos ligos genetiškai artimoje suomių populiacijoje, Suomijos statistikos duomenimis (angl. *Statistics Finland*), 2019 m. taip pat buvo viena pagrindinių mirties priežasčių (33,9 %). Nacionalinės Suomijos sveikatos 2018 m. surengtos apklausos metu nustatyta, kad diabeto paplitimas tarp 30 m. ir vyresnių asmenų siekė 7,8 proc. [138], iš jų dauguma asmenų (85 %) sirgo II tipo diabetu [139]. Taigi, nors tirtoje bendrojoje lietuvių populiacijos grupėje nustatyta poveikio genomo variantų, galimai apsaugančių nuo minėtų ligų, sergamumui ir mirtingumui nuo koronarinės širdies ligos turi įtakos gyvenamos ir kiti aplinkos veiksniai.

Apibendrinant galima teigti, kad aptartuoju tyrimo etapu nustatyti trys poveikio genomo variantai, galimai lemiantys protekcinį poveikį II tipo diabetui ir koronarinei širdies ligai. Dviejų iš jų (*SLC30A8* ir *ZC3HCl* genuose) alelių dažniai tarp lyginamųjų populiacijų skyrėsi statistiškai reikšmingai. Tyrimas atskleidė ir dažnų genomo variantų bei jų poveikio nustatymo komplikuotumą ir svarbą. Tikslus genomo variantų poveikio apibrėžimas galėtų prisidėti prie patikimesnių duomenų bazių kūrimo ir padėtų išvengti prieštaringos varianto poveikį apibūdinančios informacijos, kaip minėtu nustatyto *ADHIC* geno varianto atveju. Poveikio genomo variantus svarbu nustatyti, siekiant geriau suvokti ligų etiopatogenezės mechanizmus ir mikroevoliucinius procesus.

Daugelyje mokslinių tyrimų analizuojami reti genomo variantai koduojančiose genų srityse, todėl svarbios sąveikos tarp koduojančių ir

nekoduojančių genomo variantų lieka per mažai išnagrinėtos [26]. Dėl šios priežasties, išanalizavus genetinę populiacijos struktūrą, svarbu vykdyti tolesnę analizę, orientuojantis į pailgintų genotipų mechanizmus ir sąsajų aiškinimąsi, siekiant išskirti ligas lemiančius ar sveikatai palankius genomo veiksnius.

3.3. Lietuvos Černobylio katastrofos likviduotojų poveikio genomo variantų asociacijų analizė

Šiuo tyrimo etapu, naudojant VNP žymenų genotipavimo duomenis, atlikta LČKL ir bendrosios lietuvių populiacijos (LITGEN) tiriamųjų grupių asociacijų analizė. Kryptingai analizuoti atrinkti poveikio genomo variantai, įtraukti į sudarytą poveikio genomo variantų katalogą.

3.3.1. Lietuvos Černobylio katastrofos likviduotojų ir bendrosios lietuvių populiacijos (LITGEN) tiriamųjų grupių poveikio genomo variantų asociacijų analizė

Pirmoji analizės dalis įgyvendinta nagrinėjant LČKL grupės duomenis, juos lyginant su bendrosios lietuvių populiacijos (LITGEN) visos tiriamųjų grupės (vyrų ir moterų) duomenimis. Po kokybės kontrolės etapų, tolesnę analizę taikyta 91 LČKL ir 421 bendrosios lietuvių populiacijos grupių tiriamiesiems, 46 poveikio genomo variantams. Genotipų dažniu statistiškai reikšmingai skyrėsi, palyginti su bendrosios lietuvių populiacijos genotipų dažniais, poveikio genomo variantas rs3825942 (NM_005576.4:c.458G>A; NP_005567.2:p.(Gly153Asp)) *LOXLI* gene (3.3 lentelė). Alelių dažnių palyginimas nebuvo statistiškai reikšmingas.

3.3 lentelė. Poveikio genomo varianto rs3825942 *LOXLI* gene apibendrinta statistika

Tiriamoji grupė	AA genotipų skaičius	GA genotipų skaičius	GG genotipų skaičius	Retojo alelio dažnis	Fišerio tikslinio kriterijaus p reikšmė	Galimybių santykis (AA genotipas)	Galimybių santykis (GA genotipas)
LČKL	1	35	55	0,203	p = 0,019	0,34 [95 % PI 0,04–2,63]	1,84 [95 % PI 1,14–2,95]
LITGEN	16	104	301	0,162			

PI – pasikliautinis intervalas.

LČKL grupę sudarė vyrai, tačiau šiame tyrimo etape, siekiant padidinti atliekamo tyrimo statistinę galią, pasirinkta įtraukti bendrosios lietuvių populiacijos (LITGEN) tiriamųjų grupės vyrus ir moteris. Siekiant nustatyti,

ar nėra populiacijos stratifikacijos dėl lyties, atliktas *LOXLI* geno varianto χ^2 testas. Rezultatai parodė, kad tyrimo rezultatų skirtumas, atsižvelgiant į lytį (kontrolinėje grupėje – tik vyrai arba vyrai ir moterys), bendrosios lietuvių populiacijos (LITGEN) tiriamųjų grupėje nėra statistiškai reikšmingas ($\chi^2 = 1,103$, $p = 0,576$).

LOXLI genas koduoja fermentą, reikalingą elastino biogenezei, dalyvaujantį kolageno struktūros sudarymo procese. Dėl šios priežasties fermentas laikomas vienu svarbiausių elementų elastinėms skaiduloms formotis ir remodeluoti [140]. Svarbiausią biologinį vaidmenį *LOXLI* genas atlieka elastino homeostazėje ir remodeliuojant matriksą traumų metu, fibrozės ir vėžinių procesų vystymosi procesuose. Padidėjusi *LOXLI* raiška lemia fibrozines ligas (pavyzdžiui, plaučių fibrozę), o sumažėjusi – eksfoliacijos sindromą [141]. R. P. Igo ir bendraautoriai [142] nustatė, kad geno variantai šiame gene (variantas nustatytas ir atliekant disertacijoje pristatomą tyrimą) lemia padidėjusią (galimybių santykis ~20) eksfoliacijos sindromo riziką. Vis dėlto minėtina, kad yra mokslinių tyrimų, prieštaraujančių nustatyto poveikio geno varianto žalingam vaidmeniui ir galimai patvirtinančių protekcinės įtakos fenotipui hipotezę.

Poveikio geno variantui rs382594 J. M. Butlerio ir bendraautorių [19] nustatyta neigiama integruoto haplotipų įverčio reikšmė (–0,453) rodo teigiamos gamtinės atrankos poveikį. Tai reiškia, kad šis variantas gali būti potencialiai naudingas fenotipui ir galimai apsaugo nuo eksfoliacijos sindromo LČKL grupėje, taip lemdamas geresnę tiriamųjų adaptaciją. LČKL statistinė genotipų analizė atskleidė ir sumažėjusią galimybių santykio reikšmę, kai LČKL turi homozigotinį poveikio geno variantų genotipą ($GS(AA) = 0,34$, kai $GS(GA) = 1,84$). Tai tam tikra nuoroda į galimai protekcinį šio varianto poveikį.

Kitas argumentas, nurodantis rs3825942 poveikio geno varianto protekcinę funkciją, susijęs su JS. Žinoma, kad JS sukuria laisvuosius radikalus, kurie organizmui sukelia oksidacinį stresą. Stresas sukelia ląstelinį antioksidantų (jų kiekį sumažina) ir prooksidantų (jų kiekį padidina) disbalansą. Tai lemia dvigrandžius DNR trūkius [143, 144]. Sumažėjęs antioksidantų kiekis ir protėvių kilmės rs3825942 varianto alelis yra du reikšmingi komponentai eksfoliacijos sindromo etiopatogenezeje [145, 146]. Atlikto tyrimo rezultatai šią hipotezę pagrindžia. Likviduojant Černobylio elektrinės katastrofos padarinius, LČKL patyrė didesnes negu įprasta JS dozes. Dėl šios priežasties jie patyrė didesnę oksidacinį stresą, o jų organizmuose galimai sumažėjo antioksidantų kiekis. Eksfoliacijos sindromas tiesiogiai susijęs su glaukoma, ji diagnozuota 3,22 proc. tiriamųjų (glaukoma nustatyta iš tiriamųjų ligų istorijų). Minėtina, kad glaukoma sirgo tik

rs3825942 geno varianto homozigotinį protėvių kilmės alelį turintys tiriamieji. Lietuvos Černobylio katastrofos likviduotojai, tirtos poveikio geno varianto (išvestinio alelio) nešiotojai (tiek homozigotinis, tiek heterozigotinis genotipai), glaukoma nesirgo. Tai atskleidžia poveikio geno varianto rs3825942 retojo alelio vaidmenį vykdant natūralią (adaptyvią) glaukomos ir eksfoliacijos sindromo prevenciją. Be to, nustatyta, kad *LOXLI* geno raiška, veikiant oksidaciniam stresui, aktyvinama, o tai pagrindžia šio geno funkcinį vaidmenį esant ląstelės streso atsakui ir eksfoliacijos sindromo etiopatogenezei [141]. Vis dėlto minėto geno variantai sukelia sumažėjusią geno raišką [147], todėl gauti rezultatai papildė hipotezę apie poveikio geno variantų apsaugą nuo eksfoliacijos sindromo ir glaukomos oksidacinio streso aplinkoje.

Galiausiai, rs3825942 poveikio geno varianto aleliui būdinga dažnių kaita tarp skirtingų populiacijų (angl. *allele flipping*) [140, 148]. Tai procesas, kai poveikio geno varianto retasis alelis laikomas žalingu vienoje populiacijoje ir protekcinu kitose. Alelių dažnio ir poveikio variacija tarp skirtingų populiacijų leidžia daryti prielaidą, kad rs3825942 poveikio geno variantas yra ne vienintelis svarbus ligos molekuliniam mechanizmui, nors asociacija yra stipri [147, 149]. Minėtina, kad šis geno variantas nustatomas didelėje bendrosios populiacijos dalyje, kai retojo alelio dažnis, remiantis *dbSNP* duomenų baze, siekia 20–30 proc. Poveikio geno variantas nustatytas 98 proc. eksfoliacijos sindromu sergančių asmenų, tačiau tas pats variantas nustatytas ir 85 proc. nesergančių asmenų [150]. Didelis retojo alelio dažnis skirtingose populiacijose patvirtina galimą protekcinę šio poveikio geno varianto funkciją, tačiau reikia atkreipti dėmesį ir į populiacijai būdingus išorės veiksnius. Tokie veiksniai, kaip hipoksija, oksidacinis stresas, ultravioletinės spinduliuotės kiekis, net žemas folatų ir aukštas kofeino vartojimas, keičia *LOXLI* geno raišką [141, 151] ir, tikėtina, prisideda prie eksfoliacijos sindromo pasireiškimo.

Nustatytas poveikio geno variantas yra ne vienintelis geno veiksnys, darantis įtaką eksfoliacijos sindromui, todėl, siekiant išsiaiškinti naujus su rs3825942 poveikio geno variantu susijusius ir ligos molekulinuose mechanizmuose dalyvaujančius genus, atliktos nepusiausvyros sankibos ir epistazės analizės. Šios analizės gali padidinti tikimybę nustatyti naujas genų asociacijas, nedidelio reikšmingumo žymenis ir paaiškinti dalį trūkstamos paveldimumo informacijos [152]. Nustatyta, kad *LOXLI* variantas rs3825942 yra nepusiausvyroje sankiboje ir formuoja haplotipų bloką su kitais variantais (rs2165241, rs1078967, rs4886776, rs8041685, rs8042039, rs4886782, rs2304719, rs750460) pačiame gene. Kiti geno variantai taip pat yra nepusiausvyroje sankiboje (3.4 lentelė).

Epistazės analizė leidžia nustatyti varianto poveikio priklausomybę nuo kitų pokyčių buvimo ar nebuvimo kituose genuose (vadinamieji *modifikuojantieji variantai*, angl. *modifier variants*). Atlikta epistazės analizė identifikavo rs1500733 ir rs6502043 variantus, kurių veikimas gali būti priklausomas nuo rs3825942 varianto *LOXLI* gene.

3.4 lentelė. Nepusiausvyros sankibos ir epistazės analizės rezultatai *LOXLI* geno rs3825942 poveikio geno variantui

Genomo varianto identifikacinis numeris	Genas	Analizės tipas	Statistinis įvertis ir reikšmė
rs13243476	<i>LHFPL3</i>	Nepusiausvyra sankiba	$r^2 = 0,204$
rs10876168	<i>GALNT6</i>	Nepusiausvyra sankiba	$r^2 = 0,250$
rs7462	<i>PIH1D1</i>	Nepusiausvyra sankiba	$r^2 = 0,222$
rs1500733	<i>ANKS1B</i>	Epistazė	$p = 6,533 \times 10^{-5}$; GS = 7,824 (95 % PI 2,85–21,46)
rs6502043	<i>METRNL</i>	Epistazė	$p = 9,828 \times 10^{-5}$; GS = 0,168 (95 % PI 0,07–0,41)

GS – galimybių santykis, PI – pasikliautinis intervalas.

Atliekant minėtas analizes, nustatyti nauji kandidatiniai genai, galimai turintys įtakos eksfoliacijos sindromo etiopatogenezei. Šie genai anksčiau nebuvo siejami su eksfoliacijos sindromu. Genų biologinių sąsajų analize, naudojant *GeneMANIA* genų sąveikos įrankį, nustatytos netiesioginės sąsajos, tačiau mokslinėje literatūroje šios sąsajos nėra aprašytos. Tai patvirtina gautų rezultatų mokslinį naujumą ir tolesnių *in silico* ir funkcinių tyrimų svarbą, siekiant išaiškinti bendruosius molekulinis mechanizmus.

Apibendrinant, šiuo tyrimo etapu įrodyta, jog poveikio variantas rs3825942 *LOXLI* gene, iki šiol sietinas su didesne eksfoliacijos sindromo ir glaukomos rizika, galimai lemia LČKL apsaugą nuo šių ligų ir geresnę tiriamųjų adaptaciją. Poveikio geno variantų analizė naudinga aiškinantis daugiaveiksnių ligų ir fenotipinių požymių genetinę architektūrą. Kai kurie pavieniai poveikio geno variantai gali lemti mažesnę ligos riziką, tačiau egzistuoja ir sudėtiniai genetiniai mechanizmai, todėl labai svarbu tirti genų sąveikas ir reguliacinius elementus. Minėti elementai svarbūs analizuojamų

variantų genų raiškai, jie gali turėti tiesioginės įtakos ligoms pasireikšti. Taigi reikalingi tolesni tyrimai, analizuojantys ne tik atskirus geno variantus ar jų įvairovę, bet ir gilinantis funkcinio lygmens žinias. Remiantis pristatomo tyrimo pavyzdžiu, galima manyti, kad retų, konservatyviose geno srityse esančių *LOXLI* geno variantų funkcinė analizė skirtingose etninėse populiacijose būtų reikšmingas žingsnis siekiant išaiškinti eksfoliacijos sindromo ir glaukomos biologiją. Transkriptomo analizė ląstelės ar audinio lygmeniu galėtų padėti identifikuoti genų sąveikos tinklus, o juos patvirtinus, galėtų būti nustatytas genų sąveikos tinklas. Realūs žalingi ir protekciniai geno variantai galėtų būti taikomi kaip terapeutiniai taikiniai asmeninėje (personalizuotoje) medicinoje.

3.3.2. Lietuvos Černobylio katastrofos likviduotojų ir bendrosios lietuvių populiacijos (LITGEN) vyrų, suskirstytų pagal amžiaus grupes, poveikio geno variantų asociacijų analizė

Antruoju tyrimo analizės etapu LČKL grupės duomenys lyginti su bendrosios lietuvių populiacijos (LITGEN) vyrų, suskirstytų į amžiaus grupes, duomenimis. LČKL grupės tiriamųjų gene alelių dažniai statistiškai reikšmingai skyrėsi nuo bendrosios lietuvių populiacijos (LITGEN) 65 m. ir vyresnių vyrų amžiaus grupėje. Retojo alelio dažnis buvo didesnis, palyginti su bendrosios lietuvių populiacijos tiriamųjų grupe.

Minėtina, kad vyrų amžius varijavo tarp 50 ir 78 m., todėl, siekiant kuo labiau suvienodinti tiriamas populiacijas, iš bendrosios lietuvių populiacijos (LITGEN) grupės atrinkti to paties amžiaus vyrai. Analizuotos trys amžiaus grupės: 50–78 m. (imties analizė), 50–64 m. ir 65 m. ir vyresnio amžiaus tiriamieji (tiriamųjų grupė suskirstyta į dvi dalis, kai individų amžius mažesnis ar didesnis negu vidutinis 64 m. tiriamųjų grupės amžius). Tyrimo metu nustatyta, kad du poveikio geno variantų rs3732378 (NM_001337.4:c.839C>T) ir rs3732379 (NM_001337.4:c.745G>A) *CX3CR1* gene alelių dažniai statistiškai reikšmingai skiriasi nuo bendrosios lietuvių populiacijos (LITGEN) 65 m. ir vyresnių vyrų amžiaus grupėje. Retojo alelio dažnis yra didesnis, palyginti su bendrosios lietuvių populiacijos tiriamųjų grupe (3.5 lentelė). Abu minėti variantai yra 94 bazių porų atstumu vienas nuo kito, todėl yra nepusiausvyroje sankiboje ir paveldimi kartu [153].

3.5 lentelė. Poveikio genomo variantų alelių dažnio palyginimas rs3732378 ir rs3732379 CX3CRI gene

Genomo varianto identifikacinis numeris	Retojų alelio dažnis LČKL	Retojų alelio dažnis LITGEN	χ^2 kriterijus	p reikšmė	Galimybių santykis	95 % PI
rs3732378	0,275	0,110	7,146	0,008	3,077	1,317–7,189
rs3732379	0,338	0,171	5,958	0,015	2,474	1,182–5,179

LČKL – Lietuvos Černobylio katastrofos likviduotojai, LITGEN – bendroji lietuvių populiacija, PI – pasikliautinis intervalas. Imčių dydžiai – 40 LČKL ir 41 LITGEN tiriamųjų, 65 m. ir vyresnio amžiaus vyrų.

CX3CRI genas koduoja chemokino receptorių ir dalyvauja imuninio atsako, uždegiminiuose, ląstelių adhezijos ir chemotaksio procesuose, daugiausia smegenų ir kraujagyslių audiniuose bei imuninės sistemos ląstelėse. *OMIM* duomenų bazės duomenimis, šis genas siejamas su geltonosios dėmės degeneracija, greitesne žmogaus imunodeficitu viruso raida ir žemesne koronarinės ligos rizika. Nustatyti poveikio genomo variantai siejami su išaugusia vėlyvos pradžios Alzheimerio ligos rizika [154], padidėjusiu arterinio kraujo tūriu smegenyse per mikroglijos ir smulkiųjų kapiliarų endotelio ląstelių sąveikas [155]. Įdomu tai, kad tie patys poveikio genomo variantai nurodomi ir kaip protekciniai – jie mažina koronarinės širdies ligos riziką. Protekcinė minėtų genomo variantų funkcija įrodyta ne vienu moksliniu tyrimu [153, 156, 157].

CX3CRI genas veikia kaip uždegiminio proceso reguliatorius aterogenezės (aterosklerozinių plokštelių formavimosi proceso) metu. Šis genas veikia per makrofagų ir monocitų sutelkimą į aterosklerozinių plokštelių formavimosi vietą, kurioje vyksta uždegiminiai procesai, taip skatindamas arterijų ląstelių išgyvenamumą. Nustatyti poveikio variantai mažina ligando ir receptoriaus jungimąsi, taip sumažindami koronarinės širdies ligos riziką. Minėti rezultatai buvo įrodyti ir funkciniais tyrimais, panaudojant laboratorinių pelių modelius. Pelės, kurioms buvo išjungta šio geno raiška, rečiau sirgo ateroskleroze [157].

Apibendrinant aptartojo tyrimo etapo rezultatus, galima teigti, kad LČKL genomuose nustatyti du poveikio genomo variantai rs3732378 ir rs3732379 *CX3CRI* gene. Šie variantai galimai yra protekciniai, lemiantys geresnę LČKL adaptaciją ir potencialiai apsaugantys nuo koronarinės širdies ligos. Nepaisant to, širdies ir kraujagyslių ligos yra vienos iš labiausiai paplitusių

LČKL grupėje. Koronarinė širdies liga yra daugiaveiksnių etiologijos, todėl derėtų nepamiršti ir gyvensenos veiksnių, turinčių įtakos ligai. Pagrindiniai koronarinės širdies ligos rizikos veiksniai – rūkymas, riebaus maisto vartojimas, mažas fizinis aktyvumas, stresas ir depresija [158]. Remiantis anketinės apklausos duomenimis, matyti, kad LČKL grupėje dauguma šių rizikos veiksnių yra dažni: rūko ar rūkė 60 proc., mažai fiziškai aktyvūs 15 proc. (kai *neaktyvus* darbas ir laisvalaikis, 27 proc. – kai *neaktyvus* darbas ir *vidutiniškai* aktyvus laisvalaikis), psichikos sutrikimų turi 40 proc. tiriamųjų. Taigi protekcinę funkciją atliekančių genomo variantų vaidmuo yra reikšmingas, tačiau ši reikšmingumą gali nustelbti daug rizikos veiksnių turintis gyvenimo būdas.

3.3.3. Lietuvos Černobylio katastrofos likviduotojų ir bendrosios lietuvių populiacijos (LITGEN) 50–78 m. vyrų visų genotipavimo duomenų asociacijų analizė

Trečioji analizės dalis – LČKL grupės duomenys lyginti su bendrosios lietuvių populiacijos (LITGEN) 50–78 m. vyrų duomenimis, naudojant visus gautus genotipavimo duomenis. Atliekant didelio kiekio duomenų analizę, pavyzdžiui, atliekant didelio skaičiaus genomo variantų ir genų asociacijų analizę, naudojami daugybinių skaičiavimų (angl. *multiple testing*) korekcijos metodai, tokie kaip Bonferroni, Holmsio ir kt. Pritaikius šiuos metodus, statistiškai reikšmingų genomo variantų nebuvo nustatyta, tačiau tai nebūtinai reiškia, kad reikšmingų asociacijų nėra. Vienas iš pagrindinių plataus masto genomo asociacijos tyrimų (GWAS) trūkumų – daugybiniai skaičiavimai. Dėl šios priežasties prarandamas asociacijų reikšmingumas ir galimai svarbūs duomenys [159, 160]. Tai rodo *kryptingų* asociacijų analizių svarbą, nes dominančios genomo sritys konkretizuojamos ir analizės atliekamos tik toms konkrečioms sritims. Kryptinga analizė padidina tikimybę nustatyti statistiškai ir biologiškai reikšmingus genomo variantus ir sritis. Visa tai atskleidžiama ankstesniuose šios disertacijos skyriuose.

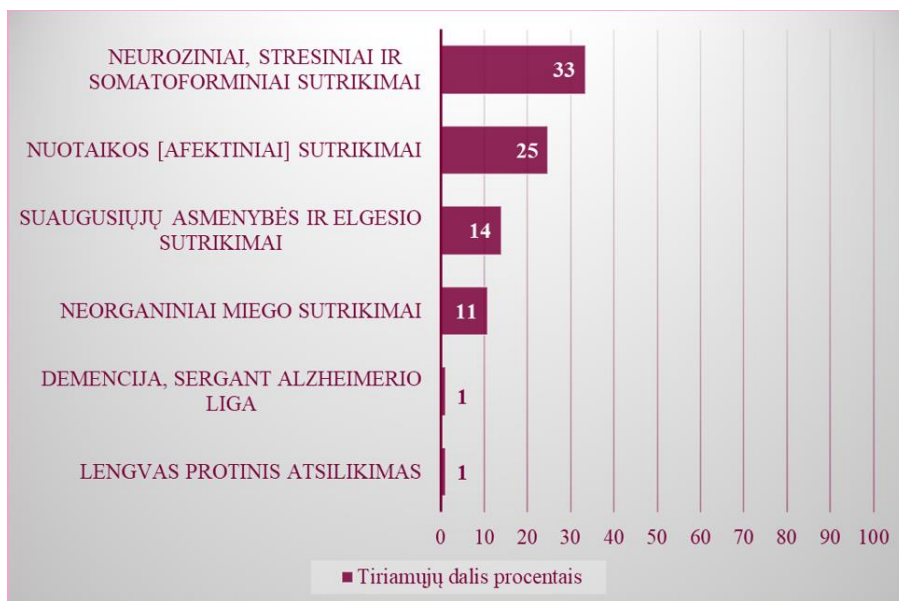
3.4. Lietuvos Černobylio katastrofos likviduotojų ir bendrosios lietuvių populiacijos (LITGEN) 50–78 m. vyrų kryptinga, su ligomis siejamų genų analizė

Šiuo tyrimo analizės etapu atlikta kryptinga, su ligomis siejamų genų tiriamųjų grupių asociacijų analizė. Ligų grupės pasirinktos remiantis dideliu analizuotų ligų dažniu LČKL grupėje bei atsižvelgiant į patirtų rizikos veiksnių (JS, stresas) sąsają su analizuotomis ligomis.

3.4.1. Psichikos būklių asociacijų analizė

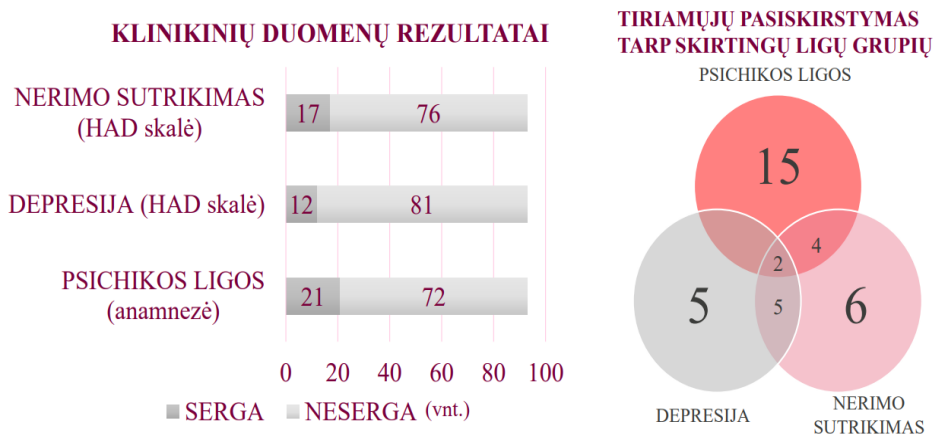
Nuo Černobylio katastrofos praėjo daugiau negu 35 m., tačiau šio įvykio padariniai vis dar juntami ir šiandien aktualūs. Likę gyvi LČKL išgyveno ekstremalias sąlygas ir prisitaikė prie likvidavimo darbų metu patirtų išgyvenimų psichologinių padarinių, besitęsiančių visą gyvenimą. Mokslinėje literatūroje minimas dalyvavimo tokio masto katastrofose poveikis asmenų psichinei sveikatai. Dalis likusių gyvų LČKL iki šiol kenčia nuo nerimo, potrauminio streso sindromo ir depresijos [161]. Potencialiai prisitaikymą ar neprisitaikymą gali lemti unikali genomo variacija, todėl vienas iš šio tyrimo analizės etapo tikslų – nustatyti genetinius veiksnius, siejamus su psichikos sutrikimais LČKL grupėje ir palyginti juos su bendrosios lietuvių populiacijos duomenimis.

Sujungus HAD skalės psichologinio klausimyno ir ligų istorijų duomenis, nustatyta, kad dominuojančios psichikos ligos LČKL grupėje yra nerimas (neurozinių, stresinių ir somatoforminių sutrikimų grupė), depresija (nuotaikos sutrikimų grupė), emocinis nestabilumas (suaugusiųjų asmenybės ir elgesio sutrikimų grupė), miego sutrikimai (neorganinių miego sutrikimų grupė) (3.3 pav.).



3.3 pav. Psichikos sutrikimų paplitimas LČKL grupėje. Dažniausi sutrikimai – nerimas (neurozinių, stresinių ir somatoforminių sutrikimų grupė), depresija (nuotaikos sutrikimų grupė), emocinis nestabilumas (suaugusiųjų asmenybės ir elgesio sutrikimų grupė), miego sutrikimai (neorganinių miego sutrikimų grupė). Daliai tiriamųjų skirtingi keli psichikos sutrikimai (žr. 3.4 pav.).

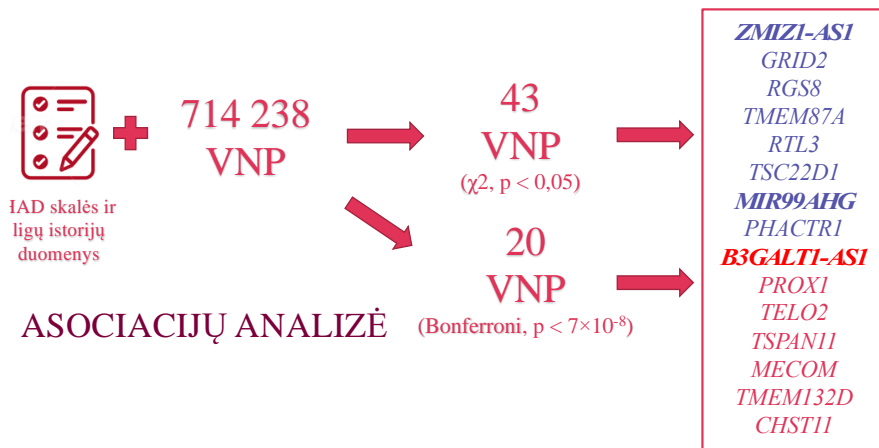
Remiantis ligų istorijų duomenimis, Lietuvos Černobylio katastrofos likviduotojų grupėje skirtingos psichikos būklės (tarp jų – depresija ir nerimo sutrikimas) nustatytos 21 tiriamajam. HAD skalės psichologinio klausimyno duomenys leido nustatyti depresiją 12 tiriamųjų, o nerimo sutrikimas nustatytas 17 tiriamųjų (kai kurie tiriamieji patenka į kelias iš šių grupių) (3.4 pav.).



3.4 pav. Klinikinių duomenų apibendrinimas ir tiriamųjų pasiskirstymas tarp skirtingų ligų grupių. Stulpelinėje diagramoje vaizduojamas sergančių asmenų skaičius, priklausomai nuo klinikinių duomenų šaltinio – HAD skalės ar ligos istorijos. Daugiausia tiriamųjų, remiantis ligų istorijos duomenimis, priskirti prie sergančiųjų psichikos ligomis grupės. Veno diagrama iliustruoja tiriamųjų pasiskirstymą tarp skirtingų ligų grupių. Persidengiančios sritys vaizduoja tiriamuosius, kurie priklausė dviem ar trimis grupėms – 5 tiriamieji sirgo tiek nerimo sutrikimu, tiek depresija; 2 kaip sergantys identifikuoti abiem klinikinių požymių vertinimo įrankiais, t. y. remiantis klinicine apžiūra ir taikant HAD skalę; 4 identifikuoti kaip sergantys nerimo sutrikimu (HAD skalės duomenimis) ir psichikos ligomis (remiantis ligos istorijomis).

Net 37 iš 93 LČKL (40 %) galimai serga psichikos sutrikimais (3.4 pav., Veno diagrama). Remiantis 2019 m. Lietuvos oficialios statistikos portalo duomenimis [162], bendrojoje lietuvių populiacijoje depresija nustatoma 5 proc. vyrų. Tai ženklus skirtumas tarp tiriamųjų grupių. Kita vertus, 60 proc. tiriamųjų (56 LČKL) šios būklės nebūdingos. Tokį atsaką potencialiai gali lemti unikali likviduotojų genomo variacija.

Atlikus trijų kryptių asociacijų tyrimą, visiems genotipavimo luste esantiems VNP (detalesnė aptarta darbo metodinėje dalyje, 2.4.2.3 poskyryje) nustatyta reikšmingų asociacijų (3.5 pav.). Statistiškai reikšmingi genomo variantai ir jų statistikos pateiktos 3.6 lentelėje.



3.5 pav. LČKL psichikos būklių asociacijų analizės rezultatų schema. 15-oje koduojančių genų nustatyti 43 genomo variantai, pasiekę statistinį reikšmingumą (kai $p < 0,05$). Atlikus Bonferroni korekciją ($p < 7 \times 10^{-8}$), statistiškai reikšmingi buvo 20 VNP (raudona spalva pažymėtuose genuose). Paryškintu šriftu išskirti RNR genai.

3.6 lentelė. LČKL psichikos būklių asociacijų analizės statistiškai reikšmingi poveikio genomo variantai koduojančiose genomo srityse. Statistiškai reikšmingi rezultatai gauti skirtingose tiriamųjų grupėse. LČKL grupės: sergantys 1) psichikos sutrikimais, 2) nerimo sutrikimu, 3) depresija ir 4) šių būklių neturintys, t. y. sveiki tiriamieji. Lyginamoji populiacija – LITGEN (bendroji lietuvių populiacija), vidupopuliacinė – lyginta su LČKL, neturinčiais šių būklių, t. y. analizė kohortos viduje. Kai reikšmingumas nustatytas ne vienoje grupėje, pateikti statistiniai duomenys, kurių p reikšmė mažiausia. Imčių dydžiai nurodyti metodinės dalies 2.4 paveiksle.

Genomo varianto identifikacinis numeris	Genas	χ^2	p	GS	Grupė, kurioje pasiektas statistinis reikšmingumas	Bonferroni p reikšmė
rs399887	<i>PROXI</i>	43,9	$3,50 \times 10^{-11}$	NA	Psichikos LITGEN	$2,10 \times 10^{-05}$
rs680277	<i>RGS8</i>	27,4	$1,65 \times 10^{-07}$	11,9	Psichikos LITGEN	$p > 0,05$
rs16854364	<i>B3GALTI-ASI</i>	32,2	$1,40 \times 10^{-08}$	NA	Nerimas LITGEN	$8,6 \times 10^{-03}$
rs9848313	<i>MECOM</i>	33	$9,30 \times 10^{-09}$	51,9	Depresija LITGEN, nerimas LITGEN	$5,74 \times 10^{-03}$
rs10516933	<i>GRID2</i>	23,3	$1,42 \times 10^{-06}$	6	Psichikos vidupopuliacinė	$p > 0,05$
rs6902050	<i>PHACTR1</i>	12,3	$4,45 \times 10^{-04}$	4,9	Nerimas vidupopuliacinė, nerimas LITGEN	$p > 0,05$
rs4980014	<i>ZMIZI-ASI</i>	27,1	$1,95 \times 10^{-07}$	22,2	Psichikos vidupopuliacinė	$p > 0,05$

Genomo variantų identifikacinis numeris	Genas	χ^2	p	GS	Grupė, kurioje pasiektas statistinis reikšmingumas	Bonferroni p reikšmė
rs16936717	ZMIZ1-AS1	28,4	$9,68 \times 10^{-08}$	5,53	Visose sveikųjų grupėse, vidupopuliacinė	$p > 0,05$
rs3897406	ZMIZ1-AS1	23,5	$1,27 \times 10^{-06}$	19,4	Psichikos vidupopuliacinė	$p > 0,05$
rs1837303	ZMIZ1-AS1	23,5	$1,27 \times 10^{-06}$	19,4	Psichikos vidupopuliacinė	$p > 0,05$
rs1343928	CHST11	32,4	$1,28 \times 10^{-08}$	NA	Nerimas LITGEN	$7,88 \times 10^{-03}$
rs3850004	TMEM132D	33	$9,30 \times 10^{-09}$	51,9	Depresija vidupopuliacinė, depresija LITGEN	$5,74 \times 10^{-03}$
rs2075364	TSPAN11	27,6	$1,50 \times 10^{-07}$	NA	Depresija vidupopuliacinė, depresija LITGEN	$3,29 \times 10^{-06}$
rs7309016	TSPAN11	20,6	$5,71 \times 10^{-06}$	NA	Depresija vidupopuliacinė, depresija LITGEN	$5,74 \times 10^{-03}$
rs1408875	TSC22D1	20,5	$5,83 \times 10^{-06}$	2,5	Nerimas sveiki LITGEN	$p > 0,05$
rs2277533	TMEM87A	26,3	$2,94 \times 10^{-07}$	3,9	Psichikos sveiki, nerimas sveiki, depresija sveiki LITGEN	$p > 0,05$
rs2294604	TELO2	36,8	$1,30 \times 10^{-09}$	51,9	Psichikos LITGEN	$7,78 \times 10^{-04}$
rs239022	MIR99AHG	27,4	$1,68 \times 10^{-07}$	42,4	Depresija, vidupopuliacinė	$p > 0,05$
rs4077512	RTL3	23	$1,60 \times 10^{-06}$	8,6	Psichikos sveiki, nerimas sveiki, depresija sveiki LITGEN	$p > 0,05$

Chr – chromosoma, GS – galimybių santykis, NA – nėra duomenų (vieno iš genotipų skaitinė vertė lygi 0).

Siekiant įvertinti galimą nustatytų variantų ir juos koduojančių genų funkcinę reikšmę, *in silico* įrankiais atlikta šių genų raiškos analizė audiniuose ir sudaryti genų sąveikos tinklai. Tyrimais įrodyta, kad, pavyzdžiui, depresijos atveju, smegenų pokyčiai yra ryškiausi migdoliniame kūne, hipokampe ir gumbure [163]. Visos šios sritys tiesiogiai susijusios su atsaku į stresą, emocijomis ir jausmais. Tyrimo metu, naudojantis *Humanbase in silico* įrankiu, nustatytų genetinių sričių koduojamų baltymų raiškos audiniuose grafikas (angl. *heat-map*, V priedas) padėjo atskleisti, kad didžiausia nustatytų

genų raiška vyksta būtent centrinėje nervų sistemoje. Tai pagrindžia hipotezę, kad šios genetinės sritys gali būti susijusios su psichikos sutrikimų etiopatogeneze. Remiantis šia analize, *RTL3* ir *MECOM* genų raiška audiniuose yra mažiausia centrinės nervų sistemos struktūrose. Dėl šios priežasties minėti genai į tolesnę analizę nebuvo įtraukti.

Siekiant įvertinti nustatytų genų tarpusavio sąveikų potencialius mechanizmus, naudojantis *Humanbase in silico* įrankiu sugeneruoti centrinei nervų sistemai ir depresijai paveikiausioms smegenų sritims būdingi funkciniai genų sąveikos tinklai (VI priedas). Centrinėje nervų sistemoje nustatytas nedidelis genų sąsajos statistinis reikšmingumas (angl. *interaction confidence*) (nustatytos genų sąsajos reikšmingumo reikšmės buvo tarp 0,1 ir 0,2, kai mažiausia reikšmė – 0, o didžiausia – 1), tačiau žinoma, kad psichikos sutrikimai yra daugiaveiksnės etiologijos ligos, o genomo pokyčių suminis poveikis gali būti reikšmingas psichikos sutrikimų patogenezei [164].

Atskirų depresijos metu paveiktų smegenų sričių analizė leido nustatyti potencialius genų sąveikos tinklus (VI priedas). Migdoliniame kūne nustatyti statistiškai reikšmingų genų *PROX1*, *TMEM87A*, *TSC22D1*, *ZMIZ1*, kurie sudaro netolimus funkcinis ryšius, variantai. Sugeneravus gumburui specifinį genų tinklą, iš tyrimo metu gauto statistiškai reikšmingų genų sąrašo, lyginant jį su migdolinio kūno genų sąveikos tinklu, papildomai įtrauktas *TELO2* genas. Iš nustatytų statistiškai reikšmingų genų sąrašo hipokampe tam tikrą funkciją atlieka *ZMIZ1* ir *TSC22D1*. Šie du genai, nepriklausomai nuo smegenų srities, tendencingai kartojasi kiekviename sugeneruotame tinkle, todėl jie gali būti reikšmingiausi iš nustatytų genų ir svarbūs psichikos sutrikimų etiopatogenezei.

ZMIZ1 priklauso ilgų nekoduojančių RNR genų klasei. Šio tipo genų didžiausia raiška vyksta nerviniame audinyje ir sėklidėse [165]. *ZMIZ1* siejamas su išsėtine skleroze per vitamino D receptorių, o pastarasis faktas patvirtina *ZMIZ1* vaidmenį reguliuojant CNS [166]. Be to, nustatytas *ZMIZ1* geno variantas rs16936717 yra statistiškai reikšmingas (χ^2 kriterijus) nerimo ($p = 9,68 \times 10^{-8}$, GS = 5,53), depresijos ($p = 8,45 \times 10^{-07}$, GS = 4,85) ir psichikos ligų neturinčių ($p = 8,35 \times 10^{-6}$, GS = 4,61) LČKL grupėse. Atlikus Bonferroni korekciją, statistinis reikšmingumas nebuvo nustatytas, tačiau nerimą jaučiančių tiriamųjų grupėje jis buvo artimas reikšmingumo slenkstinei reikšmei ($p = 0,058$). Tai, kad tas pats variantas vyrauja visose nagrinėtose sveikų LČKL grupėse, liudija apie galimą protekcinę šio varianto funkciją.

TSC22D1 koduoja transkripcijos veiksnį ir priklauso didelei ankstyvos raiškos genų klasei. Šis genas potencialiai veikia kaip transkripcijos veiksnys *CNP* genui, kuris koduoja natriuretinį peptidą, dalyvaujantį osifikacijos, reprodukcijos ir centrinės bei periferinės nervų sistemos funkcionavimo ir

vystymosi procesuose [167]. Genas labiausiai išnagrinėtas vėžio atvejais, nes koduojamas baltymas gali atlikti svarbų tumorosupresinį vaidmenį, sukeldamas vėžinių ląstelių apoptozę. Naujuose tyrimuose taip pat pabrėžiama šio geno įtaka ir su nervų sistema siejamų ligų, tokių, kaip Alzheimerio liga, etiopatogenezei, o tai patvirtina, kad šis genas gali turėti reikšmės LČKL grupės rezultatams [168]. Minėtina ir tai, kad asociacijų analizės metu šio geno varianto rs1408875 statistinis reikšmingumas (χ^2 kriterijus) apskaičiuotas neturinčių nerimo sutrikimo LČKL grupėje ($p = 5,8 \times 10^{-6}$, $GS = 2,492$). Tai rodo galimą protekcinę šio varianto funkciją, nes variantas yra dažnesnis nesergančių nerimo sutrikimu grupėje (retojo alelio dažnis – 0,55), palyginti su bendrąja lietuvių populiacija (retojo alelio dažnis – 0,33).

Remiantis *STRING in silico* analizės įrankiu, dėmesys atkreiptinas ir į *GRID2* ir *RGS8* genus, kurie susiję molekulinio lygmeniu. *GRID2* koduoja jonotropinius glutamato receptorius, esančius neuronų ir glijos ląstelių paviršiuje, smegenyse jis veikia kaip neurosiuntiklis ir yra susijęs su neurologiniais sutrikimais [169]. *RGS8* atlieka reguliacinę funkciją G-baltymo signalo perdavimo kelyje (veikia kaip slopiklis). Raiška vyksta daugelyje audinių, o jos sutrikimas siejamas su daugelio ligų patogeneze, taip pat ir su depresija [170]. Taigi, tikėtina, kad šie genai taip pat turi tiesioginę reikšmę depresijos etiopatogenezei.

3.4.2. Burnos ertmės ligų asociacijų analizė

Remiantis surinktais anketinės apklausos ir ligų istorijų duomenimis, galima konstatuoti, kad net 70 proc. LČKL (65 iš 93 tiriamųjų) būdingi virškinimo sistemos sutrikimai (3.1 pav.). Didžioji dalis šios grupės sveikatos sutrikimų – burnos ertmės ligos, ypač periodontitas ir dantų netekimas. Burnos ertmės ligų turi beveik pusė LČKL (48 %). Remiantis Lietuvos higienos instituto 2020 m. duomenimis, matyti, kad bendrojoje lietuvių populiacijoje, 45 m. amžiaus ir vyresnių vyrų grupėje, dantų ligomis sirgo apie 12 proc. tiriamųjų [127]. Taigi burnos ertmės ligos yra keturis kartus dažnesnės LČKL grupėje, palyginti su bendrąja lietuvių populiacija. Didelis sergančiųjų skaičius gali būti tiesiogiai susijęs su tiriamųjų Černobylio katastrofos likvidavimo metu patirta didesne negu įprasta JS doze. Vaikų, gimusių po Černobylio katastrofos, ir suaugusiųjų, patyrusių dideles JS dozes, mokslinių ir klinikinių tyrimų duomenimis, JS padidina dantų karieso riziką ir sumažina dantų pulpos vaskuliarizaciją. Prastesnė vaskuliarizacija yra susijusi su audinių fibroze ir atrofija, periodontine disfunkcija, o tai smarkiai padidina dantų netekimo ir osteoradionekrozės riziką [171–173]. Žinoma, kad radioaktyvūs nuklidai ^{90}Sr

ir ^{137}Cs veikia kietuosius dantų audinius. Radioaktyvus stroncio izotopas (^{90}Sr) yra kalcio analogas kaulų metabolizmo procese, todėl ^{90}Sr pakeičia kalcį dantų emalyje ir taip paveikia emalio struktūrą [174]. Atliktos apklausos duomenimis, dideliame radioaktyviajam užterštumui (po Černobylio katastrofos) būdingų teritorijų gyventojai nurodo akivaizdų dantų ir burnos ertmės sveikatos būklės pablogėjimą [171].

Šiuo tyrimo analizės etapu pasirinkta atlikti kryptingą su burnos ertmės ligomis siejamų VNP asociacijų analizę. Remiantis moksline literatūra ir duomenų bazėmis, sudarytas 909-ių VNP sąrašas 59-iuose genuose (IV priedas). Sąrašas panaudotas tolesnei asociacijų analizei: analizuoti duomenys LČKL, kuriems diagnozuota burnos ertmės ligų, ir LČKL, kuriems burnos ertmės ligų nediragnozuota (analizė atlikta kohortos viduje). Analizės metu nustatytos su burnos ertmės ligomis mažai sietos *FGF1*, *FGF2*, *FGF7* ir *BMP2* genų asociacijos su padidėjusia LČKL burnos ertmės ligų rizika (3.7 lentelė).

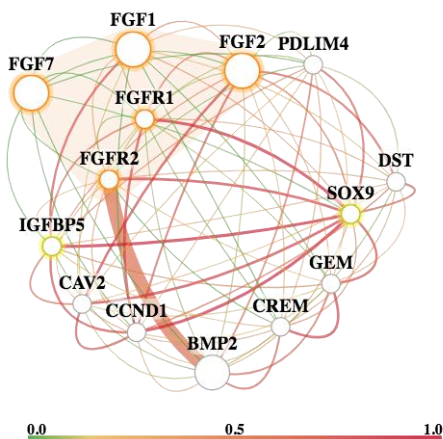
3.7 lentelė. Genų, siejamų su padidėjusia burnos ertmės ligų rizika, apibendrinta statistika

Genomo varianto identifikacinis numeris	Genas	Retojo alelio dažnis sergantųjų LČKL grupėje	Retojo alelio dažnis sveikų LČKL grupėje	χ^2 kriterijus	p reikšmė	Galimybių santykis	95 % PI
rs34022	<i>FGF1</i>	0,457	0,261	7,796	0,005	2,389	1,288–4,429
rs308388	<i>FGF2</i>	0,415	0,217	8,374	0,004	2,553	1,342–4,857
rs16962440	<i>FGF7</i>	0,553	0,337	8,798	0,003	2,436	1,346–4,41
rs3178250	<i>BMP2</i>	0,223	0,076	7,891	0,005	3,493	1,405–8,685

LČKL – Lietuvos Černobylio katastrofos likviduotojai, PI – pasikliautinis intervalas. Imčių dydžiai: 45 LČKL, sergantys burnos ertmės ligomis, ir 48 LČKL, burnos ertmės ligomis nesergantys.

Visi nustatyti genomo variantai yra nekoduojančiose geno dalyse, tačiau jie gali būti ne mažiau svarbūs, nes nurodo genomo sritį, kuri gali būti potencialiai reikšminga požymio ar būklės fenotipinei išraiškai. Tuo remiantis, tolesniame etape, naudojantis *Humanbase in silico* analizės įrankiu, atlikta nustatytų variantų genų sąveikos analizė kauliniame audinyje.

Nustatytas galimas šių genų funkcinių sąveikų kauliniame audinyje tinklas (3.6 pav.).



3.6 pav. *FGF1*, *FGF2*, *FGF7*, *BMP2* genų galimų funkcinių sąveikų kauliniame audinyje tinklas. Sąveikų stiprumą rodo spalvinis sąsajų žymėjimas – žalia spalva pažymėtos mažiausiai reikšmingos, raudona – reikšmingiausios sąsajos.

Remiantis sudarytu tinklu, nustatytos reikšmingos genų funkcinės sąveikos, patvirtinančios hipotezę, kad šie genai gali turėti įtakos LČKL burnos sveikatai. Mokslinėje literatūroje aprašytas reikšmingas fibroblastų augimo veiksnių ir kaulų morfogeninio baltymo (atitinkamai – FGF ir BMP) genų šeimų vaidmuo dantų audinių vystymuisi [175]. Mažiausiai nustatyta *FGF7* geno reikšmingų sąsajų, tačiau mokslinės literatūros analizė parodė, kad šis genas, kartu su kitais fibroblastų augimo veiksniais (tarp jų ir *FGF1* bei *FGF2*), taip pat svarbus dantų audinių vystymuisi [176].

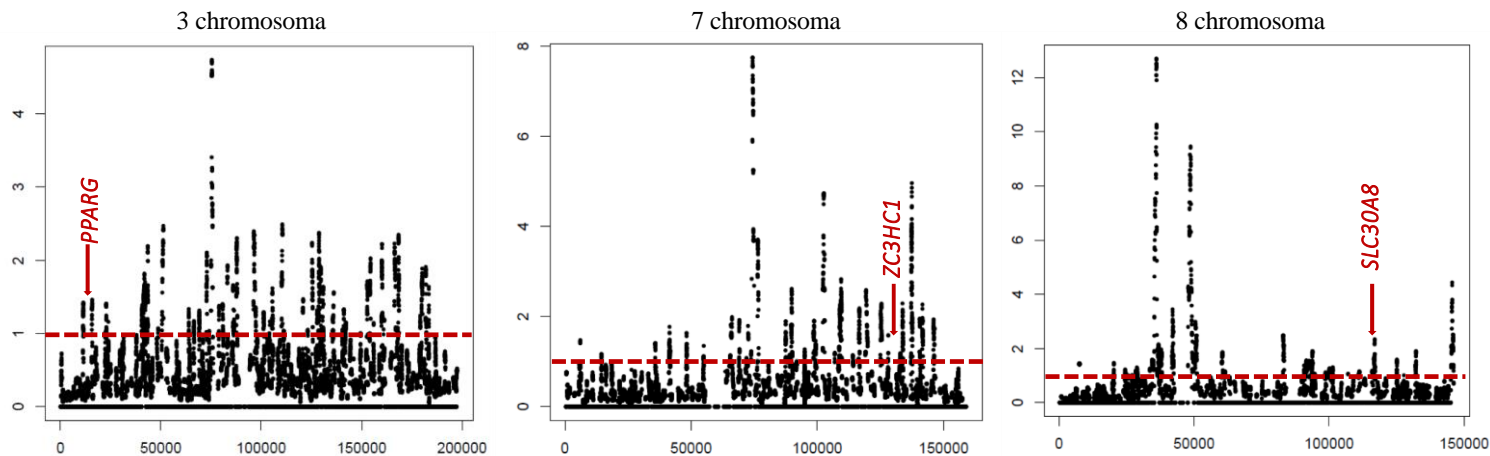
Tirtoje LČKL grupėje taip pat statistiškai reikšmingi trys variantai (rs2004330, rs1180341, rs1180343) *BMP8A* gene. *BMP8A* genas reikšmingas kremzlių ir kaulų formavimosi procesuose, kalcio reguliacijoje, palaikant kaulų homeostazę, todėl tiesiogiai susijęs ir su dantų formavimusi [115, 177]. Įdomu tai, kad, skirtingai negu kiti šiame skyriuje aptarti genomo variantai, šių variantų galimybių santykio reikšmės varijavo tarp 0,395 ir 0,43, t. y. buvo žemesnės negu 1. Tai reiškia, kad minėti genomo variantai yra dažnesni LČKL grupėje be analizuojamo požymio ir galimai mažina galimybę sirgti burnos ertmės ligomis bei turi potencialių protekcinį savybių. Šią hipotezę patvirtina ir mokslinėje literatūroje aprašomas *BMP8A* protekcinis vaidmuo gliukokortikoidų sukeltos kaulų apoptozės procese. Gliukokortikoidai, veikdami apoptoziškai, turi įtakos osteoporozei, o padidėjusi *BMP8A* raiška lemia lėtesnius apoptozės procesus [178], todėl tiesiogiai turi reikšmę kaulų homeostazei.

Apibendrinant galima teigti, kad aptartuoju tyrimo etapu nustatyti geno variantai ir sritys, potencialiai galintys turėti įtakos LČKL psichikos būklių ir burnos ertmės ligų molekuliniais mechanizmais. Nustatyti geno variantai ir sritys gali būti naudingi analizuojant minėtų daugiaveiksnių būklių genetinę architektūrą, siekiant pagilinti žinias apie etiopatogenezės mechanizmus.

3.5. Bendrosios lietuvių populiacijos ir Lietuvos Černobylio katastrofos likviduotojų teigiamos gamtinės atrankos veikiamų geno sričių analizė

3.5.1. Teigiamos gamtinės atrankos įtakos įvertinimas nustatytiems poveikio geno variantams bendrojoje lietuvių populiacijoje

Siekiant patvirtinti ar paneigti trijų aptartų (3.2 skyriuje) poveikio geno variantų *PPARG*, *SLC30A8* ir *ZC3HCl* genuose galimai protekcinį poveikį bendrosios lietuvių populiacijos (LITGEN) tiriamųjų grupėje, atlikta teigiamos atrankos poveikio analizė, naudojant genotipavimo duomenis. Pagrindinė šio tyrimo etapo idėja tokia: jei nustatyti statistiškai reikšmingi geno variantai yra naudingi tiriamosios grupės adaptacijai ir atlieka protekcinę funkciją, jie turėtų būti veikiami teigiamos gamtinės atrankos, o tiriamieji turėtų įgyti pranašumą kitų populiacijos asmenų atžvilgiu ir taip variantai, laikui bėgant, turėtų įsitvirtinti populiacijoje. Atliekant analizę, skaičiuota μ statistika, kuri tiesiogiai koreliuoja su teigiamos gamtinės atrankos poveikiu geno sričiai – μ statistikos taškų išsibarstymo grafikai chromosomoms, kuriose yra *PPARG*, *SLC30A8* ir *ZC3HCl* genai, pateikiami 3.7 paveiksle. Nustatytas statistiškai reikšmingas teigiamos gamtinės atrankos signalas *SLC30A8* gene ($\mu = 2,34$, $p < 0,05$). Nors nustatytas signalas, palyginti su signalais kitose srityse, 8 chromosomoje yra nedidelis, tai vis tiek galimai rodo šio geno ir nustatyto varianto protekcinę funkciją. Remiantis atlikta analize, *PPARG* ir *ZC3HCl* genai nėra veikiami teigiamos gamtinės atrankos (nustatytos μ reikšmės nepasiekė didžiausių 5% μ reikšmių slenkstinės vertės, $p \geq 0,05$). tačiau abu genai yra santykinai netolimu atstumu nuo gamtinės atrankos veikiamų geno sričių (*PPARG* – 800 Kbp atstumu nuo chr3:11240480-11461626, $\mu = 1,42$, $p < 0,05$; *ZC3HCl* – 2 Mbp atstumu nuo chr7:128015754-128104838, $\mu = 1,59$, $p < 0,05$). Tai reiškia, kad šios geno sritys gali būti potencialūs tolesnių teigiamos gamtinės atrankos tyrimų objektai.

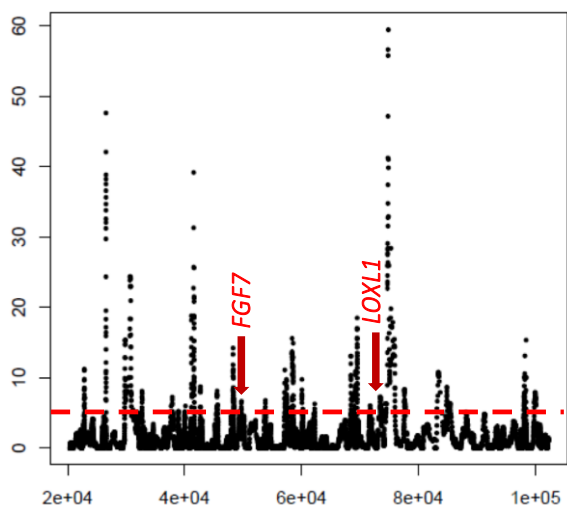


3.7 pav. μ statistikos taškų išsibarstymo grafikai 3, 7 ir 8 chromosomoms. X ašyje vaizduojama genomine pozicija (Kb), Y ašyje – μ reikšmės įvertis. Punktyrinė linija vaizduoja slenkstinę μ reikšmę, virš jos – μ reikšmės, didesnės negu 5 % visų μ reikšmių toje chromosomoje ($p < 0,05$). Aukščiausios μ statistinės reikšmės nurodo stipriausias teigiamos gamtinės atrankos veikiamas genomo sritis. *PPARG* (3 chromosoma), *ZC3HC1* (7 chromosoma) ir *SLC30A8* (8 chromosoma) kandidatinių protekcinųjų genų pozicija ir teigiamos gamtinės atrankos įtaka pažymėta rodyklėmis.

3.5.2. Teigiamos gamtinės atrankos įtakos įvertinimas nustatytiems poveikio genomo variantams Lietuvos Černobylio katastrofos likviduotojų grupėje

Ankstesnių tyrimo analizės etapų rezultatai parodė, kad LČKL grupės genuose tiriamųjų sveikatai ir adaptacijai įtakos turi genai, susiję su eksfoliacijos sindromu (*LOXLI*), koronarine širdies liga (*CX3CRI*), psichikos sutrikimais (*ZMIZ1*, *TSC22D1*), burnos ertmės ligomis (*FGF1*, *FGF2*, *FGF7*, *BMP2*, *BMP8A*). Tyrimo metu nustatyta, kad kai kurie variantai išvardytuose genuose galimai lemia apsaugą nuo minėtų ligų ir geresnę LČKL išgyvenamumą, taip pat adaptacines savybes (variantai *LOXLI*, *CX3CRI*, *ZMIZ1*, *TSC22D1*, *BMP8A* genuose). Nustatyta, kad kiti variantai galimai susiję su blogesne LČKL grupės burnos sveikata (variantai *FGF1*, *FGF2*, *FGF7*, *BMP2* genuose). Siekiant pagrįsti nustatytų genų reikšmę LČKL fenotipui, atlikta teigiamos gamtinės atrankos analizė. Ši analizė turi tiesioginį ryšį su genomines srities naudingumu tiriamojo sveikatai ir adaptacijai: kuo naudingesnė genomo sritis, tuo ji daugiau veikiama teigiamos gamtinės atrankos. Šiame tyrimo etape detaliau nagrinėtas minėtų genų teigiamos gamtinės atrankos poveikis.

LOXLI geno sričiai (chr15:73926462-73952136) nustatytas nedidelis teigiamos gamtinės atrankos poveikis ($\mu = 1,14$) (3.8 pav.), tačiau nereikšmingas ($p \geq 0,05$). Kita vertus, šis genas patenka į regioną (chr15:73049046-73980207), kurio teigiamos gamtinės atrankos įtaka yra statistiškai reikšminga (aukščiausia reikšmė – $\mu = 7,30$, $p < 0,05$). Be to, visai greta šio regiono yra kitas regionas, turintis didžiausią μ reikšmę ir reikšmingiausias chromosomoje (chr15:74491439-75982958, $\mu = 59,5$, $p < 0,05$). Minėtos genomines sritys, veikiamos teigiamos gamtinės atrankos, patvirtina šiose genomo srityse esančio *LOXLI* geno ir tirta rs3825942 varianto protekcinio vaidmens eksfoliacijos sindromo patogenezėje hipotezę.



3.8 pav. μ statistikos taškų išsibarstymo grafikas 15 chromosomai. X ašyje vaizduojama genominei pozicija (Kb), Y ašyje – μ reikšmė. Punktyrinė linija vaizduoja slenkstinę μ vertę ($\mu = 5,53$), virš jos – μ reikšmės, aukštesnės negu 5 % visų μ reikšmių toje chromosomoje ($p < 0,05$). Didžiausios μ statistinės reikšmės nurodo stipriausias teigiamos gamtinės atrankos veikiamas genomo sritis. Rodyklėmis pažymėtos analizuotos genomo sritys.

FGF7 geno sričiai (chr15: 49423237-49488775) nustatytas nedidelis (palyginti su kitais teigiamos atrankos veikiamais genomo regionais 15-oje chromosomoje) teigiamos atrankos poveikis ($\mu = 4,99$) (3.8 pav.). Ši reikšmė geno srityje nepasiekė statistinio reikšmingumo ($p \geq 0,05$), bet yra tarp dviejų reikšmingų gamtinės atrankos veikiamų genomo sričių (chr15: 48232971-48435566, $\mu = 11,17$; chr15: 49692283-49719324, $\mu = 6,64$; $p < 0,05$). Kaip buvo aptarta ankstesniame skyriuje, asociacijų analizės metu nustatyta, kad *FGF7* geno poveikio genomo variantas rs16962440 yra dažnesnis burnos sveikatos problemų turinčių LČKL grupėje, o tai reiškia, kad galimai lemia padidėjusią šių sveikatos problemų riziką. Kita vertus, šio tyrimo etapo rezultatai nurodo galimai protekcinę tirtą *FGF7* geno poveikio varianto funkciją, nes *FGF7* geno regionas yra galimas teigiamos gamtinės atrankos taikynys. Nevienareikšmiški rezultatai rodo, kad dažnai išvados apie poveikio genomo variantų reikšmę ir poveikį fenotipui negali būti daromos vien tik atliekant asociacijų ir *in silico* analizes. Reikšmingoms išvadoms formuluoti būtina atlikti keletą tyrimų, sujungiant skirtingus analizės metodus, o to ir siekta šiuo tyrimu. Minėtina, kad, siekiant patvirtinti poveikio genomo variantų reikšmę, būtų reikšmingi funkciniai genomo tyrimai, naudojant modelinius organizmus ar tiriant genų raišką audinių ar ląstelių kultūrose.

FGF7 atveju neatmestinos ir geno sąsajos ne tik su burnos ertmės ligomis, bet ir su kitomis būklėmis. Daugiaveiksnių būklių molekulinuose mechanizmuose veiklūs genai dažnai atlieka tam tikrą funkciją ne vienos būklės etiopatogenezeje. Atlikta genų sąsajų tinklų analizė atskleidė, kad *FGF7* genas turi mažiausiai sąsajų su burnos ertmės ligomis, palyginti su kitais tirtais genais, todėl neatmestina, kad šis genas siejamas su kitu požymiu ir galimai protekcinio poveikiu tam požymiui LČKL grupėje. Pavyzdžiui, nustatyta *FGF7* sąsaja su geresne kepenų regeneracija per kepenų pirmtakių ląstelių reguliavimo mechanizmą [179]. Vis dėlto, atlikus sergančiųjų kepenų ligomis LČKL asociacijų analizę rs16962440 poveikio genomo variantui LČKL kohortos viduje, lyginant sergančiuosius kepenų ligomis (n = 5) su LČKL, šiomis ligomis nesergančiais (n = 88), nustatyta, kad retojo alelio dažnis statistiškai reikšmingai nesiskiria ($\chi^2 = 1,011$, $p = 0,315$). Išlieka tikimybė, kad šis genas reikšmingas kitoms LČKL adaptacinėms savybėms.

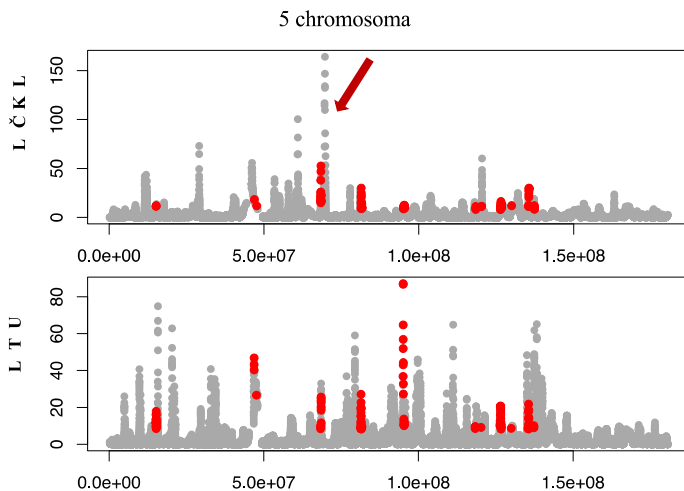
Remiantis aptartąja tyrimo analize, galima teigti, kad *CX3CRI* (chr3:39263495-39281735), *ZMIZ1* (chr10:78943328-79067895), *TSC22D1* (chr13:44432143-44577147), *FGF1* (chr5:142592178-142698070), *FGF2* (chr4:122826708-122898236) *BMP2* (chr20:6767686-6780246), *BMP8A* (chr1: 39491636-39529869) genai nepatenka į reikšmingas teigiamos gamtinės atrankos veikiamas sritis. Atsižvelgiant į teigiamos gamtinės atrankos vertinimą, matyti, kad šios genomo sritys neapsaugo nuo koronarinės širdies, psichikos ir burnos ertmės ligų, tačiau tikėtina, kad atlieka tam tikrą vaidmenį šių ligų etiopatogenezeje, todėl reikalingi tolesni funkciniai šių genų tyrimai. Tyrimo metu nustatyta, kad minėtos sritys nėra veikiamos teigiamos gamtinės atrankos, tačiau, remiantis UCSC genomo naršykle ir Varsome analizės įrankiu, matyti, kad kai kurių šių sričių tirti poveikio genomo variantai yra konservatyviuose genomo regionuose (*CX3CRI* geno *GERP++* įvertis 2,49, *PhyloP100way* įvertis 1,89; *TSC22D1* *GERP++* įvertis 2,37, *PhyloP100way* įvertis 1,29; *BMP2* *GERP++* įvertis 1,82, *PhyloP100way* įvertis 0,865). Teigiamos šių įverčių reikšmės rodo regionų konservatyvumą. Nustatytas sričių konservatyvumas iš dalies pagrindžia ir minėtų genomo sričių reikšmę etiopatogenezei.

Šiame tyrimo etape ne tik nustatytos potencialiai protekcinės genomo sritys, lemiančios geresnę LČKL adaptaciją. Tyrimui panaudota inovatyvi tyrimo strategija, variantų ir jų genų funkcinės reikšmės įvertinimui panaudojant teigiamos gamtinės atrankos rodiklius.

3.5.3. Teigiamos gamtinės atrankos veikiamų genomo sričių nustatymas ir reikšmės adaptacijai įvertinimas Lietuvos Černobylio katastrofos likviduotojų grupėje

Šiuo tyrimo etapu pirmą kartą atliktas LČKL grupės teigiamos gamtinės atrankos poveikio vertinimas, panaudojant viso genomo sekoskaitos duomenis. Atliekant tyrimą, siekta identifikuoti teigiamos gamtinės atrankos labiausiai veikiamas autosomų genomo sritis, išskirtinai būdingas LČKL grupei, palyginti su bendrąja lietuvių populiacija (ANELGEMIA). Gautais rezultatais siekta įrodyti šių genomo sričių reikšmę LČKL grupės adaptacinėms savybėms.

Lietuvos Černobylio katastrofos likviduotojų genuose didžiausią teigiamos atrankos poveikį chromosomos atžvilgiu turėjo regionai 5 ($\mu = 164,2$, 3.9 pav.), 11 ($\mu = 114$) ir 12 ($\mu = 161,1$) chromosomose (3.8 lentelė, VII priedas). Nustatytuose teigiamos gamtinės atrankos veikiamuose regionuose 9, 13, 18 ir 21 chromosomose nėra genų, todėl šie regionai detalčiau nebuvo analizuojami.



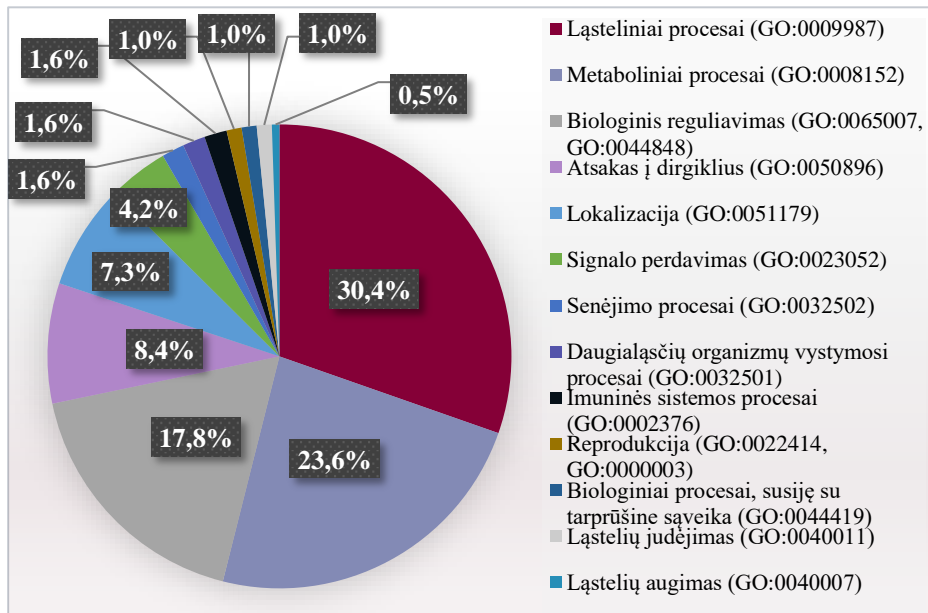
3.9 pav. Lietuvos Černobylio katastrofos likviduotojų genomų teigiamos atrankos poveikio analizė 5 chromosomoje. Lygintos Lietuvos Černobylio katastrofos (LČKL) ir bendrosios lietuvių populiacijos (LTU; ANELGEMIA) tiriamųjų grupės. X ašyje – genomine pozicija (Kb), Y ašyje – μ įverčio reikšmė. Rodykle pažymėta teigiamos gamtinės atrankos veikiama, unikali LČKL tiriamųjų grupei ir didžiausią μ reikšmę turinti reikšmė. Raudonai pažymėtos genomo sritys – *tapačios* tarp LČKL ir LTU tiriamųjų grupių, pilkos – *netapačios*, tiriamųjų grupių *unikalios* teigiamos gamtinės atrankos veikiamos genomo sritys. Kitų chromosomų teigiamos gamtinės atrankos poveikio analizės rezultatai pateikiami VII priede.

**3.8 lentelė. Nustatytos teigiamos gamtinės atrankos veikiamos autosomų geno-
sritys ir genai šiose srityse**

Chromo- soma	Genominės srities pradžia	Genominės srities pabaiga	Aukščiausia μ reikšmė srityje	Genai nustatytoje srityje
1	50024718	51586980	49,96	<i>AGBL4, ELAVL4, DMRTA2, FAF1, CDKN2C</i>
2	189198631	189731344	86,85	<i>GULP1, DIRC1</i>
3	68678446	68990461	77,44	<i>TAF4</i>
4	146494457	147354697	72,27	<i>MMAA, C4orf51, ZNF827, LSM6, SLC10A7</i>
5	68316218	69932025	164,20	<i>SLC30A5, CCNBI, CENPH, MRPS36, CDK7, CCDC125, AK6, TAF9, MARVELD2, OCLN, GTF2H2C, SERF1B, SMN2, NAIP</i>
6	64047472	64615929	56,06	<i>PTP4A1, PHF3, EYS</i>
7	151965978	152132980	28,44	<i>KMT2C</i>
8	104888906	107027425	36,35	<i>RIMS2, DCSTAMP, DPYS, LRP12, ZFPM2</i>
9	17876558	18329195	54,64	–
10	122646955	123081831	41,73	<i>WDR11</i>
11	101273279	101740156	114,00	<i>TRPC6</i>
12	109191005	111322770	161,10	<i>SSH1, DAO, USP30, ALKBH2, UNG, ACACB, MYO1H, KCTD10, MMAB, UBE3B, MVK, TRPV4, GIT2, GLTP, ANKRD13A, C12orf76, IFT81, ATP2A2, ANAPC7, GPN3, ARPC3, FAM216A, VPS29, PPTC7, RAD9B, TCTN1, HVCN1, PPP1CC, CCDC63</i>
13	82488223	83468360	30,54	–
14	73733462	73867298	61,43	<i>PAPLN, NUMB, AX747833</i>
15	74491439	75982958	59,46	<i>CCDC33, CYP11A1, SEMA7A, UBL7, ARID3B, CLK3, EDC3, CYP1A1, CSK, ULK3, SCAMP2, FAM219B, MPI, COX5A, RPP25, SCAMP5, PPCDC, C15orf39, GOLGA6C, MAN2C1, COMMD4, NEIL1, SIN3A, PTPN9, SNUPN</i>
16	74347008	74795167	29,69	<i>NPIPBI5,</i>

Chromosoma	Genominės srities pradžia	Genominės srities pabaiga	Aukščiausia μ reikšmė srityje	Genai nustatytoje srityje
				<i>CLEC18B, GLG1, RFWD3, MLKL, FA2H</i>
17	49795141	50541571	27,98	<i>CA10</i>
18	68939522	68991817	29,70	–
19	22058047	23006378	72,65	<i>ZNF208, ZNF257, ZNF676, ZNF729, ZNF98, ZNF492, ZNF99</i>
20	14021037	14066814	22,27	<i>MACROD2</i>
21	23193282	23598744	41,25	–
22	21806118	22653232	45,96	<i>TMEM191C, PI4KAP2, UBE2L3, CCDCI16, SDF2L1, PPIL2, YPEL1, MAPK1, PPM1F, TOP3B</i>

Iš viso LČKL grupėje nustatyti 123 genai, veikiami teigiamos gamtinės atrankos. Šių genų biologinės funkcijos analizė, remiantis *PANTHER* klasifikacijos sistema [123] ir taikant genų ontologijos naršyklę (angl. *Gene Ontology Browser*) [124], atskleidė, kad daugiausia genų dalyvauja ląsteliniuose, metaboliniuose ir biologinio reguliavimo procesuose (3.10 pav.).



3.10 pav. Teigiamos gamtinės atrankos veikiamų genomo sričių genų vaidmuo biologiniuose procesuose. Procesai pateikti mažėjimo tvarka pagal genų skaičių.

Į ląstelinių procesų kategoriją įtraukti ląstelės gyvybingumui svarbūs procesai, tokie kaip ląstelės dalijimasis, ciklo reguliavimas, žūtis ir kt. Metaboliniai procesai apima ląstelėse vykstančias chemines reakcijas ir jose dalyvaujančius genus, tačiau šiai kategorijai taip pat priklauso ir makromolekuliniai procesai, tokie kaip DNR pažeidimų reparacija ar replikacija. Biologinio reguliavimo kategorija labai plati, apimanti ne tik tam tikrų audinių ir organų biologinių procesų reguliavimą, bet ir reguliaciją molekulinio lygmeniu, pavyzdžiui, DNR stabilumo palaikymą ar transkripciją. Kai kurie nustatyti genai atlieka kelias biologines funkcijas, todėl priskiriami kelioms kategorijoms. Atlikta analizė įrodė, kad LČKL genomuose teigiama gamtinė atranka veikia tas genomo sritis, kurios vykdo ląstelių išgyvenamumui (ląstelių senėjimo, ląstelės ciklo reguliavimo, apoptozės) ir adaptacijai (DNR pažeidimų reparacijos) svarbius procesus.

Lietuvos Černobylio katastrofos likviduotojai patyrė didesnes negu įprasta JS dozes ir susidūrė su ekstremaliomis situacijomis. Tai, kad LČKL išgyveno ir prisitaikė prie Černobylio katastrofos likvidavimo darbų sukeltų padarinių, didžiausią reikšmę galimai turi genai, dalyvaujantys DNR pažeidimų reparacijos mechanizmuose ir reguliuojantys ląstelės ciklą. Tyrimo metu LČKL grupėje nustatyti genai, veikiami teigiamos gamtinės atrankos ir dalyvaujantys minėtuose procesuose. Tai: *CDKN2C*, *ZNF827*, *CENPH*, *CCNB1*, *CDK7*, *TAF9*, *RAD17*, *GTF2H2C*, *PTP4A1*, *WDR11*, *ALKBH2*, *UNG*, *ARID3B*, *COMMD4*, *NEIL1*, *RFWD3*, *MACROD2*, *MAPK1*, *PPM1F*, *TOP3B* (3.9 lentelė). Kaip jau minėta, šie genai svarbūs DNR pažeidimų reparacijos ir ląstelės ciklo reguliavimo mechanizmams. Jie, veikiami teigiamos gamtinės atrankos, formuoja unikalias LČKL adaptacines savybes.

3.9 lentelė. Lietuvos Černobylio katastrofos likviduotojų adaptacijai reikšmingi genai ir jų funkcija ir (ar) reikšmė fenotipui. Funkcija ar reikšmė fenotipui nustatyta remiantis UCSC genomo naršyklėje, OMIM ir *Human DNA Repair Genes* duomenų bazėse pateikiamais duomenimis ir mokslinių publikacijų analize.

Chromosoma	Genas	Funkcija ir (ar) reikšmė fenotipui
1	<i>CDKN2C</i>	Ląstelės ciklo G1 fazės kontrolė, tumorosupresinis poveikis, siejama su vėžinėmis ligomis, endokrinine displazija.
4	<i>ZNF827</i>	Telomerų reparacija, siejama su kepenų liga, susijusia su alkoholio vartojimu.
5	<i>CENPH</i>	Formuoja centromeros-kinetochoro kompleksą, siejama su vėžinėmis ligomis.
5	<i>CCNB1</i>	Būtinai ląstelės ciklui pereinant iš G2 į M fazę, siejama su vėžinėmis ligomis.

Chromo- soma	Genas	Funkcija ir (ar) reikšmė fenotipui
5	<i>CDK7</i>	Ląstelių ciklo ir transkripcijos reguliatorius, dalyvauja nukleotidų ekscizinės reparacijos procese.
5	<i>TAF9</i>	Dalyvauja ląstelės cikle, apoptozėje, DNR transkripcijos procesuose.
5	<i>RAD17</i>	Dalyvauja nuo DNR pažeidimų priklausomos rekombinacijos procese, taisant dvigrandžius DNR trūkius. DNR pažeidimų jutiklis (angl. <i>DNA damage sensor</i>).
5	<i>GTF2H2C</i>	DNR reparacija.
6	<i>PTP4A1</i>	Ląstelės ciklo reguliacija (mitozės G1-S fazijų stimuliacija), siejama su vėžiu, sisteminė skleroze, priklausomybe nuo alkoholio.
10	<i>WDR11</i>	Ląstelės ciklo progresija, signalų perdavimas, apoptozė, genų raiškos reguliavimas. Kandidatinis genas tumoro supresorijos mechanizmuose, siejamas su nutukimu, intelektine negalia, širdies anomalijomis, vėžiu.
12	<i>ALKBH2</i>	Vaidmuo tiesioginėje DNR reparacijoje (angl. <i>direct reversal of damage</i>).
12	<i>UNG</i>	Telomerų pažeidimų reparacija, bazių ekscizinė reparacija.
15	<i>ARID3B</i>	Ląstelių linijų determinaciją lemiančių genų reguliacija, ląstelės ciklo kontrolė, transkripcijos reguliavimas, galimas vaidmuo modifikuojant chromatiną.
15	<i>COMMD4</i>	Dvigrandžių DNR trūkių reparacija.
15	<i>NEIL1</i>	Bazių ekscizinė reparacija (pirimidinų).
16	<i>RFWD3</i>	DNR pažeidimų reparacija ląstelės ciklo G1 fazijoje.
20	<i>MACROD2</i>	Reparacijos funkcija nėra aiški, tačiau žinoma, kad DNR pažeidimų metu koduojamas baltymas persikelia iš branduolio į citoplazmą, siejama su vėžinėmis ligomis.
22	<i>MAPK1</i>	Ląstelių proliferacija, diferenciacija, vystymasis, transkripcijos reguliacija.
22	<i>PPM1F</i>	Neigiamas reguliatorius, esant ląstelės atsakui į stresą.
22	<i>TOP3B</i>	Genomo stabilumo palaikymas, DNR rekombinacija, ląstelių senėjimas.

Siekiant išsamiau išnagrinėti nustatytų teigiamos gamtinės atrankos veikiamų sričių poveikį, naudotasi *PopHumanScan* duomenų baze [122], kurioje pateikiama susisteminta informacija apie jau žinomas, skirtingose populiacijose nustatytas teigiamos gamtinės atrankos veikiamas genomo sritis. Penkiems genams (*WDR11*, *NEIL1*, *RFWD3*, *MACROD2*, *MAPK1*),

dalyvaujantiems pažaidų reparacijoje ir ląstelės cikle, teigiamos gamtinės atrankos poveikis nustatytas ir kitose populiacijose (3.10 lentelė).

3.10 lentelė. Reikšmingas teigiamos gamtinės atrankos poveikis nustatytiems pažaidų reparacijos ir ląstelės ciklo genams kitose populiacijose

Genas	Populiacija, kurioje nustatytas teigiamos gamtinės atrankos poveikis	Naudotas metodas, nustatyta reikšmė
<i>WDR11</i>	Europos (Anglijos britai ir škotai)	Integruotas haplotipų įvertis, 3,144
	Pietų Azijos (Gudžarato indai ir pandžabai)	Integruotas haplotipų įvertis, 3,027–3,557
<i>NEIL1</i>	Pietų Azijos (telugų indai)	Fu ir Li D, –6,959
	Afrikos (jorubai)	Fu ir Li D, –6,017 Fu ir Li F, –5,002
<i>RFWD3</i>	Pietų Azijos (bengalai)	Integruotas haplotipų įvertis, 3,189
	Rytų Azijos (kinai)	Integruotas haplotipų įvertis, 3,267
<i>MACROD2</i>	Europos (suomiai)	Tajima D, –2,547 Fu ir Li D, –6,011 Fu ir Li F, –5,179
<i>MAPK1</i>	Rytų Azijos (vietnamiečiai)	Fay ir Wu H, –7,194 Tajima D, –2,687
	Rytų Azijos (kinai)	Išplėstinis haplotipų homozigotiškumas, –2,642

Tai įrodo šių genomo sričių reikšmingumą žmogaus adaptacijai bendraja prasme, nepriklausomai nuo populiacijos. Kita vertus, kitiems analizuotiems genams teigiamos gamtinės atrankos įtaka kitose pasaulio populiacijose ir bendrojoje lietuvių populiacijoje (ANELGEMIA) nebuvo nustatyta arba buvo statistiškai nereikšminga. Tai rodo, kad LČKL galimai būdingi specifiniai adaptaciniai mechanizmai, lemiantys jų išgyvenamumą ir prisitaikymą prie visą gyvenimą besitęsiančių Černobylio katastrofos padarinių.

Nustatytos ir kitos teigiamos gamtinės atrankos veikiamos genomo sritys, galimai turinčios įtakos LČKL adaptacinėms savybėms. *DAO* genas yra LČKL 12 chromosomos teigiamos gamtinės atrankos veikiamoje genomo srityje. Šis genas koduoja peroksisominį fermentą, reikalingą neuronų diferenciacijai, dopamino sintezei [180]. Taip pat, šis genas siejamas su senėjimu. Mokslinėse publikacijose aprašoma, kad geno koduojamas fermentas skaido D-serino aminorūgštį, kurios sumažėjęs kiekis susijęs su galvos smegenų neuronų žūtimi ir vyresniame amžiuje pasireiškiančių neurologinių ligų, tokių kaip Alzheimerio liga, pasireiškimu. Didesnis D-

serino kiekis ir mažesnis *DAO* fermento aktyvumas turi neuroprotekcinių efektą [181]. Genas siejamas ir su oksidacinio streso atsaku [182], kurį patiria aukštas JS spinduliuotes dozes gaunantys asmenys, todėl galimai turi įtakos ir LČKL grupės atsakui į oksidacinį stresą bei lemia geresnę adaptaciją. Toje pačioje teigiamos gamtinės atrankos srityje yra ir *PPTC7* genas, siejamas su atsparumu aplinkos cheminiams toksinams [183]. Šio geno sričiai nustatyta teigiamos gamtinės atrankos įtaka Pietų Azijos (indų) populiacijoje (Fu ir Li D įvertis $-6,825$, $p < 0,05$) [122]. Indija – trečia pasaulyje šalis pagal užterštumą [184], todėl natūralu, kad žmogaus organizmas adaptuojasi prie esančių kenksmingų sąlygų. Skirtingas toksines sąlygas išgyvenusių asmenų grupėse nustatytas teigiamos gamtinės atrankos poveikis *PPTC7* gene rodo šios srities įtaką žmonėms prisitaikyti prie toksinių sąlygų.

Kitos LČKL adaptacijai galimai svarbios teigiamos gamtinės atrankos veikiamos genomo sritys susijusios su rezultatais, gautais šio tyrimo ankstesniuose etapuose. Nustatyti *FAF1*, *UBL7*, *PI4KAP2* genai, veikiami teigiamos gamtinės atrankos. Šie genai atitinkamai siejami su gretutiniu depresiniu sindromu (angl. *comorbid depressive syndrome*) [185], ekfoliacijos sindromu [186] ir potrauminio streso sindromu [187]. Gauti rezultatai papildo prieš tai nustatytas ekfoliacijos sindromui ir psichikos sveikatos būklėms būdingas genų asociacijas ir nurodo galimą bendrą šių genų veikimo mechanizmą.

Taip pat nustatyta, kad daugelis šiuo tyrimo etapu identifikuotų teigiamos gamtinės atrankos veikiamų genų siejami su vėžinėmis ligomis. Tai, kad šių genų sritys veikiamos teigiamos atrankos, o tikroje LČKL grupėje vėžinės ligos retos, parodo šių teigiamos gamtinės atrankos veikiamų genomo sričių galimą protekcinę funkciją vėžinių ligų kontekste ir įsitraukimą į adaptacinius mechanizmus, lemiančius LČKL išgyvenamumą.

Tyrimo metu naudota asociacijų analizės, teigiamos gamtinės atrankos poveikio vertinimo ir funkcinių *in silico* analizės metodų kombinavimo strategija leido nustatyti reikšmingas asociacijas ir bendrajai lietuvių populiacijai ir LČKL adaptacijai bei išgyvenamumui funkciškai galimai reikšmingus genomo variantus ir sritis. Gautus rezultatus patikslintų ir tyrimo strategiją praturtintų funkciniai tyrimai ląstelių kultūrose, audiniuose ar modeliniuose organizmuose. Taip pat būtų naudingas LČKL teigiamos gamtinės atrankos poveikio genomo sritims įvertinimas, naudojant tradicinius teigiamos gamtinės atrankos poveikio nustatymo metodus (pavyzdžiui, F_{ST} , integruoto haplotipų įverčio ir kt. metodus), ir gautų rezultatų palyginimas ne tik populiacijos viduje, bet ir su kitomis populiacijomis. Tai galėtų būti ateities tyrimų kryptis.

IŠVADOS

1. Plataus masto genotipavimo metodu atliktas bendros lietuvių populiacijos genomo charakterizavimas poveikio genomo variantų kontekste leido nustatyti lietuvių populiacijai būdingus genomo variantus potencialiai lemiančius sumažintą polinkį sirgti II tipo diabetu (NM_001172815.2(SLC30A8):c.826C>T) bei koronarine širdies liga (NM_001282190.1(ZC3HC1):c.1025G>A), lyginant su Europos (bendraja Europos, suomių iš Suomijos, Šiaurės ar Vakarų Europos kilmės Jutos gyventojų) populiacijomis.
2. Plataus masto genotipavimo metodu nustatyta unikali, Lietuvos Černobylio katastrofos likviduotojų viso genomo specifinė DNR sekos pokyčių variacija, statistiškai reikšmingai besiskirianti nuo bendros lietuvių populiacijos. Nustatyti poveikio genomo variantai, galimai turintys įtakos tiriamųjų adaptacijai.
 - 2.1. Palyginti su bendrąja lietuvių populiacija, nustatytas statistiškai reikšmingas genomo variantas NM_005576.4(*LOXLI*):c.458G>A ($p = 0,019$), Lietuvos Černobylio katastrofos likviduotojus galimai apsaugantis nuo eksfoliacijos sindromo ir glaukomos. Nepusiausvyros sankibos ir epistazės analizės metu nustatyti nauji kandidatiniai genai (*LHFPL3*, *GALNT6*, *PIH1D1*, *ANKS1B*, *METRNL*), galimai turintys įtakos eksfoliacijos ir glaukomos sindromo etiopatogenezei.
 - 2.2. Palyginti su vyresnių negu 65 m. amžiaus bendrosios lietuvių populiacijos vyrų grupe, nustatyti du statistiškai reikšmingi poveikio genomo variantai NM_001337.4:c.839C>T ($p = 0,008$) ir NM_001337.4:c.745G>A ($p = 0,015$) *CX3CRI* gene, galimai apsaugantys nuo koronarinės širdies ligos. Nepaisant to, širdies ir kraujagyslių ligos yra vienos labiausiai paplitusių Lietuvos Černobylio katastrofos likviduotojų tiriamųjų grupėje – tam įtakos gali turėti Lietuvos Černobylio katastrofos likviduotojų gyvensenos veiksniai.
 - 2.3. Kryptingos, su ligomis siejamų genų asociacijų analizės metu nustatyti statistiškai reikšmingi ($p < 1 \times 10^{-6}$) 43 VNP 15-oje koduojančių genų (stipriausi genai kandidatai – *ZMIZ1* ir *TSC22D1*), galimai turintys įtakos dideliame neigiamų psichikos būklių dažniui Lietuvos Černobylio katastrofos likviduotojų grupėje. Taip pat nustatytos naujos *FGF1*, *FGF2*, *FGF7* ir *BMP2* genų asociacijos su padidėjusia Lietuvos Černobylio katastrofos

likviduotojų burnos ertmės ligų rizika. Nustatytų genų sąveikos galimai turi įtakos likviduotojų fenotipui.

3. Siekiant įvertinti asociacijų analizės metu nustatytų poveikio genomo variantų protekcinę funkciją, atliktas bendrosios lietuvių populiacijos ir Lietuvos Černobylio katastrofos likviduotojų teigiamos gamtinės atrankos įtakos nustatytiems poveikio genomo variantams vertinimas. Nustatyti reikšmingi teigiamos gamtinės atrankos signalai *SLC30A8* gene (bendrojoje lietuvių populiacijoje, $\mu = 2,34$, $p < 0,05$), *LOXLI* (Lietuvos Černobylio katastrofos likviduotojų tiriamųjų grupėje, $\mu = 7,30$, $p < 0,05$) patvirtina šių asociacijų analizės metu nustatytų genų ir poveikio genomo variantų protekcinį vaidmenį. Pasirinkta kompleksinė tyrimo analizės strategija, sujungianti statistinius, gamtinės atrankos poveikio ir bioinformatikos funkcinės reikšmės įvertinimo metodus, leidžia išsamiau vertinti varianto funkcinį reikšmingumą.
4. Pirmą kartą, panaudojant viso genomo sekoskaitos duomenis, nustatytos 22 Lietuvos Černobylio katastrofos likviduotojams reikšmingiausios, palyginti su bendrąja lietuvių populiacija, unikalios teigiamos gamtinės atrankos veikiamos genomo sritys. Šiose srityse nustatyti genai *CDKN2C*, *ZNF827*, *CENPH*, *CCNB1*, *CDK7*, *TAF9*, *RAD17*, *GTF2H2C*, *PTP4A1*, *WDR11*, *ALKBH2*, *UNG*, *ARID3B*, *COMMD4*, *NEIL1*, *RFWD3*, *MACROD2*, *MAPK1*, *PPM1F*, *TOP3B* dalyvauja DNR pažaidų reparacijos mechanizmuose, reguliuoja ląstelės ciklą, palaiko homeostazę. Minėti genai svarbūs Lietuvos Černobylio katastrofos likviduotojų adaptacijai prie Černobylio katastrofos padarinių.

LITERATŪROS SĄRAŠAS (REFERENCES)

1. Rees JS, Castellano S, Andrés AM. The genomics of human local adaptation. *Trends in Genetics* 2020; 36(6): 415–428. Available at: <<https://doi.org/10.1016/j.tig.2020.03.006>>.
2. Peter BM, Huerta-Sanchez E, Nielsen R. Distinguishing between selective sweeps from standing variation and from a de novo mutation. *PLoS Genetics* 2012; e1003011. Available at: <<https://doi.org/10.1371/journal.pgen.1003011>>.
3. De Filippo C, Key FM, Ghirotto S, et al. Recent selection changes in human genes under long-term balancing selection. *Molecular Biology and Evolution* 2016; 33(6): 1435–1447. Available at: <<https://doi.org/10.1093/molbev/msw023>>.
4. Jeong C, Balanovsky O, Lukianova E, et al. The genetic history of admixture across inner Eurasia. *Nat Ecol Evol* 2019; 3: 966–976. Available at: <<https://doi.org/10.1038/s41559-019-0878-2>>.
5. Mathieson I, Lazaridis I, Rohland N, et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 2015; 528: 499–503. Available at: <<https://doi.org/10.1038/nature16152>>.
6. Pierron D, Heiske M, Razafindrazaka H, et al. Strong selection during the last millennium for African ancestry in the admixed population of Madagascar. *Nature Communications* 2018; 9(1): 1–9. Available at: <<https://doi.org/10.1038/s41467-018-03342-5>>.
7. Schrider DR, Kern AD. S/HIC: robust identification of soft and hard sweeps using machine learning. *PLoS Genetics* 2016; 12(3): e1005928. Available at: <<https://doi.org/10.1371/journal.pgen.1005928>>.
8. Nielsen R, Williamson S, Kim Y, et al. Genomic scans for selective sweeps using SNP data. *Genome Research* 2005; 15(11): 1566–1575. Available at: <<https://doi.org/10.1101/gr.4252305>>.
9. Tishkoff SA, Reed FA, Ranciaro A, et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics* 2007; 39(1): 31–40. Available at: <<https://doi.org/10.1038/ng1946>>.
10. Berg JJ, Coop G. A population genetic signal of polygenic adaptation. *PLoS Genetics* 2014; 10(8): e1004412. Available at: <<https://doi.org/10.1371/journal.pgen.1004412>>.
11. Daub JT, Hofer T, Cutivet E, et al. Evidence for polygenic adaptation to pathogens in the human genome. *Molecular Biology and Evolution* 2013; 30(7): 1544–1558. Available at: <<https://doi.org/10.1093/molbev/mst080>>.
12. White L, Romagné F, Müller E, et al. Genetic adaptation to levels of dietary selenium in recent human history. *Molecular Biology and*

- Evolution 2015; 32(6): 1507–1518. Available at: <<https://doi.org/10.1093/molbev/msv043>>.
13. Le Corre V, Kremer A. The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology* 2012; 21(7): 1548–1566. Available at: <<https://doi.org/10.1111/j.1365-294X.2012.05479.x>>.
 14. Racimo F, Schraiber JG, Casey F, et al. *Encyclopedia of Evolutionary Biology: Directional Selection and Adaptation*. Elsevier, 2016, p. 444–452. Available at: <<https://doi.org/10.1016/B978-0-12-800049-6.00028-7>>.
 15. Merilä J, Sheldon BC, Kruuk LEB. Explaining stasis: microevolutionary studies in natural populations. *Genetica* 2001; 112(1): 199–222. Available at: <<https://doi.org/10.1023/A:1013391806317>>.
 16. Hancock AM, Witonsky DB, Gordon AS, et al. Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genetics* 2008; 4(2): e32. Available at: <<https://doi.org/10.1371/journal.pgen.0040032>>.
 17. Dahlbäck B. Advances in understanding pathogenic mechanisms of thrombophilic disorders. *Blood* 2008; 112(1): 19–27. Available at: <<https://doi.org/10.1182/blood-2008-01-077909>>.
 18. Stengård JH, Zerba KE, Pekkanen J, et al. Apolipoprotein E polymorphism predicts death from coronary heart disease in a longitudinal study of elderly Finnish men. *Circulation* 1995; 91(2): 265–269. Available at: <<https://doi.org/10.1161/01.CIR.91.2.265>>.
 19. Butler JM, Hall N, Narendran N, et al. Identification of candidate protective variants for common diseases and evaluation of their protective potential. *BMC Genomics* 2017; 18(1): 575. Available at: <<https://doi.org/10.1186/s12864-017-3964-3>>.
 20. Nédélec Y, Sanz J, Baharian G, et al. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell* 2016; 167(3): 657–669. Available at: <<https://doi.org/10.1016/j.cell.2016.09.025>>.
 21. Davy T, Castellano S. The genomics of selenium: Its past, present and future. *Biochimica et Biophysica Acta (BBA) – General Subjects* 2018; 1862(11): 2427–2432. Available at: <<https://doi.org/10.1016/j.bbagen.2018.05.020>>.
 22. Bigam AW, Lee FS. Human high-altitude adaptation: forward genetics meets the HIF pathway. *Genes & Development* 2014; 28(20): 2189–2204. Available at: <<https://doi.org/10.1101/gad.250167.114>>.
 23. MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 2012; 335: 823–828. Available at: <<https://doi.org/10.1126/science.1215040>>.
 24. Maher MC, Uricchio LH, Torgerson DG, et al. Population genetics of rare variants and complex diseases. *Human Heredity* 2012; 74(3–4): 118–128. Available at: <<https://doi.org/10.1159/000346826>>.

25. Lee S, Abecasis GR, Boehnke M, et al. Rare-variant association analysis: study designs and statistical tests. *The American Journal of Human Genetics* 2014; 95(1): 5–23. Available at: <<https://doi.org/10.1016/j.ajhg.2014.06.009>>.
26. Kido T, Sikora-Wohlfeld W, Kawashima M, et al. Are minor alleles more likely to be risk alleles? *BMC Medical Genomics* 2018; 11(1): 1–11. Available at: <<https://doi.org/10.1186/s12920-018-0322-5>>.
27. Harper AR, Nayee S, Topol EJ. Protective alleles and modifier variants in human health and disease. *Nature Reviews Genetics* 2015; 16(12): 689–701. Available at: <<https://doi.org/10.1038/nrg4017>>.
28. Rossi R, Fang M, Zhu L, et al. Calculating and comparing codon usage values in rare disease genes highlights codon clustering with disease- and tissue-specific hierarchy. *PloS One* 2022; 17(3): e0265469. Available at: <<https://doi.org/10.1371/journal.pone.0265469>>.
29. Chattopadhyay A, Lu TP. Gene-gene interaction: the curse of dimensionality. *Annals of Translational Medicine* 2019; 7(24). Available at: <<https://doi.org/10.21037/atm.2019.12.87>>.
30. Li Y, Cho H, Wang F, et al. Statistical and Functional Studies Identify Epistasis of Cardiovascular Risk Genomic Variants From Genome Wide Association Studies. *Journal of the American Heart Association* 2020; 9(7): e014146. Available at: <<https://doi.org/10.1161/JAHA.119.014146>>.
31. Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* 2014; 23(13): 3133–3157. Available at: <<https://doi.org/10.1111/mec.12796>>.
32. Luikart G, Kardos M, Hand BK, et al. *Population Genomics: Advancing Understanding of Nature*. Springer 2018; 3–12. Available at: <https://doi.org/10.1007/13836_2018_60>.
33. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989; 123: 585–595. Available at: <<https://doi.org/10.1093/genetics/123.3.585>>.
34. Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. *Genetics* 2000; 155: 1405–1413. Available at: <<https://doi.org/10.1093/genetics/155.3.1405>>.
35. Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics* 1993; 133: 693–709. Available at: <<https://doi.org/10.1093/genetics/133.3.693>>.
36. Sabeti PC. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002; 419: 832–837. Available at: <<https://doi.org/10.1038/nature01140>>.
37. Voight BF, Kudravalli S, Wen X, et al. A map of recent positive selection in the human genome. *PLoS Bio* 2006; 4: e72. Available at: <<https://doi.org/10.1371/journal.pbio.0040072>>.

38. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. *Genetics* 1992; 132(2): 583–589.
39. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST} . *Nature Rev Genet* 2009; 10: 639–650. Available at: <<https://doi.org/10.1038/nrg2611>>.
40. Shriver MD, et al. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum Genom* 2004; 1: 274–286. Available at: <<https://doi.org/10.1186/1479-7364-1-4-274>>.
41. Sabeti PC, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature* 2007; 449: 913–918. Available at: <<https://doi.org/10.1038/nature06250>>.
42. Grossman SR, et al. A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* 2010; 327: 883–886. Available at: <<https://doi.org/10.1126/science.1183863>>.
43. Lohmueller KE, et al. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLoS Genet* 2011; 7: e1002326. Available at: <<https://doi.org/10.1371/journal.pgen.1002326>>.
44. Price AL, et al. Long-range LD can confound genome scans in admixed populations. *Am J Hum Genet* 2008; 83: 132–135. Available at: <<https://doi.org/10.1016/j.ajhg.2008.06.005>>.
45. Jukes TH, Cantor CR. Evolution of protein molecules. *Mammalian Protein Metabolism* 1969; 3: 21–132.
46. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* 1986; 3(5): 418–426.
47. Watterson GA. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 1975; 7(2): 256–276. Available at: <[https://doi.org/10.1016/0040-5809\(75\)90020-9](https://doi.org/10.1016/0040-5809(75)90020-9)>.
48. Tajima F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 1983; 105(2): 437–460. Available at: <<https://doi.org/10.1534/genetics.116.195271>>.
49. Tajima F. The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. *Genetics* 1996; 143(3): 1457–1465. Available at: <<https://doi.org/10.1093/genetics/143.3.1457>>.
50. Koropoulis A, Alachiotis N, Pavlidis P. Detecting Positive Selection in Populations Using Genetic: Statistical Population Genomics. *Methods in Molecular Biology* 2020; 2090. Available at: <https://doi.org/10.1007/978-1-0716-0199-0_5>.

51. Pavlidis P, Živković D, Stamatakis A, et al. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol Biol Evol* 2013; 30(9): 2224–2234. Available at: <<https://doi.org/10.1093/molbev/mst112>>.
52. Alachiotis N, Pavlidis P. RAI_{SD} detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Commun Biol* 2018; 1(1): 79. Available at: <<https://doi.org/10.1038/s42003-018-0085-8>>.
53. Vidal CM, Lane CS, Asrat A, et al. Age of the oldest known Homo sapiens from eastern Africa. *Nature* 2022; 601(7894): 579–583. Available at: <<https://doi.org/10.1038/s41586-021-04275-8>>.
54. Jeong C, Di Rienzo A. Adaptations to local environments in modern human populations. *Current Opinion in Genetics & Development* 2014; 29: 1–8. Available at: <<https://doi.org/10.1016/j.gde.2014.06.011>>.
55. Bustamante CD, Fledel-Alon A, Williamson S, et al. Natural selection on protein-coding genes in the human genome. *Nature* 2005; 437(7062): 1153–1157. Available at: <<https://doi.org/10.1038/nature04240>>.
56. Nielsen R, Bustamante C, Clark AG, et al. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biology* 2005; 3(6): e170. Available at: <<https://doi.org/10.1371/journal.pbio.0030170>>.
57. Rockman MV, Hahn MW, Soranzo N, et al. Positive selection on MMP3 regulation has shaped heart disease risk. *Current Biology* 2004; 14(17): 1531–1539. Available at: <<https://doi.org/10.1016/j.cub.2004.08.051>>.
58. Thompson EE, Kuttub-Boulos H, Witonsky D, et al. CYP3A variation and the evolution of salt-sensitivity variants. *The American Journal of Human Genetics* 2004; 75(6): 1059–1069. Available at: <<https://doi.org/10.1086/426406>>.
59. Nakajima T, Wooding S, Sakagami T, et al. Natural selection and population history in the human angiotensinogen gene (AGT): 736 complete AGT sequences in chromosomes from around the world. *The American Journal of Human Genetics* 2004; 74(5): 898–916. Available at: <<https://doi.org/10.1086/420793>>.
60. Fullerton SM, Bartoszewicz A, Ybazeta G, et al. Geographic and haplotype structure of candidate type 2 diabetes-susceptibility variants at the calpain-10 locus. *The American Journal of Human Genetics* 2002; 70(5): 1096–1106. Available at: <<https://doi.org/10.1086/339930>>.
61. Zhou G, Zhai Y, Dong X, et al. Haplotype structure and evidence for positive selection at the human IL13 locus. *Molecular Biology and Evolution* 2004; 21(1): 29–36. Available at: <<https://doi.org/10.1093/molbev/msg231>>.
62. Sakagami T, Witherspoon DJ, Nakajima T, et al. Local adaptation and population differentiation at the interleukin 13 and interleukin 4 loci.

- Genes & Immunity 2004; 5(5): 389–397. Available at: <<https://doi.org/10.1038/sj.gene.6364109>>.
63. Akey JM, Eberle MA, Rieder MJ, et al. Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biology* 2004; 2(10): e286. Available at: <<https://doi.org/10.1371/journal.pbio.0020286>>.
 64. Biswas S, Akey JM. Genomic insights into positive selection. *TRENDS in Genetics* 2006; 22(8): 437–446. Available at: <<https://doi.org/10.1016/j.tig.2006.06.005>>.
 65. Kasperavičiūtė D, Kučinskas V, Stoneking M. Y chromosome and mitochondrial DNA variation in Lithuanians. *Ann Hum Genet* 2004; 68: 438–452. Available at: <<https://doi.org/10.1046/j.1529-8817.2003.00119.x>>.
 66. Uktverytė I. Lietuvos etnolingvistinių grupių genetinės struktūros analizė remiantis informatyviais genomo žymenimis. Daktaro disertacija. Vilnius: Vilniaus universitetas, 2014. Prieiga internete: <<https://epublications.vu.lt/object/elaba:2184808/>>.
 67. Umikytė A. Lietuvos populiacijos genetinės struktūros ir evoliucinių veiksmų analizė, remiantis plataus masto genotipavimo duomenimis: praeitis ir dabartis. Daktaro disertacija. Vilnius: Vilniaus universitetas, 2018. Prieiga internete: <<https://epublications.vu.lt/object/elaba:33050072/MAIN>>.
 68. Urnikyte A, Flores-Bello A, Mondal M, et al. Patterns of genetic structure and adaptive positive selection in the Lithuanian population from high-density SNP data. *Scientific Reports* 2019; 9(1): 1–11. Available at: <<https://doi.org/10.1038/s41598-019-45746-3>>.
 69. Bernstein C, Prasad AR, Nfonsam V, et al. New Research Directions in DNA Repair: DNA damage, DNA repair and cancer. *Intech Open* 2013; 413–465.
 70. Gorbunova V, Seluanov A, Mao Z, et al. Navigation Changes in DNA repair during aging. *Nucleic Acids Research* 2007; 35(22): 7466–7474. Available at: <<https://doi.org/10.1093/nar/gkm756>>.
 71. Chatterjee N, Walker GC. Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis* 2017; 58(5): 235–263. Available at: <<https://doi.org/10.1002/em.22087>>.
 72. Prise KM, Folkard M, Michael DB. A review of the bystander effect and its implications for low-dose exposure. *Radiation Protection Dosimetry* 2003; 104(4): 347–355. Available at: <<https://doi.org/10.1093/oxfordjournals.rpd.a006198>>.
 73. Desouky O, Ding N, Zhou G. Targeted and non-targeted effects of ionizing radiation. *Journal of Radiation Research and Applied Sciences* 2015; 8(2): 247–254. Available at: <<https://doi.org/10.1016/j.jrras.2015.03.003>>.
 74. Borrego-Soto G, Ortiz-López R, Rojas-Martínez A. Ionizing radiation-induced DNA injury and damage detection in patients with breast cancer.

- Genetics and Molecular Biology 2015; 38: 420–432. Available at: <<https://doi.org/10.1590/S1415-475738420150019>>.
75. Evans HH, DeMarini DM. Ionizing radiation-induced mutagenesis: radiation studies in *Neurospora* predictive for results in mammalian cells. *Mutation Research / Reviews in Mutation Research* 1999; 437(2): 135–150. Available at: <[https://doi.org/10.1016/S1383-5742\(99\)00080-0](https://doi.org/10.1016/S1383-5742(99)00080-0)>.
 76. Zander A, Paunesku T, Woloschak G. Radiation databases and archives—examples and comparisons. *International Journal of Radiation Biology* 2019; 95(10): 1378–1389. Available at: <<https://doi.org/10.1080/09553002.2019.1572249>>.
 77. Pierce DA, Shimizu Y, Preston DL, et al. Studies of the mortality of atomic bomb survivors. Report 12, Part I. Cancer: 1950–1990. *Radiation Research* 1996; 146(1): 1–27. Available at: <<https://doi.org/10.2307/3579391>>.
 78. Mabuchi K, Fujiwara S, Preston DL, et al. Atomic-bomb survivors: long-term health effects of radiation. Philadelphia, PA: Wolter Kluwer / Lippincott Williams and Wilkins, 2011, p. 89–113.
 79. Kamiya K, Ozasa K, Akiba S, et al. Long-term effects of radiation exposure on health. *The Lancet* 2015; 386(9992): 469–478. Available at: <[https://doi.org/10.1016/S0140-6736\(15\)61167-9](https://doi.org/10.1016/S0140-6736(15)61167-9)>.
 80. Nakamura N. Genetic effects of radiation in atomic-bomb survivors and their children: past, present and future. *Journal of Radiation Research* 2006; 47(Suppl B): B67–B73. Available at: <<https://doi.org/10.1269/jrr.47.B67>>.
 81. Horai M, Mishima H, Hayashida C, et al. Detection of de novo single nucleotide variants in offspring of atomic-bomb survivors close to the hypocenter by whole-genome sequencing. *Journal of Human Genetics* 2018; 63(3): 57–363. Available at: <<https://doi.org/10.1038/s10038-017-0392-9>>.
 82. Kong A, Frigge ML, Masson G, et al. Rate of de novo mutations and the importance of father’s age to disease risk. *Nature* 2012; 488: 471–475. Available at: <<https://doi.org/10.1038/nature11396>>.
 83. Moysich RJ, Menezes AM, Michalek M. Chernobyl-related ionising radiation exposure and cancer risk: an epidemiological review. *Lancet* 2002; 3: 269–279. Available at: <[https://doi.org/10.1016/S1470-2045\(02\)00727-1](https://doi.org/10.1016/S1470-2045(02)00727-1)>.
 84. Korsakov AV, Geger EV, Lagerev DG, et al. De novo congenital malformation frequencies in children from the Bryansk region following the Chernobyl disaster (2000–2017). *Heliyon* 2020; 6(8): e04616. Available at: <<https://doi.org/10.1016/j.heliyon.2020.e04616>>.
 85. Annex J. Exposures and Effects of the Chernobyl Accident. In: Sources and Effects of Ionizing Radiation, vol. II. New York: United Nations, 2000. Available at: <<http://hprevot.fr/Chernobyl%20Report%20UNSCEAR%202011.pdf>>.

86. Minenko VV, Drozdovich SS, Tretiakovich VE, et al. Determination of annual total effective exposure doses for the residents of Belarus populated areas radionuclide-contaminated after the Chernobyl accident. *Accompl Med Sci* 1998; 3: 23–45.
87. Sevankaev AV, Lloyd DC, Potetnya OI, et al. Chromosome aberrations in Lymphocytes of Residents of Areas Contaminated by Radioactive Discharges from the Chernobyl Accident. *Radiat Protect Dosim* 1995; 58(4): 247–254. Available at: <<https://doi.org/10.1093/oxfordjournals.rpd.a082621>>.
88. Vorobtsova IE, Vorobyeva MV, Bogomazova AN, et al. Cytogenetic study of children living in St Petersburg region suffered from the Chernobyl accident. The rate of unstable chromosomes aberrations in peripheral blood lymphocytes. *Radiat Biol Radioecol* 1995; 22: 630–635.
89. Bochkov P, Bedelbaeva KA, Katosova LD, et al. Chromosome aberrations in peripheral lymphocytes of subjects living in regions with increased radiation levels. *Reports Acad Sci (Belarus)* 1991; 35(8): 745–748.
90. Stepanova EI, Misharina JA, Vdovenko VU. Remote cytogenetic effects in children exposed in prenatal period after the Chernobyl accident. *Radiat Biol Radioecol* 2002; 42(6): 700–703.
91. Mikhalevich FA, De Zwart GA, Perepetskaya NV, et al. Tates Radiation effects in lymphocytes of children living in a Chernobyl contaminated region of Belarus. *Int J Rad Biol* 2000, 76(10): 1377–1385. Available at: <<https://doi.org/10.1080/09553000050151655>>.
92. Frenzilli E, Bosco A, Antonelli G, et al. DNA damage evaluated by alkaline single cell gel electrophoresis (SCGE) in children of Chernobyl, 10 years after the disaster. *Mutat Res* 2001; 491: 139–149. Available at: <[https://doi.org/10.1016/S1383-5718\(01\)00136-X](https://doi.org/10.1016/S1383-5718(01)00136-X)>.
93. Møller AP, Mousseau TA. Biological consequences of Chernobyl: 20 years on. *Trends in Ecology & Evolution* 2006; 21(4): 200–207. Available at: <<https://doi.org/10.1016/j.tree.2006.01.008>>.
94. Yeager M, Machiela MJ, Kothiyal P, et al. Lack of transgenerational effects of ionizing radiation exposure from the Chernobyl accident. *Science* 2021; 372(6543): 725–729. Available at: <<https://doi.org/10.1126/science.abg2365>>.
95. Weinberg HS, Korol AB, Kirzhner VM, et al. Very high mutation rate in offspring of Chernobyl accident liquidators. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 2001; 268(1471): 1001–1005. Available at: <<https://doi.org/10.1098/rspb.2001.1650>>.
96. Ladygienė R, Pilkytė L, Tankevičiūtė A, et al. Ilgalaikis Černobylio atominės elektrinės avarijos poveikis Lietuvos aplinkai ir gyventojų apšvitai. In: Konferencijos „Černobylio atominės elektrinės avarija: pamokos ir pasekmės (po

- 30 metų)“ medžiaga. Vilnius: Radiacinės saugos centras, 2016. Prieiga internete: <<https://www.hi.lt/uploads/pdf/visuomenes%20sveikata/2016.priedas1/Vs%20priedas%202016%20Nr1%20Visas.pdf>>.
97. Bromet EJ. Mental health consequences of the Chernobyl disaster. *Journal of Radiological Protection* 2012; 32(1): N71. Available at: <<http://dx.doi.org/10.1088/0952-4746/32/1/N71>>.
 98. National Research Council (US) Committee on the Biological Effects of Ionizing Radiation (BEIR V). *Health Effects of Exposure to Low Levels of Ionizing Radiation: Beir V Genetic. Effects of Radiation*, National Academies Press (US): Washington DC, 1990. Available at: <<https://www.ncbi.nlm.nih.gov/books/NBK218706>>.
 99. Rahu K, Auvinen A, Hakulinen T, et al. Chernobyl cleanup workers from Estonia: follow-up for cancer incidence and mortality. *Journal of Radiological Protection* 2013; 33(2): 395. Available at: <<https://doi.org/10.1088/0952-4746/33/2/395>>.
 100. Lazutka JR. Chromosome aberrations and rogue cells in lymphocytes of Chernobyl clean-up workers. *Mutation Re-search / Fundamental and Molecular Mechanisms of Mutagenesis* 1996, 350(2): 315–329. Available at: <[https://doi.org/10.1016/0027-5107\(95\)00170-0](https://doi.org/10.1016/0027-5107(95)00170-0)>.
 101. Lazutka JR, Rimdeika GJ. Influence of the Chernobyl accident on the frequency of chromosomal damage and health status of Lithuanian clean-up workers. *Biologija* 2006; 1.
 102. Durante M, Formenti SC. Radiation-induced chromosomal aberrations and immunotherapy: Micronuclei, cytosolic DNA, and interferon-production pathway. *Frontiers in Oncology* 2018; 8: 192. Available at: <<https://doi.org/10.3389/fonc.2018.00192>>.
 103. Bjelland I, Dahl AA, Haug TT, et al. The validity of the Hospital Anxiety and Depression Scale: an updated literature review. *Journal of Psychosomatic Research* 2002; 52(2): 69–77. Available at: <[https://doi.org/10.1016/S0022-3999\(01\)00296-3](https://doi.org/10.1016/S0022-3999(01)00296-3)>.
 104. Domarkienė I. Lietuvos populiacijos genetinės struktūros tyrimas remiantis vieno nukleotido polimorfizmų asociacijos su liga analize. *Daktaro disertacija*. Vilnius: Vilniaus universitetas, 2014. Prieiga internete: <<https://epublications.vu.lt/object/elaba:2184673/>>.
 105. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018; 46: D1062–1067. Available at: <<https://doi.org/10.1093/nar/gkx1153>>.
 106. OMIM database. Johns Hopkins University, USA. 1966–2021. Available at: <<https://omim.org/>>.

107. MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 2012; 335: 823–828. Available at: <<https://doi.org/10.1126/science.1215040>>.
108. Illumina. Genome Studio Genotyping Module v2 User Guide. 2016. Available at: <https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/genomestudio/genomestudio-2-0/genomestudio-genotyping-module-v2-user-guide-11319113-01.pdf>.
109. Illumina. Infinium genotyping data analysis – a guide for analyzing infinium genotyping data using the genome studio genotyping module. 2010. Available at: <https://www.illumina.com/Documents/products/technotes/technote_infinium_genotyping_data_analysis.pdf>.
110. Guo Y, Ye F, Sheng Q, et al. Three-stage quality control strategies for DNA re-sequencing data. *Briefings in Bioinformatics* 2014; 15(6): 879–889. Available at: <<https://doi.org/10.1093/bib/bbt069>>.
111. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015; 526(7571): 68–74. Available at: <<https://doi.org/10.1038/nature15393>>.
112. Sherry ST, Ward M, Sirotkin K. dbSNP-Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. *Genome Res* 1999; 9: 677–679. Available at: <<http://www.ncbi.nlm.nih.gov/SNP/>>.
113. Kopanos C, Tsiolkas V, Kouris A, et al. VarSome: The Human Genomic Variant Search Engine. *Bioinformatics* 2019; 35(11): 1978. Available at: <<https://doi.org/10.1093/bioinformatics/bty897>>.
114. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* 2019; 47(D1): D506–D515. Available at: <<https://doi.org/10.1093/nar/gky1049>>.
115. Yates AD, Achuthan P, Akanni W, et al. Ensembl 2020. *Nucleic Acids Research* 2020; 48(D1): D682–D688. Available at: <<https://doi.org/10.1093/nar/gkz966>>.
116. Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research* 2010; 38: W214–W220. Available at: <<https://doi.org/10.1093/nar/gkq537>>.
117. Szklarczyk D, Gable AL, Nastou KC, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* 2021; 49(D1): D605–12. Available at: <<https://doi.org/10.1093/nar/gkaa1074>>.
118. Greene CS, Krishnan A, Wong AK, et al. Understanding multicellular function and disease with human tissue-specific networks. *Nature Genetics* 2015; 47: 569–576. Available at: <<https://doi.org/10.1038/ng.3259>>.

119. Raudvere U, Kolberg L, Kuzmin I, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research* 2019; 47(W1): W191–198. Available at: <<https://doi.org/10.1093/nar/gkz369>>.
120. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *Giga Science* 2021; 10(2): giab008. Available at: <<https://doi.org/10.1093/gigascience/giab008>>.
121. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res* 2002; 12(6): 996–1006. Available at: <<https://genome.ucsc.edu>>.
122. Murga-Moreno J, Coronado-Zamora M, Bodelón A, et al. PopHumanScan: the online catalog of human genome adaptation. *Nucleic Acids Research* 2019; 47(D1): D1080–D1089. Available at: <<https://doi.org/10.1093/nar/gky959>>.
123. Mi H, Ebert D, Muruganujan A, et al. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucl Acids Res* 2020; 49(D1): D394–D403. Available at: <<http://doi.org/10.1093/nar/gkaa1106>>.
124. Smith CL, Eppig JT. The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Wiley Interdiscip Rev Syst Biol Med* 2009; 1(3): 390–399. Available at: <<https://doi.org/10.1002/wsbm.44>>.
125. Wood RD, Mitchell M, Lindahl T. Human DNA repair genes. *Mutation Research / Fundamental and Molecular Mechanisms of Mutagenesis* 2005; 577(1–2): 275–283. Available at: <<https://doi.org/10.1016/j.mrfmmm.2005.03.007>>. In: <<https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html#Human%20DNA%20Repair%20Genes>>.
126. The Chernobyl Forum: 2002–2005. Chernobyl’s legacy: Health, Environmental and Socio-Economic impacts and Recommendations to the Governments of Belarus, the Russian Federation and Ukraine. Austria: IAEA, 2006. Available at: <<https://www.iaea.org/sites/default/files/chemobyl.pdf>>.
127. Lithuanian Health Information Centre of Institute of Hygiene. Available at: <https://hi.lt/html/en/health_statistic.html>.
128. Tolstrup JS, Nordestgaard BG, Rasmussen S, et al. Alcoholism and alcohol drinking habits predicted from alcohol dehydrogenase genes. *The Pharmacogenomics Journal* 2008; 8(3): 220–227. Available at: <<https://doi.org/10.1038/sj.tpj.6500471>>.
129. Edenberg HJ. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Research*

- & Health 2007; 30(1): 5. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3860432/>.
130. Palmer RH, McGeary JE, Heath AC, et al. Shared additive genetic influences on DSM-IV criteria for alcohol dependence in subjects of European ancestry. *Addiction* 2015; 110: 1922–1931. Available at: <https://doi.org/10.1111/add.13070>.
 131. Walters RK, Polimanti R, Johnson EC, et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat Neurosci* 2018; 21: 1656–1669. Available at: <https://doi.org/10.1038/s41593-018-0275-1>.
 132. Sanghera DK, Blackett PR. Type 2 diabetes genetics: beyond GWAS. *Journal of Diabetes & Metabolism* 2012; 3(198). Available at: <https://doi.org/10.4172/2155-6156.1000198>.
 133. Sarhangi N, Sharifi F, Hashemian L, et al. PPARG (Pro12Ala) genetic variant and risk of T2DM: a systematic review and meta-analysis. *Scientific Reports* 2020; 10(1): 1–18. Available at: <https://doi.org/10.1038/s41598-020-69363-7>.
 134. Flannick J, Thorleifsson G, Beer NL, et al. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet* 2014; 46(4): 357–363. Available at: <https://doi.org/10.1038/ng.2915>.
 135. Brunke-Reese D, Ssentongo P, Ssentongo AE, et al. The Role of Genetic Variant rs13266634 in SLC30A8/ZnT8 in Postoperative Hyperglycemia After Major Abdominal Surgery. *The Journal of Clinical Endocrinology & Metabolism* 2019; 104(9): 3877–3892. Available at: <https://doi.org/10.1210/jc.2018-02588>.
 136. Shadrina AS, Shashkova TI, Torgasheva AA, et al. Prioritization of causal genes for coronary artery disease based on cumulative evidence from experimental and in silico studies. *Sci Rep* 2020; 10: 10486. Available at: <https://doi.org/10.1038/s41598-020-67001-w>.
 137. Lithuanian Health Information Centre of Institute of Hygiene. Causes of Death 2019. Available at: https://www.hi.lt/uploads/pdf/leidiniai/Statistikos/Mirties_priezastys/Mirties_priezastys_2019.pdf.
 138. Koponen P, Borodulin K, Lundqvist A, et al. Terveys, toimintakyky ja hyvinvointi Suomessa: FinTerveys 2017-tutkimus. 2018. Available at: <https://www.julkari.fi/handle/10024/136223>.
 139. Sund R, Koski S. FinDM II. On the register-based measurement of the prevalence and incidence of diabetes and its long-term complications. 2009. Available at: https://www.diabetes.fi/en/finnish_diabetes_association/dehko/publications.

140. Schlötzer-Schrehardt U, Zenkel M. The role of lysyl oxidase-like 1 (LOXL1) in exfoliation syndrome and glaucoma. *Exp Eye Res* 2019; 189: 107818. Available at: <<https://doi.org/10.1016/j.exer.2019.107818>>.
141. Greene AG, Eivers SB, Dervan EW, et al. Lysyl Oxidase Like 1: Biological roles and regulation. *Exp Eye Res* 2020; 193: 107975. Available at: <<https://doi.org/10.1016/j.exer.2020.107975>>.
142. Igo RP, Bailey JNC. Genetic risk scores in complex eye disorders. *Genetics and Genomics of Eye Disease*. Elsevier, 2020, p. 259–275. Available at: <<https://doi.org/10.1016/B978-0-12-816222-4.00016-2>>.
143. Rahal A, Kumar A, Singh V, et al. Oxidative Stress, Prooxidants, and Antioxidants: The Interplay. *BioMed Res Int* 2014; 2014: 1–19. Available at: <<https://doi.org/10.1155/2014/761264>>.
144. Koc U, Cam I. Radiation and oxidative stress. *Toxicology*. Elsevier BV, 2021, p. 233–241. Available at: <<https://doi.org/10.1016/B978-0-12-819092-0.00024-8>>.
145. Abu-Amero KK, Kondkar AA, Mousa A, et al. Decreased total antioxidants status in the plasma of patients with pseudoexfoliation glaucoma. *Mol Vis* 2011; 17: 2769–2775. Available at: <<http://www.molvis.org/molvis/v17/a300/>>.
146. Nucci C, Di Pierro D, Varesi C, et al. Increased malondialdehyde concentration and re-duced total antioxidant capacity in aqueous humor and blood samples from patients with glaucoma. *Molecular Vision* 2013; 19: 1841–1846. Available at: <<http://www.molvis.org/molvis/v19/1841>>.
147. Rimoin DL, Pyeritz RE, Korf B. Glaucoma. *Emery and Rimoin’s essential medical genetics*. Elsevier, 2013, p. 498–499.
148. Rautenbach RM, Bardien S, Harvey J, et al. An Investigation Into LOXL1 Variants in Black South African Individuals With Exfoliation Syndrome. *Arch Ophthalmol* 2011; 129: 206–210. Available at: <<https://doi.org/10.1001/archophthalmol.2010.349>>.
149. Wang L, Yu Y, Fu S, Zhao W, Liu P. LOXL1 Gene Polymorphism With Exfoliation Syndrome / Exfoliation Glaucoma. *J Glaucoma* 2016; 25: 62–94. Available at: <<https://doi.org/10.1097/ijg.000000000000128>>.
150. Thorleifsson G, Magnusson KP, Sulem P, et al. Common Sequence Variants in the LOXL1 Gene Confer Susceptibility to Exfoliation Glaucoma. *Sci* 2007; 317: 1397–1400. Available at: <<https://doi.org/10.1126/science.1146554>>.
151. Pasquale LR, Kang JH, Fan B, et al. LOXL1 Polymorphisms: Genetic Biomarkers that Presage Environmental Determinants of Exfoliation Syndrome. *J Glaucoma* 2018; 27: S20–S23. Available at: <<https://doi.org/10.1097/ijg.0000000000000915>>.

152. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nat Cell Bio* 2009; 461: 747–753. Available at: <<https://doi.org/10.1038/nature08494>>.
153. Moatti D, Faure S, Fumeron F, et al. Polymorphism in the fractalkine receptor CX3CR1 as a genetic risk factor for coronary artery disease. *Blood: The Journal of the American Society of Hematology* 2001; 97(7): 1925–1928. Available at: <<https://doi.org/10.1182/blood.v97.7.1925>>.
154. López-López A, Gelpi E, Lopategui DM, et al. Association of the CX3CR1-V249I variant with neurofibrillary pathology progression in late-onset Alzheimer’s disease. *Molecular Neurobiology* 2018; 55(3): 2340–2349. Available at: <<https://doi.org/10.1007/s12035-017-0489-3>>.
155. Sakai M, Takeuchi H, Yu Z, et al. Polymorphisms in the microglial marker molecule CX3CR1 affect the blood volume of the human brain. *Psychiatry and Clinical Neurosciences* 2018; 72(6): 409–422. Available at: <<https://doi.org/10.1111/pcn.12649>>.
156. Nasibullin TR, Yagafarova LF, Yagafarov IR, et al. Combinations of polymorphic markers of chemokine genes, their receptors and acute phase protein genes as potential predictors of coronary heart diseases. *Acta Naturae* 2016; 8(1): 111–116.
157. McDermott DH, Fong AM, Yang Q, et al. Chemokine receptor mutant CX3CR1-M280 has impaired adhesive function and correlates with protection from cardiovascular disease in humans. *The Journal of Clinical Investigation* 2003; 111(8): 1241–1250. Available at: <<https://doi.org/10.1172/JCI16790>>.
158. Torpy JM, Burke AE, Glass RM. Coronary heart disease risk factors. *Jama* 2009; 302(21): 2388–2388. Available at: <<https://doi.org/10.1001/jama.302.21.2388>>.
159. Tam V, Patel N, Turcotte M, et al. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics* 2019; 20(8): 467–484. Available at: <<https://doi.org/10.1038/s41576-019-0127-1>>.
160. Goldstein DB. Common genetic variation and human traits. *New England Journal of Medicine* 2009; 360(17): 1696.
161. Gailienė D. Gyvenimas po lūžio. Vilnius: Eugrimas, 2015, p. 143–166. Prieiga internete: <https://www.fsf.vu.lt/dokumentai/Naujienos_Mokslo_pasiekimai/2015/Gyvenimas_po_luzio_i_interneta.pdf>.
162. Lietuvos oficialiosios statistikos portalas. Prieiga internete: <<https://osp.stat.gov.lt/>>.

163. Pandya M, Altinay M, Malone DA, Anand A. Where in the brain is depression? *Current Psychiatry Reports* 2012; 14(6): 634–642. Available at: <<https://doi.org/10.1007/s11920-012-0322-7>>.
164. Zhang FF, Peng W, Sweeney JA, et al. Brain structure alterations in depression: Psychoradiological evidence. *CNS Neuroscience & Therapeutics* 2018; 24(11): 994–1003. Available at: <<https://doi.org/10.1111/cns.12835>>.
165. Anderson DM, Anderson KM, Chang CL, et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* 2015; 160(4): 595–606. Available at: <<https://doi.org/10.1016/j.cell.2015.01.009>>.
166. Gregson A, Thompson K, Tsirka SE, et al. Emerging small-molecule treatments for multiple sclerosis: Focus on B cells. *F1000 Research* 2019; 8. Available at: <<https://doi.org/10.12688/f1000research.16495.1>>.
167. Sellitti DF, Koles N, Mendonça MC. Regulation of C-type natriuretic peptide expression. *Peptides* 2011; 32(9): 1964–1971. Available at: <<https://doi.org/10.1016/j.peptides.2011.07.013>>.
168. Vargas DMD, De Bastiani MA, Zimmer ER, Klamt F. Alzheimer’s disease master regulators analysis: search for potential molecular targets and drug repositioning candidates. *Alzheimer’s Research & Therapy* 2018; 10(1): 1–12. Available at: <<https://doi.org/10.1186/s13195-018-0394-7>>.
169. Zhang Q, Huang X, Chen XZ, et al. Association of Gene Variations in Ionotropic Glutamate Receptor and Attention-Deficit / Hyperactivity Disorder in the Chinese Population: A Two-Stage Case–Control Study. *Journal of Attention Disorders* 2021; 25(10): 1362–1373. Available at: <<https://doi.org/10.1177/1087054720905089>>.
170. Kobayashi Y, Takemoto R, Yamato S, et al. Depression-resistant phenotype in mice overexpressing regulator of G protein signaling 8 (RGS8). *Neuroscience* 2018; 383: 160–169. Available at: <<https://doi.org/10.1016/j.neuroscience.2018.05.005>>.
171. Afanasyev DE, Liubarets SF. Odontological effects of ionizing radiation (review). *Probl Radiac Med Radiobiol* 2020; 25: 18–55. Available at: <<https://doi.org/10.33145/2304-8336-2020-25-18-55>>.
172. Sadetzki S, Chetrit A, Sgan-Cohen HD, et al. Long-term effects of exposure to ionizing irradiation on periodontal health status – the tinea capitis cohort study. *Frontiers in Public Health* 2015; 226. Available at: <<https://doi.org/10.3389/fpubh.2015.00226>>.
173. Lu H, Zhao Q, Guo J, et al. Direct radiation-induced effects on dental hard tissue. *Radiation Oncology* 2019; 14(1): 1–11. Available at: <<https://doi.org/10.1186/s13014-019-1208-1>>.

174. Барер ГМ. Концепция патогенеза лучевого поражения зубов. *Стоматология* 1991; 5: 8–12.
175. Bei M, Maas R. FGFs and BMP4 induce both Msx1-independent and Msx1-dependent signaling pathways in early tooth development. *Development* 1998; 125(21): 4325–4333. Available at: <<https://doi.org/10.1242/dev.125.21.4325>>.
176. Du W, Du W, Yu H. The role of fibroblast growth factors in tooth development and incisor renewal. *Stem Cells International*, 2018. Available at: <<https://doi.org/10.1155/2018/7549160>>.
177. Lee HS, Lee J, Kim SO, et al. Comparative gene-expression analysis of the dental follicle and periodontal ligament in humans. *PloS One* 2013; 8(12): e84201. Available at: <<https://doi.org/10.1371/journal.pone.0084201>>.
178. Kósa JP, Kis A, Bácsi K, et al. The protective role of bone morphogenetic protein-8 in the glucocorticoid-induced apoptosis on bone cells. *Bone* 2011; 48(5): 1052–1057. Available at: <<https://doi.org/10.1016/j.bone.2011.01.017>>.
179. Takase HM, Itoh T, Ino S, et al. FGF7 is a functional niche signal required for stimulation of adult liver progenitor cells that support liver regeneration. *Genes & Development* 2013; 27(2): 169–181. Available at: <<https://doi.org/10.1101/gad.204776.112>>.
180. Huang X, Kong H, Tang M, et al. D-Serine regulates proliferation and neuronal differentiation of neural stem cells from postnatal mouse forebrain. *CNS Neurosci Ther* 2012; 18: 4–13. Available at: <<https://doi.org/10.1111/j.1755-5949.2011.00276.x>>.
181. Lin CH, Yang HT, Chiu CC, Lane HY. Blood levels of D-amino acid oxidase vs. D-amino acids in reflecting cognitive aging. *Scientific Reports* 2017; 7(1): 1–10. Available at: <<https://doi.org/10.1038/s41598-017-13951-7>>.
182. Lu JM, Gong N, Wang YC, Wang YX. D-Amino acid oxidase-mediated increase in spinal hydrogen peroxide is mainly responsible for formalin-induced tonic pain. *British Journal of Pharmacology* 2012; 165: 1941–1955. Available at: <<https://doi.org/10.1111/j.1476-5381.2011.01680.x>>.
183. Zhu J, Dubois A, Ge Y, et al. Application of human haploid cell genetic screening model in identifying the genes required for resistance to environmental toxicants: Chlorpyrifos as a case study. *Journal of Pharmacological and Toxicological Methods* 2015; 76: 76–82. Available at: <<https://doi.org/10.1016/j.vascn.2015.08.154>>.
184. World Population Review. Most Polluted Countries 2022. Available at: <<https://worldpopulationreview.com/country-rankings/most-polluted-countries>>.

185. Edwards AC, Aliev F, Bierut LJ, et al. Genome-wide association study of comorbid depressive syndrome and alcohol dependence. *Psychiatric Genetics* 2012; 22(1): 31. Available at: <<https://doi.org/10.1097/YPG.0b013e32834acd07>>.
186. Hirbo JB, Pasutto F, Gamazon ER., et al. Analysis of Genetically Determined Gene Expression Suggests Role of Inflammatory Processes in Exfoliation Syndrome. *bioRxiv*, 2020. Available at: <<https://doi.org/10.1101/2020.12.17.423318>>.
187. Kuan PF, Yang X, Clouston S, et al. Cell type-specific gene expression patterns associated with posttraumatic stress disorder in World Trade Center responders. *Translational Psychiatry* 2019; 9(1): 1–11. Available at: <<https://doi.org/10.1038/s41398-018-0355-8>>.

MOKSLINIO DARBO APROBACIJA

Darbo rezultatai paskelbti Lietuvos ir užsienio moksliniuose leidiniuose, pristatyti nacionalinėse ir tarptautinėse mokslinėse konferencijose.

Disertacijos tema paskelbtų straipsnių sąrašas:

1. **Žukauskaitė** G, Domarkienė I, Matulevičienė A, Vaitėnienė EM, Arasimavičius J, Smailytė G, Kučinskas V, Ambrozaitytė L. Possible Protective Effect of LOXL1 Variant in the Cohort of Chernobyl Catastrophe Clean-Up Workers. *Genes* 2021; 12(8): 1231. DOI: 10.3390/genes12081231.
2. **Žukauskaitė** G, Domarkienė I, Matulevičienė A, Kučinskas V, Ambrozaitytė L. Identifying Genomic Signatures of Positive Selection to Predict Protective Genomic Loci in the Cohort of Lithuanian Clean-Up Workers of the Chornobyl Nuclear Disaster. *Current Issues in Molecular Biology* 2023; 45(4), 2972-2983. DOI: 10.3390/cimb45040195.

Disertacijos tema paruoštų ir pateiktų recenzuoti straipsnių sąrašas:

1. **Žukauskaitė** G, Domarkienė I, Rančelis T, Kavaliauskienė I, Baronas K, Kučinskas V, Ambrozaitytė L. Puatative protective genomic variation in the Lithuanian population. 2022.
2. Kazlauskas E, Smailyte G, Domarkienė I, Kučinskas V, Matulevičienė A, Elklit A, **Žukauskaitė** G, Ambrozaitytė L. Psychological distress 35 years after the Chornobyl accident in the Lithuanian clean-up workers. 2023.

Žodiniai pranešimai:

1. **Žukauskaitė** G (pranešėja), Ambrozaitytė L, Domarkienė I, Rančelis T, Kučinskas V. Protekcinį genomo variantų analizė lietuvių populiacijoje. In: Lietuvos mokslų akademijos Biologijos, medicinos ir geomokslų skyriaus 12-oji jaunųjų mokslininkų konferencija „Bioateitis: gamtos ir gyvybės mokslų perspektyvos“, 2019 m. gruodžio 11 d., Kaunas, Lietuva.
2. **Žukauskaitė** G (pranešėja), Domarkienė I, Matulevičienė A, Arasimavičius J, Kučinskas V, Ambrozaitytė L. Černobylio katastrofos likviduotojų genomo tyrimai: variantai, galimai lemiantys glaukomos prevenciją. In: Lietuvos mokslų akademijos Biologijos, medicinos ir geomokslų skyriaus 13-oji jaunųjų mokslininkų konferencija „Bioateitis: gamtos ir gyvybės mokslų perspektyvos“, 2020 m. gruodžio 20 d., virtualus renginys.
3. **Žukauskaitė** G (pranešėja), Domarkienė I, Matulevičienė A, Vaitėnienė EM, Arasimavičius J, Kučinskas V, Ambrozaitytė L. Genome association study of the Lithuanian Chernobyl catastrophe liquidators indicates

potentially protective alleles for multifactorial diseases. In: Devintoji tarptautinė radiacijos konferencija (angl. Conference on Radiation in Various Fields of Research, RAD2021), 2021 m. birželio 14–18 d., Herceg Novis, Juodkalnija.

4. **Žukauskaitė** G (pranešėja), Domarkienė I, Matulevičienė A, Arasimavičius J, Kučinskas V, Ambrozaitytė L. Psichikos sutrikimų plataus masto genetinės asociacijos tyrimas Lietuvos Černobylio katastrofos likviduotojų grupėje. In: Lietuvos mokslų akademijos Biologijos, medicinos ir geomokslų skyriaus 14-oji jaunųjų mokslininkų konferencija „Bioateitis: gamtos ir gyvybės mokslų perspektyvos“, 2021 m. lapkričio 25 d., Kaunas, Lietuva. Gautas III vietos apdovanojimas (Medicinos sekcija).
5. **Žukauskaitė** G (pranešėja), Domarkienė I, Kučinskas V, Ambrozaitytė L. Genomic Signatures of Adaptation in the Cohort of Chernobyl Catastrophe Clean-up Workers from Lithuania. In: Tarptautinė mokslinė konferencija „Population Genetics from the Baltic Perspective: Microevolutionary Processes of the Lithuanian Genomes“, 2022 m. gegužės 20–21 d., Vilnius, Lietuva.
6. Domarkienė I (pranešėja), **Žukauskaitė** G, Kučinskas V, Ambrozaitytė L. Inferring Whole-genome Sequencing Structural Variation Data in the Lithuanian Population. In: Tarptautinė mokslinė konferencija „Population Genetics from the Baltic Perspective: Microevolutionary Processes of the Lithuanian Genomes“, 2022 m. gegužės 20–21 d., Vilnius, Lietuva.
7. **Žukauskaitė** G (pranešėja), Domarkienė I, Matulevičienė A, Kučinskas V, Ambrozaitytė L. Teigiamos gamtinės atrankos veikiamų genomo sričių analizė Lietuvos Černobylio katastrofos likviduotojų grupėje. In: Lietuvos mokslų akademijos Biologijos, medicinos ir geomokslų skyriaus 15-oji jaunųjų mokslininkų konferencija „Bioateitis: gamtos ir gyvybės mokslų perspektyvos“, 2022 m. lapkričio 24 d., Vilnius, Lietuva. Gautas geriausio pranešimo I vietos apdovanojimas (Medicinos sekcija).

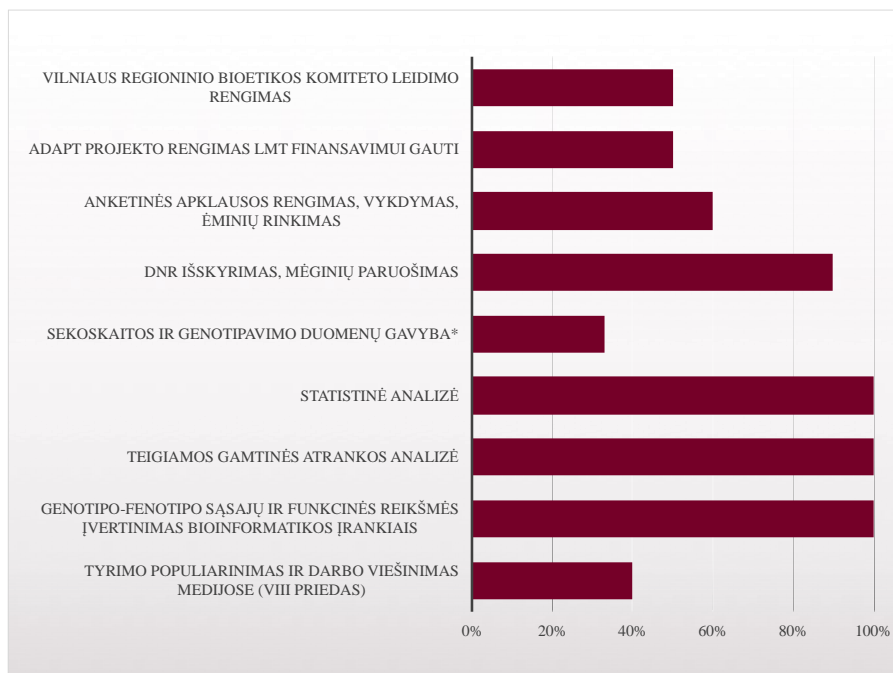
Elektroniniai pranešimai:

1. **Žukauskaitė** G, Ambrozaitytė L, Domarkienė I, Rančelis T, Kučinskas V. Frequency analysis of potentially protective genome variants in the Lithuanian population. In: Europos žmogaus genetikos draugijos konferencija (ESHG), 2020 m. birželio 6–9 d., virtualus renginys. (European Journal of Human Genetics 2020; 28(1).) Nr. P19.080.C.
2. Domarkienė I, **Žukauskaitė** G, Matulevičienė A, Arasimavičius J, Kučinskas V, Ambrozaitytė L. ADAPT – genomic initiative to analyse adaptation in the group of Lithuanian Chernobyl liquidators. In: Devintoji tarptautinė radiacijos konferencija (angl. Radiation in Various Fields of

- Research, RAD 2021), 2021 m. birželio 14–18 d., Herceg Novis, Juodkalnija.
3. **Žukauskaitė** G, Domarkienė I, Matulevičienė A, Vaitenienė EM, Arasimavičius J, Kučinskas V, Ambrozaitytė L. LOXL1 risk variant exhibits protective effect for exfoliation syndrome and glaucoma in Lithuanian Chernobyl catastrophe liquidators' cohort. In: Europos žmogaus genetikos draugijos konferencija (ESHG), 2021 m. rugpjūčio 28–31 d., virtualus renginys. Nr. P24.048.C.
 4. Matulevičienė A, **Žukauskaitė** G, Domarkienė I, Kučinskas V, Ambrozaitytė L. Association of Genomic Factors for Oral Health in the Cohort of the Lithuanian Chernobyl Catastrophe Liquidators. In: Europos žmogaus genetikos draugijos konferencija (ESHG), 2021 m. rugpjūčio 28–31 d., virtualus renginys. Nr. P13.003.D.
 5. Domarkienė I, **Žukauskaitė** G, Matulevičienė A, Urnikytė A, Kučinskas V, Ambrozaitytė L. CNV characteristics in Chernobyl power plant catastrophe clean-up workers from Lithuania suggest unique genetic variation structure. In: Amerikos žmogaus genetikos draugijos konferencija (ASHG), 2021 m. spalio 18–22 d., virtualus renginys.
 6. **Žukauskaitė** G, Domarkienė I, Matulevičienė A, Vaitenienė EM, Arasimavičius J, Kučinskas V, Ambrozaitytė L.. Detection of recent positive selection signatures in the cohort of Lithuanian Chernobyl catastrophe liquidators. In: Europos žmogaus genetikos draugijos konferencija (ESHG), 2022 m. birželio 11–14 d., Viena, Austrija. Nr. EP20.002.
 7. Domarkienė I, **Žukauskaitė** G, Urnikytė A, Pranckėnienė L, Dauengauer-Kirlienė S, Molytė A, Matulevičienė A, Pilypienė I, Ambrozaitytė L, Kučinskas V. Whole-genome sequencing data reveals higher number of structural variants in Chernobyl catastrophe clean-up workers from Lithuania. In: Europos žmogaus genetikos draugijos konferencija (ESHG), 2022 m. birželio 11–14 d., Viena, Austrija. Nr. P.14.010.A.
 8. Samaitė K, Domarkienė I, **Žukauskaitė** G, Kučinskas V, Ambozaitytė L. Possibly protective genomic variants against ionizing radiation in DNA homologous recombination repair genes. In: Europos žmogaus genetikos draugijos konferencija (ESHG), 2022 m. birželio 11–14 d., Viena, Austrija. Nr. EP25.007.
 9. **Žukauskaitė** G, Domarkienė I, Kučinskas V, Ambrozaitytė L. Genomic positive selection signature analysis in the cohort of the Lithuanian Chernobyl catastrophe clean-up workers. In: Jungtinė EAA-ISGA-ICEM konferencija, 2022 m. rugpjūčio 24–27 d., Vilnius, Lietuva. Nr. 128.

PRIEDAI

I priedas. Disertacijos autorės atliktų darbų indėlis



*Atlikta darbo laboratorijoje dalis (Lietuvos Černobylio katastrofos likviduotojų genotipavimas).

II priedas. Lietuvos Černobylio katastrofos likviduotojų apklausos anketa

Apsauginiai DNR mechanizmai – plataus masto Lietuvos Černobylio katastrofos likviduotojų viso genomo tyrimas (*ADAPT*)

Identifikacinis numeris:.....

BENDROJI INFORMACIJA

1. Anketos pildymo data: _____
Vieta: _____
2. Vardas, pavardė: _____
3. Lytis:
1 vyras
2 moteris
4. Gimimo data (metai, mėnuo, diena): _____
5. Ūgis (cm): Svoris (kg):
6. Tautybė:
1 lietuvis, -ė
2 rusas, -ė
3 lenkas, -ė
4 kita _____
7. Išsilavinimas:
1 pradinis
2 vidurinis
3 specialusis vidurinis
4 aukštasis
5 kita _____
8. Gyvenamoji vieta:
1 miestas
2 kaimas
3 Apskritis
4 Savivaldybė
5 Miestas ar kaimas _____
9. Šeimyninė padėtis:
1 vedęs
2 nevedęs
3 išsiskyręs
4 našlys

10. Užimtumas:

- 1 darbininkas,-ė
- 2 tarnautojas,-ė
- 3 pensininkas,-ė
- 4 namų šeimininkas; -ė
- 5 bedarbis, -ė
- 6 kita _____

11. Kontaktiniai duomenys:

el.paštas: _____ tel.nr. _____

INFORMACIJA APIE ŠEIMĄ**12. Užpildykite žemiau esančią lentelę apie kiekvieną šeimos (biologinį) narį:**

Ryšys	Gimimo metai	Sveikatos problemos/ sutrikimai	Amžius, kai mirė (jei mirė)	Mirties priežastis
Močiutė (mamos mama)				
Senelis (mamos tėvas)				
Močiutė (tėvo mama)				
Senelis (tėvo tėvas)				
Motina				
Tėvas				
Jūsų seserys				
1				
2				
3				
4				
Jūsų broliai				

1				
2				
3				
4				
Jūsų sūnūs				
1				
2				
3				
4				
Jūsų dukros				
1				
2				
3				
4				

13. Kiek Jums gimė vaikų po Černobylio katastrofos (įrašykite):

ČERNOBYLIO KATASTROFOS LIKVIDAVIMAS

14. Ar turite dokumentą/dokumento kopiją, patvirtinantį, kad dirbote 30 km zonoje:

1 taip

2 ne

15. Kiek kartų buvote komandiruotas į Černobylį (jei buvote komandiruotas daugiau nei vieną kartą, 21- 25 klausimuose pildykite atitinkamai tiek sekcijų, kiek buvote komandiruotas):

1 vieną

2 du

3 tris

4 daugiau (nurodykite)

16. Nurodykite savo atvykimo į Černobylį datą (metai, mėnuo, diena):

- 1 _____
- 2 _____
- 3 _____

17. Nurodykite savo išvykimo iš Černobylio datą (metai, mėnuo, diena):

- 1 _____
- 2 _____
- 3 _____

18. Ar turite dokumentą/dokumento kopiją, nurodanti komandiruotės į Černobylį trukmę:

- 1 taip
- 2 ne

19. Ar turite dokumentą/dokumento kopiją, patvirtinanti apšvitos dozę:

- 1 taip
- 2 ne

Jei pažymėjote „taip“, prašome pateikti šią informaciją:

1 Dokumento išdavimo data (metai, mėnuo, diena):

2 Laikotarpis, per kurį buvo gauta dozė (metai, mėnuo, diena):

- 1 nuo _____ iki _____
- 2 nuo _____ iki _____
- 3 nuo _____ iki _____

3 Dozė (įrašyti): _____

matavimo vienetai (įrašyti): _____

4 Ar tai suminė dozė:

- 1 taip
- 2 ne
- 3 nežinau

20. Ar dirbote pamaininiu būdu:

- 1 taip
- 2 ne

21. Prašome nurodyti, kur ir kurią laiko dalį (valandomis) dirbote 30 km zonoje:

<p>1. Atvirame lauke. 1 taip, trukmė _____ 2 ne</p> <p>2 pastatuose 1 taip, trukmė _____ 2 ne</p> <p>3 transporto priemonėse 1 taip, trukmė _____ 2 ne</p> <p>4 kita (įrašykite)</p>	<p>5 atvirame lauke 1 taip, trukmė _____ 2 ne</p> <p>6 pastatuose 1 taip, trukmė _____ 2 ne</p> <p>7 transporto priemonėse 1 taip, trukmė _____ 2 ne</p> <p>8 kita (įrašykite)</p>
<p>9 atvirame lauke 1 taip, trukmė _____ 2 ne</p> <p>10 pastatuose 1 taip, trukmė _____ 2 ne</p> <p>11 transporto priemonėse 1 taip, trukmė _____ 2 ne</p> <p>12 kita (įrašykite)</p>	<p>13 atvirame lauke 1 taip, trukmė _____ 2 ne</p> <p>14 pastatuose 1 taip, trukmė _____ 2 ne</p> <p>15 transporto priemonėse 1 taip, trukmė _____ 2 ne</p> <p>16 kita (įrašykite)</p>

22. Prašome nurodyti vietovę (-es), kur ir kiek valandų dirbote toje zonoje:

<p>1 Vietovė: _____</p> <p>2 Laikotarpis, kurį dirbote toje zonoje (metai, mėnuo, diena): nuo _____ iki _____</p> <p>3 Kiek valandų per dieną dirbote toje zonoje: _____</p>
<p>4 Vietovė: _____</p> <p>5 Laikotarpis, kurį dirbote toje zonoje (metai, mėnuo, diena): nuo _____ iki _____</p> <p>6 Kiek valandų per dieną dirbote toje zonoje: _____</p>

7 Vietovė: _____

8 Laikotarpis, kurį dirbote toje zonoje (metai, mėnuo, diena):

nuo _____ iki _____

9 Kiek valandų per dieną dirbote toje zonoje: _____

10 Vietovė: _____

11 Laikotarpis, kurį dirbote toje zonoje (metai, mėnuo, diena):

nuo _____ iki _____

12 Kiek valandų per dieną dirbote toje zonoje: _____

23. Prašome nurodyti vietovę (-es), kur ir kiek valandų gyvenote toje zonoje:

1 Vietovė: _____

2 Laikotarpis, kurį dirbote toje zonoje (metai, mėnuo, diena):

nuo _____ iki _____

3 Kiek valandų per dieną būdavote toje zonoje:

4 Kokiam pastate gyvenote:

- 1 palapinėje
- 2 mediniame pastate
- 3 mūriniame pastate
- 4 kita (įrašykite)

5 Vietovė: _____

6 Laikotarpis, kurį dirbote toje zonoje (metai, mėnuo, diena):

nuo _____ iki _____

7 Kiek valandų per dieną būdavote toje zonoje:

8 Kokiame pastate gyvenote:

- 1 palapinėje
- 2 mediniame pastate
- 3 mūriniame pastate
- 4 kita (įrašykite)

9 Vietovė: _____

10 Laikotarpis, kurį dirbote toje zonoje (metai, mėnuo, diena):

nuo _____ iki _____

11 Kiek valandų per dieną būdavote toje zonoje:

12 Kokiame pastate gyvenote:

- 1 palapinėje
- 2 mediniame pastate
- 3 mūriniame pastate
- 4 kita (įrašykite)

24. Ar apšvitos dozė buvo matuojama:

1 taip 2 ne 3 nežinau	4 taip 5 ne 6 nežinau
7 taip 8 ne 9 nežinau	10 taip 11 ne 12 nežinau

25. Koku metodu buvo matuojama apšvitos dozė:

<p>1 asmeniniu dozimetru</p> <p>1 taip</p> <p>2 ne</p> <p>2 grupinės dozimetrijos metodu</p> <p>1 taip</p> <p>2 ne</p> <p>3 pagal maršrutą</p> <p>1 taip</p> <p>2 ne</p> <p>4 kita (įrašykite): _____</p>	<p>5 asmeniniu dozimetru</p> <p>1 taip</p> <p>2 ne</p> <p>6 grupinės dozimetrijos metodu</p> <p>1 taip</p> <p>2 ne</p> <p>7 pagal maršrutą</p> <p>1 taip</p> <p>2 ne</p> <p>8 kita (įrašykite): _____</p>
<p>9 asmeniniu dozimetru</p> <p>1 taip</p> <p>2 ne</p> <p>10 grupinės dozimetrijos metodu</p> <p>1 taip</p> <p>2 ne</p> <p>11 pagal maršrutą</p> <p>1 taip</p> <p>2 ne</p> <p>12 kita (įrašykite): _____</p>	<p>13 asmeniniu dozimetru</p> <p>1 taip</p> <p>2 ne</p> <p>14 grupinės dozimetrijos metodu</p> <p>1 taip</p> <p>2 ne</p> <p>15 pagal maršrutą</p> <p>1 taip</p> <p>2 ne</p> <p>16 kita (įrašykite): _____</p>

26. Ar teko dirbti šiuos darbus:

1	Dalyvavimas sarkofago statyboje prie pažeisto reaktoriaus	taip	ne
2	Grafito gabalų ir kitų labai užterštų radioaktyviomis medžiagomis fragmentų šalinimas nuo elektrinės pastatų ir aikštelių šalia ventiliacijos angų	taip	ne
3	Elektrinės patalpų ir įrenginių jų viduje deaktyvavimas	taip	ne
4	Elektrinės teritorijos ir įrenginių, esančių ne pastatų viduje, deaktyvavimas	taip	ne
5	Radiacinė žvalgyba	taip	ne
6	Transporto priemonių deaktyvavimas specialiuose punktuose (PUSO)	taip	ne
7	Elektrinės įrenginių remontas ir aptarnavimas	taip	ne

8	Kiti darbai elektrinės teritorijoje (įrašykite): _____	taip	ne
9	Radioaktyvių atliekų laidojimas ir deaktyvavimo darbai ne elektrinės teritorijoje	taip	ne
10	Kelių tiesimas 30 km zonoje	taip	ne
11	Transporto priemonės vairavimas	taip	ne
12	Elektrinės objektų arba 30 km zonos saugojimas	taip	ne
13	Kiti darbai ne elektrinės teritorijoje (įrašykite): _____	taip	ne

27. Ar naudojotės apsaugos priemonėmis darbo metu 30 km zonoje:

1	Respiratorius, medicininė kaukė	taip	ne
2	Dujokaukė	taip	ne
3	Pirštinės	taip	ne
4	Apsauginiai akiniai	taip	ne
5	Apsauginiai rūbai (chalatas, kombinezonas)	taip	ne
6	Švininė prijuostė	taip	ne
7	Jodo preparatai	taip	ne
8	Maudymasis duše po darbo	taip	ne
9	Speciali transporto priemonių apsauga (švininiais skydai)	taip	ne
10	Kita (įrašykite): _____	taip	ne

28. Ar teko dirbti elektrinės teritorijoje:

1 taip

2 ne

29. Ar naudojotės apsaugos priemonėmis ne darbo metu:

1	Respiratorius, medicininė kaukė	taip	ne
2	Pirštinės	taip	ne
3	Apsauginiai akiniai	taip	ne
4	Apsauginiai rūbai (chaldas, kombinezonas)	taip	ne
5	Švininė prijuostė	taip	ne
6	Speciali transporto priemonių apsauga (švininiais skydai)	aip	ne
7	Kita (įrašykite): _____	taip	ne

30. Nurodykite priežastis, dėl kurių išvykote iš Černobylio:

- 1 gauta apšvitos dozė buvo didesnė už leistiną
- 2 gauta apšvitos dozė pasiekė leistiną
- 3 baigėsi komandiruotė
- 4 dėl ligos (įrašykite)
- 5 kita (įrašykite)

31. Jūsų asmeninis požiūris į radiaciją – apšvitos dozę, matavimo metodus ir kita:

JŪSŪ MITYBOS IPROČIAI

32. Kiek kartų per savaitę valgote (valgėte) šviežias daržoves:

Iki darbo Černobylyje	Darbo Černobylyje metu		Dabar
1 nė karto	Atvežtas	Vietines	13 nė karto
2 1-2 kartus	5 nė karto	9 nė karto	14 1-2 kartus
3 3-5 kartus	6 1-2 kartus	10 1-2 kartus	15 3-5 kartus
4 6-7 kartus	7 3-5 kartus	11 3-5 kartus	16 6-7 kartus
	8 6-7 kartus	12 6-7 kartus	

33. Kiek kartų per savaitę valgote (valgėte) virtas ar troškintas daržoves (išskyrus bulves):

Iki darbo Černobylyje	Darbo Černobylyje metu		Dabar
1 nė karto	Atvežtas	Vietines	13 nė karto
2 1-2 kartus	5 nė karto	9 nė karto	14 1-2 kartus
3 3-5 kartus	6 1-2 kartus	10 1-2 kartus	15 3-5 kartus
4 6-7 kartus	7 3-5 kartus	11 3-5 kartus	16 6-7 kartus
	8 6-7 kartus	12 6-7 kartus	

34. Kiek kartų per savaitę valgote (valgėte) šviežius vaisius, uogas:

Iki darbo Černobylyje	Darbo Černobylyje metu		Dabar
1 nė karto	Atvežtus	Vietinius	13 nė karto
2 1-2 kartus	5 nė karto	9 nė karto	14 1-2 kartus
3 3-5 kartus	6 1-2 kartus	10 1-2 kartus	15 3-5 kartus
4 6-7 kartus	7 3-5 kartus	11 3-5 kartus	16 6-7 kartus
	8 6-7 kartus	12 6-7 kartus	

35. Kiek kartų per savaitę valgote (valgėte) virtus ar konservuotus vaisius, uogas:

Iki darbo Černobylyje	Darbo Černobylyje metu	Dabar

1 n� kartu	Atve�tus	Vietinius	13 n� kartu
2 1-2 kartus	5 n� kartu	9 n� kartu	14 1-2 kartus
3 3-5 kartus	6 1-2 kartus	10 1-2 kartus	15 3-5 kartus
4 6-7 kartus	7 3-5 kartus	11 3-5 kartus	16 6-7 kartus
	8 6-7 kartus	12 6-7 kartus	

36. Kaip da nai vartojate (vartojote) gr dinius produktus (duona, gr dų ko es ir kt.):

Iki darbo �ernobylyje	Darbo �ernobylyje metu		Dabar
1 nevartojau	Atve�tus	Vietinius	16 nevartuju
2 1-2 k./sav.	6 nevartojau	11 nevartojau	17 1-2 k./sav.
3 3-5 k./sav.	7 1-2 k./sav.	12 1-2 k./sav.	18 3-5 k./sav.
4 kasdien�	8 3-5 k./sav.	13 3-5 k./sav.	19 kasdien�
5kelis kartus per dieną	9 kasdien�	14 kasdien�	20kelis kartus per dieną
	10 kelis kartus per dieną	15kelis kartus per dieną	

37. Kaip da nai vartojate (vartojote)  uvį ir  uvies produktus:

Iki darbo �ernobylyje	Darbo �ernobylyje metu		Dabar
1 nevartojau	Atve�tus	Vietinius	13 nevartuju
2 1-2 k./sav.	5 nevartojau	9 nevartojau	14 1-2 k./sav.
3 3-5 k./sav.	6 1-2 k./sav.	10 1-2 k./sav.	15 3-5 k./sav.
4 kasdien�	7 3-5 k./sav.	11 3-5 k./sav.	16 kasdien�
	8 kasdien�	12 kasdien�	

38. Kaip da nai vartojate (vartojote) m s  ir m sos produktus:

Iki darbo �ernobylyje	Darbo �ernobylyje metu		Dabar
1 nevartojau	Atve�tus	Vietinius	13 nevartuju
2 1-2 k./sav.	5 nevartojau	9 nevartojau	14 1-2 k./sav.
3 3-5 k./sav.	6 1-2 k./sav.	10 1-2 k./sav.	15 3-5 k./sav.
4 kasdien�	7 3-5 k./sav.	11 3-5 k./sav.	16 kasdien�
	8 kasdien�	12 kasdien�	

39. Kaip dažnai vartojate (vartojote) kiaušinius kaip atskirą patiekalą:

Iki darbo Černobylyje	Darbo Černobylyje metu		Dabar
1 nevartojau 2 1-2 k./sav. 3 3-5 k./sav. 4 kasdieną	Atvežtus 5 nevartojau 6 1-2 k./sav. 7 3-5 k./sav. 8 kasdieną	Vietinius 9 nevartojau 10 1-2 k./sav. 11 3-5 k./sav. 12 kasdieną	13 nevartojau 14 1-2 k./sav. 15 3-5 k./sav. 16 kasdieną

40. Kaip dažnai vartojate (vartojote) pieną ir pieno produktus:

Iki darbo Černobylyje	Darbo Černobylyje metu		Dabar
1 nevartojau 2 1-2 k./sav. 3 3-5 k./sav. 4 kasdieną	Atvežtus 5 nevartojau 6 1-2 k./sav. 7 3-5 k./sav. 8 kasdieną	Vietinius 9 nevartojau 10 1-2 k./sav. 11 3-5 k./sav. 12 kasdieną	13 nevartojau 14 1-2 k./sav. 15 3-5 k./sav. 16 kasdieną

41. Kokius riebalus dažniausiai vartojate (vartojote) maistui gaminti (kepti, virti, troškinti):

Iki darbo Černobylyje	Darbo Černobylyje metu		Dabar
1 sviestą 2 margariną 3 augalinį aliejų 4 gyvulinius taukus	<u>Atvežtus</u> 5 sviestą 6 margariną 7 augalinį aliejų 8 gyvulinius taukus	<u>Vietinius</u> 9 sviestą 10 margariną 11 augalinį aliejų 12 gyvulinius taukus	13 sviestą 14 margariną 15 augalinį aliejų 16 gyvulinius Taukus

42. Ar geriate (gėrėte) kavą:

Iki darbo Černobylyje	Darbo Černobylyje metu	Dabar
1 taip 2 ne	3 taip 4 ne	5 taip 6 ne

43. Ar saldinate (saldinote) kavą:

Iki darbo Černobylyje	Darbo Černobylyje metu	Dabar
1 taip 2 ne	3 taip 4 ne	5 taip 6 ne

44. Kokią geriate (gėrėte) kavą:

Iki darbo Černobylyje	Darbo Černobylyje metu	Dabar
1 plikytą 2 filtruotą 3 tirpią 4 be kofeino	5 plikytą 6 filtruotą 7 tirpią 8 be kofeino	9 plikytą 10 filtruotą 11 tirpią 12 be kofeino

45. Kiek per dieną kavos puodelių geriate (gėrėte):

Iki darbo Černobylyje	Darbo Černobylyje metu	Dabar
1 mažiau nei vieną puodelį 2 Vieną 3 du 4 tris 5 keturis 6 penkis ir daugiau	7 mažiau nei vieną puodelį 8 Vieną 9 du 10 tris 11 keturis 12 penkis ir daugiau	13 mažiau nei vieną puodelį 14 Vieną 15 du 16 tris 17 keturis 18 penkis ir daugiau

46. Ar geriate (gėrėte) žalią arbatą:

Iki darbo Černobylyje	Darbo Černobylyje metu	Dabar
1 taip 2 ne	3 taip 4 ne	5 taip 6 ne

47. Ar saldinate (saldinote) arbatą:

Iki darbo Černobylyje	Darbo Černobylyje metu	Dabar
1 taip 2 ne	3 taip 4 ne	5 taip 6 ne

48. Kiek per dieną arbatos puodelių geriate (gėrėte):

Iki darbo Černobylyje	Darbo Černobylyje metu	Dabar
1 mažiau nei vieną puodelį	7 mažiau nei vieną puodelį	13 mažiau nei vieną puodelį
2 vieną	8 vieną	14 vieną
3 du	9 du	15 du
4 tris	10 tris	16 tris
5 keturis	11 keturis	17 keturis
6 penkis ir daugiau	12 penkis ir daugiau	18 penkis ir daugiau

49. Ar valgote (valgėte) visada tuo pačiu laiku:

Iki darbo Černobylyje	Darbo Černobylyje metu	Dabar
1 taip	3 taip	5 taip
2 ne	4 ne	6 ne

50. Kiek kartų valgote (valgėte) per dieną:

Iki darbo Černobylyje	Darbo Černobylyje metu	Dabar
1 vieną	5 vieną	9 vieną
2 du	6 du	10 du
3 tris	7 tris	11 tris
4 keturis ir daugiau	8 keturis ir daugiau	12 keturis ir daugiau

51. Ar valgote (valgėte) papildomai tarp pagrindinių valgymų:

Iki darbo Černobylyje	Darbo Černobylyje metu	Dabar
1 niekada	5 niekada	9 niekada
2 kai kada	6 kai kada	10 kai kada
3 dažnai	7 dažnai	11 dažnai
4 visada	8 visada	12 visada

FIZINĖ VEIKLA IR JUDEJIMAS

52. Žemiau pateiktose grupėse aprašyti pavyzdžiai, kuriuose kalbama apie dabartinį Jūsų darbą arba tą darbą, kurį Jūs dirbote prieš gaudami nedarbingumo lapelį arba išeidami į pensiją. Pažymėkite tik 1 atsakymą.

1 grupė • Darbas yra iš esmės sėdimas, ir darbo metu Jūs daug nevaikštote. Pvz.: darbas prie rašomojo stalo, paprastų detalių surinkimas ir panašiai.

2 grupė • Dirbant Jūsų darbą reikia gana daug vaikščioti, bet nereikia nešioti arba kilnoti sunkių daiktų. Pvz.: klientų aptarnavimas, lengvas pramoninis darbas, pardavėjo darbas, dėstyimas, kai daug vaikštoma.

3 grupė • Jūs daug vaikštote, o be to, kilnojate daiktus arba tenka daug laiptoti aukštyn ir žemyn. Pvz.: paštininko darbas, sunkesnės pramonės darbai, statybos, sunkesnis darbas restorane.

4 grupė • Jūs dirbate sunkų fizinį darbą, kilnojate sunkius daiktus ir dirbate fiziškai intensyviai. Pvz.: slaugos darbai, darbai, reikalaujantys kilnoti sunkius objektus ir atlikinėti sunkius tvarkymo darbus.

53. Žemiau pateiktose grupėse aprašyti pavyzdžiai, kuriuose kalbama apie tai, kiek Jūs pastaruosius kelerius metus judate laisvalaikio metu. Jeigu šis rodiklis skiriasi žiemą ir vasarą, pamėginkite nustatyti vidurkį. Pažymėkite tik 1 atsakymą.

1 grupė • Laisvalaikį Jūs daugiausiai skiriate skaitymui, televizoriaus žiūrėjimui, kino teatro lankymui ir kitai fiziškai neaktyviai veiklai.

2 grupė • Mažiausiai 4 kartus per savaitę Jūs einate pasivaikščioti, važinėjate dviračiu ar kitaip judate. Čia įeina ir ėjimas arba važiavimas dviračiu į darbą bei atgal, sekmadieniniai pasivaikščiojimai, nesunkus darbas sode, žvejyba, stalo tenisas, žaidimas kėgliais.

3 grupė • Laisvalaikį skiriate bėgiojimui, plaukimui, lauko tenisui, badmintonui, aerobikai ir kitoms panašioms judrioms sporto šakoms. Šiai grupei priklauso sunkesni sodo darbai. Tokia veikla turi būti užsiiminėjama bent 3 kartus per savaitę.

4 grupė • Laisvalaikį skiriate sunkioms treniruotėms bei dalyvavimui reguliariai vykstančioms bėgimo, orientavimosi, slidinėjimo, plaukimo, futbolo, rankinio varžybose porą kartų per savaitę.

ŽALINGI VEIKSNIAI

54. Ar vartojate (vartojote) alkoholinius gėrimus:

Iki darbo Černobylyje	Darbo Černobylyje metu	Dabar
1 taip, dažnai 2 taip, retkarčiais 3 ne	4 taip, dažnai 5 taip, retkarčiais 6 ne	7 taip, dažnai 8 taip, retkarčiais 9 ne

55. Jei vartojate (vartojote) alkoholinius gėrimus, kokius ir kiek mililitrų per dieną:

Iki darbo Černobylyje	Gėrimas		Kiekis (ml)
	alus	1 taip, nurodykite kiekį: 2 ne	
	vynas, šampanas	3 taip, nurodykite kiekį: 4 ne	
	likeriai	5 taip, nurodykite kiekį: 6 ne	
	degtinė	7 taip, nurodykite kiekį: 8 ne	
	kita (įrašykite) _____	9 taip, nurodykite kiekį: 10 ne	
	Darbo Černobylyje metu	alus	11 taip, nurodykite kiekį: 12 ne
vynas, šampanas		13 taip, nurodykite kiekį: 14 ne	
likeriai		15 taip, nurodykite kiekį: 16 ne	
degtinė		17 taip, nurodykite kiekį: 18 ne	

	vietinės gamybos alkoholiniai gėrimai (vynas, degtinė ar kt.)	19 taip, nurodykite kiekį: 20 ne	
	kita (įrašykite) _____	21 taip, nurodykite kiekį: 22 ne	
Dabar	alus	23 taip, nurodykite kiekį: 24 ne	
	vynas, šampanas	25 taip, nurodykite kiekį: 26 ne	
	likeriai	27 taip, nurodykite kiekį: 28 ne	
	degtinė	29 taip, nurodykite kiekį: 30 ne	
	kita (įrašykite) _____	31 taip, nurodykite kiekį: 32 ne	

56. Kaip dažnai vartojate alkoholinius gėrimus šiuo metu:

- 1 niekada
- 2 labai retai
- 3 kartą per mėnesį ir rečiau
- 4 2-3 kartus per mėnesį
- 5 kartą per savaitę
- 6 keletą kartų per savaitę
- 7 kasdien

57. Ar pasikeitė alkoholinių gėrimų vartojimas po Jūsų grįžimo iš Černobylio:

- 1 vartoju alkoholinius gėrimus dažniau
- 2 vartoju alkoholinius gėrimus rečiau
- 3 iš viso nebevartoju alkoholinių gėrimų
- 4 pradėjau vartoti alkoholinius gėrimus tik po grįžimo iš Černobylio
- 5 alkoholinių gėrimų vartojimas nepasikeitė

58. Ar esate vartojęs narkotinių medžiagų:

- 1 taip (įrašykite) _____
 2 ne

59. Ar rūkote (rūkėte):

Iki darbo Černobylyje	Darbo Černobylyje metu	Dabar
1 taip, kasdien 2 taip, kartais 3 ne	4 taip, kasdien 5 taip, kartais 6 ne	7 taip, kasdien 8 taip, kartais 9 ne

60. Jei rūkote (rūkėte), kiek cigarečių per dieną (įrašykite):

Iki darbo Černobylyje	Darbo Černobylyje metu	Dabar

61. Jei vartojate, kokį tabaką renkatės:

- 1 nevartuju
 2 rūkomą
 3 kramtomą
 4 uostomą

62. Kiek metų rūkėte (įrašykite, jei rūkėte): _____**63. Jei Jums kada nors teko dirbti su šiomis medžiagomis:**

	Medžiaga	Darbo trukmė (metais)		
1	benzinas		taip	ne
2	organiniai tirpikliai		taip	ne
3	pesticidai		taip	ne
4	mineralinės trąšos		taip	ne
5	asbestas		taip	ne
6	amoniakas		taip	ne
7	suvirinimo aerozoliai		taip	ne

	Medžiaga	Darbo trukmė (metais)		
8	dervos		taip	ne
9	klijai		taip	ne
10	cementas		taip	ne
11	asfaltas		taip	ne
12	bitumas		taip	ne
13	nafta		taip	ne
14	žibalas		taip	ne
15	organiniai dažai		taip	ne
16	pigmentai		taip	ne
17	Kita (įrašykite): _____		taip	ne

64. Ar Jums teko kada nors dirbti darbą susijusį su radiacija *(pavyzdžiui, medicina, atominė energetika, pramoninė radiografija, tarnyba armijoje su radiacijos šaltiniais ar kt.)* **be darbo Černobylyje:**

- 1 taip (įrašykite) _____
- 2 ne
- 3 nežinau

65. Ar Jums teko kada nors dirbti su sveikatai kenksmingomis cheminėmis medžiagomis be darbo Černobylyje *(pavyzdžiui, naftos perdirbimo gamykloje, chemijos gamykloje, gumos fabrike, žemės ūkyje, karinėje tarnyboje ar kt.):*

- 1 taip (įrašykite) _____
- 2 ne
- 3 nežinau

66. Perskaitykite kiekvieną teiginį ir pažymėkite atsakymą, kuris labiausiai atspindi Jūsų savijautą per praėjusią savaitę. Ilgai nesvarstykite, nes pirmą reakciją į kiekvieną klausimą, tiksliau atspindi Jūsų savijautą nei ilgai apgalvotas atsakymas.

<p>Aš jaučiu įtampą ir nerimą 1 Beveik visą laiką, pastoviai 2 Didelę laiko dalį, dažnai 3 Laikas nuo laiko, retkarčiais 4 Niekada nejaučiu</p>	<p>Man linksma 5 Niekada 6 Labai retai 7 Kartais 8 Didžiąją laiko dalį</p>	<p>Jaučiu, kad nerimstu vietoje 9 Labai stipriai 10 Gana stipriai 11 Truputį 12 Visiškai ne</p>
<p>Mane ir dabar džiugina tai, kas teikė džiaugsmą anksčiau 13 Visiškai tiek pat 14 Mažiau nei anksčiau 15 Žymiai mažiau 16 Beveik visai nedžiugina</p>	<p>Aš galiu ramiai sėdėti ir atsipalaiduoti 17 Visada 18 Dažnai 19 Retai 20 Niekada</p>	<p>Iš gyvenimo aš laikiu kažko malonaus 21 Tiek pat kiek visada 22 Mažiau nei anksčiau 23 Žymiai mažiau nei anksčiau 24 Visiškai nelaukiu</p>
<p>Aš jaučiu baimę, lyg kažkas siaubingo turėtų atsitikti 25 Labai aiškiai ir stipriai 26 Taip, bet nestipriai 27 Nežymiai, bet tai manęs nejaudina 28 Visiškai ne</p>	<p>Aš jaučiuosi užslopintas ir sulėtėjęs 29 Beveik visą laiką 30 Labai dažnai 31 Kartais, nežymiai 32 Visiškai ne</p>	<p>Mane staiga apima didelis nerimas ir baimė 33 Tikrai labai dažnai 34 Pakankamai dažnai 35 Retai 36 Visiškai ne</p>
<p>Aš galiu juoktis ir suprasti humorą 37 Taip kaip ir anksčiau 38 Mažiau ir sunkiau 39 Žymiai mažiau ir sunkiau 40 Visiškai negaliu</p>	<p>Mane apima baimė lydima vidinio virpulių ir spaudimo po krūtine 41 Visiškai ne 42 Kartais 43 Gana dažnai 44 Labai dažnai, pastoviai</p>	<p>Man suteikia džiaugsmo gera knyga, radijo ar TV laida 45 Dažnai 46 Kartais 47 Retai 48 Labai retai</p>

<p>Mane vargina neramios mintys ir rūpesčiai 49 Didžiąją laiko dalį, pastoviai 50 Daug laiko, dažnai 51 Laikas nuo laiko, bet ne dažnai 52 Tik retkarčiais</p>	<p>Aš nustojau rūpintis savo išvaizda 53 Pradėjau visiškai Nesirūpinti 54 Nesirūpinu tiek, kiek reikėtų 55 Rūpinuosi, bet mažiau nei anksčiau 56 Rūpinuosi tiek pat, kiek visuomet</p>	
---	---	--

KLINIKINIAI DUOMENYS pagal TLK-10-AM klasifikaciją
(pagrindinė liga bei lydinčios ligos)

67-83 klausimus pildo gydytojas - genetikas ar medicinos gydytojas

67. Navikai (pagal TLK-10-AM kodą C00-D48; diagnozės nustatymo data):

Kodas							
-------	--	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

68. Kraujo ir kraujodaros organų ligos bei tam tikri sutrikimai, susiję su imuniniais mechanizmais (pagal TLK-10-AM kodą D50-D89)

Kodas							
-------	--	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

69. Endokrininės, mitybos ir medžiagų apykaitos ligos (pagal TLK-10-AM kodą E00-E89)

Kodas							
-------	--	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

70. Psichikos ir elgesio sutrikimai (pagal TLK-10-AM kodą F00-F99)

Kodas							
-------	--	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

71. Nervų sistemos ligos (pagal TLK-10-AM kodą G00-G99)

Kodas						
-------	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

72. Akies ir jos priedinių organų ligos (pagal TLK-10-AM kodą H00-H59)

Kodas						
-------	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

73. Ausies ir speninės ataugos ligos (pagal TLK-10-AM kodą H60-H95)

Kodas						
-------	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

74. Kraujotakos sistemos ligos (pagal TLK-10-AM kodą I00-I99)

Kodas						
-------	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

75. Kvėpavimo sistemos ligos (pagal TLK-10-AM kodą J00-J99)

Kodas						
-------	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

76. Virškinimo sistemos ligos (pagal TLK-10-AM kodą K00-K93)

Kodas						
-------	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

77. Odos ir poodžio ligos (pagal TLK-10-AM kodą L00-L99)

Kodas						
-------	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

78. Jungiamojo audinio ir raumenų bei skeleto ligos (pagal TLK-10-AM kodą M00-M99)

Kodas							
-------	--	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

79. Lyties ir šlapimo sistemos ligos (pagal TLK-10-AM kodą N00-N99)

Kodas							
-------	--	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

80. Simptomai, požymiai ir nenormalūs klinikiniai bei laboratoriniai radiniai, neklasifikuojami kitur (pagal TLK-10-AM kodą R00-R99)

Kodas							
-------	--	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

81. Sužalojimai, apsinuodijimai ir tam tikri išorinių poveikių padariniai

(pagal TLK-10-AM kodą S00-T98)

Kodas							
-------	--	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

82. Išorinės sergamumo ir mirtingumo priežastys (pagal TLK-10-AM kodą U50-U73, U90, V00-Y98)

Kodas							
-------	--	--	--	--	--	--	--

_____ m. _____ mėn. _____ d.

83. Papildoma klinikinė informacija:

84. Ar Jums kada nors buvo taikytas žemiau išvardintas gydymas dėl onkologinės ligos:

1 taip (pažymėkite tinkamą) 2 ne

1 spindulinė terapija

2 chemoterapija

- 3 operacinis gydymas
- 4 kita _____

85. Ar Jums kada nors buvo atlikti žemiau išvardinti radiologiniai instrumentiniai tyrimai ir kiek kartų:

1 danties rentgenograma

- 1 taip kiek kartų (įrašykite): _____
- 2 ne
- 3 nežinau

2 panoraminė dantų rentgenograma

- 1 taip kiek kartų (įrašykite): _____
- 2 ne
- 3 nežinau

3 krūtinės ląstos rentgenograma

- 1 taip kiek kartų (įrašykite): _____
- 2 ne
- 3 nežinau

4 kaulų ar sąnarių rentgenograma

- 1 taip kiek kartų (įrašykite): _____
- 2 ne
- 3 nežinau

5 širdies arba kraujagyslių radiologiniai tyrimai

- 1 taip kiek kartų (įrašykite): _____
- 2 ne
- 3 nežinau

6 skrandžio radiologiniai tyrimai su/be kontrastine medžiaga

- 1 taip kiek kartų (įrašykite): _____
- 2 ne
- 3 nežinau

7 žarnyno radiologiniai tyrimai su/be kontrastine medžiaga

- 1 taip kiek kartų (įrašykite): _____
- 2 ne
- 3 nežinau

8 kaukolės rentgenograma

1 taip kiek kartų (įrašykite): _____

2 ne

3 nežinau

9 galvos smegenų kompiuterinė tomografija

1 taip kiek kartų (įrašykite): _____

2 ne

3 nežinau

10 skydliaukės tyrimas su radioaktyviu jodu

1 taip kiek kartų (įrašykite): _____

2 ne

3 nežinau

11 kitas aukščiau nepaminėtas radiologinis tyrimas (įrašykite):

86. Ar vartojate maisto papildus, įrašykite (vitaminų, mineralinių medžiagų preparatus ar kt.):

1 beveik kasdien _____

2 niekada

3 kartais _____

4 1-2 mėnesius per metus _____

5 3-4 mėnesius per metus _____

6 nežinau

87. Kokius vaistinius preparatus vartojate? Įrašykite vaistinio preparato pavadinimą arba kokiam sveikatos sutrikimui jis yra skirtas.

88. Ar esate alergiškas:

1 taip (įrašykite) _____ 2 ne

89. Ar esate alergiškas kokiems nors vaistiniams preparatams:

1 taip (įrašykite) _____ 2 ne

DĖKOJAME JUMS UŽ SKIRTĄ LAIKĄ!

III priedas. Poveikio genomo variantų katalogas. Poveikio variantai buvo išrinkti remiantis skirtingais moksliniais šaltiniais.

1 šaltinis – Butler JM, et al. Identification of candidate protective variants for common diseases and evaluation of their protective potential. BMC Genomics 2017; 18(1): 575.

2 šaltinis – Harper AR, et al. Protective alleles and modifier variants in human health and disease. Nature Reviews Genetics 2015; 16(12): 689–701.

3 šaltinis – *ClinVar* ir (ar) *OMIM* duomenų bazės (protekcinę funkciją atliekantys genomo variantai).

Rs – identifikacinis genomo varianto kodas pagal *dbSNP* duomenų bazę, REF – referentinis alelis, ALT – alternatyvus alelis.

Šaltinis	Genas	rs (dbSNP)	REF	ALT	Siejama liga ar požymis
1.	<i>IL7R</i>	rs6897932	C	T	Išsėtinė sklerozė
	<i>COL11A1</i>	rs3753841	G	A	Glaukoma
	<i>SLC30A8</i>	rs13266634	C	T	II tipo diabetas
	<i>TYK2</i>	rs12720356	A	C	Psoriazė
	<i>KCNE4</i>	rs12621643	T	G	Ūmi limfoblastinė leukemija
	<i>IL17REL</i>	rs5771069	G	A	Opinis kolitas
	<i>PLCE1</i>	rs3765524	T	C	Stemplės ir skrandžio vėžys
	<i>TNFSF13</i>	rs3803800	A	G	IgA nefropatija
	<i>WDFY4</i>	rs7097397	G	A	Sisteminė raudonoji vilkligė
	<i>PLCE1</i>	rs2274223	G	A	Stemplės vėžys
	<i>CFB</i>	rs641153	G	A	Su amžiumi susijusi geltonosios dėmės degeneracija
	<i>CD6</i>	rs11230563	C	T	Uždegiminė žarnyno liga
	<i>KLRC4</i>	rs2617170	T	C	Behceto liga
	<i>ZC3HC1</i>	rs11556924	C	T	Koronarinė širdies liga
<i>TYK2</i>	rs2304256	C	A	I tipo diabetas	

Šaltinis	Genas	rs (dbSNP)	REF	ALT	Siejama liga ar požymis
	<i>GIPR</i>	rs1800437	G	C	Nutukimas
	<i>C2</i>	rs9332739	G	C	Su amžiumi susijusi geltonosios dėmės degeneracija
	<i>PPARG</i>	rs1801282	C	G	II tipo diabetas
	<i>TUBD1</i>	rs1292053	G	A	Uždegiminė žarnyno liga
	<i>MMP17</i>	rs6598163	G	A	Migrena
	<i>LOXL1</i>	rs3825942	G	A	Glaukoma, eksfoliacijos sindromas
	<i>IL23R</i>	rs11209026	G	A	Uždegiminė žarnyno liga
	<i>MPV17L2</i>	rs874628	A	G	Išsėtinė sklerozė
	<i>RTKN2</i>	rs3125734	T	C	Reumatinis artritas
	<i>LRRC34</i>	rs6793295	C	T	Intersticinė plaučių liga
	<i>SH2B1</i>	rs7498665	G	A	Nutukimas
	<i>ANKLE1</i>	rs8100241	G	A	Krūties vėžys
	<i>BANK1</i>	rs10516487	G	A	Sisteminė raudonoji vilkligė
	<i>ANKLE1</i>	rs2363956	T	G	Kiaušidžių vėžys
	<i>NFKBIE</i>	rs2233434	G	A	Reumatinis artritas
	<i>LPL</i>	rs328	C	G	Hipertenzija
	<i>WFS1</i>	rs1801214	T	C	II tipo diabetas
	<i>THADA</i>	rs7578597	T	C	II tipo diabetas
	<i>IL13</i>	rs20541	G	A	Psoriazė
	<i>MLPH</i>	rs2292884	G	A	Prostatos vėžys
2.	<i>PCSK9</i>	rs11591147	G	T	Hipercholesterolemija
	<i>PCSK9</i>	rs67608943	C	G	Hipercholesterolemija

Šaltinis	Genas	rs (dbSNP)	REF	ALT	Siejama liga ar požymis
	<i>PCSK9</i>	rs72646508	C	T	Hipercholesterolemija
	<i>PCSK9</i>	rs28362263	G	A	Hipercholesterolemija
	<i>SCN9A</i>	rs202152511	C	T	Paveldimi skausmo sutrikimai
	<i>SCN11A</i>	rs483352920	T	C	Paveldimi skausmo sutrikimai
	<i>DYNC2LI1</i>	rs769975073	G	A	Sklerosteozė
	<i>DYNC2LI1</i>	rs201948500	C	G	Sklerosteozė
	<i>LPA</i>	rs41272114	G	A	Širdies ir kraujagyslių ligos
	<i>LPA</i>	rs10455872	T	C	Širdies ir kraujagyslių ligos
	<i>LPA</i>	rs3798220	T	C	Širdies ir kraujagyslių ligos
	<i>TYK2</i>	rs34536443	C	G	Sisteminė raudonoji vilkligė, reumatinis artritas
	<i>TYK2</i>	rs35018800	C	T	Sisteminė raudonoji vilkligė, reumatinis artritas
	<i>TYK2</i>	rs12720356	T	G	Sisteminė raudonoji vilkligė, reumatinis artritas
3.	<i>ACKR1</i>	rs2814778	T	C	Atsparumas <i>Plasmodium Vivax</i>
	<i>ADH1B</i>	rs2066702	C	T	Priklausomybė nuo alkoholio
	<i>ADH1B</i>	rs1229984	T	C	Priklausomybė nuo alkoholio
	<i>ADH1C</i>	rs698	A	G	Priklausomybė nuo alkoholio
	<i>ADH1C</i>	rs1693482	G	A	Priklausomybė nuo alkoholio
	<i>ALDH2</i>	rs671	G	A	Priklausomybė nuo alkoholio
	<i>APOC3</i>	rs140621530	G	T	Hiperalfalipoproteinemija, koronarinė širdies liga
	<i>APOC3</i>	rs147210663	G	A	Hiperalfalipoproteinemija, koronarinė širdies liga
	<i>APOC3</i>	rs138326449	G	A	Hiperalfalipoproteinemija, koronarinė širdies liga
	<i>APOC3</i>	rs76353203	C	T	Hiperalfalipoproteinemija, koronarinė širdies liga
	<i>APP</i>	rs63750847	G	A	Alzheimerio liga

Šaltinis	Genas	rs (dbSNP)	REF	ALT	Siejama liga ar požymis
	<i>C2</i>	rs547154	G	T	Su amžiumi susijusi geltonosios dėmės degeneracija
	<i>CASP8</i>	rs3834129	AGTAAG	–	Plaučių vėžys
	<i>CCR2</i>	rs1799864	G	A	ŽIV
	<i>CCR5/LOC102724297</i>	rs1800940	G	T	ŽIV
	<i>CCR5/LOC102724297</i>	rs1799987	A	G	ŽIV
	<i>CCR5/LOC102724297</i>	rs1800560	T	A	ŽIV
	<i>CCR5/LOC102724297</i>	rs333	ACAGTCAGTAT CAATTCTGGAA GAATTTCCAGA	–	ŽIV, hepatitas C
	<i>CX3CR1</i>	rs3732378	G	A	ŽIV, koronarinė širdies liga, geltonosios dėmės degeneracija
	<i>CX3CR1</i>	rs3732379	C	T	ŽIV, koronarinė širdies liga, geltonosios dėmės degeneracija
	<i>CXCL12</i>	rs387906400	G	A	ŽIV
	<i>FCGR2B</i>	rs1050501	T	C	Sisteminė raudonoji vilkligė, maliarija
	<i>GABBR2</i>	rs1435252	G	A	Priklausomybė nuo nikotino
	<i>GABBR2</i>	rs3780422	C	T	Priklausomybė nuo nikotino
	<i>GABBR2</i>	rs2779562	T	C	Priklausomybė nuo nikotino
	<i>GABBR2</i>	rs3750344	T	C	Priklausomybė nuo nikotino
	<i>HBB</i>	rs334	A	T	Beta-talasemija, Heinz kūnelių anemija, Hb SS liga, maliarija
	<i>HBB</i>	rs33950507	G	A	Beta-talasemija, Heinz kūnelių anemija, Hb SS liga, maliarija

Šaltinis	Genas	rs (dbSNP)	REF	ALT	Siejama liga ar požymis
	<i>HBB</i>	rs33930165	G	A	Beta-talasemija, Heinz kūnelių anemija, Hb SS liga, maliarija
	<i>IGF1R</i>	rs33958176	G	A	Atsparumas <i>insulin-like growth factor 1</i>
	<i>IGF1R</i>	rs121912429	G	A	Atsparumas <i>insulin-like growth factor 1</i>
	<i>IGF1R</i>	rs121912428	C	T	Atsparumas <i>insulin-like growth factor 1</i>
	<i>IGF1R</i>	rs121912427	A	C	Atsparumas <i>insulin-like growth factor 1</i>
	<i>IGF1R</i>	rs121912426	G	A	Atsparumas <i>insulin-like growth factor 1</i>
	<i>IL4R</i>	rs1805015	T	C	Atsparumas atopijai
	<i>IL4R</i>	rs1805010	A	G	Atsparumas atopijai, imunonepakankamumo sindromas
	<i>KCNMB1/KCNIP1</i>	rs11739136	G	A	Atsparumas hipertenzijai
	<i>MC4R</i>	rs52820871	A	C	Nutukimas, monogeninis diabetas
	<i>MT-ND3</i>	rs2853826	A	G	Atsparumas Parkinsono ligai
	<i>ODC1</i>	rs2302615	G	A	Sumažėjusi storosios žarnos adenomos pasikartojimo rizika
	<i>PKHD1</i>	rs137852944	C	T	Apsauga nuo policistinės inkstų ligos, žarnyno vėžio
	<i>PRNP</i>	rs267606980	G	T	Apsauga nuo Kuru
	<i>PRSS2/TRB</i>	rs61734659	G	A	Apsauga nuo pankreatito
	<i>SLC30A8</i>	rs587777582	AAGATCA	–	II tipo diabetas
	<i>SLC30A8</i>	rs200185429	C	T	II tipo diabetas
	<i>SLC4A1</i>	rs769664228	CGGCAGCCAG GACCTGGGGGC TGAATGC	–	Atsparumas maliarijai

Šaltinis	Genas	rs (dbSNP)	REF	ALT	Siejama liga ar požymis
	<i>SSTR5</i>	rs121917877	C	T	Atsparumas somatostatino analogui
	<i>TIRAP</i>	rs8177374	C	T	Atsparumas invazyviai pneumokokinei ligai, bakteriozei, maliarijai, <i>Mycobacterium tuberculosis</i>
	<i>TLR1</i>	rs5743618	G	T	Atsparumas raupsams
	<i>TLR5</i>	rs5744168	C	T	Atsparumas legioneliozei, sisteminė raudonoji vilkligė, mieloidozė
	<i>TOP2A</i>	rs267607133	G	A	Atsparumas DNR topoizomerazės II, inhibicijai
	<i>BDNF/BDNF-AS</i>	rs6265	G	A	Apsauga nuo obsesinio kompulsinio sindromo
	<i>FUT2/LOC105447645</i>	rs601338	G	A	Atsparumas <i>Norwalk</i> virusui
	<i>TLR3</i>	rs3775291	C	T	ŽIV
	<i>TOR1A</i>	rs1801968	G	C	Distonija
	<i>NOD2</i>	rs104895421	T	A	Blau sindromas, Krono liga, sarkoidozė
	<i>NOD2</i>	rs104895422	C	T	Blau sindromas, Krono liga, sarkoidozė
	<i>NOD2</i>	rs104895427	C	T	Blau sindromas, Krono liga, sarkoidozė
	<i>NOD2</i>	rs104895430	G	A	Blau sindromas, Krono liga, sarkoidozė
	<i>NOD2</i>	rs104895456	G	A	Blau sindromas, Krono liga, sarkoidozė
	<i>NOD2</i>	rs104895467	A	G	Blau sindromas, Krono liga, sarkoidozė
	<i>NOD2</i>	rs104895486	C	T	Blau sindromas, Krono liga, sarkoidozė
	<i>CFH</i>	rs1061170	C	T	Glomerulonefritas, atipinis hemolitinis-ureminis sindromas, su amžiumi susijusi geltonosios dėmės degeneracija
	<i>IL2RA</i>	rs11594656	T	A	Diabetas
	<i>EGFR</i>	rs121434569	C	T	Plaučių vėžys

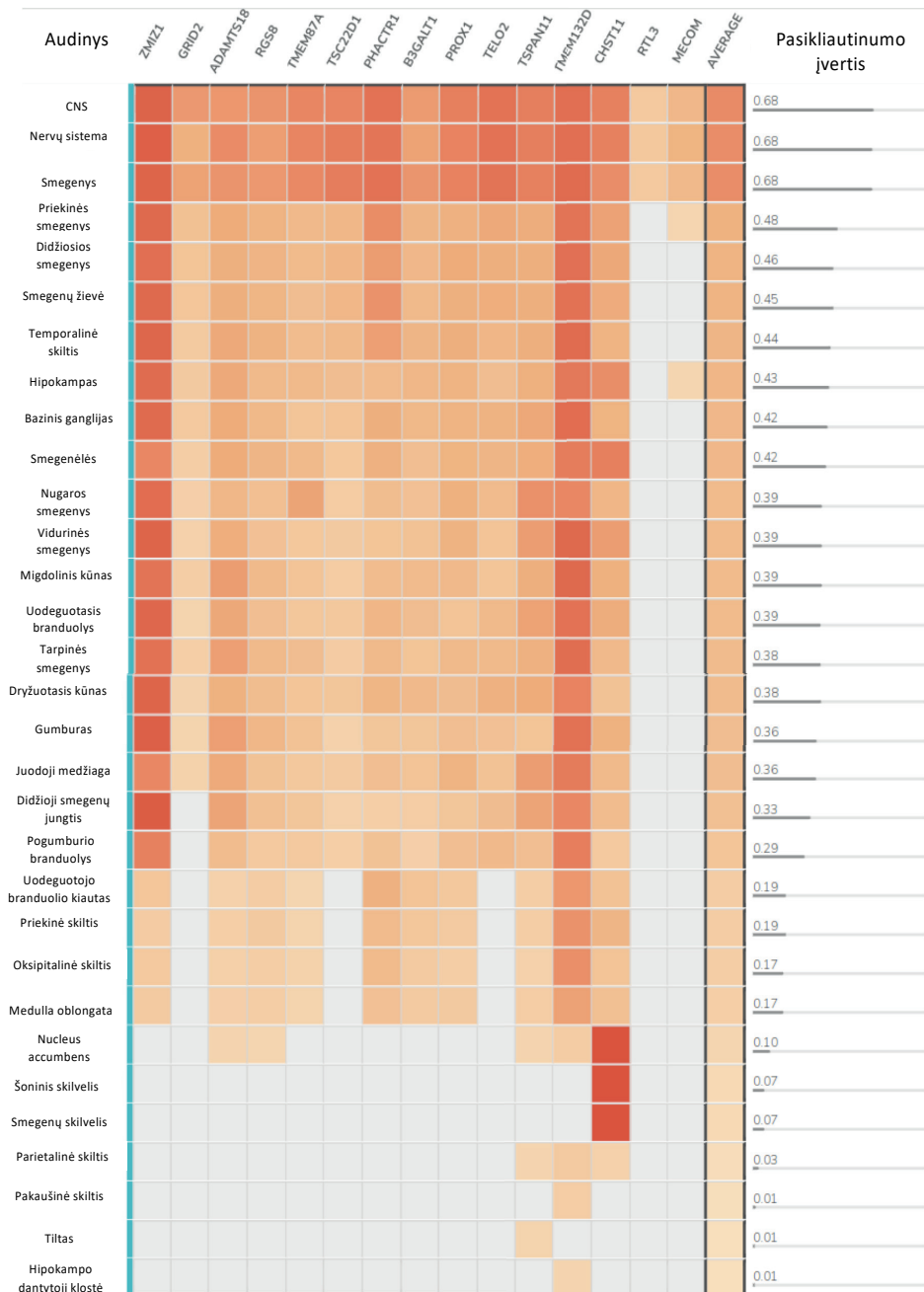
Šaltinis	Genas	rs (dbSNP)	REF	ALT	Siejama liga ar požymis
	<i>IGF1R</i>	rs1253103806	G	A	Atsparumas <i>insulin-like growth factor 1</i>
	<i>UMOD</i>	rs13333226	A	G	Hipertenzija, diabetas
	<i>IGF1R</i>	rs1409058783	G	T	Atsparumas <i>insulin-like growth factor 1</i>
	<i>IRF5</i>	rs1432329681	T	–	Sisteminė raudonoji vilkligė, reumatinis artritas
	<i>PRSS1</i>	rs143909348	G	A	Paveldimas pankreatitas
	<i>PCSK9</i>	rs1553137557	GC	AA	Šeiminė hipercholesterolemija
	<i>GYPC</i>	rs1553470034	CAGAGCCTGAT CCAGGGATGTC TGGATGGCCGG ATGGCAGAATG GAGACCTCCAC CCCCACCATAA TGGACATTGTC GTCATTG	–	Atsparumas maliarijai
	<i>IGF1R</i>	rs1555434208	G	A	Atsparumas <i>insulin-like growth factor 1</i>
	<i>IGF1R</i>	rs1555460945	G	T	Atsparumas <i>insulin-like growth factor 1</i>
	<i>SLC45A2</i>	rs16891982	C	G ar C	Melanoma
	<i>COL1A1</i>	rs1800012	C	A	Kaulų tankis
	<i>NQO1</i>	rs1800566	G	A	Alkoholio vartojimas, vėžys
	<i>IL10</i>	rs1800872	T	G	Atsparumas <i>graft-versus-host</i> ligai, ŽIV
	<i>FCGR2A</i>	rs1801274	A	T ar G	Krono liga, uždegiminė žarnyno liga, sisteminė raudonoji vilkligė, Kawasaki liga, opinis kolitas
	<i>TG</i>	rs180195	A	G	Autoimuninė skydliaukės liga

Šaltinis	Genas	rs (dbSNP)	REF	ALT	Siejama liga ar požymis
	<i>IRS2</i>	rs1805097	C	T	II tipo diabetas
	<i>NOS3</i>	rs2070744	C	T	Koronarinės arterijos spazmas
	<i>MC4R</i>	rs2229616	C	T	Diabetas, nutukimas
	<i>CLEC1A</i>	rs2306894	G	C	Aspergiliozė
	<i>TAB2</i>	rs237025	G	A	Diabetas
	<i>UGT1A1</i>	rs3064744	TA	TA[9]	Crigler-Najjar sindromas, Lucey-Driscoll sindromas, Gilberto sindromas
	<i>CTLA4</i>	rs3087243	G	A	Hashimoto tiroiditas, celiakija
	<i>HBB</i>	rs33945705	G	A ar T	Hemoglobino struktūriniai pokyčiai
	<i>HBB</i>	rs33946267	C	T ar G ar A	Beta-talasemija, Hb SS liga
	<i>TLR1</i>	rs4833095	T	C	Astma, raupsai
	<i>SCN8A</i>	rs587780455	A	G	Traukuliai, epilepsija
	<i>LAMA3</i>	rs61751706	C	T	Jungtinė epidermolizė <i>bullosa gravis</i> Herlitz sindromas
	<i>FGG</i>	rs75848804	C	G	Fibrinogenas Osaka V
	<i>LDLR</i>	rs771917370	C	T ar G	Hipercholesterolemija
	<i>SOD1</i>	rs80265967	A	C	Amiotrofinė lateralinė sklerozė

IV priedas. Genų, siejamų su burnos ertmės ligomis, sąrašas

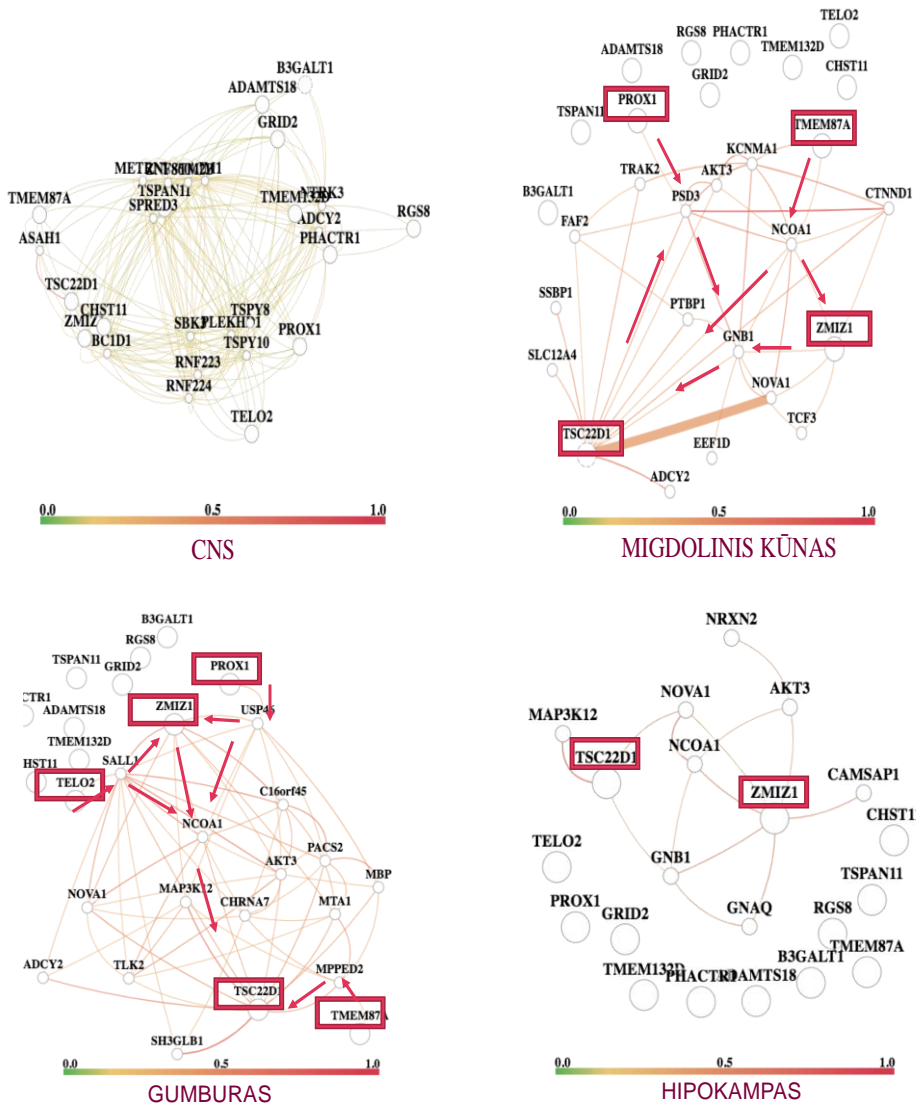
<i>SPARC</i>	<i>APEX1</i>	<i>PDGFD</i>	<i>BMP10</i>	<i>FGF5</i>	<i>FGF20</i>
<i>MYC</i>	<i>IL1</i>	<i>BMP1</i>	<i>BMP11</i>	<i>FGF6</i>	<i>FGF21</i>
<i>HSD17B2</i>	<i>IL10</i>	<i>BMP2</i>	<i>BMP15</i>	<i>FGF7</i>	<i>FGF22</i>
<i>IFNK</i>	<i>IGF1</i>	<i>BMP3</i>	<i>EGF</i>	<i>FGF8</i>	<i>FGF23</i>
<i>TGFA</i>	<i>IGF2</i>	<i>BMP4</i>	<i>EGFL6</i>	<i>FGF9</i>	<i>TNFSF11</i>
<i>TGFB1</i>	<i>PDGFB</i>	<i>BMP5</i>	<i>EGFR</i>	<i>FGF10</i>	<i>TNFRSF11B</i>
<i>TGFB2</i>	<i>PDGFRA</i>	<i>BMP6</i>	<i>FGF1</i>	<i>FGF16</i>	<i>IBSP</i>
<i>TGFB3</i>	<i>PDGFRB</i>	<i>BMP7</i>	<i>FGF2</i>	<i>FGF17</i>	<i>AMELY</i>
<i>VEGF</i>	<i>PDGFC</i>	<i>BMP8a</i>	<i>FGF3</i>	<i>FGF18</i>	<i>AMELX</i>
<i>XRCC1</i>	<i>PDGFA</i>	<i>BMP8b</i>	<i>FGF4</i>	<i>FGF19</i>	

V priedas. Su psichikos sveikatos būklėmis susijusių, analizės metu nustatytų genetinių sričių koduojamų baltymų raiškos audiniuose grafikas. Parengtas pagal Humanbase in silico įrankį. RTL3 ir MECOM baltymų raiška smegenų

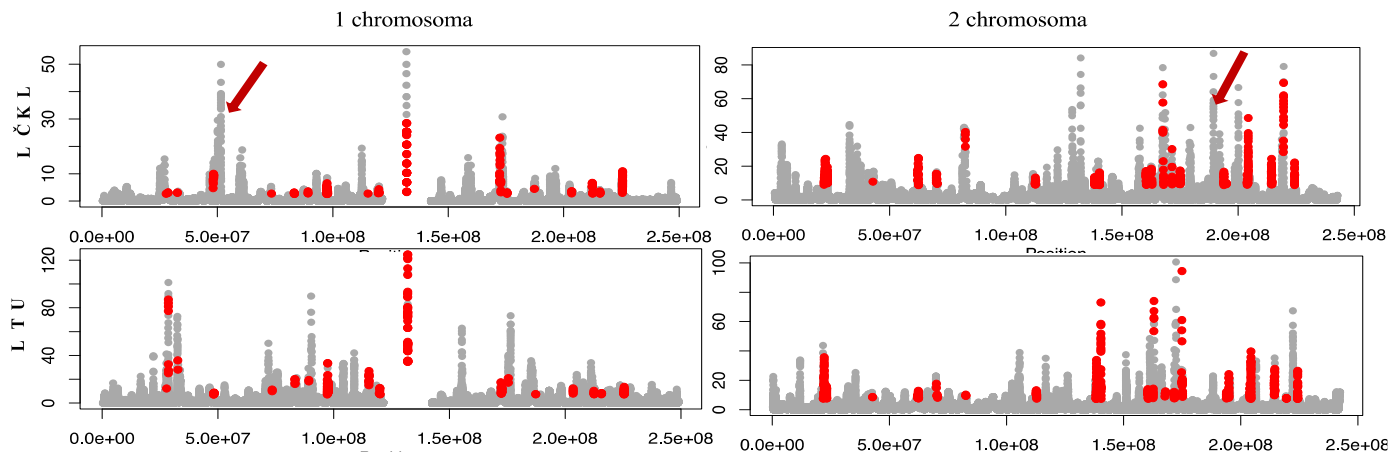


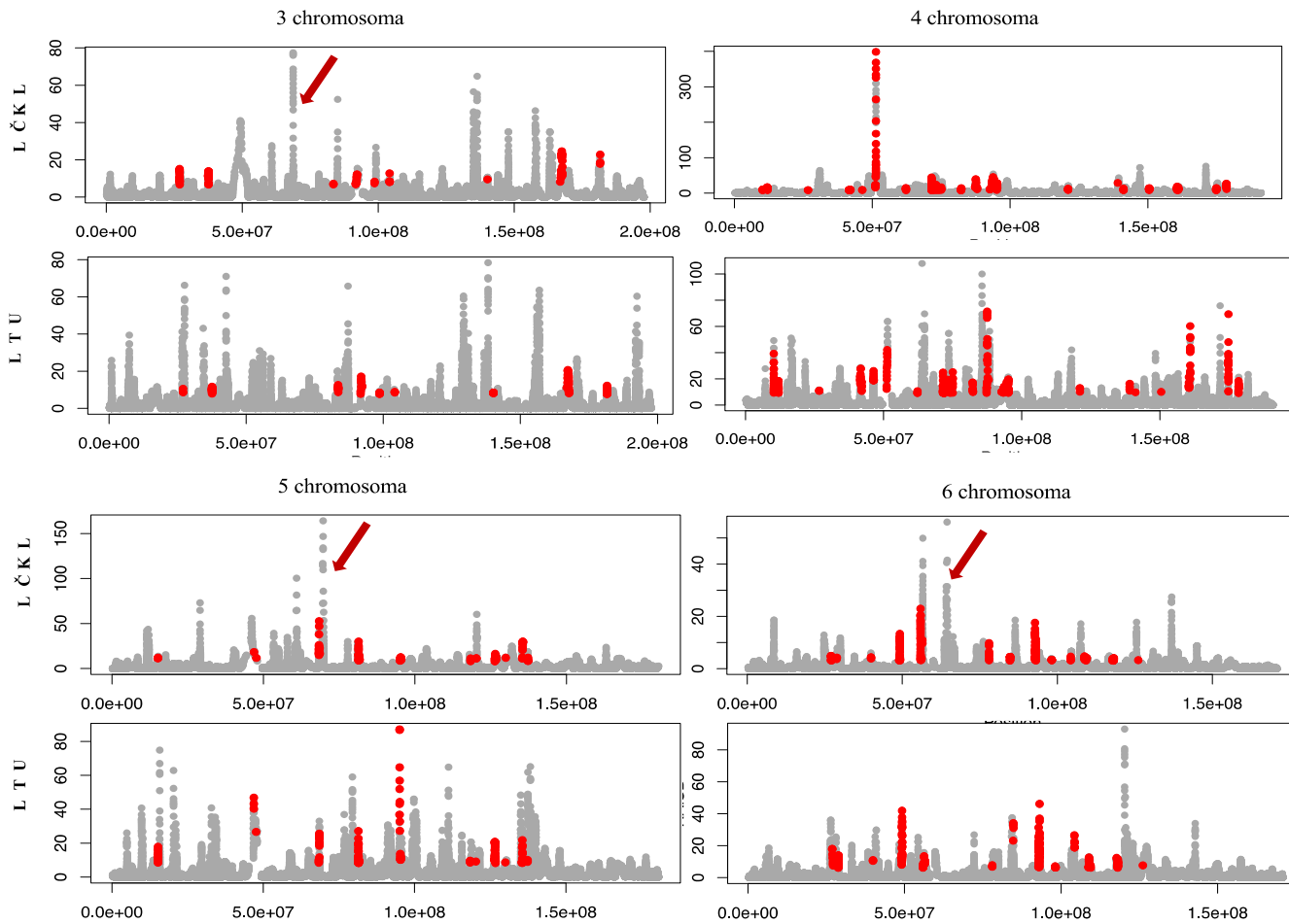
Flatiron Institute - Simons Foundation

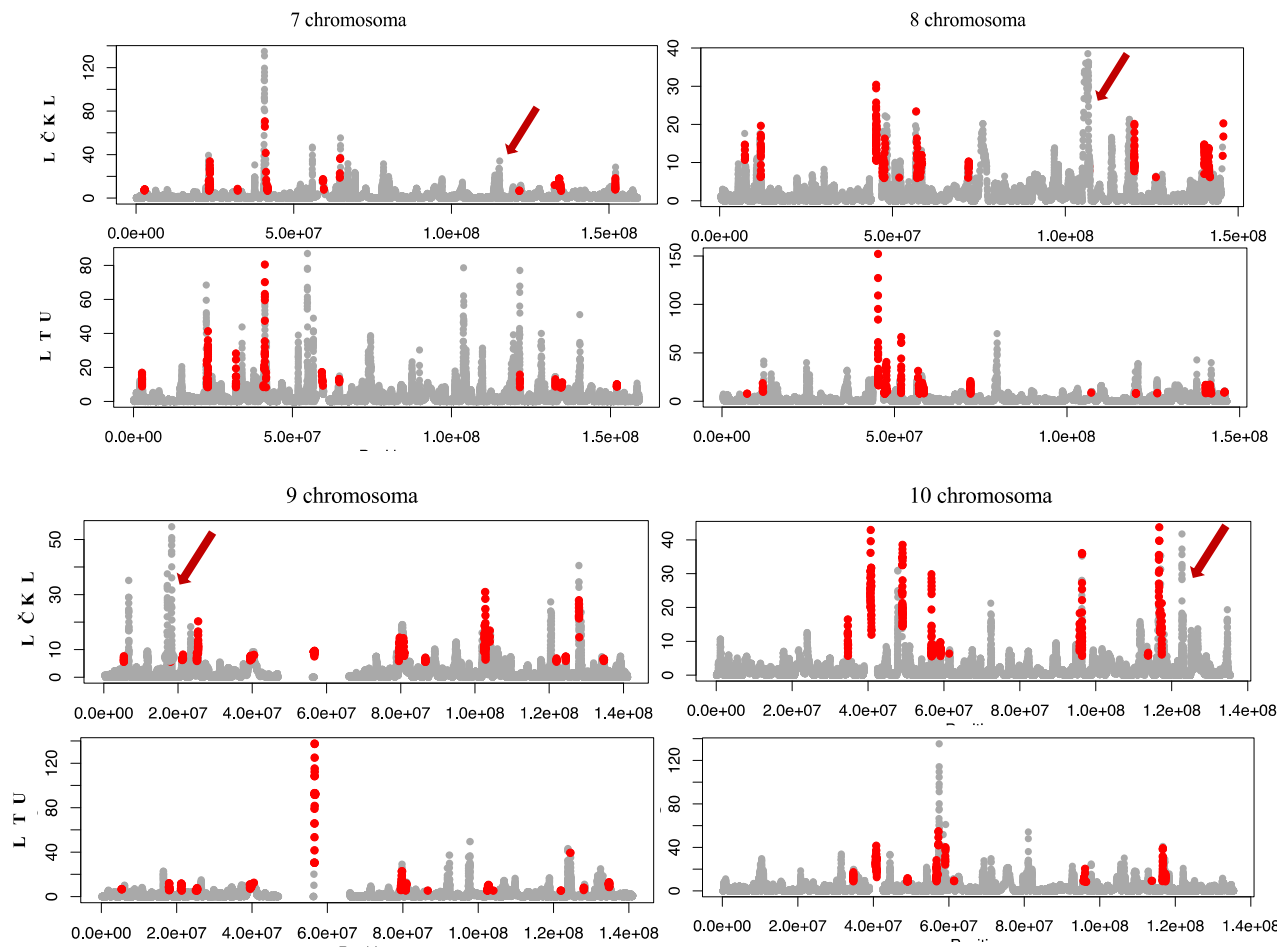
VI priedas. Centrinei nervų sistemai ir depresijai paveikiausioms smegenų sritims būdingi funkciniai genų sąveikos tinklai, sudaryti naudojantis *Humanbase in silico* įrankiu. Pažymėti genai – asociacijų analizės metu nustatyti genai. Genų sąsajos reikšmingumas žymimas spalviškai: žalia spalva reiškia nereikšmingą sąsają, raudona spalva – reikšmingą sąsają.

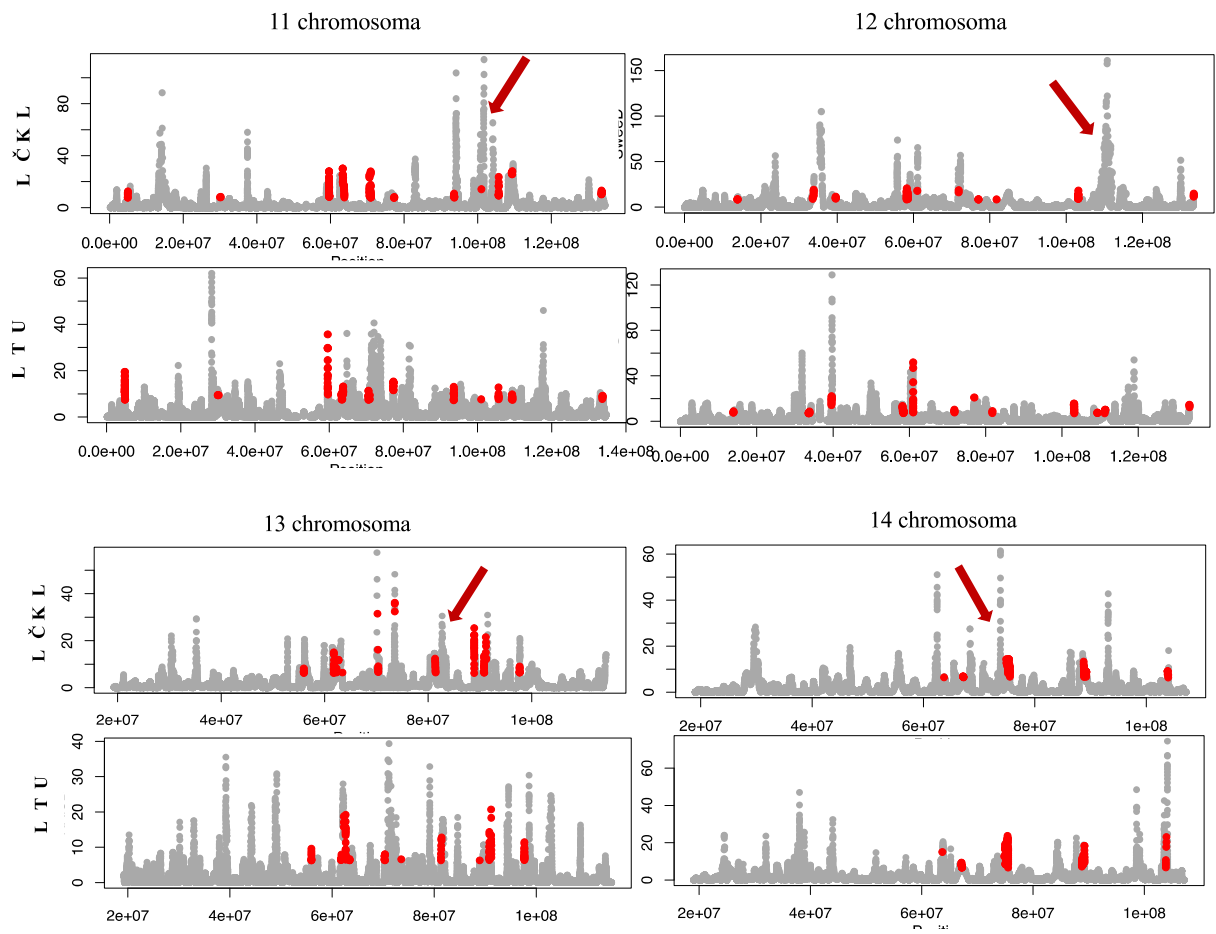


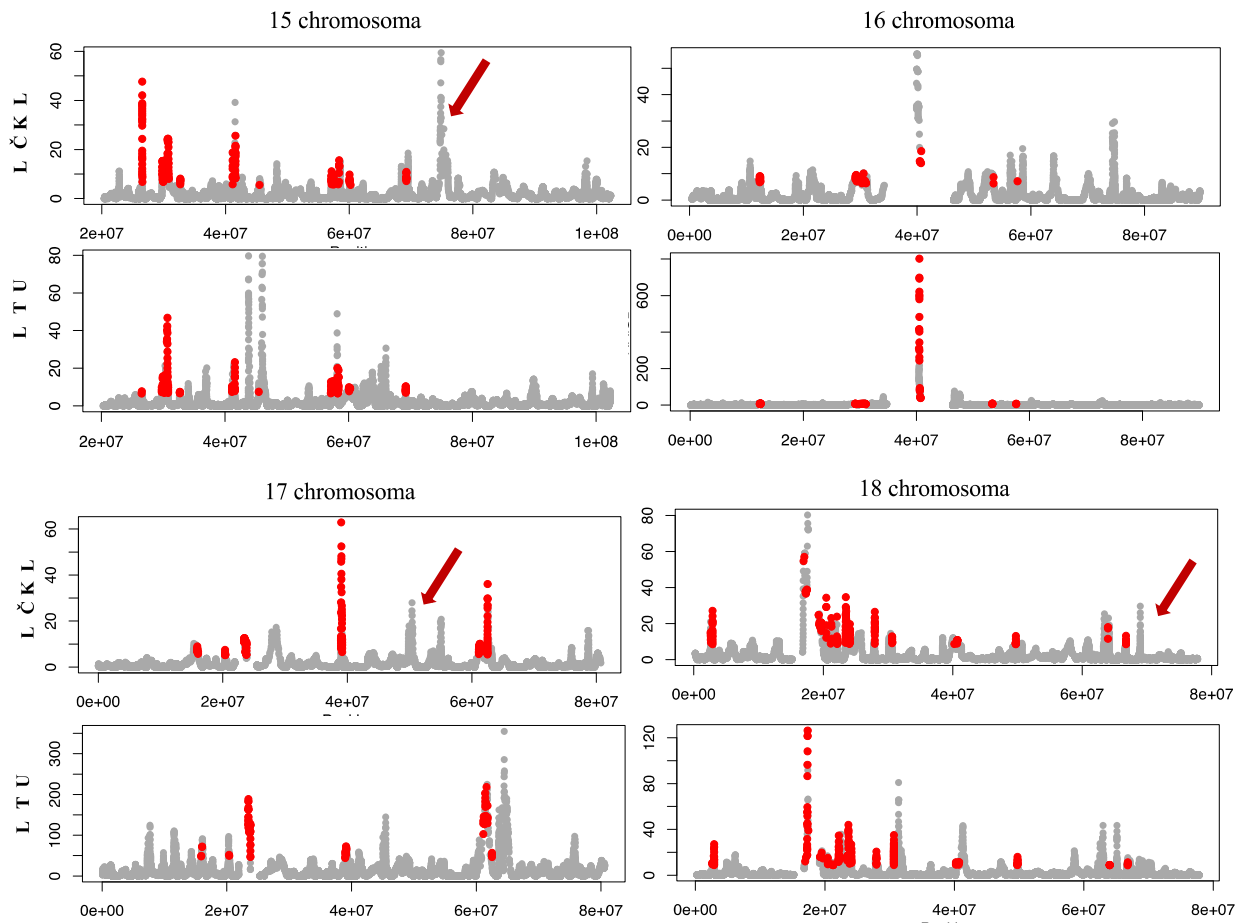
VII priedas. Lietuvos Černobylio katastrofos likviduotojų genomų teigiamos atrankos poveikio analizė autosomose. Lygintos Lietuvos Černobylio katastrofos (LČKL) ir bendrosios lietuvių populiacijos (LTU; ANELGEMIA duomenys) tiriamųjų grupės. X ašyje – genomine pozicija, Y ašyje – μ įverčio reikšmė. Rodyklėmis pažymėtos LČKL grupei būdingos teigiamos gamtinės atrankos veikiamos, unikalios ir aukščiausių μ įvertį turinčios reikšmės. Raudonai pažymėtos genomo sritys – tapačios tarp LČKL ir LTU tiriamųjų grupių, pilkos – netapačios teigiamos gamtinės atrankos veikiamos genomo sritys. 4 ir 16 chromosomose dėl dirbtinai padidinto centromerinio signalo scheminis vaizdavimas nebuvo galimas, taigi unikalios sritys nustatytos rankiniu būdu, naudojant *RAiSD Report* failus.

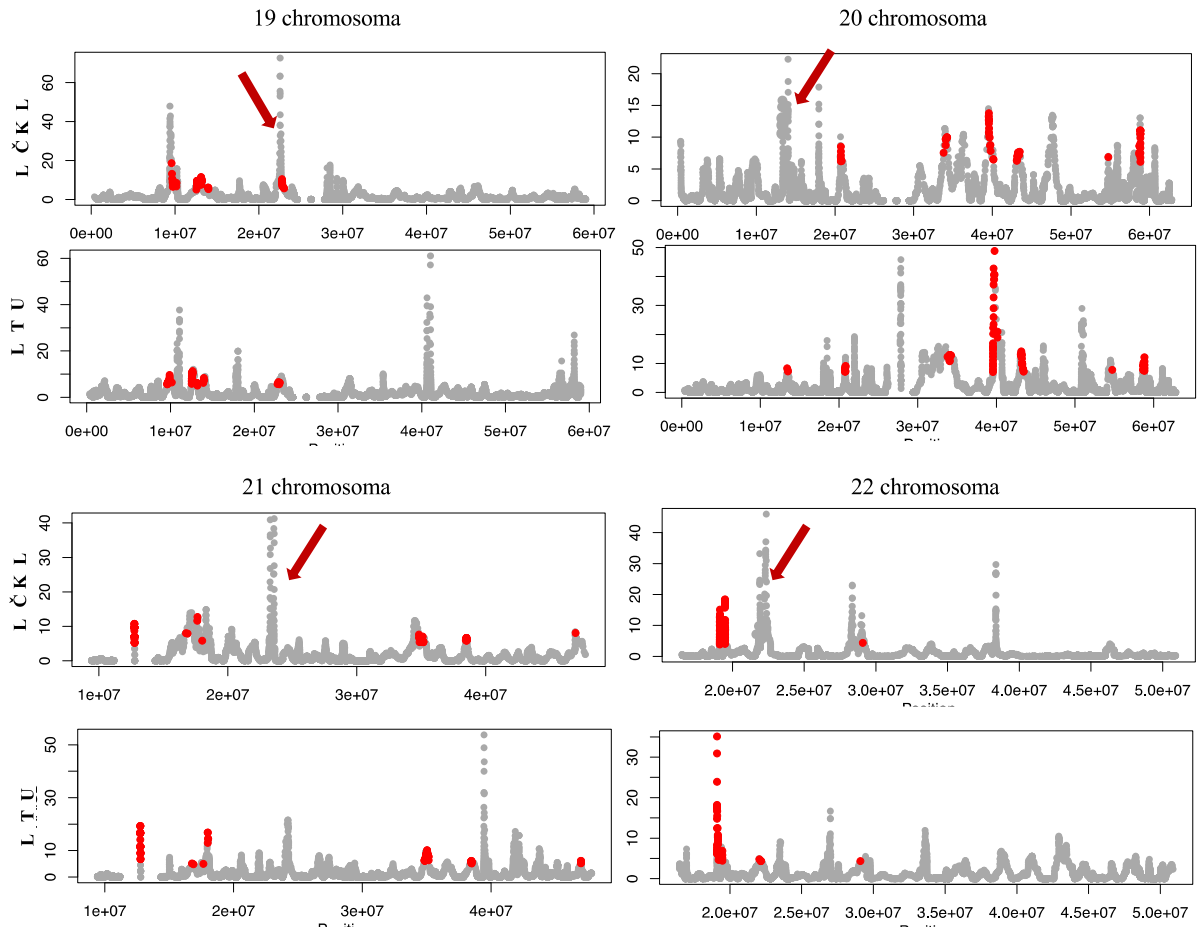












VIII priedas. Tyrimo populiarinimas ir darbo viešinimas medijose

Žodiniai pranešimai mokslo populiarinimo renginiuose

1. **Žukauskaitė G** (pranešėja). „Žmogaus genomas ir adaptacija. Lietuvos Černobylio katastrofos likviduotojų genomo tyrimai“. Mokslo populiarinimo renginys „Erdvėlaisis žemė 2020“. Vilnius, Lietuva.

Internetiniai tinklalapiai ir straipsniai

1. Straipsnis Vilniaus universiteto mokslo populiarinimo žurnale „Spectrum“ (2021; 1(33)): „Lietuvos Černobylio katastrofos likviduotojų genomo tyrimai paneigė populiarų mitą“. Straipsnis perspausdintas *LRT.lt* portale „Mokslas ir IT“ rubrikoje (2021 m. balandžio 8 d.) ir Lietuvos judėjimo „Černobylis“ laikraštyje „Černobylietis“ (2021; 1(117)).
2. Straipsnis *delfi.lt* portale: „Černobylio likviduotojų genomo tyrimai sukėlė įdomių klausimų: jiems nebūdingos infekcinės ligos“, 2021 m. gegužės 2 d.

Televizija

1. Interviu „Černobylio likviduotojai liudija Lietuvos mokslininkų tyrime“ LRT televizijos tiesioginėje laidoje „Labas rytas, Lietuva“, ved. Ignas Krupavičius, 2021 m. balandžio 21 d. Pranešėjos – dr. Laima Ambrozaitytė ir dr. Ingrida Domarkienė.
2. Interviu „Černobylio pasekmės žmonijai“ *lnk.lt* laidoje „Info diena“, ved. Liuda Kudinova, 2021 m. gegužės 6 d. Pranešėja – dr. Laima Ambrozaitytė.

Radio laidos

1. Interviu „Kaip atrodo Černobylio likviduotojų genai?“, *Žinių radijo* laida „Ateitis priklauso jiems“, ved. Ignas Klėjus, 2021 m. balandžio 22 d. Pranešėja – dr. Laima Ambrozaitytė.
2. Interviu „Černobylio likviduotojų genomo ypatumai“, LRT radijo *Klasika* laida „Kosmoso departamentas“, ved. Sigita Vegytė, 2021 m. gegužės 28 d. Pranešėja – **Gabrielė Žukauskaitė**.

SUMMARY

1. INTRODUCTION

Research problem and its relevance

Survival and adjustment to changing environmental conditions are important components of adaptation in a population. Adaptation is directly dependent on the forces of natural selection; therefore, natural selection is a key factor in evolution that has the power to alter the genetic architecture of a population. In this way, past populations were and future populations are formed by acting on the phenotypes, adaptive properties, and structure of populations at the molecular level. Specific regions of the genome and its variation differ between populations, possibly due to different microevolutionary processes taking place, and this results in different characteristics of populations. In order to ascertain the genome factors that determine the uniqueness of the population, it is important to identify and study (beneficial and harmful) genome variants in the Lithuanian population as well. It is also important to identify population-specific genome loci affected by positive selection, which may determine the adaptability of the population. By ascertaining the genomic determinants of uniqueness of the population, the results obtained could be used for population-specific diagnosis and treatment in the era of personalised medicine.

In the face of a changing environment and cataclysmic events like ecological disasters, genome variants that once were protective may become deleterious. Microevolutionary processes are therefore ongoing and lead to transformations in the genetic architecture of the population that is adapting. During cataclysmic events, natural selection pressures can change the characteristics of a population in a relatively short period of time, as the population must adapt quickly to the threat posed. For this reason, populations that have survived ecological disasters are particularly suitable for studies of adaptation. Lithuania is unique as a population because it has experienced the consequences of the Chernobyl¹ nuclear power plant disaster and participated in its clean-up. Among the surviving Lithuanian clean-up workers of the Chernobyl nuclear disaster (LCWC) there are not only ill individuals, but also who are aging relatively healthy. They are valuable as an object of research because they survived extreme conditions, adapted to the lifelong effects of ionizing radiation (IR) and the consequences of the disaster itself. Potentially, these properties may depend on their unique genome variation. Studies of

¹ Other correct name is Chornobyl (in Ukrainian language); in this study, the more widespread name Chernobyl (in Russian language) has been chosen.

genome variation and positive selection loci of these and other similar cohorts would help to identify the genomic loci relevant to survival and adaptation, and provide a better understanding of ongoing microevolutionary processes.

This study evaluated the adaptive effects (morbidity, response to experienced stress and effects of IR) of genome variants that may be both beneficial and harmful – the so-called *effect* genome variants – in the cohorts of the general Lithuanian population and LCWC. In order to evaluate the possible protective effect of the genome variants identified during the association analysis in the general Lithuanian population cohort, it was chosen to use not only *in silico* analysis tools, but also to perform positive selection analysis of variant-specific genome loci. It is a new research strategy that combines association with functional *in silico* and positive natural selection analyses. In this study, genome loci under positive selection were identified in the unique cohort of the LCWC for the first time. These loci potentially determine the survival rates of nuclear disaster clean-up workers and their adaptive qualities to extreme conditions as well as the lifelong effects of the Chernobyl nuclear disaster.

Novelty of the study

In this study, we have performed a targeted analysis of effect genome variants in the cohorts of the general Lithuanian population and LCWC using genotyping data of genome-wide SNPs and exome sequencing data. We have analysed the effect of identified, statistically-significant genome variants in these cohorts that may affect their adaptive qualities. During this study, an innovative research strategy was applied using not only *in silico* analysis tools, but also a method of positive natural selection loci detection to evaluate the functional significance of identified variants.

In order to determine the genome loci important for human adaptation and survival under extreme environmental conditions, the genome sequencing data of LCWC and the general Lithuanian population was used for the positive natural selection analysis for the first time. The method chosen for the study to assess the impact of positive natural selection is relatively new. It covers several important types of positive natural selection sweep signatures and analyses them simultaneously, thus providing a comprehensive assessment of the impact of positive natural selection. It is important to emphasize that the genomes of LCWC were investigated at the molecular level for the first time ever. To date, very little has been done to study the genomes of individuals who have received high doses of IR at the high-resolution level. A new, deeper level of genome-wide analysis, as well as the identification of regions of positive selection that potentially lead to a better survival and adaptation of

individuals, provide a better understanding of ongoing microevolutionary processes.

Continuity of the study

The collection of DNA samples, the genome-wide genotyping and whole genome sequencing data can be re-used; this allows the research to be further developed. After the identification of potentially protective genome variation and loci of positive selection in the cohort of LCWC and evaluation of the significance of these loci using *in silico* tools, it will be possible to extend the analysis and clarify their significance using functional genome analysis methods.

The aim of the dissertation

To evaluate the significance of the effect genome variants and loci under positive natural selection for human adaptation, using the data of genome-wide genotyping and whole genome sequencing in the cohorts of the general Lithuanian population and Lithuanian clean-up workers of Chernobyl nuclear disaster.

Main tasks of the study:

1. To compile a catalogue of candidate effect genome variants, to determine and compare the frequencies of alleles and genotypes in the cohorts of the general Lithuanian population and the Lithuanian clean-up workers of the Chernobyl nuclear disaster. Also, to identify effect genome variants which are important for adaptation in both cohorts using genome-wide SNPs genotyping and exome sequencing data.
2. To identify unique genome loci under positive natural selection in the cohort of the Lithuanian clean-up workers of the Chernobyl nuclear disaster in comparison with the general Lithuanian population, using whole genome sequencing data.
3. To evaluate positive natural selection signatures in the genome loci of the identified effect genome variants in the cohorts of the Lithuanian clean-up workers of the Chernobyl nuclear disaster and the general Lithuanian population.
4. To evaluate the functional significance of the identified effect genome variants and the loci under positive natural selection for the phenotype using bioinformatical tools and scientific literature analysis.

Statements to be defended

1. Effect genome variation shapes phenotypic and adaptive characteristics, specific to the cohorts of the general Lithuanian population and Lithuanian clean-up workers of the Chernobyl nuclear disaster.

2. The functional significance of effect genome variants can be assessed using not only traditional *in silico* analysis tools but also methods of positive natural selection effect assessment.

3. Due to the extreme environmental pressures experienced, the genomes of the Lithuanian clean-up workers of Chernobyl nuclear disaster contain unique loci under positive selection that possibly determine the health and adaptive properties of the survivors.

2. MATERIALS AND METHODS

2.1. Research process

This study has been performed in 2018–2022. It is part of the project “Adaptive genetic mechanisms – a comprehensive study of whole genome variation in the group of the Lithuanian Chernobyl catastrophe liquidators” acronym ADAPT (project no. S-MIP-20-35), which was performed at the Department of Human and Medical Genetics Institute of Biomedical Sciences, Faculty of Medicine, Vilnius University (VU MF BMI DHMG). The bioethical approval of the Vilnius Regional Biomedical Research Ethics Committee (No. 2019/4-1119-612) was issued for the research. Two data sets of the general Lithuanian population group were used for the study, which were obtained from:

1) VU MF BMI DHMG 2011–2015 project “Genetic diversity and structural changes of the Lithuanian population related to evolution and the most common diseases” acronym LITGEN (project no. VP1-3.1-ŠMM-07-K-01-013, bioethical approval No. 158200-05-329-79);

2) VU MF BMI DHMG 2020–2022 project “Ancient and new alleles in the Lithuanian genome: mutations, natural selection and adaptation”, the acronym ANELGEMIA (Project No. S-MIP-20-34, bioethical approval No. 2020/6-1243-724).

All individuals signed consents to participate in this biomedical research.

This study consists of three parts:

1) the search and analysis of effect genome variants in the general Lithuanian population (LITGEN) group using the data of genome-wide SNP genotyping and exome sequencing;

2) the search and analysis of effect genome variants in the group of LCWC and comparison with the general Lithuanian population (LITGEN) using genome-wide SNP genotyping data;

3) analysis of genome loci under positive natural selection in the LCWC group and comparison with the general Lithuanian population group (ANELGEMIA) using whole genome sequencing data.

The reference genome of *Homo sapiens* GRCh37 (hg19) was used in the study. The research strategy and workflow are presented in Figure 2.1.

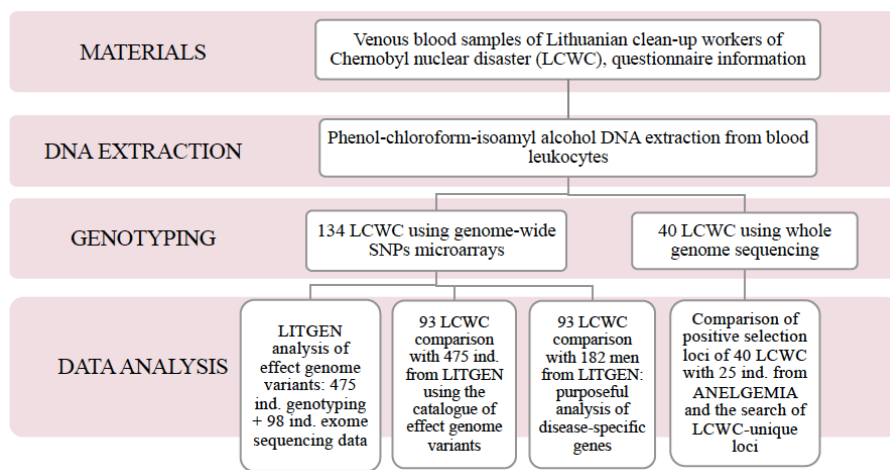


Fig. 2.1. Research strategy and workflow. LCWC – Lithuanian clean-up workers of Chernobyl nuclear disaster, LITGEN – study group of the general Lithuanian population from LITGEN project, ANELGEMIA – study group of the general Lithuanian population from ANELGEMIA project. Ind. – individuals.

2.2. Cohort of the study

During the study, a total of 134 venous blood samples of LCWC were collected. The study group included 128 men and 6 women who were involved in the elimination of the Chernobyl nuclear disaster’s consequences. All individuals are residents of the Republic of Lithuania. Due to the fact that the number of individuals in the study cohort increased after the start of the data analysis stage for 93 individuals and there was no possibility to include new individuals without major changes to the study strategy, 93 men of the LCWC cohort were included in the further analysis. At the time of the study, the average age of LCWC was 64 years. Two datasets (LITGEN and ANELGEMIA) of the general Lithuanian population group were used as control groups for the study. LITGEN dataset included genotyping and exome sequencing data of 475 individuals (239 women and 236 men). The average

age of the individuals was 53 years. ANELGEMIA dataset included whole genome sequencing data of 25 men. The average age of the individuals was 36 years.

2.3. Analysis strategy of the questionnaire data of the Lithuanian clean-up workers of Chernobyl nuclear disaster

All participants had to fill out the questionnaire (it is in the Lithuanian language, available as Appendix II of the main dissertation document), which allowed to collect information about the individuals' roles in the clean-up process of the Chernobyl nuclear disaster, working and living conditions in the Chernobyl, their lifestyle habits, psychological condition, and clinical data. The collected data were processed and descriptive statistical analysis was performed with *Microsoft Office Excel* software. *Rstudio v3.5.2* software was used to assess statistical significance. The analysis of questionnaire and clinical data allowed to identify the most common disease groups in the group of LCWC. Groups of oral and mental illnesses were selected for further analysis due to the biological association of these illnesses with exposure to IR, the psychological experiences of LCWC, and the high incidence of these illnesses in the LCWC cohort.

2.4. Genome-wide SNPs based genotyping using microarrays

The genomic DNA of the LCWC was extracted from blood leukocytes using the phenol-chloroform-isoamyl alcohol extraction method. Concentration and purity of the extracted DNA were determined on a *NanoDrop® ND-1000* spectrophotometer. The DNA extraction, concentration and purity measurement protocols used for the study were described in detail by Ingrida Domarkienė, PhD [104, Appendix I]. The genotyping of LCWC was performed using *Infinium OmniExpress-24v1.3* microarrays (714 238 SNPs) and the *Illumina HiScanSQ* system. The protocol used for genotyping and data quality control is described in detail by Ingrida Domarkienė, PhD [104, Appendix II]. Genotyping of the general Lithuanian population (LITGEN) was performed by VU MF BMI DHMG LITGEN project scientists and only the data collected were used for this study. After quality control, 91 LCWC genotypes were set for further analysis.

2.5. Analysis of the genotyping data

2.5.1 Catalogue of effect genomic variants

A catalogue of 144 genomic variants from the *ClinVar* [105] and *OMIM* [106] databases and scientific publications was compiled. The main criteria to include a variant from the databases were 1) clinical significance review status (protective or uncertain) and 2) count of submissions (more than 1). The inclusion criteria for a variant from scientific publications were 1) influence of the variant on gene function (i.e., the variant was expected to alter gene function; mostly loss-of-function) and 2) the frequency of the variant (i.e., rare or previously rare alleles that increased in frequency possibly because of positive effects on the phenotype). This catalogue [Appendix III] was used for the targeted gene-set association analysis.

2.5.2. Primary analysis of genotyping data (preparation)

The *GenomeStudio 2.0* software was used for genotyping data preparation [108]. The main stages of primary analysis are as follows:

1) Creation of a new analysis project: uploading raw genotyping data and additional files (sample sheet file in *.csv* format, SNP manifest file in *.bpm* format).

2) SNP grouping (clustering) of genotyping data: the initial stage of data quality control, during which the quality parameters of samples and SNP markers are calculated.

3) Removal of low-quality SNP markers. The main parameters indicating the quality of SNP markers are:

a) the *Call Frequency* value was selected >0.97 ;

b) the *GenTrain* value as close to 1 as possible;

c) the *ClusterSep* value >0.4 [109]. SNP markers with a *ClusterSep* value of less than 0.27 were removed. Markers with *ClusterSep* values between 0.27 and 0.4 were reclustered and included in further analysis.

4) Removal of poor quality samples: samples are evaluated according to the *Call Rate* value. This value was chosen to be >97 percent [110]. Due to the poor *Call Rate* value, 2 samples were excluded from the LCWC group of 93 individuals, and 10 samples were excluded from the general Lithuanian population (LITGEN) group of 475 individuals.

5) Creation of input files required for further analysis in *.ped* and *.map* formats using *GenomeStudio* software plug-in *PLINK Input Report Plug-in v2.1.4*.

2.5.3. Secondary analysis of the genotyping data

PLINK v1.90 software was used for the secondary analysis. First, data quality control of the general Lithuanian population (LITGEN) group (465 individuals) was performed. The main steps were:

- 1) combining the data of the general Lithuanian population (LITGEN) from different microarrays;
- 2) removal of chromosomes 0, 23, 24, 25;
- 3) quality control according to estimates: *--mind* 0.1 (removal of individuals missing >10% of the genotype); *--gene* 0.01 (removal of SNPs missing >1% of genotype); *--hwe* 0.001 (elimination of SNPs that do not comply with the Hardy-Weinberg equilibrium law, with a p value $<10^{-3}$);
- 4) filtering of effect genome variants from genotyping data using the effect genome variant catalogue in *.txt* file format and using the *--extract* command to select effect genome variants from genotyping data;
- 5) association analysis with other European populations was performed.

An analogous analysis was performed for the genotyping data of the LCWC group.

2.5.4. Statistical analysis strategy of the genotyping data

The statistical analysis is divided into two main stages:

- 1) allele and genotype frequency analysis of the general Lithuanian population (LITGEN) by combining the exome sequencing data of 98 individuals and comparing it with certain European populations;
- 2) association analysis of effect genome variants in LCWC and comparison with LITGEN cohort. A detailed scheme of statistical analysis is presented in Figure 2.2.

During the first stage of the statistical analysis, effect variant allele and genotype frequencies were calculated and compared with common European (EUR) and individual European populations (Utah residents of Northern or Western European Ancestry, CEU) and the Finnish in Finland (FIN)) using *1000G project* data [111] available in the *NCBI dbSNP* database [112]. These individual European populations were selected on the basis of a study that found that the Lithuanian population is evolutionary and genetically consistent with the aforementioned populations, and the analysis of loci under positive selection confirms these findings [68]. Calculations were performed using *PLINK v1.90* and *Rstudio v3.5.2* software *Rcmdr* package ($\alpha = 0.05$).

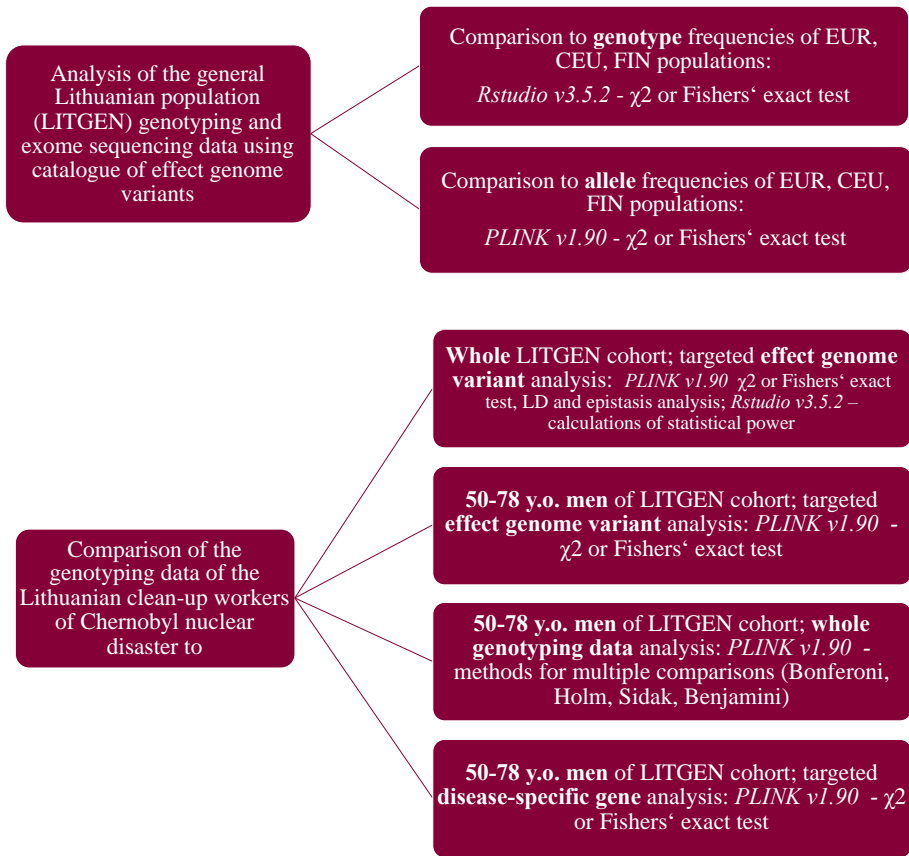


Fig. 2.2. Scheme of statistical analysis of genotyping data.

2.5.5. Functional evaluation strategy of identified statistically significant effect genome variants using *in silico* tools

To assess the effect of identified statistically significant effect genome variants based on genome and phenotype, variants were analysed using *in silico* analysis tools and scientific databases: *Varsome* [113], *Uniprot* [114], *Ensembl* [115], *ClinVar* [105], *OMIM* [106]. Functional molecular relationships between genes were sought and functional gene networks were generated using *GeneMANIA* [116], *STRING* [117], *Humanbase* [118], *G:Profiler* [119] *in silico* analysis tools.

Also, the Ensembl database was used to compare statistically significant effect alleles identified in the general Lithuanian population (LITGEN) with different primate species (*Gorilla gorilla*, *Pongo abelii*, *Theropithecus gelada*, and *Chlorocebus sabaeus*) to determine whether the effect allele was

of derived origin (and potentially leading to a protective function). A positive natural selection analysis with *RAiSD* software was performed to substantiate or negate the potential protective role of the identified effect genome variants.

2.6. Exome and whole genome sequencing

2.6.1. Exome sequencing of the general Lithuanian population (LITGEN) group

Exome sequencing was performed for 98 individuals of Lithuanian origin using the 5500 series SOLiD™ system [66]. Sequencing was performed by the scientists of the LITGEN project. In this study, only exome sequencing data were used for the association analysis, combining it with genotyping data.

2.6.2. Whole genome sequencing of the general Lithuanian population (ANELGEMIA) group and Lithuanian clean-up workers of Chernobyl nuclear disaster

Whole genome sequencing was performed for 25 and 40 individuals of the general Lithuanian population (ANELGEMIA) group and LCWC group, respectively. Selection criteria for genome sequencing of LCWC group were:

- 1) the amount of IR dose experienced in the study group (doses should have been documented and as high as possible);
- 2) individuals worked and (or) lived within the 30 km zone around the Chernobyl nuclear power plant and had the supporting documents to prove it;
- 3) type of tasks performed during the clean up process (this study included individuals who worked on the most dangerous tasks and closest to the Chernobyl nuclear power plant).

Genome sequencing and primary quality control of raw data files (.fastq) were performed at the *Center for Genomics and Transcriptomics (CeGaT)* in Germany using the *Illumina NovaSeq 6000* sequencing system, under contract with Vilnius University.

2.6.3. Analysis strategy of genomic loci affected by positive natural selection of the general Lithuanian population (ANELGEMIA) group and Lithuanian clean-up workers of Chernobyl nuclear disaster

Positive natural selection leads to adaptive processes in the population. Identifying the loci of the genome under positive selection could help to answer the question of which loci are beneficial for the survival of individuals and their adaptive properties. In order to prove or deny the protectiveness of

the identified effect genome variants and their impact on the study populations, as well as to identify genomic loci under positive selection which are unique to LCWC, an analysis of loci under positive selection was performed using the freely available RAI_iSD v2.9 (*Raised Accuracy in Sweep Detection*) software [52]. This software was installed and used for the analysis of the general Lithuanian population group and LCWC for the first time.

The main advantage of this software is the combination (μ statistic) of three main positive selection sweep signatures, which describes the effect of positive selection on genomic loci. These main signatures of positive natural selection are: 1) local reduction at the polymorphism level, 2) particular shift in the site frequency spectrum toward low- and high-frequency derived variants, and 3) a localized pattern of linkage disequilibrium (LD) levels, characterized by high LD on each side of a beneficial mutation and low LD between loci that are located on different sides of the beneficial allele. Other similar tools are designed to identify each of these signatures separately. In addition to this, other advantages of RAI_iSD are the negligible computing memory requirements and the processing of large amounts of data in a relatively short period of time. The software makes it possible to determine the effects of positive natural selection through genomic variation that existed in genomes before the selection pressure (in the case of LCWC, before the catastrophe).

Whole genome sequencing data from the *.vcf.gz* file format were used for the analysis. The initial step also requires the *bcftools v1.8* [120] software tool, which is used to merge different individuals' *.vcf* files into a single group (LCWC or ANELGEMIA) file. The following command line was used to do this (further examples include analysis of the general Lithuanian population (ANELGEMIA) group):

```
bcftools merge -l ANELGEMIAmerge.txt -O v -o mergedANELGEMIA.vcf
```

where the *.txt* file contains the directories of the *.vcf.gz* files to be merged. The same command was repeated with LCWC data.

Further analysis was performed for each autosome separately due to excessive LCWC and ANELGEMIA data file size. From the combined files generated in the first step, separate files were generated for each chromosome (a total of 22 files per study group). This was done using the command line:

```
grep -w chr1 mergedANELGEMIA.vcf > ANELGEMIAchr1.vcf
```

The resulting file of chromosome 1 must have the header (first) line of the file changed, so the following command lines are used:

```
head -1 ANELGEMIAchr1.vcf – command shows the first line of the file.  
Then it is copied and pasted to the second command between the  $\langle \rangle$  characters.
```

```
sed -i '1s / ## contig = <ID = chr1, length = 249250621> //' ANELGEMIAchr1.vcf
```

A *header.vcf* file is prepared that stores the correct header line of the *.vcf* file. This file is merged with the first chromosome file:

```
cat ANELGEMIAchr1.vcf >> header.vcf
```

 – this command line transfers information from *ANELGEMIAchr1.vcf* to the *header.vcf* file, the latter being renamed *ANELGEMIAchr1H.vcf*.

Finally, R*AiSD* software was operated using the command line:

```
RAiSD -n ANELGEMIAchr1 -I ANELGEMIAchr1H.vcf -X centromeros.txt -R -P.
```

 On this command line, the *-X* argument removes selected genomic regions from the analysis that are listed in the *.txt* file. In this study, the centromeric regions of the autosomes studied were excluded from the analysis because the μ value is artificially increased in these regions due to the potentially small variation in the genome and because of the technical features of μ computation. For this reason, the adjacent regions of the centromeres are not interpreted in the further analysis either. Due to a software bug, the centromere regions on chromosomes 4 and 16 were not removed and were analysed manually using the generated *RAiSD_Report* files.

The specified command line generates three files, *RAiSD_Info.ANELGEMIAchr1*, *RAiSD_Report.ANELGEMIAchr1*, and *RAiSD_Plot.ANELGEMIAchr1.pdf*. The *Report* file stores information about the genomic positions of loci under positive selection, the estimates of the three main signatures of positive natural selection separately, and the values of the combined μ parameter. The *Plot* file graphically displays the distribution of the three main signatures of the positive natural selection and the μ estimates (y-axis) in a context of a genomic position in a chromosome (x-axis). All these steps, discussed in this example of the analysis of the data of the general Lithuanian population (ANELGEMIA) group for the chromosome 1, are repeated for all chromosomes and LCWC data. A total of 44 sets of files were generated (22 chromosomes for ANELGEMIA and LCWC study groups, respectively). The effect of a positive natural selection is considered to be found when $\mu > 0$. The higher the value of μ , the stronger the effect of positive selection in the analysed genomic loci. Also, in order to identify the most significant loci of the genome affected by positive natural selection, it was assumed that the majority of the genome is neutral, i.e. is not affected by positive selection. Therefore, genomic loci were considered as significant when they reached top 5% of the highest μ estimates. This value of 5% of the highest estimates corresponds to the value of *p* and is used as the significance threshold [52].

In order to compare the genomic loci under positive natural selection in the LCWC and the general Lithuanian population (ANELGEMIA) groups and to find LCWC-specific loci that may affect the adaptation of the of this study group, the following command line was used:

RAiSD -n commonoutlierschr1 -CO RAiSD_Report.LCKLchr1 1 7 RAiSD_Report.ANELGEMIAchr1 1 7 -COT 0.05 -COD 5000 – 5% of the highest rated μ values (COT argument; $\alpha = 0.05$) were chosen to determine the most significant loci. This command line is designed to find and single out common regions under positive natural selection. Four files are generated for each chromosome analysed, while *RAiSD_CommonOutlierPlot.pdf* is the most important for this analysis. This file displays two graphs comparing the values of μ of the particular chromosome between the two study groups. The regions of the chromosome marked in *red* shows *common* positive natural selection loci in both study groups. The unique regions of the chromosome under positive natural selection for each study group are shown in *gray* (Figure 3.6). The *RAiSD_Report file* is used to determine the genomic position of the exact region. It was chosen to include unique genomic loci affected by positive natural selection of the LCWC group with the highest μ values for the further analysis.

The *RAiSD_Report* file was also used to evaluate the effect of positive natural selection on effect genome variants, which were identified during the association analysis performed on the genome-wide SNPs genotyping data of LCWC and the general Lithuanian population (LITGEN) group. Based on the genomic position of the variant, the corresponding region can be found in the file, which includes information about the value of μ . This value is then interpreted in the context of other μ values in the chromosome and the significance of μ in relation to the threshold value.

The *UCSC genome browser* [121] was used for the analysis of genomic loci, and *Ensembl* [115], *OMIM* [109], *GeneMANIA* [116], *PopHumanScan* [122], *PANTHER* [123], *Gene Ontology Browser* [124], *Human DNA Repair Genes* [125] databases and tools were used to assess gene functionality.

3. RESULTS AND DISCUSSION

3.1. Descriptive statistical analysis of the questionnaire data on Lithuanian clean-up workers of the Chernobyl nuclear disaster

Ninety-three LCWC analysed in this study were 50–78-year-old men. They were sent to the highest radioactivity zone (about 30 km from the nuclear reactor) for various tasks that were supposed to prevent the spread of

radioactive contamination. The most health-damaging activities were the manual collection of granite pieces, the management of the nuclear power plant site, the handling and disinfection of equipment inside or near the nuclear power plant. Other, more common activities were driving vehicles to and from the Chernobyl nuclear power plant or road maintenance.

According to available official data [126], the doses of IR experienced by clean up workers in the years 1986–1989 were up to 100 mSv. This data confirm our own data, which show that as many as 43% of LCWC experienced doses up to 100 mSv (Figure 3.2). Data on experienced IR doses were collected during a questionnaire survey and were not always documented (e.g., the dose was not recorded on a military ticket), so inaccuracies are possible. Doses higher than 200 mSv were detected in 16% of LCWC.

According to the anamnesis, LCWC are mainly affected by diseases of the circulatory system, the musculoskeletal system and connective tissue, and the digestive system. Figure 3.1 shows the distribution of all characteristic disease groups in the cohort of LCWC.

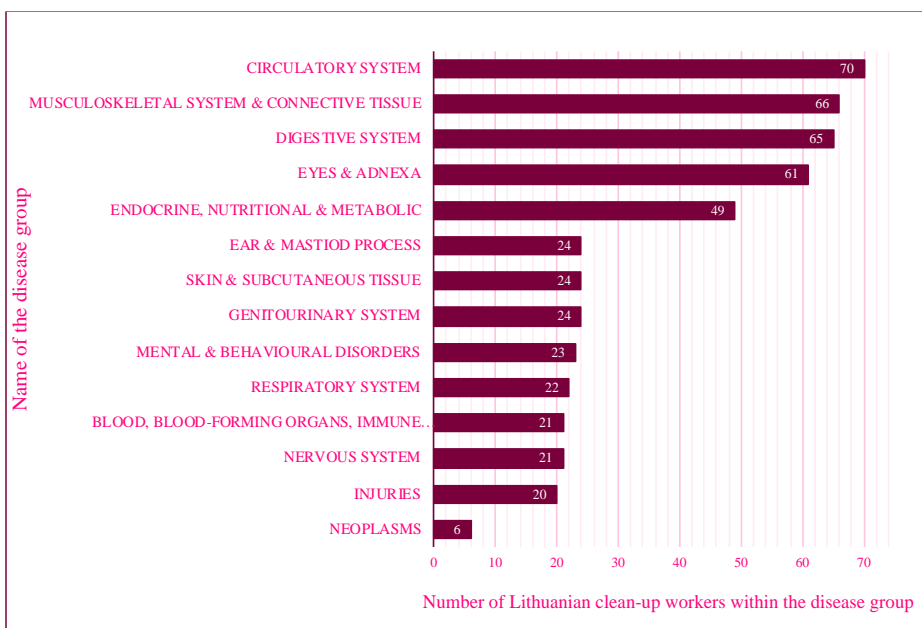


Fig 3.1. Distribution of the most common disease groups in the cohort of LCWC. The x-axis shows the number of individuals in a disease group. Some individuals belong to several disease groups. The y-axis shows disease groups.

This distribution of the most common disease groups almost coincides with the distribution in the general Lithuanian population, according to the data of the Lithuanian Institute of Hygiene [assessed on January 2021, 127]. Most of these diseases are age-related. Interestingly, there were only a few cases of

tumors, although IR is known for its carcinogenic effect. In addition, neoplasms are one of the most common causes of death in the general Lithuanian population aged 45 and over [127]. This discrepancy may have been due to the unique, adaptive qualities of the LCWC genomes or due to the structure of the LCWC group (it is likely that most of the individuals who had tumours are already deceased).

In the further part of this analysis, the aim was to characterize the LCWC by examining which disease groups are the most common in groups of different experienced IR doses (Figure 3.2.).

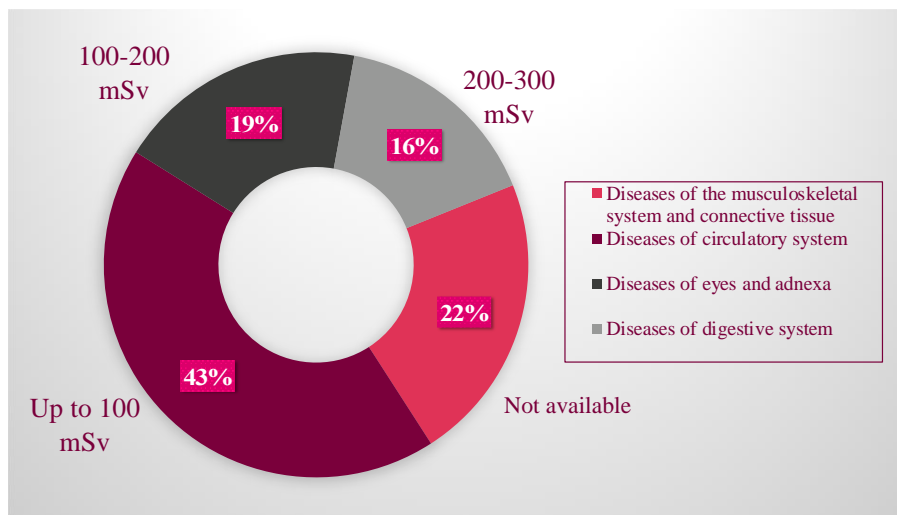


Fig 3.2. Distribution of doses of ionizing radiation (IR) experienced by LCWC. The percentages indicate the proportion of individuals depending on the exposure to different doses of IR. The diagram shows the most common groups of diseases in different groups of IR doses experienced.

3.2. Analysis of the effect genome variants of the general Lithuanian population (LITGEN)

The aim of the first stage of the analysis was to elucidate the effect genome variants characteristic of the general Lithuanian population. This was achieved using the compiled catalogue of selected effect genome variants for the genome-wide SNPs genotyping and exome sequencing data. Of the 144 effect genome variants in the general Lithuanian population group (LITGEN), frequencies of alleles of four genome variants were statistically significantly different as compared to Utah Residents with Northern and Western European Ancestry (CEU), Finns from Finland (FIN), and the general European population (EUR) (Table 3.1). The result of the comparison of the allele frequencies of the two variants (in the *PPARG* and *ADH1C* genes) with the

general European population was close to statistical significance. The calculated p-value may vary depending on the sample size, i.e. increasing the sample size may make the result statistically significant or non-significant. Thus, it was chosen to include potentially statistically significant (close to statistical significance) results in the further analysis. No statistically significant differences were found between genotype frequencies by comparing these cohorts.

Table 3.1. Comparison of frequencies of effect variants in the Lithuanian and European populations. Distribution of the effect allele genotypes and the statistics for the evaluation of differences in frequencies of effect variants in the Lithuanian and European populations. Only significant or borderline significant (*PPARG* and *ADH1C*) results are shown. CEU – Utah residents (CEPH) with Northern and Western European Ancestry; EUR – European population; FIN – Finnish in Finland; LTU – Lithuanian population; MAF – minor allele frequency.

Genome variant ID	Gene	Change *	EP	HOMO ALT		HETERO		HOMO REF		MAF (LTU)	MAF	χ^2	p
				LTU	EP	LTU	EP	LTU	EP				
rs1801282	<i>PPARG</i>	NM_0013546682: c.34C>G	EUR	4	5	46	111	118	387	0.161	0.120	3.632	0.05
rs13266634	<i>SLC30A8</i>	NM_0011728152: c.826C>T	FIN		18		43		38	0.318	0.399	4.839	0.03
			CEU	50	6	195	36	219	57				
rs11556924	<i>ZC3HCl</i>	NM_0012821901: c.1025G>A	FIN	77	5	225	40	163	54	0.408	0.253	16.642	<0.01
rs2274223	<i>PLCE1</i>	NM_0011659792: c.4856A>G	EUR	65	59	223	222	175	222	0.381	0.338	3.919	0.04
rs7498665	<i>SH2B1</i>	NM_0011458121: c.1450A>G	EUR	19	52	36	227	43	224	0.378	0.329	7.015	0.03
rs698	<i>ADH1C</i>	NM_000669.5: c.1048A>G	EUR	25	93	48	221	25	189	0.5	0.405	5.878	0.05

* All effect alleles were minor alleles.

EP – the population to which the comparison is made, CEU – Utah residents (CEPH) with Northern and Western European ancestry; EUR – general European population; FIN – Finnish in Finland; HETERO – heterozygous genotype count, HOMO ALT – homozygous alternative genotype count, HOMO REF – homozygous reference genotype count LTU – Lithuanian population; MAF – minor allele frequency.

According to the scientific literature, these amino acid-modifying variants potentially protect against alcohol dependence (*ADH1C*, rs698), type II diabetes (*PPARG*, rs1801282; *SLC30A8*, rs13266634), coronary heart disease (*ZC3HCl*, rs115524), obesity (*SH2B1*, rs7498665), and oesophageal cancer (*PLCE1*, rs2274223).

After the allele origin, functional *in silico* and scientific literature analyses, three effect genome variants (*PPARG*, rs1801282; *SLC30A8*, rs13266634, *ZC3HCl*, rs115524) have been identified as potentially protective against type II diabetes and coronary heart disease. Also, the study highlighted the complexity and importance of identifying common genome variants and their

effects. A more precise definition of the effect genome variants can contribute to the development of more reliable databases and avoid conflicting information describing the effects of the variant. In general, the identification of effect genome variants is important for a better understanding of the disease etiopathogenetic mechanisms and microevolutionary processes.

3.3. Association analysis of the effect genome variants in the cohort of Lithuanian clean-up workers of the Chernobyl nuclear disaster

During this stage of the study, an association analysis was performed for the LCWC and the general Lithuanian population (LITGEN) cohorts using genome-wide SNPs genotyping data. Targeted analysis of preselected effect genome variants from the compiled catalogue was performed.

The first part of the analysis was performed for the LCWC cohort, comparing genotyping data with the general Lithuanian population (LITGEN) cohort. After data quality control, 91 LCWC, 421 individuals from the general Lithuanian population groups, and 46 effect genome variants were set for further analysis. The frequency of genotypes statistically significantly differed between two cohorts for the genome variant rs3825942 (NM_005576.4:c.458G>A; NP_005567.2:p.(Gly153Asp)) in the *LOXLI* gene (Table 3.2). A comparison of allele frequencies did not show any statistical significance.

Table 3.2. Summary statistics of the variant rs3825942 in the gene *LOXLI*, which showed statistical significance. LCWC – Lithuanian clean-up workers of Chernobyl nuclear disaster, LITGEN – the general Lithuanian population from LITGEN project, CI – confidence interval.

Study group	AA genotype count	GA genotype count	GG genotype count	Minor Allele Frequency	Fisher's exact test, p-value	Odds Ratio (AA genotype)	Odds Ratio (GA genotype)
LCWC	1	35	55	0.203	p=0.019	0.34 [95% CI 0.04-2.63]	1.84 [95% CI 1.14-2.95]
LITGEN	16	104	301	0.162			

The *LOXLI* gene has important roles in maintaining cell homeostasis. Also, the incidence of exfoliation syndrome and glaucoma is associated with *LOXLI* risk and (or) protective genome variants. This analysis showed that the possible influence of recent positive selection [19], allele-flipping phenomenon (alternative allele is evaluated as a risk allele in some ethnic populations and as protective in others) [140, 148], and the fact that only individuals with a homozygous reference allele have glaucoma in the LCWC cohort suggests that the identified rs3825942 effect genome variant may be protective against exfoliation syndrome in this cohort. Because the identified

genomic variant is not the only genomic factor influencing the onset of exfoliation syndrome, LD and epistasis analyses have been performed to elucidate new genes involved in the molecular mechanisms of the disease together with the rs3825942 variant. Thus, these analyses have identified new candidate genes *LHFPL3*, *GALNT6*, *PIH1D1*, *ANKS1B*, *METRNL* potentially involved in the etiopathogenesis of exfoliation syndrome.

In the second part of the analysis, a comparison of LCWC with the men of the general Lithuanian population (LITGEN) was performed, dividing them by age groups. The study found that the two effect genome variants in the *CX3CR1* gene rs3732378 (NM_001337.4:c.839C>T, $\chi^2 = 7.146$, $p = 0.008$) and rs3732379 (NM_001337.4:c.745G>A, $\chi^2 = 5.958$, $p = 0.015$) statistically significantly differ from the general Lithuanian population (LITGEN) in the age group of 65 and older men. The frequency of the rare alleles in LCWC cohort is higher compared to the general Lithuanian population. Both of these variants are 94 base pairs apart, are in LD and inherited together [153]. These effect genome variants are known as protective, reducing the risk of coronary heart disease, and their protective effect has been demonstrated in a number of studies [153, 156, 157]. These effect genome variants are potentially protective in the LCWC cohort as well, leading to better adaptation qualities and potentially protecting against coronary heart disease. Nevertheless, cardiovascular diseases are one of the most common in the LCWC group. Coronary heart disease is multifactorial, so it is important to keep in mind the environmental factors (e.g. the individual's lifestyle) that influence the onset of disease. The main risk factors for coronary heart disease are smoking, fatty foods, low physical activity, stress, and depression [158]. According to the data of the questionnaire, in the group of LCWC most of these risk factors are common – 60% of LCWC are or were smoking, low physical activity was observed in LCWC (15% of individuals reported being engaged in sedentary work and leisure, 27% – in sedentary work and moderately active leisure). Mental disorders are observed in 40% of the individuals. Although the role of protective genome variants is significant, this significance may be overshadowed by a high-risk lifestyle in the LCWC cohort.

The third part of the analysis was performed for the LCWC cohort, comparing it with men aged 50–78 years from the general Lithuanian population (LITGEN) for all genome-wide SNPs genotyping data. Multiple testing correction methods, such as Bonferroni, Holms, and others were used for the association analysis for a large number of genome variants, and genes. No statistically significant differences between two cohorts and genome variants were identified using these methods. This does not necessarily indicate that there are no significant associations. One of the major drawbacks

of genome-wide association studies is multiple testing. As a result, the significance of associations may be lost, and potentially important data can be lost as well [159, 160]. This demonstrates the importance of targeted association analyses, where genomic regions of interest are specified and analyses are performed only for those specific regions.

3.4. Targeted, disease-associated gene analysis in the cohort of Lithuanian clean-up workers of the Chernobyl nuclear disaster

At this stage of the analysis, a targeted association analysis of disease-associated genes was performed by comparing the LCWC and the general Lithuanian population (LITGEN) cohorts. The disease groups were selected for the analysis based on their high frequency in LCWC group and taking into account the biological association of experienced risk factors (IR) with the analysed diseases.

First, association analysis with mental disorders was performed. Questionnaire data analysis showed that the predominant mental disorders in the LCWC group included anxiety, depression, emotional instability, and sleep disorders. As many as 37 out of 93 LCWC (40%) may have mental disorders. Meanwhile, according to the data of the Lithuanian Official Statistics Portal [assessed on March 2021, 162], depression was found in 5% of men in the general Lithuanian population in 2019, which is a significant difference between the study groups. On the other hand, 60% of the individuals did not have these conditions. Such results could potentially be driven by unique genome variation of LCWC. Association analysis was performed for individuals with and without mental disorders for all genome-wide SNPs genotyping data, and significant associations were identified. Significant SNPs were identified in 15 genes (Figure 3.3).

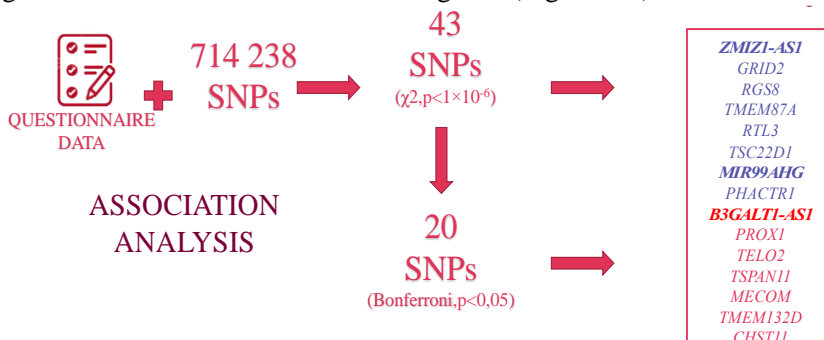


Fig 3.3. Scheme of the results of the association analysis with mental disorders in LCWC cohort. 43 genome variants were identified in 15 coding genes, reaching statistical significance ($p < 0.05$). After Bonferroni correction ($p < 7 \times 10^{-8}$), 20 SNPs were statistically significant (in the genes marked red). RNA genes are marked in bold.

In order to evaluate the possible functional significance of the identified variants and the genes encoding them, the *Humanbase* [118] *in silico* tool was used to analyse the gene expression in different tissues of the central nervous system (Appendix V) and to create gene interaction networks (Appendix VI). This analysis helped to identify genes *ZMIZ1* and *TSC22D1*, which may be involved in the etiopathogenesis of depression and other mental disorders. Also, these genes may have a protective effect, based on scientific literature analysis [165–168] and the fact that variants in these genes reached statistical significance in the LCWC cohort without mental disorders.

In the second part of this analysis, oral diseases were chosen as a disease group of interest. According to the collected data of the questionnaire, as many as 70% of LCWC (65 out of 93 individuals) have gastrointestinal disorders. The majority of health problems in this group were oral diseases, especially periodontitis and tooth loss. Almost half of LCWC (48%) have oral diseases, while approximately 12% of men aged 45 and over had dental diseases in 2020 according to the Lithuanian Institute of Hygiene [assessed on January 2022, 127]. Oral diseases are four times more common in the LCWC group compared to the general Lithuanian population. Such a high number might be related to the higher than usual dose of IR experienced during the Chernobyl nuclear disaster, because IR has an effect on bone homeostasis [171–174].

At this stage of the analysis of the study, a targeted association analysis of SNPs associated with oral diseases was performed. Based on the scientific literature and databases, a list of 909 SNPs in 59 genes was compiled (Appendix IV), which was used for further association analysis between LCWC with oral diseases and those without (analysis within the cohort). The analysis identified associations of *FGF1*, *FGF2*, *FGF7* and *BMP2* genes with a high risk of oral diseases. Three variants (rs2004330, rs1180341, rs1180343) in the gene *BMP8A* also reached statistical significance in the studied LCWC group. Interestingly, unlike the other identified genome variants, the odds ratio values for aforementioned variants varied between 0.395 and 0.43, i.e., they were lower than 1. This means that these genome variants are more common in the group of LCWC without oral diseases, possibly reducing the chance of having oral diseases and having potentially protective properties. This hypothesis is further supported by the fact that the protective role of *BMP8A* in glucocorticoid-induced bone apoptosis has been described in the scientific literature [178].

3.5. Analysis of the genome loci under positive natural selection in the cohorts of the general Lithuanian population and Lithuanian clean-up workers of the Chernobyl nuclear disaster

3.5.1. Evaluation of the impact of positive natural selection on the identified effect genome variants in the general Lithuanian population (LITGEN)

In order to confirm or negate the possible protective effect of the three discussed genome variants in the *PPARG*, *SLC30A8* and *ZC3HC1* genes identified in the general Lithuanian population (LITGEN) cohort, a positive natural selection signature analysis was performed using genotyping data. The main idea of this stage of the study is that if the identified statistically significant genome variants are beneficial to the cohorts' adaptation and have a protective function, they should be exposed to positive natural selection and the individuals should gain an advantage over other individuals in the population over time. The analysis calculates μ statistics that directly correlate with the effect of positive natural selection on the genome region. Statistical μ dot plots for the chromosomes of *PPARG*, *SLC30A8*, and *ZC3HC1* genes are shown in Figure 3.4. A significant positive natural selection signal was found in the *SLC30A8* gene ($\mu = 2.34$, $p < 0.05$). Although the detected signal is low compared to the signal in other regions on chromosome 8, it still confirms the potentially protective effect of this gene and the identified variant. Based on the analysis, *PPARG* and *ZC3HC1* genes are not under positive natural selection (μ values did not reach the threshold value of top 5% estimates). However, both of these genes are relatively close to other regions of the genome which are under positive natural selection (*PPARG* – 800 Kbp from chr3:11240480-11461626, $\mu = 1.42$, $p < 0.05$; *ZC3HC1* – 2 Mbp from chr7:128015754-128104838, $\mu = 1.59$, $p < 0.05$). This means that these regions of the genome are potential targets for further studies of positive natural selection.

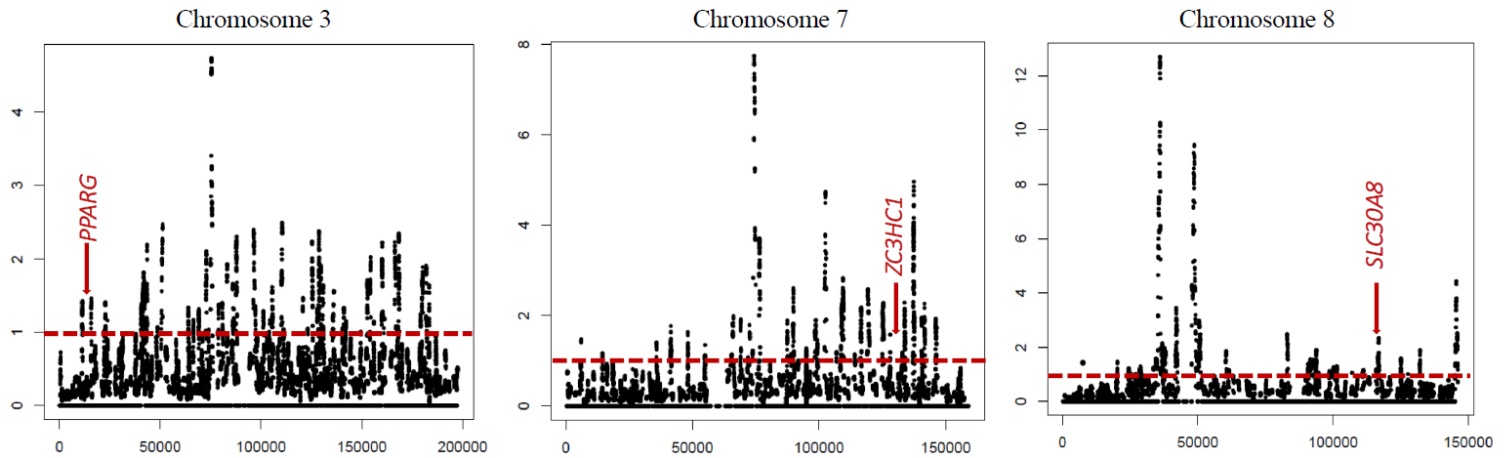


Figure 3.4. Dot plots of μ statistics for chromosomes 3, 7, and 8. The X-axis shows the genomic position and the Y-axis an estimate of the μ value. The dotted line represents the threshold value of μ , meaning that μ values above the line are higher than top 5% of all μ values on that chromosome ($p < 0.05$). The highest statistical values of μ indicate the strongest regions of the genome affected by positive natural selection. The position of the candidate protective genes *PPARG* (chromosome 3), *ZC3HC1* (chromosome 7) and *SLC30A8* (chromosome 8) and the effect of positive natural selection in that region are indicated by arrows.

3.5.2. Evaluation of the impact of positive natural selection on the identified effect genome variants in the cohort of Lithuanian clean-up workers of the Chernobyl nuclear disaster

The results discussed in the previous stages of the study showed gene associations with exfoliation syndrome (*LOXLI*), coronary heart disease (*CX3CR1*), mental disorders (*ZMIZI*, *TSC22D1*), and oral diseases (*FGF1*, *FGF1*, *FGF7*, *BMP2*, *BMP8A*). This study found that some variants in the listed genes may lead to protection against these diseases and, in turn, determine the better survival and adaptive properties of LCWC (variants in the genes *LOXLI*, *CX3CR1*, *ZMIZI*, *TSC22D1*, *BMP8A*). Other variants have been shown to be possibly associated with poor oral health in the LCWC cohort (variants in the genes *FGF1*, *FGF2*, *FGF7*, *BMP2*). In order to substantiate the effect of the identified genes and their variants for the LCWC phenotype, a positive natural selection analysis was performed. The analysis of positive natural selection has a direct relationship with the usefulness of the genomic loci for the health and adaptation of the individual – the more useful the genomic loci, the more it is affected by positive natural selection. At this stage of the study, the positive natural selection effects on these genes will be reviewed in detail.

For the *LOXLI* gene region (chr15:73926462-73952136), low positive natural selection signature value ($\mu = 1.14$) was observed (Figure 3.5), however insignificant ($p \geq 0.05$).

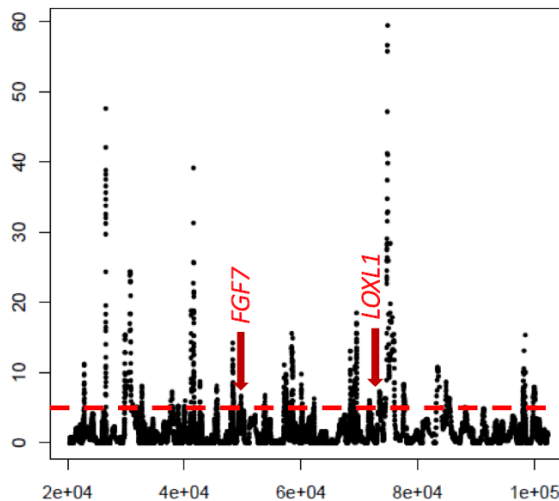


Fig 3.5. Dot plot of μ statistics for chromosome 15. The X-axis shows the genomic position and the Y-axis an estimate of the μ value. The dotted line represents the threshold value of μ ($\mu = 5.53$), meaning that μ values above the line are higher than top 5% of all μ values on that chromosome ($p < 0.05$). The highest statistical values

of μ indicate the strongest regions of the genome affected by positive natural selection. Analysed genes loci are indicated by arrows.

On the other hand, this gene belongs to a region (chr15:73049046-73980207) with a significant positive natural selection signature value (highest value $\mu = 7.30$, $p < 0.05$). In addition, other, the most significant, region with the highest μ value in the chromosome is next to the *LOXLI* gene region (chr15:74491439-75982958, $\mu = 59.5$, $p < 0.05$). The fact that these genomic regions are affected by positive natural selection confirms the hypothesis of the potentially protective role of the *LOXLI* gene and the investigated variant of rs3825942 in the etiopathogenesis of exfoliation syndrome. However, further functional analysis is needed.

For the *FGF7* gene region (chr15:49423237-49488775), a small effect ($\mu = 4.99$) of positive natural selection was observed when compared to other loci under positive selection on chromosome 15 (Figure 3.5). This positive selection signature value did not reach significance in the gene region analysed ($p \geq 0.05$). However, this region is between two other significant regions under positive natural selection (chr15:48232971-48435566, $\mu = 11.17$; chr15:49692283-49719324, $\mu = 6.64$; $p < 0.05$). As discussed in the previous sections, the association analysis found that the genomic variant of the *FGF7* gene, rs16962440, is more common in the cohort of LCWC with oral diseases, suggesting that this variant is associated with a potentially increased risk. On the other hand, the results of this stage of the study suggest a potentially protective function of the *FGF7* gene, as the region of the *FGF7* gene is a potential target for positive natural selection. Inconsistent findings suggest that conclusions about the significance of effect genome variants and their effects on the phenotype cannot be drawn by association and *in silico* analyses alone. In order to draw meaningful conclusions, several analyses are required that would combine different analytical methods, which has been the aim of this study as well. Also, functional genome studies using model organisms or studying gene expression in tissue or cell cultures is crucial to confirm the significance of effect genome variants.

In this case of *FGF7*, gene associations with not only oral diseases but also other conditions cannot be ruled out. Such genes often play a role and are active in the molecular mechanisms of several multifactorial diseases. An analysis of gene networks showed that the *FGF7* gene is the least associated with oral diseases compared to other genes studied. Therefore, it cannot be ruled out that this gene is associated with another trait and a possible protective effect on that trait in the LCWC group. For example, an association

of *FGF7* with better liver regeneration has been identified through the regulatory mechanism of liver progenitor cells [179].

Based on the analysis of this study, *CX3CR1* (chr3:39263495-39281735), *ZMIZ1* (chr10:78943328-79067895), *TSC22D1* (chr13:44432143-44577147), *FGF1* (chr5:142592178-142698070), *FGF2* (chr4:122826708-122898236), *BMP2* (chr20:6767686-6780246), *BMP8A* (chr1:39491636-39529869) genes do not fall into significant loci affected by positive natural selection. Although these regions of the genome are not protective against coronary heart disease, mental disorders, and oral health, respectively, they are likely to play a role in the etiopathogenesis of these diseases, and further, functional studies of these genes are needed. In addition, although these loci are not affected by positive natural selection, the genomic variants studied in some of these loci are in conserved genomic regions based on the *UCSC genome browser* and the *Varsome* analysis tool (*CX3CR1* gene *GERP* score 2.49, *PhyloP100way* score 1.89; *TSC22D1* *GERP* score 2.37, *PhyloP100way* score 1.29, *BMP2* *GERP* score 1.82, *PhyloP100way* score 0.865). Positive values of these estimates indicate regional conservatism. The found conserved regions also partially support the importance of these genomic regions in the etiopathogenesis of aforementioned diseases.

3.5.3. Identification of genome loci under positive natural selection and assessment of loci significance for adaptative qualities of the cohort of Lithuanian clean-up workers of Chernobyl nuclear disaster

At this stage of the study, an evaluation of the positive natural selection signatures in the LCWC cohort was performed using whole genome sequencing data for the first time. The aim was to identify genomic loci of autosomal chromosomes which have the highest positive natural selection signature values and are unique to the LCWC cohort as compared to the general Lithuanian population (ANELGEMIA). The obtained results intended to prove the significance of these genomic regions for the adaptation of the LCWC cohort.

In the genomes of LCWC, the loci with the highest positive selection signature values were identified for chromosomes 5 ($\mu = 164.2$, Figure 3.6), 11 ($\mu = 114$), and 12 ($\mu = 161.1$) (Table 3.3, Appendix VII). Chromosomes 9, 13, 18, 21 do not contain genes in the identified regions under positive natural selection; therefore, these regions have not been analysed further.

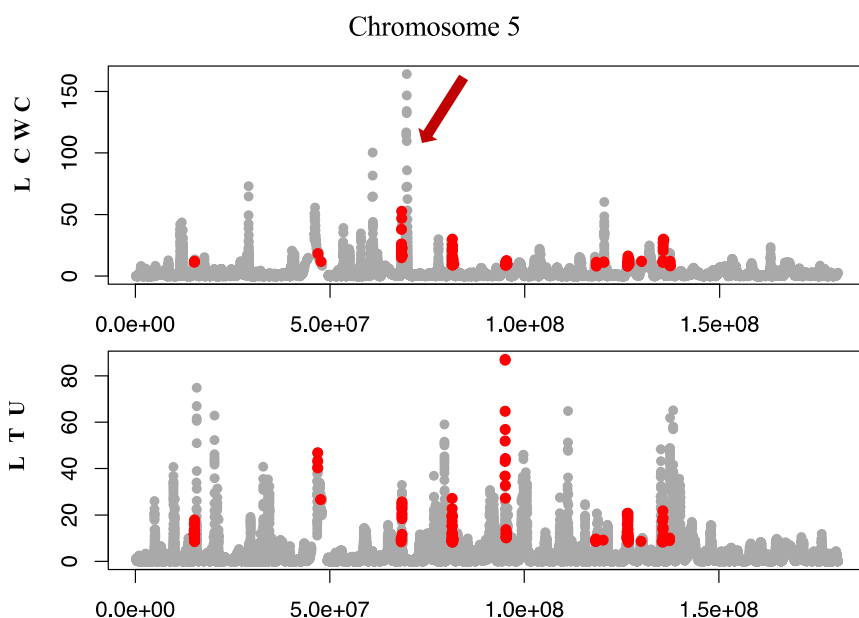


Figure 3.6. Analysis of the loci under positive selection on chromosome 5. The cohorts of the Lithuanian clean-up workers of Chernobyl nuclear disaster (LCWC) and the general Lithuanian population (LTU; ANELGEMIA data) were compared. On the X axis the genomic position and on the Y axis the positive selection signature μ value are depicted. The positive selection loci unique to each group are marked grey, and the common loci are marked red. An arrow marks the positive selection locus unique to LCWC which has highest μ value, meaning it is the most significant. The results of the analysis of positive natural selection for other chromosomes are given in Appendix VII.

Table 3.3. Autosomal regions of the genome under positive natural selection.

Chr	Genomic locus position		Highest μ value	Genes in the genomic loci
	Start	End		
1	50024718	51586980	49.96	<i>AGBL4, ELAVL4, DMRTA2, FAF1, CDKN2C</i>
2	189198631	189731344	86.85	<i>GULP1, DIRC1</i>
3	68678446	68990461	77.44	<i>TAF4A</i>
4	146494457	147354697	72.27	<i>MMAA, C4orf51, ZNF827, LSM6, SLC10A7</i>
5	68316218	69932025	164.20	<i>SLC30A5, CCNB1, CENPH, MRPS36, CDK7, CCDC125, AK6, TAF9, MARVELD2, OCLN, GTF2H2C, SERF1B, SMN2, NAIP</i>
6	64047472	64615929	56.06	<i>PTP4A1, PHF3, EYS</i>
7	151965978	152132980	28.44	<i>KMT2C</i>
8	104888906	107027425	36.35	<i>RIMS2, DCSTAMP, DPYS, LRP12, ZFPM2</i>
9	17876558	18329195	54.64	-
10	122646955	123081831	41.73	<i>WDR11</i>

Chr	Genomic locus position		Highest μ value	Genes in the genomic loci
	Start	End		
11	101273279	101740156	114.00	<i>TRPC6</i>
12	109191005	111322770	161.10	<i>SSH1, DAO, USP30, ALKBH2, UNG, ACACB, MYO1H, KCTD10, MMAB, UBE3B, MVK, TRPV4, GIT2, GLTP, ANKRD13A, C12orf76, IFT81, ATP2A2, ANAPC7, GPN3, ARPC3, FAM216A, VPS29, PPTC7, RAD9B, TCTN1, HVCN1, PPP1CC, CCDC63</i>
13	82488223	83468360	30.54	-
14	73733462	73867298	61.43	<i>PAPLN, NUMB, AX747833</i>
15	74491439	75982958	59.46	<i>CCDC33, CYP11A1, SEMA7A, UBL7, ARID3B, CLK3, EDC3, CYP1A1, CSK, ULK3, SCAMP2, FAM219B, MPI, COX5A, RPP25, SCAMP5, PPCDC, C15orf39, GOLGA6C, MAN2C1, COMM4, NEIL1, SIN3A, PTPN9, SNUPN</i>
16	74347008	74795167	29.69	<i>NPIP15, CLEC18B, GLG1, RFWD3, MLKL, FA2H</i>
17	49795141	50541571	27.98	<i>CA10</i>
18	68939522	68991817	29.70	-
19	22058047	23006378	72.65	<i>ZNF208, ZNF257, ZNF676, ZNF729, ZNF98, ZNF492, ZNF99</i>
20	14021037	14066814	22.27	<i>MACROD2</i>
21	23193282	23598744	41.25	-
22	21806118	22653232	45.96	<i>TMEM191C, PI4KAP2, UBE2L3, CCDC116, SDF2L1, PPIL2, YPEL1, MAPK1, PPM1F, TOP3B</i>

A total of 123 genes under positive natural selection, specific to the LCWC group were identified. The analysis of the biological function of these genes is based on the *PANTHER* classification system [123]; the *Gene Ontology Browser* [124] showed that most genes are involved in cellular, metabolic, and biological regulatory processes (Figure 3.7).

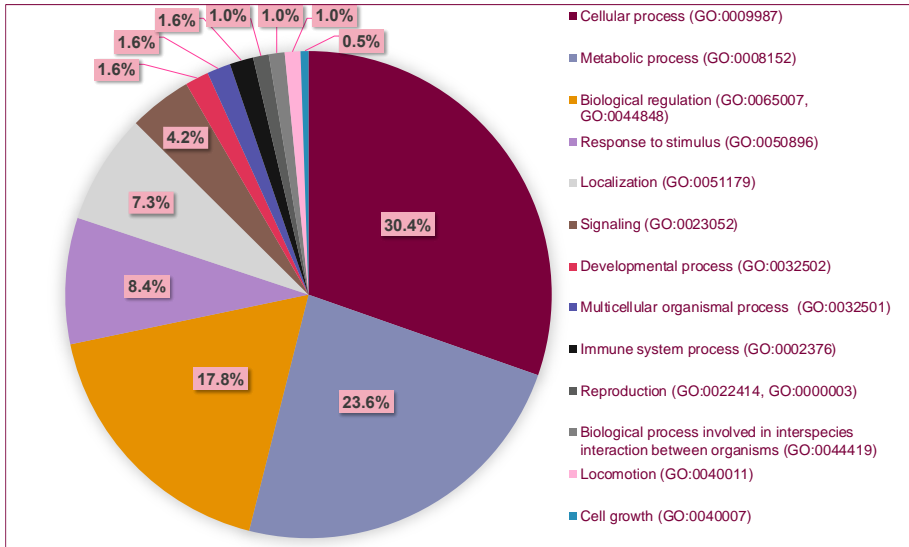


Fig. 3.7. The biological role of genes in the genomic loci under positive natural selection. Processes in the legend are listed in descending order of gene number involved.

The category of cellular processes includes processes that are important for cell viability, such as cell division, cell cycle regulation, and cell death. Metabolic processes include chemical reactions in cells and the genes involved in those reactions. Macromolecular processes such as DNA damage repair or DNA replication also fall into this category. The category of biological regulation involves a broad spectrum of processes, encompassing not only the regulation of biological processes in certain tissues and organs, but also regulation at the molecular level, such as the maintenance of DNA stability or transcription. Some identified genes perform multiple biological functions and therefore fall into several categories. In this analysis, positive natural selection in LCWC genomes has been shown to affect those regions of the genome that are involved in processes important for cell survival (cell aging, cell cycle regulation, apoptosis) and adaptation (repair of DNA damage).

The Lithuanian clean-up workers of the Chernobyl nuclear disaster suffered higher than usual doses of IR and encountered extreme and stressful situations. Due to the fact that LCWC have survived and adapted to the consequences of the Chernobyl nuclear disaster, the genes involved in the DNA damage repair and the regulation of the cell cycle are of paramount importance for their adaptation. The identified genes involved in the aforementioned processes are: *CDKN2C*, *ZNF827*, *CENPH*, *CCNB1*, *CDK7*, *TAF9*, *RAD17*, *GTF2H2C*, *PTP4A1*, *WDR11*, *ALKBH2*, *UNG*, *ARID3B*, *COMMD*, *MAPK1*, *PPM1F*, *TOP3B* (Table 3.4). These genes not only are

important in the mechanisms of DNA damage repair and cell cycle regulation but also are subject to positive natural selection and may form unique adaptive properties of LCWC.

Table 3.4. Genes important for the adaptation of LCWC, their function, and (or) impact on the phenotype. Function or impact on the phenotype was determined from data provided in the *UCSC* genome browser, *OMIM* and *Human DNA Repair Genes* databases, and analysis of scientific publications.

Chr	Gene	Function and (or) impact on phenotype
1	<i>CDKN2C</i>	Cell cycle G1 control, tumour suppression, cancer, endocrine dysplasia
4	<i>ZNF827</i>	Reparation of telomeres, associated with alcohol-related liver disease
5	<i>CENPH</i>	Forms centromere-kinetochore complex, cancer
5	<i>CCNB1</i>	Cell cycle G2-M control, cancer
5	<i>CDK7</i>	Cell cycle and transcription, nucleotide excision repair
5	<i>TAF9</i>	Cell cycle, apoptosis, DNA transcription
5	<i>RAD17</i>	DNA lesions-dependent recombination, DNA damage sensor.
5	<i>GTF2H2C</i>	DNA reparation
6	<i>PTP4A1</i>	Cell cycle G1-S control, cancer, systemic sclerosis, alcohol dependence
10	<i>WDR11</i>	Cell cycle progression, apoptosis, signalling, transcription regulation, associated with obesity intellectual disability, cancer, heart anomalies
12	<i>ALKBH2</i>	Direct reversal of DNA damage
12	<i>UNG</i>	Reparation of telomeres, base excision repair
15	<i>ARID3B</i>	Cell cycle control, transcription regulation, regulation of chromatin structure
15	<i>COMMD4</i>	Repair of double-stranded DNA breaks
15	<i>NEIL1</i>	Base excision repair
16	<i>RFWD3</i>	DNA lesions repair in G1 phase of cell cycle
20	<i>MACROD2</i>	Potential function in DNA repair, associated with cancer
22	<i>MAPK1</i>	Proliferation, differentiation, transcription regulation
22	<i>PPM1F</i>	Regulation of cellular stress response
22	<i>TOP3B</i>	Genome stability, DNA recombination, aging

The *PopHumanScan* database [122], which provides systematic information on known loci under positive selection in different populations, has been used to further investigate the effects of identified loci under positive natural selection in the LCWC cohort. For five genes (*WDR11*, *NEIL1*, *RFWD3*, *MACROD2*, *MAPK1*) involved in DNA damage repair and cell cycle, the effect of positive natural selection was also observed in other populations. This demonstrates the significance of these regions for human adaptation in general, regardless of population. On the other hand, the influence of positive natural selection for other analysed genes in other world populations and in the general Lithuanian population (ANELGEMIA) was not

determined or was insignificant. This shows that LCWC is possibly characterized by specific adaptation mechanisms that determine their survival and adaptation to the lifelong consequences of the Chernobyl nuclear disaster.

Also, other regions of the genome affected by positive natural selection that may affect the adaptive properties of LCWC have been identified. The *DAO* gene is in the genomic loci affected by positive natural selection in chromosome 12. This gene encodes a peroxisomal enzyme required for neuronal differentiation, dopamine synthesis [180]. This gene is associated with aging and may have a neuroprotective effect [181]. Also, this gene is associated with the oxidative stress response [182], which is also experienced by individuals receiving high doses of IR, and therefore may play a role in the response to oxidative stress and better adaptation in the LCWC cohort. In the same region of positive natural selection is the *PPTC7* gene, which is associated with resistance to environmental chemical toxins [183]. The influence of positive natural selection in the South Asian (Indian) population was also found in this gene locus (Fu and Li D score -6.825, $p < 0.05$) [122]. India is the third country in the world in terms of pollution [184], so it is natural for the human body to adapt to the harmful conditions that surround it. The observed effect of positive natural selection in the *PPTC7* gene in different cohorts of survivors of toxic conditions shows the influence of this locus on human adaptation to toxins.

Other genomic loci affected by natural selection that may be important for the adaptation of LCWC are related to the results obtained in the previous stages of this study. The identified genes *FAF1*, *UBL7*, *PI4KAP2* are affected by positive natural selection. These genes are associated with comorbid depressive syndrome [185], exfoliation syndrome [186], and post-traumatic stress disorder [187], respectively. This result complements the previously identified gene associations characteristic of the exfoliation syndrome and mental health conditions and suggests a possible overall mechanism of action.

It has also been found that many of the genes affected by positive natural selection identified at this stage of the study are associated with cancer. The fact that these genome regions are under positive selection, and that cancer is rare in the studied group of LCWC, suggests a possible protective function of these loci in the context of cancer and involvement in adaptation mechanisms that determine LCWC survival.

The strategy of combining association, positive natural selection signature and functional *in silico* analysis methods allowed to identify significant associations, genome variants and loci that are possibly functionally significant for the general Lithuanian population as well as important for the adaptation and survival of the LCWC. The results obtained could be refined and the research strategy could be enriched by functional *in vivo* studies (in

cell cultures, tissues or model organisms). It would also be beneficial to evaluate the effect of positive natural selection on the genomic loci of LCWC using traditional methods (for example, F_{ST} , integrated haplotype score, etc.) and comparing the obtained results not only within the population, but also to other populations. This could be the direction of future research.

4. CONCLUSIONS

1. The characterization of the genome of the general Lithuanian population in the context of effect genomic variants, using whole-genome genotyping method, allowed the identification of genomic variants characteristic of the Lithuanian population potentially lead to a reduced predisposition to type II diabetes (NM_001172815.2(*SLC30A8*):c.826C>T) and coronary heart disease (NM_001282190.1(*ZC3HC1*):c.1025G>A) compared to European (Finnish, Utah of Northern or Western European ancestry) populations.
2. A genome-wide genotyping method has been used to determine a unique genome variation of DNA sequence changes in the cohort of Lithuanian clean-up workers of the Chernobyl disaster, which is statistically significantly different from the general Lithuanian population. The effect genome variants that potentially determine clean-up workers' adaptive properties were identified.
 - 2.1. Compared to the general Lithuanian population, a statistically significant genome variant NM_005576.4(*LOXLI*):c.458G>A ($p=0.019$) was identified, which is possibly protective against the exfoliation syndrome and glaucoma in the cohort of Lithuanian clean-up workers of the Chernobyl nuclear disaster. New candidate genes (*LHFPL3*, *GALNT6*, *PIH1D1*, *ANKS1B*, *METRNL*) potentially influencing the etiopathogenesis of exfoliation syndrome and glaucoma were identified.
 - 2.2. Two statistically significant effect genome variants NM_001337.4:c.839C>T ($p=0.008$) and NM_001337.4:c.745G>A ($p=0.015$) were identified in the *CX3CR1* gene, possibly protecting against coronary heart disease in the cohort of Lithuanian clean-up workers of the Chernobyl nuclear disaster as compared to the group of the general Lithuanian population males aged over 65 years. Nevertheless, cardiovascular diseases are one of the most common disease groups in the cohort of Lithuanian clean-up workers of the Chernobyl nuclear disaster; their lifestyle habits may influence their morbidity even though the protective factors are present.

- 2.3. Targeted association analysis of disease-related genes helped to identify 43 statistically significant ($p < 1 \times 10^{-6}$) SNPs in 15 genes (the strongest candidate genes being *ZMIZ1* and *TSC22D1*), possibly influencing the high frequency of mental disorders in the cohort of Lithuanian clean-up workers of the Chernobyl nuclear disaster. Also, new gene associations of *FGF1*, *FGF2*, *FGF7* and *BMP2* have been identified with an increased risk of oral diseases for Lithuanian clean-up workers of the Chernobyl nuclear disaster. The interactions of all identified genes and their variation may affect the phenotype of the cohort, forming unique population properties.
3. In order to assess the protective function of the effect genome variants identified during the association analysis, a positive natural selection signature analysis of the identified effect genome variant loci was performed for the general Lithuanian population and Lithuanian clean-up workers of the Chernobyl nuclear disaster. Significant positive natural selection signals were identified in the *SLC30A8* gene (in the general Lithuanian population, $\mu=2.34$, $p < 0.05$), *LOXL1* (in the cohort of Lithuanian clean-up workers of the Chernobyl nuclear disaster, $\mu=7.30$, $p < 0.05$). These signals point towards the protective role of these genes and genomic variants identified. The complex research analysis strategy selected for this study, which combined association, positive natural selection signature, and *in silico* analysis methods, allows a deeper assessment of the functional significance of the variant.
4. The 22 most significant, autosomal genomic loci under positive selection have been identified using whole-genome sequencing data for the first time. These loci are unique to the cohort of Lithuanian clean-up workers of the Chernobyl nuclear disaster in comparison with the general Lithuanian population. *CDKN2C*, *ZNF827*, *CENPH*, *CCNB1*, *CDK7*, *TAF9*, *RAD17*, *GTF2H2C*, *PTP4A1*, *WDR11*, *ALKBH2*, *UNG*, *ARID3B*, *COMMD4*, *NEIL1*, *RFWD3*, *MACROD2*, *MAPK1*, *PPM1F*, *TOP3B* genes have been identified in the aforementioned loci, which are involved in the regulation of the cell cycle and are important for the adaptation to the lifelong effects of the Chernobyl nuclear disaster in the cohort of Lithuanian clean-up workers of the Chernobyl nuclear disaster.

PUBLICATIONS AND PRESENTATIONS

Publications related to the topic of the thesis:

1. **Žukauskaitė** G, Domarkienė I, Matulevičienė A, Vaitėnienė EM, Arasimavičius J, Smailytė G, Kučinskas V, Ambrozaitytė L. Possible Protective Effect of LOXL1 Variant in the Cohort of Chernobyl Catastrophe Clean-Up Workers. *Genes* 2021; 12(8): 1231. DOI: 10.3390/genes12081231.
2. **Žukauskaitė** G, Domarkienė I, Matulevičienė A, Kučinskas V, Ambrozaitytė L. Identifying Genomic Signatures of Positive Selection to Predict Protective Genomic Loci in the Cohort of Lithuanian Clean-Up Workers of the Chornobyl Nuclear Disaster. *Current Issues in Molecular Biology* 2023; 45(4), 2972-2983. DOI: 10.3390/cimb45040195.

Oral presentations:

1. **Žukauskaitė** G, Ambrozaitytė L, Domarkienė I, Rančelis T, Kučinskas V. Analysis of protective genome variants in Lithuanian population. In: „Bioateitis: gamtos ir gyvybės mokslų perspektyvos“, 2019 December 11., Kaunas, Lithuania.
2. **Žukauskaitė** G, Domarkienė I, Matulevičienė A, Arasimavičius J, Kučinskas V, Ambrozaitytė L. Genome studies of Chernobyl disaster clean-up workers: variants potentially leading to glaucoma prevention. In: „Bioateitis: gamtos ir gyvybės mokslų perspektyvos“, 2020 December 20, virtual conference.
3. **Žukauskaitė** G, Domarkienė I, Matulevičienė A, Vaitėnienė EM, Arasimavičius J, Kučinskas V, Ambrozaitytė L. Genome association study of the Lithuanian Chernobyl catastrophe liquidators indicates potentially protective alleles for multifactorial diseases. In: Conference on Radiation in Various Fields of Research, RAD2021), 2021 June 14–18, Herceg Novi, Montenegro.
4. **Žukauskaitė** G, Domarkienė I, Matulevičienė A, Arasimavičius J, Kučinskas V, Ambrozaitytė L. A large-scale genetic association study of mental disorders in a group of Lithuanian Chernobyl disaster clean-up workers. In: „Bioateitis: gamtos ir gyvybės mokslų perspektyvos“, 2021 November 25, Kaunas, Lithuania. Best presentation award.
5. **Žukauskaitė** G, Domarkienė I, Kučinskas V, Ambrozaitytė L. Genomic Signatures of Adaptation in the Cohort of Chernobyl Catastrophe Clean-up Workers from Lithuania. In: „Population Genetics from the Baltic Perspective: Microevolutionary Processes of the Lithuanian Genomes“, 2022 May 20–21, Vilnius, Lithuania.

6. Domarkienė I, **Žukauskaitė** G, Kučinskas V, Ambrozaitytė L. Inferring Whole-genome Sequencing Structural Variation Data in the Lithuanian Population. In: „Population Genetics from the Baltic Perspective: Microevolutionary Processes of the Lithuanian Genomes“, 2022 May 20–21, Vilnius, Lithuania.
7. **Žukauskaitė** G, Domarkienė I, Matulevičienė A, Kučinskas V, Ambrozaitytė L. Analysis of genomic regions affected by positive natural selection in the group of Lithuanian Chernobyl disaster clean-up workers. In: „Bioateitis: gamtos ir gyvybės mokslų perspektyvos“, 2022 November 24, Vilnius, Lithuania. Best presentation award.

Poster presentations:

1. **Žukauskaitė** G, Ambrozaitytė L, Domarkienė I, Rančelis T, Kučinskas V. Frequency analysis of potentially protective genome variants in the Lithuanian population. In: ESHG conference, 2020 June 6–9, virtual conference. (European Journal of Human Genetics 2020; 28(1).) Nr. P19.080.C.
2. Domarkienė I, **Žukauskaitė** G, Matulevičienė A, Arasimavičius J, Kučinskas V, Ambrozaitytė L. ADAPT – genomic initiative to analyse adaptation in the group of Lithuanian Chernobyl liquidators. In: Radiation in Various Fields of Research, RAD 2021, 2021 June 14–18, Herceg Novi, Montenegro.
3. **Žukauskaitė** G, Domarkienė I, Matulevičienė A, Vaitenienė EM, Arasimavičius J, Kučinskas V, Ambrozaitytė L. LOXL1 risk variant exhibits protective effect for exfoliation syndrome and glaucoma in Lithuanian Chernobyl catastrophe liquidators’ cohort. In: ESHG conference, 2021 August 28–31, virtual conference. Nr. P24.048.C.
4. Matulevičienė A, **Žukauskaitė** G, Domarkienė I, Kučinskas V, Ambrozaitytė L. Association of Genomic Factors for Oral Health in the Cohort of the Lithuanian Chernobyl Catastrophe Liquidators. In: ESHG conference, 2021 August 28–31, virtual conference. Nr. P13.003.D.
5. Domarkienė I, **Žukauskaitė** G, Matulevičienė A, Urnikytė A, Kučinskas V, Ambrozaitytė L. CNV characteristics in Chernobyl power plant catastrophe clean-up workers from Lithuania suggest unique genetic variation structure. In: ASHG conference, 2021 October 18–22, virtual conference.
6. **Žukauskaitė** G, Domarkienė I, Matulevičienė A, Vaitenienė EM, Arasimavičius J, Kučinskas V, Ambrozaitytė L.. Detection of recent positive selection signatures in the cohort of Lithuanian Chernobyl

catastrophe liquidators. In: ESHG conference, 2022 June 11–14, Vienna, Austria. Nr. EP20.002.

7. Domarkienė I, **Žukauskaitė** G, Urnikytė A, Pranckėnienė L, Dauengauer-Kirlienė S, Molytė A, Matulevičienė A, Pilypienė I, Ambrozaitytė L, Kučinskas V. Whole-genome sequencing data reveals higher number of structural variants in Chernobyl catastrophe clean-up workers from Lithuania. In: ESHG conference, 2022 June 11–14, Vienna, Austria. Nr. P.14.010.A.

8. Samaitė K, Domarkienė I, **Žukauskaitė** G, Kučinskas V, Ambozaitytė L. Possibly protective genomic variants against ionizing radiation in DNA homologous recombination repair genes. In: ESHG conference, 2022 June 11–14, Vienna, Austria. ESHG conference, 2022 June 11–14, Vienna, Austria. Nr. EP25.007.

9. **Žukauskaitė** G, Domarkienė I, Kučinskas V, Ambrozaitytė L. Genomic positive selection signature analysis in the cohort of the Lithuanian Chernobyl catastrophe clean-up workers. In: Joint EAA-ISGA-ICEM conference, 2022 August 24–27, Vilnius, Lithuania. Nr. 128.

PADEKA

Nuoširdžiai dėkoju visiems, prisidėjusiems prie šio darbo, pasidalijusiems savo patirtimi ir padėjusiems man eiti šiuo nelengvu, tačiau įdomiu, įtraukiančiu ir intriguojančiu genetikos mokslo keliu.

Akad. prof. habil. dr. Vaidučiui Kučinskui – už pasiūlytą tyrimo idėją, parodytą pasitikėjimą ir suteiktą galimybę vykdyti šį tyrimą.

Doc. dr. Laimai Ambrozaitytei – už mentorystę nuo pirmųjų žingsnių genetikos mokslo srityje, kantrybę, palaikymą, racionalumą sudėtingose situacijose ir pagalbą visokeriopais moksliniais klausimais. Nuoširdžiai dėkoju už pasitikėjimą ir suteiktą galimybę vykdyti šį tyrimą.

Dr. Ingridai Domarkienei – už visokeriopą pagalbą, dalijimąsi žiniomis, kantrybę, palaikymą ir nuoširdumą.

Dėkoju visiems tiriamiesiems, skyrusiems laiko ir sutikusiems tapti šio tyrimo dalimi.

Doc dr. Aušrai Matulevičienei ir med. gyd. Evelinai Marijai Vaitėnienei – už reikšmingą pagalbą vertinant Lietuvos Černobylio katastrofos likviduotojų sveikatos būklę.

Recenzentams: prof. dr. Loretai Cimbalsienei, prof. habil. dr. Juozui Rimantui Lazutkai, prof. dr. Audronei Jakaitienei, dr. Alinai Urnikytei – už indėlį, kad darbas būtų geresnis.

Kolegėms Vilijai Deveikienei ir Dalytei Pliaugo – už pagalbą ir atgaivintus DNR gryninimo įgūdžius.

Daliai Kuzminskienei už vertingą pagalbą sprendžiant doktorantūros studijų metu iškilusius klausimus.

Dr. Valentinai Ginevičienei – už pagalbą vykdant likviduotojų apklausą.

Dr. Tautvydui Rančeliui ir dr. Nikolaos Alachiotis – už pagalbą įsisavinant darbo bioinformatikos įrankiais ypatumus, Justui Arasimavičiui – už parengtą duomenų bazę.

Kolegoms ir bičiuliams Karoliui Baronui, Gabijai Mazur, Mildai Petkevičienei, Jolantai Savickienei, Jovitai Latvytei, dr. Eglei Mazgelytei, dr. Aušrai Morkūnienei – už nuoširdumą, palaikymą, netilstantį juoką ir galimybę dalytis iškilusiais sunkumais.

Mylimiausiems ir artimiausiems – mamai Anželai, tėčiui Arūnui, broliui Lukui, močiutei Marytei, Gotfridui, Jokūbui, Ramintai, Justei, Augustei, Justinai, Darijai, Sabinai – dėkoju už besąlygišką palaikymą ir pagalbą siekiant svajonių, bendrą juoką ir ašaras.

Be Jūsų nebūčiau ten, kur esu.

CURRICULUM VITAE

Vardas, Pavardė	Gabrielė Žukauskaitė	
Gimimo data	1992 10 18	
El. paštas	gabriele.zukauskaite@mf.vu.lt	
Išsilavinimas		
2018-2023	Vilniaus universiteto Medicinos fakultetas, Doktorantūra	
2015-2017	Vilniaus universiteto Medicinos fakultetas, Medicinos genetikos magistrantūra (<i>Cum Laude</i> diplomas)	
2011-2015	Lietuvos sveikatos mokslų universiteto Medicinos fakultetas, Medicininės ir veterinarinės genetikos bakalauras	
2007-2011	Vilniaus Simono Daukanto gimnazija	
Profesinė patirtis		
2017-dabar	Vilniaus universiteto ligoninės Santaros klinikų Medicininės genetikos centras	Medicinos genetikė
2020-dabar	Vilniaus universiteto Medicinos fakulteto Biomedicinos mokslų instituto Žmogaus ir medicininės genetikos katedra	Jaunesnioji mokslo darbuotoja
2017-2018		Lektorė
2016-2018		Laborantė
Moksliniai tyrimai		
<ol style="list-style-type: none"> 1. Functional model for the mitochondrial disease evaluation and biomarker development. TAP-LLT-02/2015. 2. Adaptive genetic mechanisms – a comprehensive study of whole genome variation in the group of the Lithuanian Chernobyl catastrophe liquidators (S-MIP-20-35, ADAPT). Lietuvos mokslo tarybos remiamą veiklos kryptį „Mokslininkų grupių projektai“. 		
Apdovanojimai		
<ol style="list-style-type: none"> 1. 14th Baltic Congress Of Laboratory Medicine 2018 – Geriausias standinis pranešimas „Evaluation of Automated Next Generation Sequencing Data Analysis Pipeline“. 2. Jaunųjų mokslininkų konferencija „Bioateitis: gamtos ir gyvybės mokslų perspektyvos“ 2021 – Geriausias žodinis pranešimas (3 vieta) „Psichikos sutrikimų plataus masto genetinės asociacijos tyrimas Lietuvos Černobylio katastrofos likviduotojų grupėje“. 3. Jungtinė EAA-ISGA-tarptautinė konferencija 2022 – Geriausio doktoranto elektroninio stendinio pranešimo apdovanojimas „Genomic positive selection signature analysis in the cohort of the Lithuanian Chernobyl catastrophe clean-up workers“. 4. Jaunųjų mokslininkų konferencija „Bioateitis: gamtos ir gyvybės mokslų perspektyvos“ 2022 – Geriausias žodinis pranešimas (1 vieta) „Teigiamos gamtinės atrankos veikiamų genomo sričių analizė Lietuvos Černobylio katastrofos likviduotojų grupėje“. 		
Kvalifikacijos kėlimas		
Mokslinė-praktinė konferencija "Retų ligų gydymo inovacijos"		2018
Konferencija „Pasaulinė Dauno sindromo diena Lietuvoje“		2018
Kongresas „14th Baltic Congress in Laboratory Medicine “		2018
Mokymai "What to do when I want to go for a research internship".		2019

Mokslinė-praktinė konferencija “Dismorfologijos ruduo 2019”	2019
Mokymai “Good Laboratory Pipetting and Maintenance, pipette calibration concepts. Hands on training in customers laboratory”	2019
Mokymai “Organizing & maximizing rare disease biological sample data in biobanks”	2019
Jaunųjų mokslininkų konferencija "Bioateitis: gamtos ir gyvybės mokslų perspektyvos”	2019
Seminaras "Modern diagnosis, treatment and monitoring of hereditary metabolic diseases"	2020
Europos žmogaus genetikos draugijos (ESHG) virtuali konferencija	2020
Mokslinė-praktinė konferencija "COVID-19 issues (1)"	2020
Jaunųjų mokslininkų konferencija "Bioateitis: gamtos ir gyvybės mokslų perspektyvos”	2020
Seminaras “Atkartojamumas Biomedicinoje”	2021
Konferencija "Sprendžiame mes" paminėti Pasaulinei Dauno sindromo dienai Lietuvoje	2021
Mokslinė-praktinė konferencija “Medical Technologies and Clinical Practice: From Infection Control Management to the Challenges of Professional Ethics”	2021
Tarptautinė konferencija „Ninth international conference on radiation in various fields of research (RAD 2021)“	2021
Kursai „Introduction to the Statistical Analysis of Genome-wide Association Studies at University of Surrey“ liepos 5-9 dienomis.	2021
Europos žmogaus genetikos draugijos (ESHG) virtuali konferencija	2021
Seminaras “Įgimtos ir/ar paveldimos genetinės ligos: diagnostikos iššūkiai“	2021
Konferencija “Pagalba moteriai krizinio nėštumo metu”	2021
Konferencija “Kraujo navikinių susirgimų ir kamieninių kraujodaros ląstelių transplantacijos aktualijos”.	2021
Jaunųjų mokslininkų konferencija "Bioateitis: gamtos ir gyvybės mokslų perspektyvos”	2021
Konferencija “Personalizuota medicina onkologijoje. Atradimai, inovacijos, technologijos”	2021
Seminaras “Genetika daugiadisciplininėje sergančiųjų retomis ligomis priežiūroje”	2022
Konferencija "Sprendžiame mes" paminėti Pasaulinei Dauno sindromo dienai Lietuvoje	2022
Europos žmogaus genetikos draugijos (ESHG) konferencija, Viena, Austrija.	2022
Mokslinė-praktinė konferencija “Dismorfologijos vasara 2022”	2022
Jungtinė EAA-ISGA-ICEM tarptautinė konferencija	2022
Jaunųjų mokslininkų konferencija "Bioateitis: gamtos ir gyvybės mokslų perspektyvos”	2022
Narystės	
2018-dabar	Lietuvos žmogaus genetikos draugija
Mokslinio domėjimosi sritys	
Populiacijų genetika, genetinių ligų molekulinė diagnostika, žmogaus nevaisingumo genetika, embriologija.	

Vilniaus universiteto leidykla
Saulėtekio al. 9, III rūmai, LT-10222 Vilnius
El. p. info@leidykla.vu.lt, www.leidykla.vu.lt
bookshop.vu.lt, journals.vu.lt
Tiražas 20 egz.