

VILNIAUS UNIVERSITETAS

Laura
PRANCKĖNIENĖ

De novo mutacijų intensyvumo
įvertinimas bendrosios Lietuvos
populiacijos ir asmenų su intelektine
negalia egzomuose

DAKTARO DISERTACIJA

Medicinos ir sveikatos mokslai,
Medicina M 001

VILNIUS 2019

Disertacija rengta 2015–2019 metais Vilniaus universitete
Mokslinius tyrimus rėmė Lietuvos mokslo taryba

Mokslinis vadovas:

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

TURINYS

SANTRUMPOS	6
ĮVADAS	13
1. Tiriamoji problema ir aktualumas	13
1.1. Tyrimo tikslas	15
1.2. Tyrimo uždaviniai	15
1.3. Darbo naujumas ir reikšmė	16
1.4. Ginamieji teiginiai.....	16
1.5. Disertantės indėlis	16
1.6. Darbo aprobacija	17
TYRIMŲ APŽVALGA.....	19
2.1. <i>De novo</i> mutacijų identifikavimas.....	20
2.2. <i>De novo</i> mutacijų susidarymo priežastys	24
2.3. <i>De novo</i> mutacijų pasiskirstymo ir genominio konteksto santykis	28
2.5. Mutacijų rinkiniai ir kritinės mutacijų susidarymo sritis	30
2.6. Tėvų amžiaus įtaka <i>de novo</i> mutacijų dažniui	31
2.7. <i>De novo</i> mutacijų kilmės ir amžiaus vertinimas	41
2.8. <i>De novo</i> mutacijų dažnio skaičiavimo metodai ir jų taikymas evoliucinės ir populiacijų genetikos tyrimuose.....	43
2.9. <i>De novo</i> mutacijos – neurovystymosi ligų priežastis	48
2.9.1. <i>De novo</i> mutacijos intelektualinės negalios asmenų grupėje	49
2.9.2. <i>De novo</i> mutacijos asmenų, sergančių autizmo spektro ligomis, grupėje	51
DARBO METODOLOGIJA	54
3.1. Tiriamieji.....	54
3.2. Tyrimo eiga ir metodai.....	56
3.2.1. DNR išgryninimas	56
3.2.2. Viso egzomo sekoskaita	57

3.2.3. NKS duomenų apdorojimas	57
3.2.4. <i>De novo</i> mutacijų identifikavimas ir tikrinimas	59
3.2.5. Biologinės tėvystės tikrinimas	60
3.2.6. Tyrimo duomenų anotavimas	60
3.2.7. <i>De novo</i> mutacijų dažnio apskaičiavimas	61
3.2.8. Tėvų amžiaus, genetinių ir epigenetinių veiksnių, galinčių turėti įtakos <i>dnM</i> intensyvumui, analizė	62
3.2.9. <i>De novo</i> mutacijų spektro sudėties ir šablonų Lietuvos, asmenų su IN ir asmenų su IN sibsų grupėse nustatymas ir vertinimas.....	64
3.2.10. Sekų su <i>de novo</i> mutacijomis santykinio tinkamumo vertinimas.....	65
3.2.11. <i>De novo</i> mutacijų patogeniškumo mechanizmų nustatymas.....	66
3.2.12. Duomenų analizė ir statistiniai metodai	67
TYRIMO REZULTATAI	69
4.1. <i>De novo</i> mutacijos, jų skaičius ir dažnis Lietuvos populiacijos genome	69
4.2. <i>De novo</i> mutacijos, jų skaičius ir dažnis asmenų su IN ir jų sibsų genomuose	73
4.3. <i>De novo</i> mutacijų spektro sudėties ir šablonų Lietuvos, asmenų su IN ir asmenų su IN sibsų grupėse analizės rezultatai	79
4.4. Baltymų sekų, kuriose nustatytos <i>de novo</i> mutacijos, santykinio tinkamumo analizės rezultatai.....	82
4.5. Funkcinė <i>de novo</i> mutacijų analizė <i>in silico</i>	89
TYRIMO REZULTATŲ aptarimas	96
5.1. Algoritmo <i>de novo</i> mutacijoms identifikuoti atranka....	96
5.2. <i>De novo</i> mutacijų spektras ir dažnis Lietuvos populiacijoje, asmenų su IN ir asmenų su IN sibsų grupėse.	98

5.3. Baltymų sekų, kuriose nustatytos <i>de novo</i> mutacijos, santykinio tinkamumo analizės rezultatų aptarimas	102
5.4. Funkcinės <i>in silico</i> analizės rezultatų aptarimas	105
IŠVADOS	110
LITERATŪROS SĄRAŠAS	112
PRIEDAI	128
PADĖKA	162

SANTRUMPOS

- $\hat{\mu}_G$ – mutacijų dažnis kartai
 $\hat{\mu}_y$ – metinis mutacijų dažnis
 N_e – efektyvus populiacijos dydis
 p_- – žalingų *de novo* mutacijų dalis
 p_0 – neutralių *de novo* mutacijų dalis
 R_P – Spirmeno koreliacijos koeficientas
 R_S – Pirsono koreliacijos koeficientas
 μ_G – *de novo* mutacijų skaičius vaiko genome
 $\hat{\alpha}$ – vidutinis alfa
1000 genomų projektas – 1000 genomų projekto visų tirtų populiacijų duomenys
5-mC – 5-metilcitozinas
A – adeninas
AARS – Alanil-TRNR sintetazė
AB – alelių balansas
AD – autosominis dominantinis paveldėjimas
Ala – alaninas
AR – autosominis recesyvus paveldėjimas
Arg – argininas
Asn – asparaginas
Asp – asparto rūgštis
ARID1B – *AT-Rich Interaction Domain 1BOL2A1* – 2-o tipo kolageno $\alpha 1$ grandinės genas (angl. *Collagen, type II, alpha-1*)
ATP – adenozino trifosfatas
AutDB – su autizmu siejamų genų duomenų bazė
BAF250 – nuo ATP priklausomas su Brahma asocijuotas veiksnys
BAM – dvejetainė sekoskaitos duomenų formato rinkmena (angl. *binary file format*)
Bp – bazių poros
bRNR – branduolio ribonukleorūgštys
C – citozinas
CADD – įrankis, žmogaus genome įvertinantis pavienių nukleotidų, intarpų ir iškritų variantų žalingumą (angl. *Combined Annotation Dependent Depletion*)
CEU – Jutos gyventojų, turinčių Šiaurės ir Vakarų Europos protėvių, populiacija

CHARGE sindromas – akies koloboma, širdies ydos, choanų atrezija, raidos atsilikimas, genitalijų anomalijos, ausų anomalijos ir (ar) kurtumas (angl.

CHARGE association: coloboma of the eye; heart anomaly; atresia, hoanal; retardation of mental and somatic development; microphallus; ear abnormalities and/or deafness)

CHD7 – transkripcijos veiksnio genas (angl. *Chromodomain helicase DNA binding protein 7*)

Chr – chromosoma

Cys – cisteinas

ClinVAR – NCBI duomenų bazė, kurioje kaupiami duomenys apie genomo variantų patogeniškumą

COSMIC – mutacijų, esančių somatinėse vėžinėse ląstelėse, katalogas (angl. *Catalogue Of Somatic Mutations In Cancer*)

CpG – citozino ir guanino dinukletoidas

CREBBP – CREB sujungiantis baltymas

CSMD1 – transkripcijos veiksnio genas (angl. *Chromodomain helicase DNA binding protein 7*)

CTR9 – CTR9 homologas, Paf1/RNR II polimerazės komplekso komponentas

d – tikėtinas analizuojamų *DNR* sekų išsiskyrimas tarp organizmų rūšių poros
 μ – skirtumų tarp sekų dažnis

dbSNP – genų variantų duomenų bazė (angl. *Database of single nucleotide polymorphisms (SNPs) and multiple small-scale variations*)

denovo-db – patogeninių, naudingų ir neaiškios klinikinės reikšmės *DNM* duomenų bazė

DHS – deoksiribonukleazei1 (*DNase1*) ypač jautrūs genomo regionai (angl. *DNase1 hypersensitivity sites*)

DLG4 – sinapsių plastiškumo genas

DNAH8 – 8 dineino aksonemalinė sunkioji grandinė

dnM – *de novo* mutacija

DNR – deoksiribonukleorūgštis

DP – referentiniui aleliui alternatyvaus alelio padengimas

ENCODE – tarptautinis funkcinių genomo elementų analizės konsorciumas (angl. *Encyclopedia of DNA Elements*)

ESPN – espino genas (angl. *Espin, mouse, homolog of*)

Ensembl – žmogaus genomo naršyklė

eQTL – raiškos lygis skirtinguose audiniuose

Esp6500siv2_all – egzomo sekoskaitos projekto (angl. *NHLBI GO Exome Sequencing Project – ESP*) visų tirtų asmenų duomenys

ESPN – espinas
ESRRG – su estrogenų susijęs γ receptoriaus
ETNPPL – etanolamino fosfato fosfoliazė
ExAC – viso egzomo duomenų kaupimo ir analizės konsorciumas (angl. *The Exome Aggregation Consortium*)
ExAC_all – ExAC konsorciumo visų populiacijų duomenys
F – tiesioginis pradmuo (angl. *forward*)
FASTKD1 – FAST kinazės 1 domenas
FATHMM – funkcinė analizė pagal Markovo modelius (angl. *Functional Analysis through Hidden Markov Models*)
FGF3 – fibroblastų augimo 3 veiksnio genas (angl. *Fibroblast Growth Factor 3*)
FGFR2 – fibroblastų augimo 2 veiksnio genas (angl. *Fibroblast Growth Factor Receptor 2*)
FMRP – lūžiosios X intelektualinės negalios baltymo genas (angl. *Fragile X Mental Retardation 1*)
G – guaninas
GABAerginis – susijęs ar veikiantis neurosiuntiklį GABA
GATK – genomo analizės įrankių rinkinys (angl. *The Genome Analysis Toolkit*)
Gb – giga bazė
GeneCards – žmogaus genų duomenų bazė
GERP++ – genomo regionų evoliucinio dažnio profiliavimas (angl. *Genomic Evolutionary Rate Profiling*)
GIGYF2 – su GRB10 sąveikaujantis GYF baltymas 2
Gly – glicinas
Gln – glutaminas
Glu – glutamatas
GQ – genotipo kokybės įvertis
GTPazė – guanozino trifosfato hidrolazė
H3K27ac – histono H3 27 pozicijoje esančio lizino modifikacija acetilo grupe
H3K4me1 – histono H3 4 pozicijoje esančio lizino modifikacija metilo grupe
H3K4me3 – histono H3 4 pozicijoje esančio lizino modifikacija trimis metilo grupėmis
HGMD – žmogaus genų patogeninių variantų duomenų bazė (angl. *The Human gene mutation database*)
His – histidinas
HRAS – proto-onkogenas (angl. *HRas Proto-Oncogene, GTPase*)

Human Splicing Finder – sukirpimo svarbių genomo variantų vertinimo duomenų bazė
IDGenetics – genų ir genų variantų, susijusių su intelektine negalia, duomenų bazė
Ile – izoleucinas
IN – intelektinė negalia
IQ – intelekto koeficientas
YRI – Ibadano Jorubos (Nigerija) populiacija
iRNR – informacinė RNR
Iškrita-intarpas – angl. *indel* tipo mutacija
Kb – kilobazė (1000 bazių porų)
kDNR – kopijinė deoksiribonukleorūgštis
KSP – kopijų skaičiaus pokytis
LCL – limfoblastoidinių ląstelių linija
Leu – leucinas
LITGEN – projekto „Lietuvos populiacijos genetinė įvairovė ir sandaros kitimai, susiję su evoliucija ir dažniausiai paplitusiomis ligomis“ akronimas
Lys – lizinas
LLP – lytinių ląstelių pirmtakai
LRT – bioinformacinės analizės tyrimo įrankis (angl. *Likelihood ratio test*)
MAF – retojo alelio dažnis (angl. *minor allele frequency*)
Mb – megabazė
Met – metioninas
MIM – pagal Mendelio dėsnius paveldimų ligų katalogas (angl. *Mendelian Inheritance in Man*)
MYOM1 – miomezinas 1
Mm – milijonai metų
MutH – DNR klaidų taisymo baltymas mutH
MutL – DNR klaidų taisymo baltymas mutL
MutS – DNR klaidų taisymo baltymas mutS
NFATC3 – aktyvintų T ląstelių barnduolio veiksnys
nkRNR – nekoduojanti ribonukleorūgštis
NKS – naujos kartos sekoskaita
NTS – netransliuojama geno seka
OMIM – pagal Mendelio dėsnius paveldimų žmogaus ligų elektroninis katalogas (angl. *Online Mendelian Inheritance in Man*®)
 p_- – žalingų *de novo* mutacijų dalis
 p_+ – naudingų *de novo* mutacijų dalis
 p_0 – neutralių *de novo* mutacijų dalis

PGR – polimerazės grandininė reakcija
Phe – fenilalaninas
PI – pasikliautinis intervalas
PhyloP – algoritmas, vertinantis genomų sekų evoliucinį konservatyvumą (angl. *phylogenetic P-values*)
POLD1 – DNR polimerazė $\delta 1$
Polyphen-2 – bioinformacinės analizės įrankis (angl. *Polymorphism Phenotyping v2*)
PPL – periplakinai
PPM1D – fosfatazės genas (angl. *Protein Phosphatase, Mg²⁺/Mn²⁺ Dependent 1D*)
pre-mRNR – preinformacinė ribonukleorūgštis
Pro – prolinas
PROVEAN – baltymų variacijos poveikio analizės įrankis (angl. *Protein Variation Effect Analyzer*)
PTPN11 – baltymų tirozino fosfatazės genas (angl. *Protein Tyrosine Phosphatase, Non-Receptor Type 11*)
PUBMED – elektroninė medicinos mokslo publikacijų, gyvybės mokslų žurnalų, elektroninių knygų duomenų bazė
PVK – *de novo* mutacijų dažnis vienai pozicijai per kartą
R – atvirkštinis pradmuo (angl. *reverse*)
R² – determinacijos koeficientas
RAC1 – Rac šeimos mažosios GTPazės genas (angl. *Rac Family Small GTPase 1*)
RAPGEF1 – Rap guanino nukleotido pakeitimo veiksnys 1
RAxML – baltymo sekos filogenetinės analizės programa, paremta atsitiktinės atrankos didžiausios tikimybės principu (angl. *Randomized Accelerated Maximum Likelihood*)
RNR – ribonukleorūgštis
Rs – vieno nukleotido varianto numeris pagal referentinę geno seką
s – atrankos koeficientas
S fazė – ląstelės dalijimosi fazė, kurios metu sintetinama DNR ir centrosomos kopija
SATB1 – specifinis AT nukleotidų gausus prisijungimo baltymas 1
SFARI gene – su autizmu siejamų genų duomenų bazė
SIFT – bioinformacinės analizės įrankis (angl. *Sorting Tolerant From Intolerant*)

SYNCRIP – sinatotagminą prijungiančio ir citoplazmoje su RNR sąveikaujančio baltymo genas (angl. *Synaptotagmin Binding Cytoplasmic RNA Interacting Protein*)

SiPhy – algoritmas, vertinantis gamtinės atrankos poveikį geonomo regionams

SYT1 – sinaptotagminas 1

SKL – spermatozonijų kamieninės ląstelės

SLC17A8 – transmembraninio nešiklio genas (angl. *Solute carrier family 17 (vesicular glutamate cotransporter), member 8*)

SLC19A2 – transmembraninio nešiklio genas (angl. *Solute carrier family 19 (thiamine transporter), member 2*)

SMAD6 – SMAD šeimos genas (angl. *SMAD Family Member 6*)

SMO – *Frizzled* klasės receptorius

SON – SON DNR prijungiančio baltymo genas (angl. *SON DNA Binding Protein*)

SOX5 – *SRY dėžutės* genas (angl. *SRY-Box 5*)

SP – standartinė paklaida

SPB – Senojo pasaulio beždžionės (angl. *Old World Monkey*)

STRING – programa, baltymų sąveikai tirti

SWISS-MODEL – baltymo struktūros atviros prieigos modeliavimo programa

SW/SNF – SWItch/sukrozės nefermentuojamas

swMutSel0 – *TdG12 site-wise mutation-selection* skaičiavimo modelis

T – timinas

t – vidutinis laikas iki bendro protėvio (išsiskyrimo laikas)

TCF20 – transkripcijos veiksnio genas (angl. *Transcription Factor 20*)

Ter – priešlaikinė baltymo sintezės baigmė

Thr – treoninas

Tyr – tirozinas

TLK2 – kinazės genas (angl. *Tousled Like Kinase 2*)

TLN – talinas 1

TpG – timino ir guanino dinukleotidas

TRIP12 – skydliaukės hormonų receptoriaus genas (angl. *Thyroid Hormone Receptor Interactor 12*)

Trp – triptofanas

UCSC – Kalifornijos Santa Kruzo universitetas, JAV (angl. *University of California Santa Cruz*)

UNIGENE – projekto „Unikalūs genomo persitvarkymai esant įgimties nervų sistemos raidos sutrikimams: kilmė, genominiai mechanizmai, funkcinės ir klinikinės pasekmės“ akronimas

UV-B – 315–280 nm ilgio ultravioletiniai spinduliai

UV-C – < 280 nm ilgio ultravioletiniai spinduliai
UvrAB – UvrAB endonukleazė
UvrC – UvrC endonukleazė
UvrD – UvrD endonukleazė
Val – valinas
VES – viso egzomo sekoskaita
VarScan – *de novo* variantų nustatymo programa (angl. *Variant detection in massively parallel Sequencing data*)
Vcf – nustatytų genomų variantų byla (angl. *variant calling format*)
VGS – viso genomo sekoskaita
VNP – vieno nukleotido pokytis
VNP – vieno nukleotido variantas
VšĮ VUL SK MGC – Viešosios įstaigos Vilniaus universitetinės ligoninės
Santaros klinikų Medicininės genetikos centras
W – adeninas arba timinas
w – santykinis tinkamumas
WWOX – transkripcijos veiksnio genas (angl. *Chromodomain helicase DNA binding protein 7*)
XD – su X chromosoma susijęs dominantinis paveldėjimas
XR – su X chromosoma susijęs recesyvus paveldėjimas
ZNF219 – transkripcijos veiksnys „cinko pirštelio“ baltymas 219 (angl. *Zinc finger protein 219*)
 β – regresijos koeficientas
 γ – gama
 δ – delta
 ε – epsilon

ĮVADAS

„Be mutacijų mes nebūtume savimi. Klaidos reikalingos tam, kad galėtumėme atverti duris šalia esančiai galimybei“

Steven Johnson

1. Tiriamoji problema ir aktualumas

Visuotinai suprantama, kad mutacijos – fundamentali evoliucijos jėga, kuri kuria genetinę įvairovę ir formuoja fenotipinį kintamumą, todėl jų dažnio ir savybių tyrimai yra svarbūs ir aktualūs tiek populiacijų kilmės, tiek ir medicininės genetikos kontekste. Vykstant apvaisinimui, kiekviena žmogaus zigota pusę genomo paveldi iš motinos oocito, o kitą pusę iš tėvo spermatozoido. Nepaisant to, kad tokiu būdu genetinė informacija perduodama iš kartos į kartą, kiekvienas iš mūsų gimsta su keliasdešimčia (vidutiniškai 20–80) naujų genetinių pokyčių – *de novo* mutacijų [1–5]. *De novo* mutacijos susidaro besiformuojant embrionui, vėliau vaisiui ir visą žmogaus gyvenimą tiek somatinėse, tiek ir lytinėse ląstelėse. Vis dėlto tik tos *de novo* mutacijos, kurios susidaro lytinėse ląstelėse, gali būti perduodamos kitai kartai ir tik iš jų formuojasi tolesnė populiacijos genetinė įvairovė [6–8].

Genomo tyrimai pradėti dar 1980 m., tačiau vien tam, kad būtų galima perskaityti pirmąjį žmogaus genomą, prireikė dar 20 metų [9]. Pirmojo žmogaus genomo analizės etapo metu buvo nustatyta daugiau nei 100 milijonų genomo variantų, tačiau tai tesudarė žinių apie genomą pamatą [10]. Išanalizavus genomo variantų sudėtį, nustatyta, kad didžioji jų dalis populiacijoje yra dažni (nustatomi > 0,5 % populiacijos), maždaug 40 000 – 200 000 variantų yra reti (nustatomi < 0,5% populiacijos), o likusieji – unikalūs pavieniams žmonėms. Dabar žinoma, kad dažni variantai sudaro didžiąją kiekvieno genomo variacijos dalį ir lemia didžiausius skirtumus tarp individų genomų [11]. Tai aiškinama faktu, kad dažni genomo variantai susidarė prieš daugelį kartų, taigi jais dalijasi net ir skirtingos populiacijos, o reti variantai susidarė ne taip seniai, todėl spėjo išplisti tik pavienėse populiacijose [11].

Tobulėjant genomo sekoskaitos technologijoms, antro ir svarbiausio žmogaus genomo analizės etapo metu buvo imta interpretuoti genomo

variantus ne tik pavieniui, bet ir atsižvelgiant į susistemintą genominių varianto kontekstą ir jo ypatybes. Deja, iki dabar geno variantų asociacija su žmogaus bruožais sėkmingai nustatyta tik nedaugeliui geno variantų, todėl didžiosios kataloguotų ir naujai nustatomų geno variantų dalies poveikis žmonių bruožams paveldėti ar polinkiui sirgti tokiomis ligomis, pavyzdžiui, intelektine negalia ar autizmu, nėra išaiškintas [12]. Nesutariama ir dėl dažnų, ir dėl retų geno variantų indėlio dydžio bruožams paveldėti.

O lytinėse ląstelėse susidariusios *de novo* mutacijos jau dešimtmečius yra siejamos su žmonių ligomis. Aneuploidijos – vienas iš dažniausių *de novo* įvykių. Geriausiai to žinomas pavyzdys yra 21 chromosomos trisomija, dar žinoma kaip Dauno sindromas, aprašytas 1959 m. [13]. Kito *de novo* mutacijų tipo – kopijų skaičiaus pokyčių (KSP) įvairovės vaidmuo ligoms [14] suprastas tik išradus genominių lustų technologiją. Ji leido išsiaiškinti, kad net esant labai mažam KSP susidarymo dažniui (0,01 ar 0,02 pokyčiai vienoje kartoje [14–16]), KSP gali būti sunkių ir ankstyvos pradžios neurodegeneracinių ligų ar įgimtų anomalijų priežastimi dėl jų neigiamo poveikio genų veiklai [17]. Vis dėlto visiškai akivaizdi *de novo* mutacijų reikšmė žmogaus ligoms tapo tik dėl naujos kartos sekoskaitos (NKS) metodų, kurie suteikė galimybę patikimai (tiesiogiai) ir prienamai nustatyti visų tipų *de novo* mutacijas [14].

NKS metodai iš esmės pakeitė požiūrį į mutacijų tyrimus, leido atlikti ne tik tiesioginius empirinius mutacijų dažnio ir molekulinį savybių vertinimus, bet ir suprasti skirtingų mutacijų tipų santykinį indėlį į nuolatinę genetinę variaciją ir evoliucinį jų likimą populiacijoje [18]. Vis dėlto pagrindinis mutacijų tyrimų iššūkis tebėra mutacijų funkcijos paradoksalumas. Nors žinoma, kad evoliucijos eigoje, vykstant adaptacijai, mutacijų variacija yra būtina, tačiau daugumos mutacijų, lemiančių fenotipo pasikeitimus, tinkamumas paprastai yra neigiamas ar net žalingas [19]. Natūraliose populiacijose, kurioms būdinga intensyvi gamtinė atranka, žalingos *de novo* mutacijos per trumpą evoliucijos periodą yra greitai eliminuojamos, todėl *de novo* mutacijų analizė yra labai ribota.

Lietuvos populiacijos genetinę struktūrą tyrė dr. I. Kavaliauskienė ir dr. Tautvydas Rančelis [20, 21]. Dr. I. Kavaliauskienė išanalizavo Lietuvos populiacijos genetinę struktūrą pagal egzomo autosomų, Y chromosomos ir mitochondrijų DNR genetinių žymenų alelių pasiskirstymą, o dr. Tautvydas Rančelis tyrė išskirtinai patogeninių geno variantų pasiskirstymą. Dėl *de novo* mutacijų Lietuvos populiacija iki šiol tirta nebuvo. Lietuvos populiacijos mastu apibendrinta, o ne pavienių atvejų aprašymais grįsta *de*

novo mutacijų analizė svarbi tuo, kad naudojant triadų (tėvas, motina ir bent vienas vaikas) tyrimo strategiją galima nustatyti ne tik skirtingų geno variantų, bet ir atskirų geno sričių *de novo* mutacijų dažnį tirtose Lietuvos populiacijos ir asmenų, turinčių intelektualinę negalią, grupėse. Viso egzomo sekoskaitos duomenų analizė triadose leidžia tiesiogiai nustatyti skirtingas ir gamtinės atrankos nepaveiktas *de novo* mutacijas, kurios gali būti asocijuotos tiek su dažnomis, tiek su retomis ligomis ar adaptyviais geno regionais [22–28]. Aktuali ir visų nustatytų geno variantų funkcijos ir konteksto analizė, leidžianti nustatyti galimus mutacijų susidarymo, o intelektualinę negalią turinčių asmenų grupėje ir patogenezės mechanizmus.

1.1. Tyrimo tikslas

De novo egzomo mutacijų pasiskirstymo, susidarymo mechanizmų ir poveikio analizė Lietuvos populiacijos ir turinčių intelektualinę negalią asmenų bei intelektualinę negalią turinčių asmenų šibų grupėse.

1.2. Tyrimo uždaviniai

1. Atrinkti optimalų algoritmą *de novo* mutacijoms nustatyti.
2. Nustatyti Lietuvos populiacijai, asmenims, turintiems intelektualinę negalią, ir intelektualinę negalią turintiems asmenims šibų grupėse būdingą *de novo* mutacijų spektrą ir palyginti jį su Jutos gyventojų, kilusių iš Šiaurės ir Vakarų Europos (CEU), Ibadano Jorubos (Nigerija) (YRI), islandų, škotų ir pietų Dakotos populiacijose nustatyta *de novo* mutacijų spektro struktūra bei nustatyti galimus *de novo* mutacijų susidarymo mechanizmus.
3. Apskaičiuoti *de novo* mutacijų intensyvumo dažnį genome ir palyginti jį su islandų, danų, olandų, kanadiečių ir amerikiečių bendrosiose populiacijose bei autizmo spektro liga sergančiųjų grupėje nustatytu dažniu.
4. Nustatyti tėvų amžiaus, genominių ir epigenetinių veiksnių įtaką *de novo* mutacijų radimosi intensyvumui.
5. *In silico* nustatyti *de novo* mutacijų funkcijos įtaką Lietuvos populiacijos ir asmenų, turinčių intelektualinę negalią, fenotipui; nustatyti galimus intelektualinės negalios patogenezės mechanizmus.

6. Įvertinti, kaip *de novo* mutacijos veikia aminorūgščių santykinį tinkamumą, ir, jei veikia, kokie santykinio tinkamumo šablonai apibūdina Lietuvos populiaciją.

1.3. Darbo naujumas ir reikšmė

1. Apibendrintai tirtas *de novo* mutacijų pasiskirstymas, apskaičiuoti jų dažniai genome, įvertintas patogeniškumas ir galimi susidarymo mechanizmai papildo visame pasaulyje atliekamų *de novo* mutacijų tyrimų duomenis ir sudaro pamatą tolesniems nekoduojančių ir reguliacinių elementų variantų tyrimams tiek bendrojoje populiacijoje, tiek ir asmenų, turinčių intelektualinę negalią, grupėje.

2. Nustatyti nauji nepublikuoti patogeniniai / galimai patogeniniai intelektualinę negalią lemiantys genomo variantai.

3. Pirmą kartą sudarytas specifinių Lietuvos populiacijos grupei *de novo* genomo variantų katalogas papildo kuriamą referentinio lietuvių genomo duomenų bazę ir yra svarbus ateities didesnių kohortų ar klinikinių Lietuvos populiacijos tyrimų pagrindas.

1.4. Ginamieji teiginiai

1. Lietuvos populiacijai būdinga savita *de novo* mutacijų spektro sudėtis, intensyvumas ir susidarymo mechanizmai.

2. Lietuvos populiacijoje susidaro tokios *de novo* mutacijos, kurių bendras santykinis tinkamumas nedaro žalingos įtakos asmens fenotipui ir sveikatai, todėl genome gali persistuoti daugelį kartų.

3. Asmenų, turinčių intelektualinę negalią, egzomuose žalingų *de novo* mutacijų yra reikšmingai daugiau nei kitose tirtose grupėse, taigi tai gali būti intelektualinės negalios priežastis.

1.5. Disertantės indėlis

Disertantė pati atliko *Illumina HiSeq 2500* genetiniu analizatoriumi nuskaitytų egzomų pirminių duomenų bioinformacinį apdorojimą, identifikavo *de novo* mutacijas, jas patikrino ir tvirtino *Sanger* sekoskaitos metodu, atliko biologinės tėvystės tikrinimą. Disertantė taip pat pati atliko visą duomenų bioinformacinę, *in silico* funkcinę ir statistinę analizę.

1.6. Darbo aprobacija

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

Disertacijos tema paskelbtų straipsnių sąrašas:

1. **Pranckėnienė L**, Kučinskas V. Trumpų *de novo* iškritų ir intarpų intensyvumo įvertinimas skirtinguose žmogaus genomo regionuose. *Laboratorinė medicina*, 2016; 18 (71): 118–122.

2. **Pranckėnienė L**, Jakaitienė A, Ambrozaitytė L, Kavaliauskienė I, Kučinskas V. Insights Into *de novo* Mutation Variation in Lithuanian Exome. *Frontiers in Genetics*. 2018;9:315. doi:10.3389/fgene.2018.00315. Impact faktorius – 4,151.

3. **Pranckėnienė L**, Preikšaitienė E, Gueneau L, Reymond A, Kučinskas V. *De novo* duplication in the *CHD7* gene associated with severe CHARGE syndrome. *Genomics Insights*. 2019. H faktorius – 5.

Stendiniai pranešimai:

1. **Pranckėnienė L**, Jakaitienė A, Kučinskas V. „An evaluation of *de novo* mutation content in the lithuanian exome“, Europos žmogaus genetikos draugijos konferencija, 2017 m. gegužės 27–30 d., Kopenhagoje, Danijoje.

2. **Pranckėnienė L**, Jakaitienė A, Kučinskas V, Utkus A. An evaluation of *de novo* mutation rate in lithuanian exome. 2017 04 4-7 EMGM konferencija (The European Mathematical Genetics Meeting), Tartu, Estijoje.

3. **Pranckėnienė L**, Jakaitienė A, Ambrozaitytė L, Kučinskas V. „Impact and rates of exonic *de novo* mutations in patients with intellectual disability“, 2018 m. gegužės 27–30 d. Europos žmogaus genetikos draugijos konferencija, Milane, Italijoje.

4. **Pranckėnienė L**, Preikšaitienė E, Kučinskas V, Reymond A, Gueneau L. „*De novo* insertion in the *CHD7* gene causing a CHARGE syndrome“, 2018 m. spalio 16–20 d. Amerikos žmogaus genetikos draugijos konferencija, San Diego, Kalifornijoje, JAV.

5. **Pranckėnienė L**, Kučinskas V. „Impact and rates of exonic *de novo* mutations in patients with intellectual disability“, 2019 m. rugpjūčio 30–31 d., Gene forum, Tartu, Estija.

Žodiniai pranešimai:

1. **Pranckėnienė L**, Kučinskas V. „An evaluation of *de novo* mutation process intensity in the different regions of human genome“. Tarptautinė konferencija „Evolutionary medicine: pre-existing mechanisms and patterns of current health issues“, 2016 m. birželio 14–17 d., Vilniuje, Lietuvoje.

2. **Pranckėnienė L**, Jakaitienė A, Ambrozaitytė L, Kučinskas V. „*De novo* mutacijų ir jų intensyvumo analizė lietuvių triadų grupėje“. 2017 12 07 LMA Jaunųjų mokslininkų konferencija „Bioateitis“. Diplomas už geriausią žodinį pranešimą, III vieta, Vilniuje, Lietuvoje.

3. **Pranckėnienė L**. „Distribution and rates of exonic *de novo* mutations in patients with intellectual disability“. 2018 05 10–12 Baltijos šalių laboratorinės medicinos (BALM) kongrese, Vilniuje, Lietuvoje.

4. **Pranckėnienė L**, Jakaitienė A, Kučinskas V. „*De novo* mutacijų ir jų intensyvumo analizė lietuvių triadų grupėje“. 2018 12 14 LMA Jaunųjų mokslininkų konferencija „Bioateitis“. Diplomas už geriausią žodinį pranešimą, III vieta, Vilniuje, Lietuvoje.

TYRIMŲ APŽVALGA

De novo mutacijos – geno variantai, nepaveldėti iš tėvų. Yra skiriami du *de novo* mutacijų (*dnM*) tipai – somatinėse ir lytinėse ląstelėse susidariusios mutacijos. Somatinės mutacijos susidaro bet kurioje organizmo ląstelėje, išskyrus lytines ląsteles besiformuojant embrionui, vėliau vaisiui ir visą žmogaus gyvenimą dėl DNR taisymo mechanizmų klaidų, aplinkos veiksnių ar kitų priežasčių. *De novo* mutacijos lytinėse ląstelėse susidaro išskirtinai tik iki apvaisinimo ir yra perduodamos palikuonims, tačiau patys tėvai šių *DNR* pokyčių neturi.

Istoriškai lytinėse ląstelėse įvykusių mutacijų dažnis buvo skaičiuojamas analizuojant genetinių ligų dažnius. Pirmą kartą 1935 m. Haldane mutacijų dažnį genetinei sričiai per vieną kartą nustatė remdamasis hemofilijos paplitimu [6, 29], o jau 2002 m. Kondrashov, pasinaudojęs sukauptomis žiniomis apie žinomas genetines ligas lemiančias genetines sritis, apskaičiavo *de novo* mutacijų dažnį visame žmogaus genome [7]. Vėliau *dnM* ir jų savybės buvo tiriamos naudojant modelinius organizmus, visiškai penetrantines, dominantiniu būdu pagal Mendelio dėsnius paveldimas ligas, taikant lyginamosios genomikos ir populiacijų genetikos metodus. Vis dėlto šių metodų galimybės yra ribotos, jie yra netiesioginiai ir veikiami kitų veiksnių, pavyzdžiui, gamtinės atrankos. Nuo maždaug 2000 m., išradus naujos kartos sekoskaitos (NKS) metodus ir sparčiai juos tobulinant, tiesioginiams visų tipų, nuo vieno nukleotido pokyčių (VNP), mažų (iki 20bp) iškritų-intarpų iki didelių struktūrinių geno pokyčių (**1 lent.**), *de novo* mutacijų paplitimo tyrimams genome buvo imta naudoti tėvų ir vaikų triadų (tėvo, motinos ir bent vieno vaiko) tyrimo metodiką. Atsižvelgiant į tai, kad kiekviena analizuojama triada yra unikali ir triados labai skiriasi, viso geno NKS tyrimų duomenimis *de novo* VNP dažnis genome svyruoja tarp 1,0 ir $1,8 \times 10^{-8}$ per nukleotidą vienai kartai [8, 30–33], todėl skaičiuojama, kad kiekvienas palikuonis vidutiniškai turi nuo 44 iki 82 *de novo* VNP, iš kurių viena dvi mutacijos yra koduojančioje geno sekoje (egzome) [1, 31–35]. Triadų genomų NKS tyrimo metodas taip pat leido nustatyti *de novo* mutacijų kilmės tipus pagal tai, kokio tipo ląstelėje (somaticinėje ar lytinėje) mutacija susidarė, jų genominių kontekstą, galimus susidarymo mechanizmus [31–33] ir kitas ypatybes, pavyzdžiui, tėvinę kilmę. Dabar yra žinoma, kad dauguma *de novo* taškinių ir nedidelių iškritų-intarpų mutacijų, susidariusių lytinėse ląstelėse, yra tėvinės kilmės ir kad didesnis tėvų (tėvo ir motinos) amžius zigotos susidarymo metu lemia, kad vaikas turi daugiau *de novo* mutacijų nei vyresnis tų pačių tėvų vaikas [1, 35–37].

Pagal patogeniškumą mutacijos skirstomos į tris pagrindines grupes – naudingos, neutralios ir patogeninės. Naudingos mutacijos, suteikiančios fenotipinį pranašumą, greitai plinta populiacijoje [38–40], o neutralios mutacijos gali plisti tik kaip genų dreifo rezultatas [41]. Patogeninės mutacijos lemia žalingus fenotipinius požymius, todėl prieš reprodukcijos fazę ir jai vykstant yra stipriai veikiamos išgryninančiosios gamtinės atrankos ir tokiu būdu yra apribojamas jų plitimas [42]. Tiek paveldėtos, tiek *de novo* mutacijos gali priklausyti bet kuriam šių tipų, tačiau įsidėmėtina, kad *dnM* yra unikalios ir genetiškai nutolusios nuo paveldėtų mutacijų, nes jos susidaro tarp dviejų kartų, dar prieš gamtinės atrankos poveikį. Populiacijos lygmeniu toks naujų požymių praradimas ar įgijimas yra rūšių evoliucijos varomoji jėga, tačiau individo lygmeniu to paties požymio praradimas ar įgijimas gali būti ir genetinės ligos priežastis.

1 lent. *De novo* genetinių variantų klasifikacija. *Pagal 1000 genomų projekto duomenis. Didelių struktūrinių persitvarkymų pavyzdžiuose juodos linijos atitinka chromosomos dalį, o raudonos ir mėlynos linijos – genetiniai regionai su DNR variantu. Adaptuota iš [43]

Varianto tipas	Variantų skaičius genome*	<i>De novo</i> variantų skaičius	Pavyzdys
Vieno nukleotido variantas (VNV)	3 mln	40-82	Referentinė seka: CTGGATGAA Variantinė seka: CTGGGTGAA
Nedidelės iki 20 bp dydžio iškritos – intarpai	360 000	2-4	Iškrita Referentinė seka: CTGGATGAA Variantinė seka: CTGG-TGAA
			Intarpas Referentinė seka: CTGGA-TGAA Variantinė seka: CTGGAGTGAA
Dideli struktūriniai persitvarkymai	16 000	0,01-0,03	Iškrita Referentinė seka: Variantinė seka:
			Duplikacija Referentinė seka: Variantinė seka:
			Intarpas Referentinė seka: Variantinė seka:

2.1. *De novo* mutacijų identifikavimas

Pirmosios DNR sekoskaitos technologijos buvo sukurtos maždaug 1970 metais. Jos reikalavo daug kruopštaus darbo, o sekvenavimas buvo lėtas procesas. Per maždaug 20 metų DNR sekoskaitos metodai buvo ištobulinti tiek, kad 1990 m. buvo pradėtas Žmogaus genomo projektas, kurio metu (per

13 metų, investavus maždaug 3 milijardus dolerių) buvo nuskaitytas pirmasis žmogaus genomus. Tuo pačiu metu DNR sekoskaitos metodai buvo tobulinami ir kuriamos naujos kartos sekoskaitos (NKS) technologijos, leidžiantys nuskaityti daugiau mėginių tuo pačiu metu per trumpesnę laiką.

Pirmą kartą NKS technologijos buvo pritaikytos 1000 genomų projekto metu. Jomis buvo nusekvenuota daugiau nei 2500 žmonių genomų iš 26 populiacijų [11]. Lyginant nuskaitytus genomus tarpusavyje ir su referentiniu genomu (pirmuoju nuskaitytu žmogaus genomu) nustatyta daugiau nei 80 milijonų geno variantų ir apskaičiuotas jų dažnis. Ši geno įvairovė sudarė maždaug 90 % visos genetinės įvairovės tarp mažiausiai 1 % asmenų iš visų 26 populiacijų [11].

Tobulinamos naujos kartos sekoskaitos technologijos atpigo, todėl jas imta taikyti ne tik dideliuose tarptautiniuose projektuose, bet ir pavienių žmonių genomų analizei – dažnų, retų ir unikalių tiek sveikų, tiek sergančių žmonių geno variantų analizei [11].

NKS technologijos yra grįstos DNR suskaidymu į 30–500 bp dydžio fragmentus ir lygiagrečiai jų sekoskaita. Daugelis tyrėjų naudoja suporuotų galų sekoskaitos metodą, kai nuskaitymi du trumpi DNR sekų fragmentai, atskirti fiksuotu (dažniausiai 500 bp) atstumu. Fragmentų ar fragmentų porų sekoskaita yra atsitiktinė, todėl būtina nuskaityti daug daugiau DNR fragmentų nei jų yra viename genome, tokiu būdu gaunant kuo daugiau sekų, kurių visuma apimtų visą genomą ir atitiktų abu haplotipus [22].


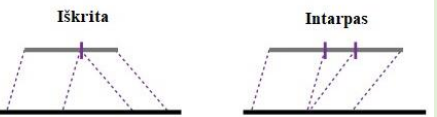
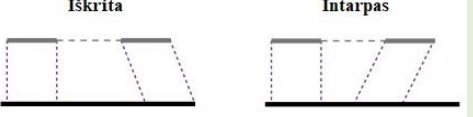

NKS technologijos klaidų dažnis nuo 0,1 % iki 1 % visų nuskaitytų bazių [44]. DNR fragmentai yra nuskaitymi nepriklausomai vienas nuo kito, todėl būtina turėti pakankamai dubliuotų tos pačios sekos fragmentų, kad būtų galima atskirti genetinę įvairovę nuo sekoskaitos klaidų. Sekoskaitos padengimas atspindi DNR sekos fragmentų, kurie persidengia su kiekviena geno baze, skaičių. Svarbu paminėti, kad priklausomai nuo taikomos technologijos ir analizuojamo regiono struktūros, atskirų geno regionų padengimas skiriasi, todėl ne visus geno regionus yra vienodai lengva nuskaityti ir analizuoti. Skaičiuojama, kad maždaug 15 % viso žmogaus geno negali būti sėkmingai nuskaityta dėl jo struktūros, todėl jis vadinamas nepasiekiamu [44–46].

Pirminis sekoskaitos rezultatas yra pavieniai po genomą išsibarstę DNR fragmentai, kuriuos pirmiausia reikia sudėlioti referentinio geno eilės tvarka. Tam dažniausiai naudojamas būdas yra išskaidytų fragmentų prilygiavimas prie referentinio geno [14, 15, 47]. Vis dėlto, kadangi genomai tarpusavyje skiriasi ne vien dėl sekoskaitos klaidų, bet ir dėl individų genetinės įvairovės, naudojami prilygiavimo prie referentinio geno

metodai turi prilygiavimo paklaidų ir leidžia prilygiuoti sekas su nederančiomis bazėmis ar sekų fragmentų tarpais, taip pat kiekvieną prilygiuotą bazę įvertina kiekybiniu patikimumo įverčiu, kuris atspindi prilygiavimo kokybę [16]. Įmanoma, kad tas pats sekos fragmentas yra tinkamas prilygiuoti prie kelių genomo vietų, kur DNR sekos yra labai panašios. Tokiu atveju dalis prilygiavimo algoritmų pateikia visus įmanomus prilygiavimo variantus [15], o kita dalis atsitiktinai parenka labiausiai tikėtiną genomo regioną, kuriam priskiria nusekvenuotą DNR fragmentą [14, 16].

Kai visi DNR fragmentai yra prilygiuoti, visi nesutapimai ir nenuoseklūs fragmentai yra vertinami kaip galimi genetiniai variantai. Kadangi dalis variantų gali būti dėl nekokybiško prilygiavimo arba sekoskaitos klaidų, NKS rezultatų analizėje genetiniai variantai yra vertinami pagal sudėtinis statistinius klaidų ir mėginių nepriklausomumo, bialelinių variantų, diploidų segregacijos ir kitus modelius [17, 22, 46]. Dėl didelės genetinių variantų įvairovės (nuo vieno VNP iki didelių struktūrinių chromosomų persitvarkymų), jiems vertinti taikomi skirtingi algoritmai, kurie priklausomai nuo varianto tipo naudoja skirtingus sekoskaitos duomenis (**2 lent.**). Nors yra daug skirtingų genetinių variantų nustatymo modelių, dauguma jų kiekvieną sekoskaitos fragmentą vertina kaip nepriklausomą stebėjimą, kurio patikimumui vertinti naudojama informacija, surinkta iš tame regione esančių visų nustatytų fragmentų duomenų. Svarbiausias vertinimo patikimumo rodiklis yra sekos fragmento padengimas. Tokiu būdu atskiriami genetiniai variantai nuo klaidų.

2 lent. *De novo* mutacijų nustatymo metodai naudojant NKS metodus. Adaptuota iš [43]

Fragmentų nesutapimai		Variantų tipai: VNP, iškritos, intarpai Variantų dydžiai: maži (paprastai perpus mažesni nei nuskaityto fragmento ilgis)
<p>Mutacijos tipas</p> <p>Sekų fragmentai</p> <p>Referentinė seka</p>		<p>Sulygiuoti DNR fragmentai yra patikrinami, ar nėra nesutapimų, lyginant nuskaitytus sekų fragmentus su referentine genomo seka. Bazių nesutapimai reiškia vieno nukleotido pokyčius (VNP), tarpai nusekvenuose ir sulygiuotose sekose atitinka iškritas, o tarpai referentinėje sekoje – intarpus.</p>
Išskaidyti fragmentai		Variantų tipai: iškritos, intarpai Variantų dydžiai: įvairūs
<p>Mutacijos tipas</p> <p>Sekų fragmentai</p> <p>Referentinė seka</p>		<p>Tie nuskaityti DNR fragmentai, kurių neįmanoma prilygiuoti prie referentinio genomo, yra išskaidomi į mažesnius fragmentus ir tada kiekvienas jų yra prilygiuojamas. Tokiu būdu atstumai tarp išskaidytų fragmentų gali būti identifikuojami kaip iškritos ar intarpai.</p>
Nederančios fragmentų poros		Variantų tipai: iškritos, intarpai, inversijos, duplikacijos, chromosomų persitvarkymai Variantų dydžiai: dideli (paprastai didesni nei 100 bp)
<p>Mutacijos tipas</p> <p>Sekų fragmentai</p> <p>Referentinė seka</p>		<p>Šis būdas gali būti taikomas tik tada, kai analizuojami suporuotų galų sekoskaitos duomenys, t. y. kai fragmentai atskirti fiksuotu atstumu, yra sekvenuojami poromis. Tada intarpai ir iškritos yra nustatomi kaip reikšmingi skirtumai tarp tikėtino ir stebimo fiksuoto atstumo dydžio. Pažymėtina, kad, analizuojant vienos krypties fragmentus, galima identifikuoti ir kompleksinius struktūrinius variantus.</p>
Horizontaliojo padengimo kokybė		Variantų tipai: iškritos, duplikacijos Variantų dydžiai: labai dideli (paprastai didesni nei 1 kb)
<p>Mutacijos tipas</p> <p>Sekų fragmentai</p> <p>Referentinė seka</p>		<p>Visame genome yra vertinamas nuskaitytų ir prilygiuotų fragmentų horizontalusis padengimas. Tuose regionuose, kur yra iškritų, bus matomas pusinis (heterozigotinio genotipo atveju) regiono padengimas arba homozigotinio genotipo atveju – regionas fragmentais išvis nebus padengtas. Duplikuoti regionai bus padengti 1,5 karto (heterozigotinio genotipo atveju) arba 2 kartus (homozigotinio genotipo atveju) daugiau nei viso tiriamo prilygiuoto regiono fonas.</p>

Nustačius genetinius variantus, pagal kiekvieno varianto genetinę sritį yra priskiriamas genotipas. Atsižvelgiant į visus sekoskaitos duomenų rodiklius, apimančius analizuojamą genomo regiono poziciją, yra nustatomas labiausiai tikėtinas genotipas. Daugelis algoritmų rezultatą pateikia genotipo patikimumo įverčiais arba genotipo tikimybėmis visiems nustatytiems genotipams. Taigi genomo variantų nustatymas yra patikimesnis ir tikslesnis, kai DNR sekos padengimas yra didesnis. Verta paminėti, kad vienas didžiausių genomo analizės iššūkių yra sekoskaitos duomenų apdorojimas, nes, pavyzdžiui, kai sekoskaitos vertikalusis padengimas yra $\times 12$ kartų, vienas genomas gali būti ~ 200 Gb dydžio, o tam reikia didelių bioinformacinių išteklių ir kompleksinių duomenų analizės metodų [23, 25]. *DnM* identifikavimą apsunkina ir metodologijos kompleksiskumas, nes vienu metu analizuojamas ne vienas, o trys (motinos, tėvo ir bent vieno vaiko) genomai.

Nustatyti genomo variantai anotojami remiantis duomenimis, surinktais iš jau nuskaitytų ir išanalizuotų genomų, pavyzdžiui, 1000 Genomų, ENCODE, dbSNP, COSMIC, ExAC duomenų bazių duomenimis.

2.2. *De novo* mutacijų susidarymo priežastys

DnM susidarymo priežastis galima skirstyti į dvi grupes: pirminę ir antrinę. Pirminę grupę sudaro *dnM*, susidaranti DNR biosintezės arba DNR klaidų taisymo metu. DNR biosintezės metu *dnM* gali atsirasti dėl klaidų vykstant DNR replikacijai, kai DNR polimerazė klaidingai įjungia nukleotidus į DNR grandinę [25]. Vykstant DNR replikacijai, daugiausia pirmąją DNR grandinę katalizuoja ϵ , o atsiliekančią – δ DNR polimerazė. Abi šios polimerazės nukleotidus į DNR grandinę įjungia ypač atrankiai, *in vitro* vidutiniškai viena neatitiktis pasitaiko kas 10^4 – 10^5 bp [26, 27]. Taip pat abi polimerazės turi po klaidas taisantį domeną, kuris įvykus replikacijai patikrina suporuotų nukleotidų geometriją, kad būtų užtikrintas teisingų heterociklinių bazių (toliau bazių) įjungimas [25].

Vienos ar keletos bazių porų nesutaptys gali lemti besireplikuojančios DNR struktūros pokyčius, tačiau jie gali būti ištaisomi ir klaidų taisymo mechanizmų [28]. Klaidų taisymo mechanizmai (**1 pav.**) yra labai efektyvūs, nes mutacijų, susidariusių vykstant DNR replikacijai, skaičius yra daug mažesnis nei polimerazės klaidų dažnis. Specifinių bazių porų pakaitų dažnis gali skirtis priklausomai nuo klaidų taisymo greičio, kuris tuo pačiu apibrėžia ir specifinių bazių porų pakaitų mutavimo dažnį [28]. Nevisiškas ištaisymas gali lemti vienos ir keletos bazių porų pakaitų ar iškritų-intarpų atsiradimą.

Be to, replikacijos metu gali būti įjungtos ir neteisingos struktūros bazės, o tai taip pat lemia bazių porų nesutaptis ir pakaitas [48].

Antrinės *dnM* susidarymo priežastys – tai spontaniniai pažeidimai dėl endogeninių ar egzogeninių veiksnių, pavyzdžiui:

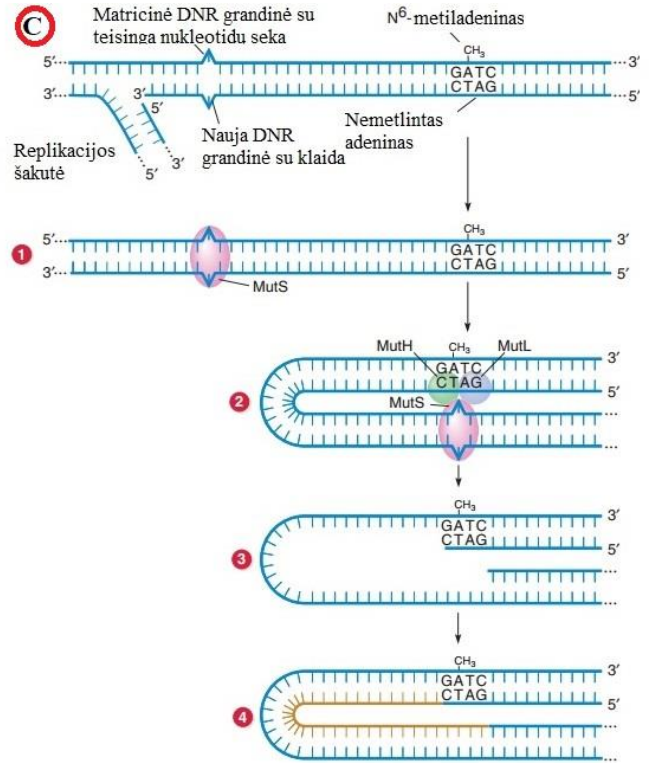
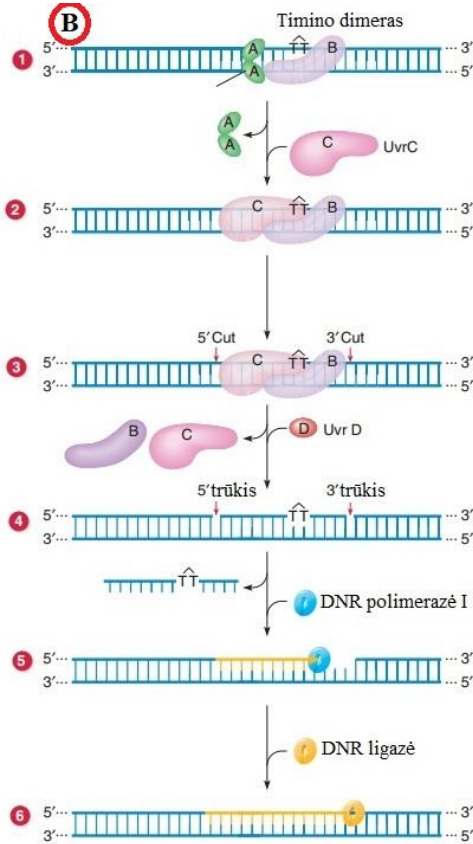
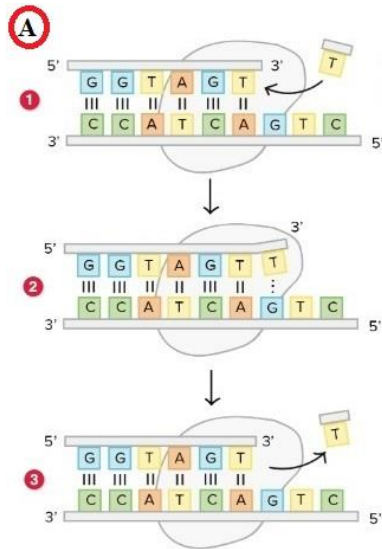
1. Ultravioletinės spinduliuotės (UV-C ir UV-B spinduliai). Dėl šios spinduliuotės poveikio susidaro pirimidino bazių dimerai – dvi gretimos pirimidino bazės sujungiamos kovalentine jungtimi;

2. Jonizuojančios spinduliuotės. γ ir rentgeno spinduliai gali sukelti viengrandžius ir dvigrandžius DNR molekulės trūkius;

3. Aktyvių cheminių medžiagų, pavyzdžiui, angliavandenilių, aflatoksinų poveikio. Alkilinantys veiksniai gali perkelti metilo ar alkilo grupę ant DNR bazės. Taip pat dėl šių medžiagų poveikio susidaro bazių, esančių toje pačioje DNR grandinėje arba skirtingose grandyse, ryšys.

Prieš replikaciją šie spontaniniai pažeidimai dažniausiai yra ištaisomi nukleotidų ir bazių ekscizinės reparacijos metodais [49]. Vis dėlto nepakankamas premutacijų ištaisymas prieš naują DNR replikacijos etapą gali lemti mutacijos įtvirtinimą vienoje ar abiejose dukterinėse ląstelėse [50]. Taigi būtent šie mutagenų žalingo poveikio DNR ir klaidų ištaisymo dažnio skirtumai leidžia susidaryti *dnM*. Tuo atveju, kai mutacijų taisymas nepavyksta ar nevyksta, DNR replikacija gali būti visiškai sustabdyta ir inicijuojama ląstelės žūtis [50].

Dažnai daroma prielaida, kad lytinėse ląstelėse, ypač spermatozoiduose ir jų pirmtakuose, *dnM* susidaro dėl DNR replikacijos klaidų gemetogenezės proceso metu. Vis dėlto nepakankamas DNR pažeidimų ištaisymas taip pat gali būti *dnM* susidarymo priežastis spermatogenezės metu, nes dėl tęstinės proliferacijos ir trumpų periodų tarp ląstelių dalijomosi ląstelės DNR klaidų taisymo mechanizmai pažeidimų gali tiesiog nespėti taisyti [48, 50]. Oogenezės metu spontaninės DNR mutacijos taip pat siejamos su nepakankamu ištaisymo mechanizmu (**1 pav.**), tačiau oocitams būdingesnės aneuploidijos nei pavienės taškinės ar iškritų-intarpų *dnM* [48]. Taigi iš esmės *dnM* dažnis tiesiogiai priklauso nuo replikacijos klaidų dažnio ir mitozijų skaičiaus arba laiko tarp mitozijų ir DNR klaidų taisymo efektyvumo [48].



1 pav. Pagrindiniai DNR klaidų taisymo mechanizmai. A – DNR klaidų taisymas, kai DNR polimerazė veikia kaip egzozonukleazė. Polimerazei 1) įjungus klaidingą nukleotidą, ji pati jį 2) identifikuoja ir 3) panaudodama 3'→5' egzozonukleazinį aktyvumą pašalina. B – nukleotidų ekskizijos reparacija. 1) Homologinės rekombinacijos proceso metu MutS baltymų kompleksas atpažįsta ir jungiasi prie DNR grandinės su klaida. 2) MutS jungiasi su MutL ir MutH kompleksais tuo pačiu ir prie DNR grandinės su klaida pritraukiant nemetilintą GATC seką. 3) MutH sukelia viengrandžius nemetilintos DNR trūkius grandinėje su klaida, o egzozonukleazė tą DNR fragmentą iškerpa. 4) DNR polimerazė III įjungia naujus teisingus nukleotidus, o DNR trūkį pašalina DNR ligazė. C– bazių ekskizijos reparacija. Nehomologinio galų sujungimo metodo metu UvrAB baltymų kompleksas atpažįsta DNR pažaidas, pavyzdžiui, imino dimerus, tada UvrA baltymai atsijungia. Prisijungia UvrC baltymai. UvrBC baltymų kompleksas suardo fosfodiesterinį ryšį 3' gale ties ketvirtu nukleotidu iki pažaidos ir 5' gale ties 7 nukleotidu po pažaidos, tada disocijuoja nuo DNR. UvrD baltymas (helikazė) jungiasi prie 5' DNR gale esančio trūkio ir DNR fragmentas su pažaida disocijuoja. DNR polimerazė I įjungia naujus teisingus nukleotidus, o DNR trūkį užtaiso DNR ligazė. Adaptuota iš [51]

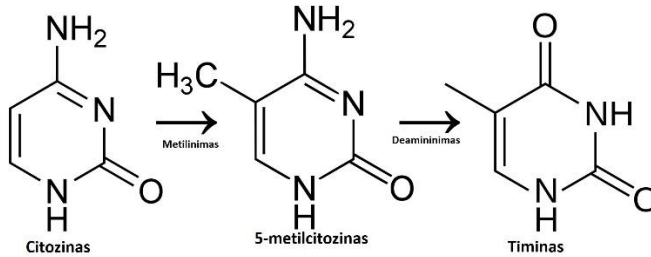
2.3. *De novo* mutacijų pasiskirstymo ir genominio konteksto santykis

Daugumoje tyrimų nustatytas *dnM* dažnis žmogaus genome yra $1,0\text{--}1,8\times 10^{-8}$ per nukleotidą vienai kartai [30–34, 52], mutagenzės procesas nėra visiškai atsitiktinis [8]. Mutabilumo įvairovė skirtinguose genomo regionuose aiškinama genomo ypatybėmis, susijusiomis su sekos sudėtimi ir genominiu kontekstu [52, 53]. Dalis genominių veiksmų, svarbių mutagenzei, manoma, yra visų tipų žmogaus ląstelėse. Tai yra vietinis bazių porų kontekstas, rekombinacijos dažnis ir replikacijos laikas [8, 33, 35]. Replikacijos laikas atspindi tvarką, pagal kurią skirtingi genomo regionai replikuojasi ląstelės ciklo S fazės metu. Tie genomo regionai, kurie replikuojasi vėlai, pasižymi didesne genetinė įvairove nei tie regionai, kurie replikuojasi anksčiau [28]. Dėl šios priežasties manoma, kad genetinę įvairovę lemia didesnis mutabilumas replikacijos pabaigoje dėl deoksiribonukleotidų išsekimo, nors tokie pokyčiai, pavyzdžiui, polimerazės aktyvumo pokytis ar sumažėjęs mutacijų klaidų taisymo mechanizmų aktyvumas, taip pat gali būti viena iš to priežasčių [25, 54, 55].

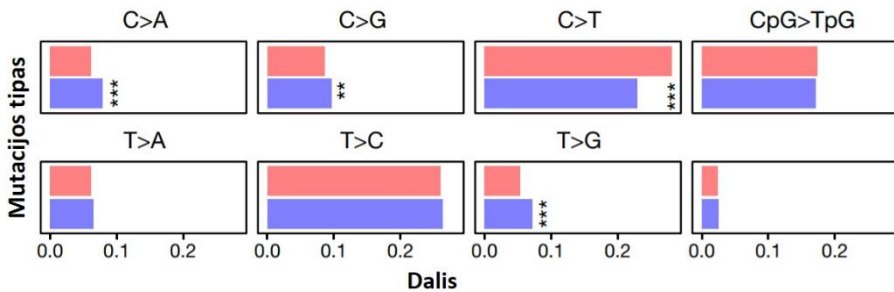
Kiti veiksniai, turintys įtakos mutagenzei, gali varijuoti priklausomai nuo ląstelės genetinės architektūros, chromatino būsenos ir transkripcinio aktyvumo [56–59], pavyzdžiui, genomo regionai, kuriuose gausu CpG dinukleotidų. Citozinas CpG dinukleotidų gausiose genomo srityse yra metilinamas penktoje iš šešių atomų žiedo padėtyje, dėl to susidaro 5-metilcitozinas (5-mC). Metilintas CpG dinukleotidas yra nestabilus dėl spontaninio 5-mC deaminavimo, kurio metu 5-mC virsta timinu, todėl *DNR* grandinėje susidaro nevienodas G:T santykis [2, 32, 59] (**2 pav.**). Skaičiuojama, kad CpG dinukleotidai yra nuo 10 iki 18 kartų mutabilesni nei kiti dinukleotidai [60]. Taigi dėl tokio aukšto CpG mutabilumo į TpG *de novo* mutacijos, kurių rezultatas yra tranzicijos, yra maždaug du kartus dažnesnės, nei tos, kurios lemia transversijas [50, 56].

Verta paminėti, kad epigenetinis metilinimo profilis embriogenezės metu kinta. Tik susidarius zigotai, nesvarbu, kokia embriono lytis, vyksta keli visos *DNR* demetilinimo ir remetilinimo etapai tokiu būdu ištrinant ir atkuriant ląstelių epigenetinę atmintį [2]. Maždaug septintą savaitę po apvaisinimo, įvykus lyties determinacijai, vyrų ir moterų embrionų ląstelių metilinimo profiliai diverguoja: vyrų embrionuose *DNR* remetilinimas vyksta dar prieš spermatogonijų diferenciaciją, o moterims labai vėlai – tik prieš pat ovuliaciją [2]. Dėl šios priežasties vyrų ląstelėse yra metilintos gerokai ilgesnį laiką, nuo septintos savaitės po zigotos susidarymo iki apvaisinimo momento, o moterų ląstelėse metilinimo procesas prasideda tik nuo ovuliacijos momento. Dėl to

skaičiuojama, kad po brandos tėvai turi apie 6,5 karto daugiau CpG>TpG *de novo* mutacijų nei motinos ir kartu maždaug du kartus daugiau transversijų tipo mutacijų (3 pav.).



2 pav. Spontaninis citozino virtimas timinu



3 pav. *De novo* mutacijų spektras pagal mutacijų tipus. Raudona spalva atitinka moterų, mėlyna – vyrų grupę. Žvaigždutėmis pažymėti reikšmingi mutacijų dažnio skirtumai pagal tikslųjį Fišerio testą. Adaptuota iš [2]

2.4. *De novo* mutacijų šablonai

Atsižvelgiant į tai, kokių būdu *dnM* susidarė (dėl DNR replikacijos klaidos, mutagenų poveikio ar sutrikusio DNR pažaidų taisymo), skiriami mutagenezės modeliai. Mutagenezės modelius apibūdina mutacijų šablonai (angl. *mutational signature*), specifiniai konkrečiai ląstelei, audiniui ar organizmui [61]. Remiantis 12 000 vėžio genomų analize, buvo identifikuota 4,9 milijono somatinių mutacijų, iš kurių išskirta 49 (seniau (2013 m.) buvo 30) mutacijų šablonai, asocijuoti su mutavimo proceso aktyvumu somatinėse ląstelėse [62]. Detalūs kiekvieno šablono aprašymai turi atvirą prieigą <http://cancer.sanger.ac.uk/cosmic/signatures>. Visos šio tyrimo metu identifikuotos mutacijos yra sugrupuotos į 96 grupes (tipus) pagal šešis galimus bazių pakaitų tipus (C>A, C>G, C>T, T>A, T>C ir T>G) ir vieną iš

keturių galimų bazių porų šalia mutacijos 3' ir 5' galuose. Kiekvienas mutacijos tipas – tai trinukleotidas, kurio vidurinėje pozicijoje yra įvykusi mutacija į specifinį nukleotidą, todėl kiekvienas mutacijos šablonas yra apibrėžiamas kaip konkrečios stebėtos mutacijos dažnis [63].

Naujausi tyrimai atskleidė, kad embrioninėse ląstelėse susidariusių *dnM* spektras geriausiai koreliuoja su pirmu ir penktu mutacijų šablonais [31, 32]. Tai leidžia daryti prielaidą, kad mutacijos procesas, susijęs su šiais mutacijų šablonais somatinėse ląstelėse, gali būti aktyvus ir embrioninėse ląstelėse, nors tai paaiškinantys mechanizmai išlieka nežinomi. Pirmas mutacijų šablonas reprezentuoja apie 25 % visų embrioninėse ląstelėse įvykusių *de novo* mutacijų. Dėl metilinto citozino deaminavimo jam būdinga didelė C>T tranzicijų dalis CpG dinukleotidų gausiuose regionuose [31, 59]. Penktas mutacijų šablonas, kuriam būdingos daugiausiai A>G tranzicijos, apibūdina likusius 75 % visų *de novo* mutacijų [31]. Šio mutacijų šablono susiformavimo mechanizmas taip pat nėra aiškus, tačiau manoma, kad šį šabloną sudarančios mutacijos gali susidaryti dėl antrinio spontaninio adenino deaminavimo iki hipoksantino, kuris genomo „mašinų“ yra laikomas guaninu [64]. Taigi apibendrintai šis *dnM* šablonas yra asocijuotas su nevienodu pakaitų dažniu komplementariose DNR grandinėse dėl tarpinių reakcijų produktų [64].

2.5. Mutacijų rinkiniai ir kritinės mutacijų susidarymo sritys

De novo mutacijos dėl atsitiktinio pasiskirstymo gali susidaryti bet kurioje genomo vietoje, tačiau kartais kelios mutacijos susidaro reikšmingai arčiau šalia viena kitos nei tikėtasi [8]. Tokios mutacijos formuoja **mutacijų rinkinius** (angl. *clusters*), o *dnM* juose išsidėsto nuo 10 iki 100 kb atstumu [7, 8, 65, 66]. Mutacijų rinkiniai pasižymi unikaliu mutacijų deriniu. Jiems būdingas mažesnis tranzicijų ir didelis C>G transversijų dažnis TCW>TTW arba TCW>TGW kontekste, čia W atitinka adeniną arba timiną [65, 66]¹. Žinoma, kad pirminis mutacijų rinkinių šaltinis yra ilgos viengrandės DNR

¹ Šis fenomenas buvo aprašytas somatinėse ląstelėse **vėžio kontekste**, dar žinomas „*kataegis*“ (kilęs iš graikų kl., liet. *griausmas*) pavadinimu. Tai mutavimo modelis, susijęs su APOBEC (apolipoproteino B mRNR redaguojančių fermentų, panašių į katalizinius polipeptidus; angl. *apolipoprotein B mRNA editing enzyme, catalytic polyptide-like*) fermentų šeima [67, 71].

molekulės, susiformavusios dėl dvigrandės DNR trūkių arba sutrikusios DNR replikacijos šakutės funkcijos [67]. Tokia viengrandė DNR gali būti sumaišyta su retroelementais ir atakuojama APOBEC fermentų, kurie konvertuoja citoziną į uracilą [64]. Tokio tipo mutacijos yra ištaisomos bazių ekscizinės reparacijos būdu pirmiausiai veikiant DNR glikozilazei ir vėliau į klaidas linkusiai polimerazei (angl. *error-prone polymerase*) [68].

Mutacijų rinkinių susidarymas taip pat gali būti aiškinamas chromosomų persitvarkymais. Įrodyta, kad *de novo* VNP mutacijų dažnis yra didesnis ir VNP rinkiniai susidaro arti *de novo* KSP trūkio taškų [4, 69]. Tikėtina, kad tai lemia mažo tikslumo KSP replikacinis mechanizmas ir į klaidas linkusi DNR polimerazė naudojama DNR klaidoms taisyti.

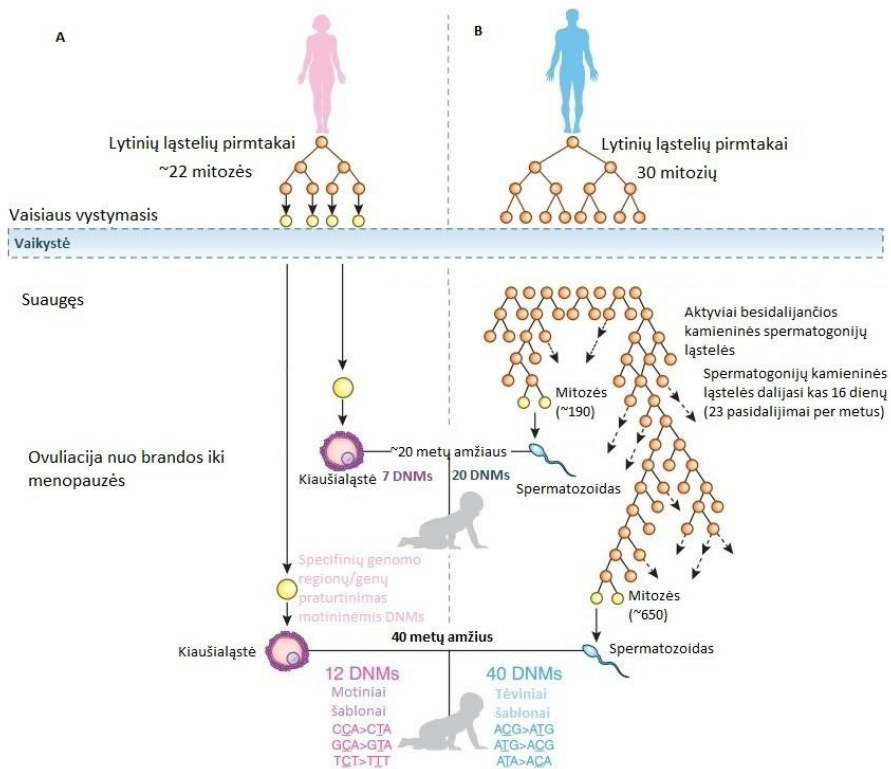
Kitaip nei mutacijų rinkiniai, kurie susidaro vieno individo genome, **kritinės mutacijų susidarymo sritys** – persidengiančios genomines sritys, aptinkamos reikšmingai dažniau nei tikėtasi grupėje individų [70]. Naujausiais viso genomo sekoskaitos tyrimais ir naudojant paslėptus Markovo modelius tokios kritinės mutacijų susidarymo sritys, išsidėsčiusios kas 1 Mb, buvo nustatytos ir patvirtintos tiek egzome [8], tiek ir visame žmogaus genome [33]. Verta paminėti, kad du motininės, bet ne tėvinės kilmės regionai su *CSMD1* ir *WWOX* genais pasižymi didesniu mutacijų dažniu. To mechanizmas dar nėra aiškus, tačiau žinoma, kad pastarasis regionas, kaip ir ribosominė DNR (rDNR) [69], genų rinkiniai [71], segmentinės duplikacijos [72], mikrosatelitai [73], yra linkę mutuoti, todėl juose nustatomas nuo trijų iki keturių kartų didesnis mutacijų dažnis, nei bendras *dnM* dažnio vidurkis visame genome [69].

2.6. Tėvų amžiaus įtaka *de novo* mutacijų dažniui

Žmonių embrionuose lytinių ląstelių pirmtakai (toliau LLP) išsivysto iš epiblastų, po apvaisinimo praėjus nuo aštuonių iki 14 ląstelių dalijimūsi [74]. Vyrų ir moterų embrionuose per pirmuosius ląstelės dalijimus mutacijų dažnis yra panašus (gametogenezės metu susidaro maždaug 0,2–0,6 mutacijų į haploidinį genomą) [31]. LLP specializuojantis vyrų embrionuose susiformuoja kamieninių spermatogonijų klonas, o moterų – pirminių oocitų [31, 74]. Nesvarbu, kokia embriono lytis, šiame vystymosi etape mutacijų dažnis išlieka panašus abiejų lyčių – per vieną ląstelių pasidalijimą susidaro maždaug nuo 0,5 iki 0,7 mutacijų į haploidinį genomą [31]. Kaip jau minėta

2.3 skyriuje, vykstant individų brandai tolesni spermatogenezės ir oogenezės procesai ima skirtis vis labiau.

Spermatogonijų kamieninės ląstelės mitozės būdu dalijasi apytiksliai kas 16 dienų. Tokiu būdu susiformuoja diferencijuotos spermatogonijų ląstelės, tačiau dalis pirminio spermatogonijų klanų išlieka nepakitusi. Iš diferencijuotų spermatogonijų mitozės būdu, po kurios eina du mejozės etapai, susidaro spermatidės, o iš šių – spermatozoidai (**4B pav.**) [75]. O moterų kiekvieno menstruacinio ciklo metu tik keli pirminiai oocitai iš viso rezervo užbaigia pirmąjį mejozės dalijimąsi. Po ovuliacijos subręsta vienas (ar keli) oocitai ir jis užbaigia antrą mejozinį dalijimąsi (**4A pav.**). Taigi po LLP pasidauginimo embriogenezės metu, kol visiškai subręsta, oocitai replikuoja tik vieną kartą, o spermatogonijų ląstelės, kol visiškai subręsta iki spermatozoidų, gali patirti šimtus DNR replikacijų ir ląstelių dalijimųsi, todėl apie 80 % visų lytinėse ląstelėse įvykusių taškinių *dnM* susidaro tėviniame alelyje. Vyresnis tėvo amžius apvaisinimo metu, manoma, yra pagrindinis veiksnys, sietinas su didesniu *dnM* skaičiumi palikuonio genome tiek populiacijos, tiek ir tos pačios šeimos lygmeniu (**6A pav.**) [1, 31, 33]. Spermatogonijų ląstelės dalijasi visą gyvenimą, o tai tikėtina sudaro galimybę kauptis *dnM* dėl klaidų DNR replikacijos metu ir dėl klaidų taisant nereplikavusią DNR tarp ląstelių dalijimųsi [76]. Taip pat su amžiumi gali sumažėti DNR taisymo mechanizmų bei endogeninės apsaugos sistemos prieš aktyvias deguonies formas efektyvumas [77, 78]. Žinoma, kad jaunesnių tėvų vaikų *dnM* **derinys**, palyginti su vyresnių tėvų vaikais, pasižymi savitais mutacijų šablonais ir lokalizacija vėlai besireplikuojančiuose genomo regionuose. Tai rodo, kad, be tėvų amžiaus, yra papildomų veiksnių, dėl kurių susidaro specifinės *dnM* [33, 43]. Danų tyrėjų apskaičiuota, kad su kiekvienu tėvo amžiaus metais lytinėse ląstelėse susidaro nuo vienos iki trijų *de novo* mutacijų daugiau, tačiau šis efektas skiriasi priklausomai nuo tiriamos šeimos [33, 43, 52]. Manoma, kad *dnM* skaičiaus įvairovę lemia individų mutagenezės dažnio nevienodumas dėl spermatogonijų kamieninių ląstelių dalijimosi dažnio skirtumų ir DNR klaidų taisymo mechanizmuose dalyvaujančių baltymų genetinės įvairovės [75, 79].

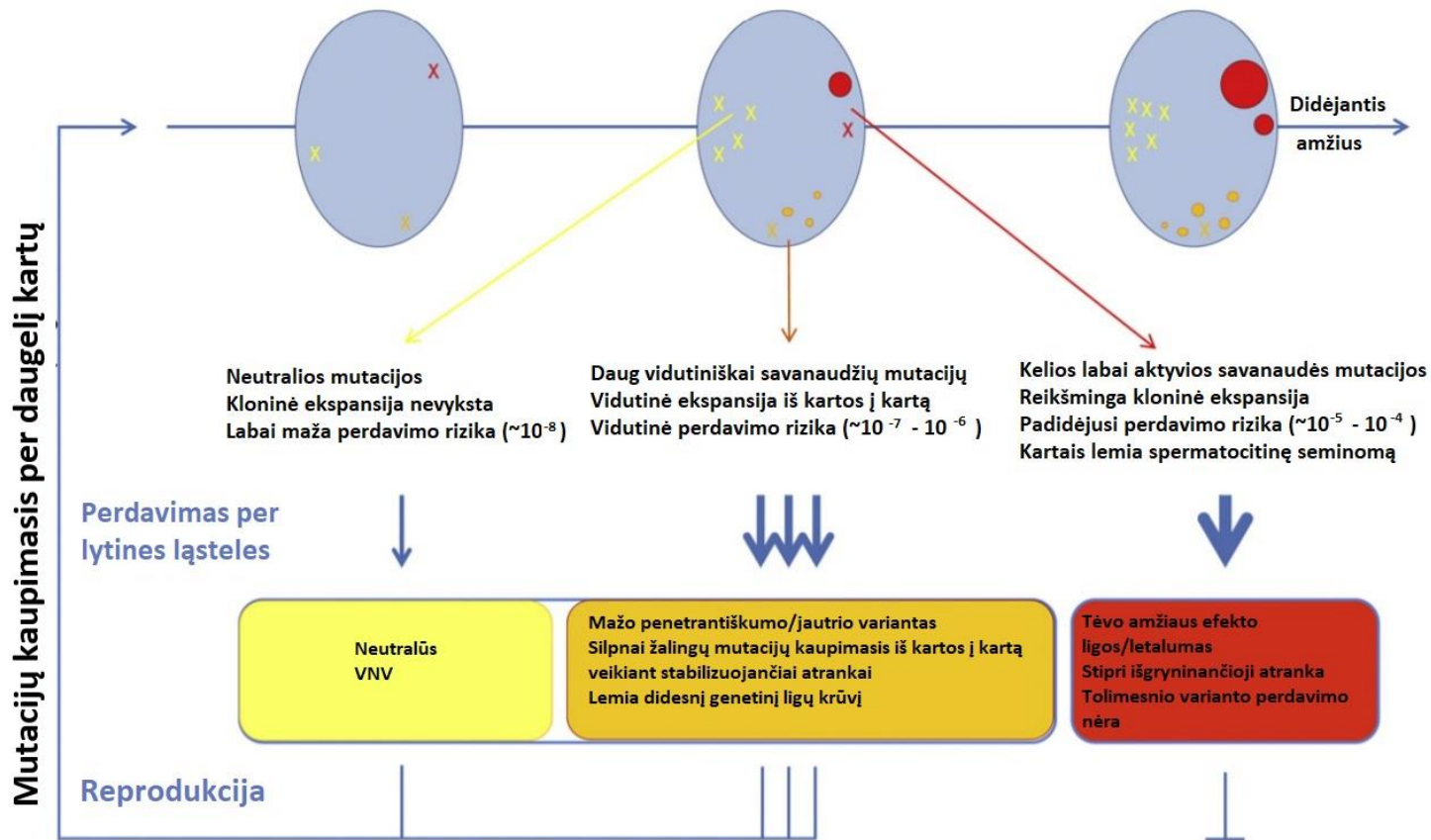


4 pav. Ląstelių dalijimasis A) oogenezės ir B) spermatogenezės metu. *DNMs* – *de novo* mutacijos. Besiformuojant kiaušialąstei per visą moters gyvenimą vyksta 24 ląstelių pasidalijimai. Vyrų ląstelių dalijimosi skaičius priklauso nuo amžiaus. Adaptuota iš [75]

Minėta, kad ryškus *de novo* mutacijų padidėjimas kartu su tėvų amžiumi buvo nustatytas mažam, tačiau specifiniam *dnM* pogrupiui, pasikartojančiam genome. Manoma, kad šios *dnM* suteikia spermatozonių kamieninėms ląstelėms pranašumo dauginantis, todėl sėklidėse dėl teigiamos gamtinės atrankos vyksta tokių mutuočių ląstelių kloninė ekspansija [81, 82]. Pavyzdžiui, RAS-MAPK kelio genuose įvykusios funkcijos įgijimo mutacijos lemia kloninę spermatozonių kamieninių ląstelių ekspansiją dėl selektyvaus jų pranašumo [80, 81]. Remiantis mutacijų karštų taškų (angl. *Mutation Hot Spot*) modeliu, tai aiškinama didesniu simetrinio ir asimetrinio mutuočių spermatozonių kamieninių ląstelių dalijimosi santykiu [81, 82]. Manoma, kad šių *dnM* susidarymas spermatozonių kamieninėse ląstelėse įvairuoja visų vyrų, tačiau vyresnių (36 m. ir vyresnių) vyrų sėklidėse

mutantinių spermatogonijų kamieninių ląstelių klonų skaičius yra didesnis [81, 83].

Įdomu, kad pirmosios nustatytos mutacijos, inicijuojančios kloninę ekspansiją spermatogonijų kamieninėse ląstelėse, lemia tokias vystymosi ligas – Noonan ar Costel sindromą (*de novo* mutacijos *PTPN11* ir *HRAS* genuose) [84], Apert, Creouzon ir Pfeiffer sindromus (*FGFR2*) [83, 85], achondroplaziją, Menkės sindromą, tanatrofinę displaziją (*FGF3*) [86] ir dauginę endokrininę neoplaziją (*RET*) [87]. Šios ligos yra žinomos kaip pasikartojančios, autosominiu dominantiniu būdu paveldimos ar tiesiog tėvinės kilmės ligos [50, 81], o mutacijos, kurių atžvilgiu spermatogonijų kamieninių ląstelių lygmeniu vyksta teigiama gamtinė atranka, tačiau organizmui jos yra patogeninės, vadinamos savanaudėmis mutacijomis [78, 88] (**5 pav.**). Dėl spermatogonijų ląstelėse susidariusių savanaudiškų *de novo* mutacijų, genetinių ligų dažnis esant vyresniam tėvų amžiui eksponentiškai didėja kartu su tėvų amžiumi [88] ir yra iki 1000 kartų didesnis nei tikėtinas bendrojoje populiacijoje [50]. Taip pat yra keliamą hipotezę, kad savanaudiškose *dnM* spermatogonijų kamieninėse ląstelėse gali būti daug dažniau aptinkamų fenotipų, tokių kaip intelektinė negalia, autizmas ar epilepsija, priepuolis [88]. Be to, pastebėta, kad esant teigiamai gamtinei atrankai, savanaudiškų *dnM* dažnai susidaro sėklidžių vėžinėse ląstelėse, todėl šių mutacijų grupės taip pat gali lemti tumorogenezę [62].

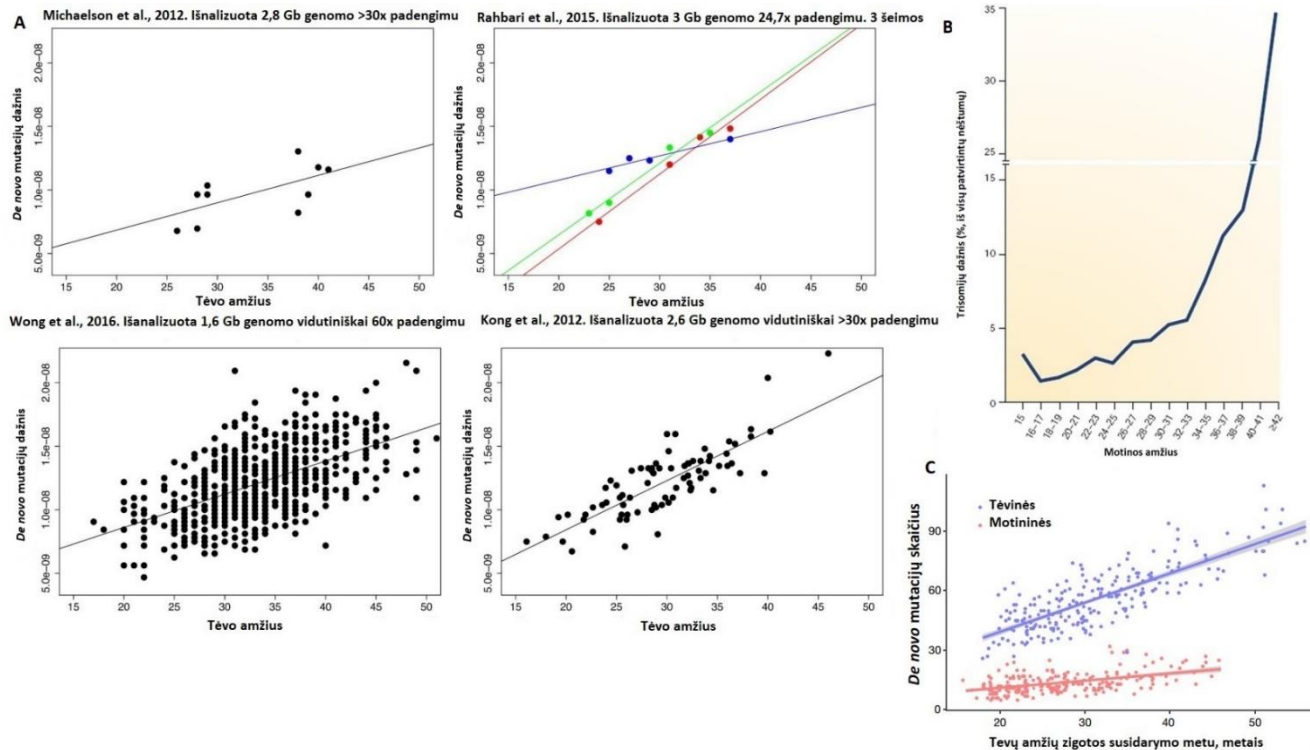


5 pav. Savanaudžių mutacijų padariniai. Mėlyni ovalai – tai sėklidės nuo brandos (kairėje) iki senatvės (dešinėje). Vykstant spermatogenezei, sėklidėse susidaro mutacijų, pažymėtų X. Priklausomai nuo mutacijų funkcinių padarinių spermatogonijų kamieninėse ląstelėse (SKL) galimi trys scenarijai. Funkciškai neutralios mutacijos (geltona spalva) nesikaupia ir yra asocijuotos su labai maža perdavimo palikuonims rizika. Raudonai pažymėtos yra dažniausiai tipinės tėvo amžiaus efekto mutacijos, būdingos vyresnio amžiaus vyrams. Jos patiria stiprią išgryninančiąją atranką, dėl to susidaro jų klonai (ovalai) ir padidėja pastarųjų perdavimo rizika (iki 1000 kartų didesnė už įprastą mutacijų dažnį) palikuonims. Retai šios mutacijos yra asocijuotos su spermatocitine seminoma. Oranžine splava pažymėtas scenarijus, kai susidaro vidutiniškai savanaudės mutacijos, pavyzdžiui, silpnos funkcijos įgijimo, kopijų skaičiaus pokyčiai ar raiškos reguliavimo pokyčiai ir išplinta nuo >1 iki 100 kartų daugiau. Daugelis tokių mutacijų gali egzistuoti per daugelį kartų, nes yra neutralios ar vidutiniškai patogeninės ir reikšmingai reprodukcinio tinkamumo (angl. *fitness*) nesumažina. Adaptuota iš [76].

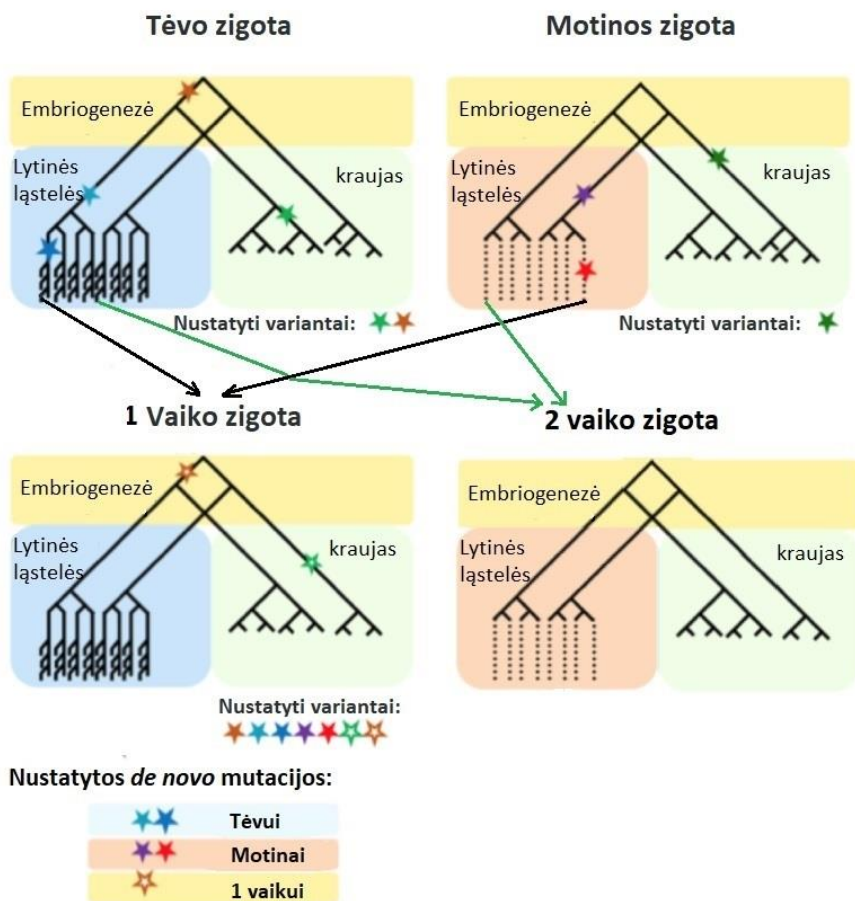
Motinos amžiaus efektas *de novo* mutacijų intensyvumui genome taip pat gerai ištirtas. Maždaug nuo 1998 m. žinoma, kad nuo motinos amžiaus kiaušialąstės apvaisinimo metu priklauso aneuploidijų tikimybė (**6B pav.**) [2, 3]. Tai lemia neteisingas chromosomų išsiskyrimas mejozės I arba mejozės II etapuose [13]. O *de novo* taškinių ar nedidelių (iki 20 bp) iškritų-intarpų motininės kilmės mutacijų priklausomybė nuo motinos amžiaus atrasta ir įrodyta amerikiečių tik 2016 metais [3]. Manoma, kad šio tipo mutacijos susidaro dėl DNR pažaidų susikaupimo ar patogeninių mutacijų pirminiuose oocituose dėl užtrukusios mejozės pirmos profazės [3]. Vis dėlto yra ir kitų paaiškinimų, pavyzdžiui, *dnM* skaičiaus didėjimo efektas gali atsirasti, jei vėliau ovuliuojantys oocitai daugiau kartų dalijasi mitozės būdu [89]. Tokiu atveju moterims maždaug nuo 17 iki 40 metų, kai nedidelių *de novo* mutacijų skaičius reikšmingai didėja, moterų oocitai mitozės būdu turėtų dalytis du kartus daugiau nei jaunesnių moterų oocitai. Vis dėlto šis scenarijus nepaaiškina tėvų ir motinų amžiaus, kai *dnM* skaičius ima didėti, santykinio stabilumo efekto [76].

DnM skaičiaus didėjimas su motinos amžiumi taip pat aiškinamas pozigotinėmis mutacijomis. Nors *dnM* įprastai apibrėžiamos kaip mutacijos, susidariusios tėvų lytinėse ląstelėse, tačiau triadų tipo tyrimuose analizuojami probando somatiniai audiniai, dažniausiai kraujas, todėl nustatytos *dnM* galėjo susidaryti tiek tėvų lytinėse ląstelėse, tiek ir ankstyvosios embriogenezės metu somatinėse ląstelėse (**7 pav.**) [90]. Verta paminėti, kad

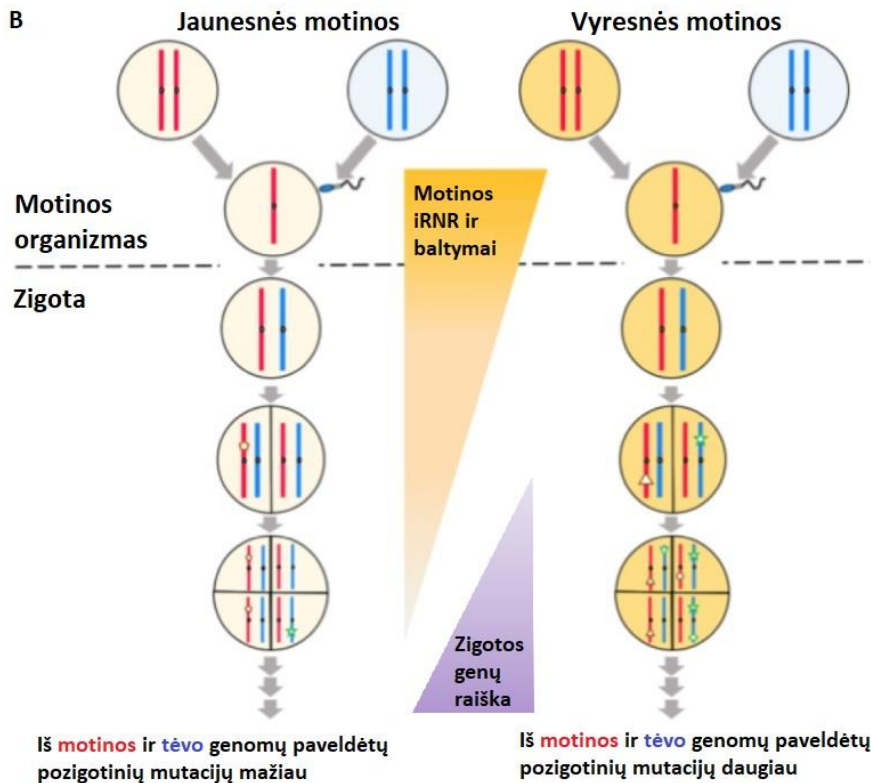
embriogenezės metu kelių pirmųjų ląstelių dalijimasis yra santykinai mutageninis [91–93], nes jų metu taisomos DNR pažaidos, susidariusios formuojantis spermatozoidams, kai neveikė ar netinkamai klaidas taisė bazių ekskizinės reparacijos mechanizmas [92] ir tik nuo kiaušialąstėje esančio baltymų ir iRNR transkriptų rezervo priklauso, kaip kokybiškai šios mutacijos yra ištaisomos iki ketvirto ląstelių dalijimosi etapo [94, 95]. Taigi vyresnių motinų replikacijos ir klaidų taisymo mechanizmai gali būti nepakankamai efektyvūs ir tai gali lemti didesnę palikuonio *dnM* skaičių ne tik tose chromosomose, kurios paveldėtos iš motinos, bet ir iš tėvo (**8 pav.**) [96, 97].



6 pav. A) Tėvų amžiaus ir *de novo* mutacijų dažnio priklausomybė. **B)** Motinos amžiaus ir trisomijų priklausomybė. Ne visos trisomijos pasireiškia vienodu dažniu. Dažniausiai nustatomos yra 13, 18 ir 21 chromosomų trisomijos. Kitų chromosomų, pavyzdžiui, 16 chromosomos, trisomijų dažnis didėjant motinos amžiui gali išlikti linijinis [98]. **C)** *De novo* taškinių ir trumpų išskirtų-intarpų skaičiaus priklausomybė nuo tėvo ir motinos amžiaus [25].



7 pav. *DnM* susidarymas gemetogenezės ir embriogenezės metu. Čia pavaizduota, kad naujausias bendras visų ląstelių protėvis yra apvaisinta kiaušialąstė. Pilnavidurės žvaigždutės vaizduoja tėvų *dnM*, o tuščiavidurės – vaiko. Analizuojant *dnM* standartiniu triadų analizės metodu, svarbu atsižvelgti į vaiko genomo alelių balansą pagal tiriamą genomo poziciją ar regioną, todėl tėvų genome identifikuotos vaiko *dnM* turi nebūti ar jos dalis tarp genomo variantų turi būti labai minimali. Tada vaiko genome identifikuojama *de novo* pozigotinė mutacija (ruda tuščiavidurė žvaigždutė), o tėvų lytinėse ląstelėse dauginasi tik nemutantiniai ląstelių klonai. Adaptuota iš [59, 95].



8 pav. Motinos amžiaus įtaka pozigotinių *de novo* mutacijų formavimuisi. Adaptuota iš [95]

Nors pagal šį modelį vyresnio amžiaus moterų dėl priešzigotinių ir pozigotinių mutacijų susikaupimo, tikėtina, motininės kilmės *dnM* turėtų susidaryti daugiau, tačiau, remiantis islandų 199 triadų viso genomo sekoskaitos duomenimis, su kiekvienais motinos amžiaus metais vaiko genome motininės ir tėvinės kilmės *de novo* taškinių ir mažų iškritų-intarpų mutacijų susidaro labai panašiai – atitinkamai po 0,34 ir 0,30 per metus [95]. Tai aiškinama tuo, kad prieš įvykstant apvaisinimui kiaušialąstė ir spermatozoidas yra sukaupę skirtingą DNR pažaidų skaičių, ir būtent spermatozoidai patiria didesnę oksidacinę stresą [96, 97], o įvykus apvaisinimui, bet prieš susiliejant spermatozoido ir kiaušialąstės haploidiniams probranduoliams, ląstelių genomai yra skirtingai epigenetiškai modifikuojami [88].

Tarp susikaupusių tėvinės ir motininės kilmės *dnM* iki 10 % visų *dnM* sudaro C>A transversijos [98]. Tarp motininės kilmės *dnM* taip pat

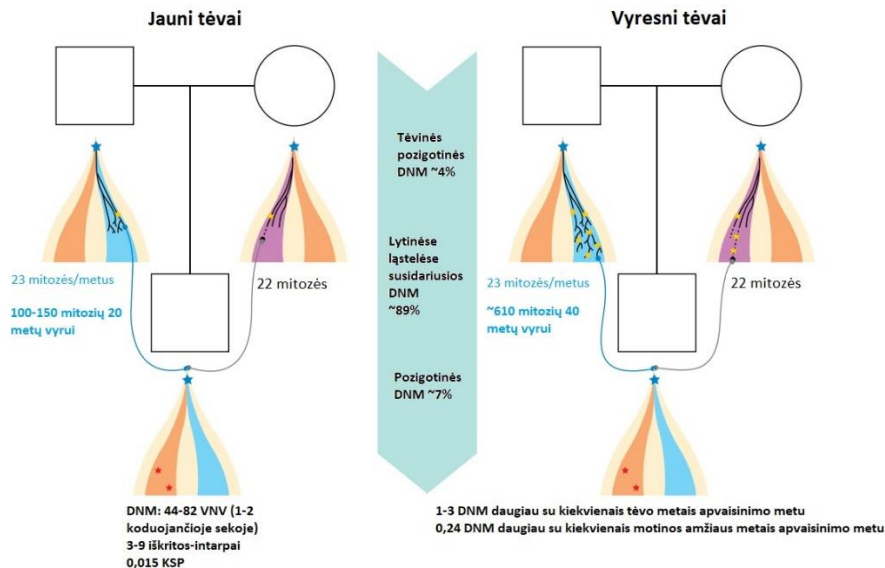
aptinkama ir savanaudžių mutacijų [74]. Manoma, kad viena jų dalis vyresnio amžiaus moterims suteikia galimybę subrandinti oocitus ir taip pailginti vaisingą moters gyvenimo etapą, o kita aptinkama genetiniuose regionuose, kuriuose yra vėžiui slopinti svarbūs genai [74].

2.7. *De novo* mutacijų kilmės ir amžiaus vertinimas

NKS metodų atradimas leido mokslininkams parodyti, kad *dnM* susidaro ne tik lytinėse, bet ir somatinėse ląstelėse ir daug dažniau, nei prieš tai buvo skaičiuojama [103]. Somatinės mutacijos, kurių atsiranda vėliau ar postnataliai, atvirksčiai, gali išlikti ir išplisti tik viename audinyje ar nedideliame somatinių ląstelių skaičiuje (**9 pav.**). Skaičiuojama, kad kraujyje didelis mozaicizmo dažnis yra pastovus maždaug 7 % *dnM*, nes, tikėtina, atsirado ankstyvoje pozigotinėje stadijoje [5, 102, 104]. Dėl šios priežasties ir kartu atsižvelgus į tai, kad embriono vystymuisi būdingas dažnas chromosomų nestabilumas ir struktūriniai genomo persitvarkymai, manoma, kad ankstyva embriogenezė yra didelio mutabilumo periodas [105, 106]. Kaip jau minėta, prieš transkripcijos iniciaciją ir transliaciją zigotoje žmogaus embriogenezė priklauso tik nuo motinos baltymų, esančių ir sąveikaujančių su oocitu [99], kurie gali lemti DNR replikacijos metu susidariusias *dnM* ir tuo pačiu daryti įtaką genomo nestabilumui [90], todėl priklausomai nuo embriogenezės etapo, kada *dnM* susidaro, *dnM* gali egzistuoti skirtinguose dauginių organų lygiuose ar būti organams specifinės [100, 101]. Somatinių žmogaus audinių tyrimas, kurio metu buvo analizuoti dauginiai mėginiai iš to paties individo, patvirtino, kad pozigotinių *dnM* paplitimas skirtingos embrioninės kilmės audiniuose somatinėse ir lytinėse ląstelėse yra skirtingas [88] ir, nors šių mutacijų neįmanoma susekti surinktuose tokių audinių, pavyzdžiui, kraujo ar burnos gleivinės epitelio, ėminiuose, tačiau jos gali būti perduodamos palikuonims kaip lytinėse ląstelėse susidarę *de novo* genomo variantai.

Somatinėse ląstelėse, manoma, kaupiasi šimtai skirtingų *dnM* mutacijų, susidariusių postnataliniame etape ir per visą žmogaus gyvenimą [110]. Didelių chromosomų anomalijų buvo nustatyta daugelyje žmogaus organų audinių [111], pavyzdžiui, kraujyje, kur šių pažeidimų su amžiumi tik daugėja [112–114]. Pavyzdžiui, Y chromosomos praradimas sėklidžių kraujo ląstelėse buvo nustatytas daugiau nei 15 % vyrų per 70 metų ir vyresniems [114, 115]. O jaunesnių nei 70 metų žmonių audiniams įprastai nustatomas mažas somatinių mutacijų mozaicizmo dažnis [115], pavyzdžiui, smegenims

[102], kraujui [103–105] ir odai apskaičiuotas somatinių mutacijų dažnis yra nuo dviejų iki šešių VNP Mb koduojančioje sekoje vienai ląstelei [106]. Manoma, kad *dnM* mutacijų dažnis yra didesnis somatinėse ląstelėse nei lytinėse dėl DNR replikacijos ir klaidų taisymo mechanizmų efektyvumo skirtumų lytinėse ir somatinėse ląstelėse, taip pat dėl mutagenų poveikio skirtumų [107].



9 pav. *DnM* kilmės ir amžiaus vertinimas. 20 metų vyro spermatogonijos mitozės būdu dalijasi 100–150 kartų, o 20 metų moterų oocitai tik 22 kartus (kairėje). *DnM* susidaro kiekviename kartoje dėl klaidų replikacijos, DNR pažaidų taisymo tėvų embriogenezės, gametogenezės ar pozigotinių procesų metu. Vyresnis tėvų amžius yra siejamas su didesniu *dnM* skaičiumi (dešinėje). Pavyzdžiui 40 metų vyro spermatogonijos dalijasi daugiau kaip 600 kartų, todėl kiekvienais metais jo vaikas turi viena–trimis *dnM* daugiau. Oogenezės metu vyksta fiksuotas mitozijų skaičius, tačiau *dnM* kaupiasi dėl bėgant laikui neįvykusių ar neefektyviai vykusių DNR pažaidų taisymo. Taigi su kiekvienais motinos amžiaus metais vaikas turi 0,24 mutacijos daugiau. Somatinės ląstelės čia pavaizduotos oranžine spalva, vyrų lytinių ląstelių linija – mėlyna, moterų – violetine. Mėlynos žvaigždutės žymi pozigotines mutacijas lytinėse ir somatinėse ląstelėse, geltonos – mutacijas, susidariusias tik lytinėse ląstelėse, raudonos – somatines mutacijas, susidariusias

embriogenezės metu ar postnataliai. Pažymėtina, kad tarp ~89 % lytinėse ląstelėse susidariusių *dnM* tėvinės ir motininės kilmės *dnM* santykis yra 4:1, tėvų amžius nesvarbu. Čia DNМ – *de novo* mutacija. Adaptuota iš [4].

2.8. *De novo* mutacijų dažnio skaičiavimo metodai ir jų taikymas evoliucinės ir populiacijų genetikos tyrimuose

Lytinėse ląstelėse įvykusių *de novo* mutacijų dažnį per kartą apibrėžia skirtumų tarp naujai susiformavusios zigotos ir gametų genomų, iš kurių zigota susiformavo, dažnis. Kol šis dydis negali būti tiksliai ir lengvai išmatuojamas dėl anksčiau minėtų priežasčių, dabar yra įmanoma suskaičiuoti tik giminingą dažniui įvertį – *de novo* mutacijų, esančių vaiko genome, bet nesančių tėvų genomuose, skaičių μ_G [30]. Šį skaičiavimo metodą pritaikė dešimtys tyrėjų grupių, tyrusių tėvų ir jų vaikų (triadų) genomus. Jų tyrimų duomenimis, apskaičiuotas μ_G įvertis yra 10^{-8} eilės į bp (**3 lent.**).

Be tiesioginio *de novo* mutacijų dažnio nustatymo pagal NKS duomenis, yra ir kitų būdų (**10 pav.**) [108]. Vienas iš populiariausių – populiacijų genetikos modelių rinkinys, pagrįstas genomo regionų, paveldėtų iš bendro protėvio, analize [109, 110]. Naudojant šį metodą, daroma prielaida, kad mutacijos yra funkciškai neutralios ir tikėtinas analizuojamų DNR sekų išsiskyrimas (t. y. tikėtinas mutacijų skaičius į bp) tarp organizmų rūšių poros (*d*) yra lygus $2\mu t$, čia μ yra vidutinis skirtumų tarp sekų dažnis per analizuojamą laiko periodą, o *t* – vidutinis laikas iki bendro protėvio (išsiskyrimo laikas). Žinant daugiau nei vienos rūšies ortologinių sekų skaičių, galima apskaičiuoti μ reikšmę ir išsiskyrimo laiką ($d/2\mu$). Dažniausiai tyrėjai nori žinoti rūšių išsiskyrimo laiką ne kartomis, o metais, todėl tam naudojamas metinis mutacijų dažnis ($\hat{\mu}_y$). Jam apskaičiuoti dažniausiai mutacijų dažnis kartai $\hat{\mu}_G$, nustatytas iš tėvų ir vaiko sekoskaitos duomenų, dalijamas iš vidutinio reprodukcinio amžiaus reikšmės (t. y. vienos kartos trukmės metais) [111]. Tokiais skaičiavimais $\hat{\mu}_y \approx 0,5 \times 10^{-9}$ bp [112, 113].

Šiuo tyrimo metodu ištyrus daugelio senovės žmonių (arba tiesiog protėvinius) branduolio genomus ir identifikavus genomų skirtumus, juos taip pat galima prilyginti evoliucijos „laikrodžio“ žymenims, kurie leidžia atskirti genomo evoliucijos laikotarpius. Remiantis $0,5 \times 10^{-9}$ bp *de novo* mutacijų dažnio reikšme apskaičiuota, kad Afrikos ir ne Afrikos (kalbama apie neandartaliečius, iš kurių kilo azijiečiai ir europiečiai) populiacijos

atsiskyrė daugiau nei prieš 100 000 metų [111, 114], o žmogus nuo šimpanzės – prieš 12 milijonų metų (mm) [111, 112, 115]. Tokie skaičiavimo rezultatai yra didesni nei skelbta anksčiau, bet neprieštarauja esamiems, vis dėlto labai ribotiems, paleontologiniams įrodymams apie hominidų kilmę [113, 114, 116, 117]. Daugiau klausimų kyla, kai lyginamas žmogaus ir orangutano ar žmogaus ir senojo pasaulio beždžionės (angl. *Old World Monkey*, toliau SPB) divergencijos laikas.

Pagal tariamai neutralias mutacijų pozicijas genome [99] apskaičiuota, kad žmogus nuo orangutano divergavo prieš 31 mm, o žmogus nuo SPB prieš 62 mm. Tai reiškia, kad žmogaus-orangutano divergencija įvyko dar oligoceno, o žmogaus-SPB prieš ar eoceno periodu. Remiantis metiniu mutacijų dažniu, apskaičiuotu pagal triadų sekoskaitos duomenis, šios rūšių divergencijų datos yra pernelyg didelės, kad jas būtų galima lengvai palyginti su dabartinėmis žiniomis apie rūšių divergenciją pagal fosilijų datavimo rodiklius [118, 119], todėl tai tik pagrindžia naudojamų populiacijos modelių netikslumą.

Kitas būdas evoliucijos įvykių chronologijai nustatyti yra $\hat{\mu}_y$ įverčių, nustatytų pagal sekoskaitos duomenis, palyginimas su $\hat{\mu}_y$ įverčiais, apskaičiuotais pagal primatų divergencijos dažnius, o šie pagal – fosilijų datavimo t rodiklius. Be abejo, tokie apskaičiuoti t rodikliai yra visiškai netiesioginiai, nes 1) fosilijų datavimo rodiklių yra sąlygiškai nedaug ir 2) pagal fosilijų požymius galima įvertinti tik mažai artimų rūšių išplitimą [120, 121]. Situaciją apsunkina ir tai, kad artimai susijusių rūšių t atspindi ne tik divergencijos laiką, bet ir vidutinį laiką iki artimiausio bendro protėvio protėvinėje populiacijoje, kuri galėjo būti didelė [111, 122]. Nors verta pažymėti, kad žmonių ir šimpanzių divergencijos laikas t , manoma, yra mažiausiai dviem milijonais metų ilgesnis nei išsiskyrimo laikas [111, 123, 124]. Taigi toks vertinimo metodas nėra patikimas. Nepaisant to, iki šiol bendru tyrėjų sutarimu naudojama \hat{t} vertė yra 6–7,5 mm žmonėms ir šimpanzėms [125, 126], 15–20 mm žmonėms ir orangutanams [126] ir 25–35 mm žmonėms ir SPB [114, 127]. Naudojant šias \hat{t} reikšmes ir anksčiau minėtą $\hat{\mu}_y = d/2\hat{t}$ formulę *de novo* mutacijų dažnis genome yra 10^{-9} eilės per metus, tai yra du kartus mažiau nei skaičiuojant $\hat{\mu}$ tik pagal sekoskaitos duomenų analizės duomenis, pateiktus **3 lent.** Kitaip tariant, bendru sutarimu naudojamos divergencijos reikšmės leidžia suprasti, kad pagal mutacijų skaičių, nustatytą, remiantis žmonių genomų sekoskaitos duomenimis, mutacijos kaupiasi greičiau nei turėtų.

3 lent. *De novo* mutacijų dažniai pagal žmogaus genomo sekoskaitos tyrimų duomenis [128]

Tyrimas	<i>dnM</i> dažnis bp kartai ($\times 10^{-8}$)	Vidutinis tėvų amžius, metais	<i>dnM</i> dažnis kai tėvo amžius 30 metų ($\times 10^{-8}$) ²	Tėvo amžiaus poveikis <i>dnM</i> skaičiui sulyg kiekvienais tėvo metais	Vertinamo genomo dydis, Gb (klaidingai neigiamų rezultatų dalis, %)	Triadų skaičius	Vidutinis sekos padengimas, kartais	CpG regionuose įvykusių tranzicijų skaičius ³
Roach 2010 [30] ^{4 5}	1,10 (0,68-1,70)	-	-	-	1,8 (5)	2	61,3 ⁷	0,178 (0,037-0,320)
Conrad 2011 (CEU) [129]	1,17 (0,86-1,62)	-	-	-	2,5 (5)	1	29,3	0,146 (0,046-0,246)
Conrad 2011 (YRI) [129]	0,97 (0,67-1,34)	-	-	-	2,5 (3,5)	1	29,2	0,114 (0,009-0,220)
Campbell 2012 [130]	0,96 (0,82-1,09)	26,3	-	-	2,2 (1,7 ⁷)	5	13,0	0,165 (0,110-0,220)
Kong 2012 [1]	1,20	29,7	1,21	2,01	2,6 (2,0)	78	30,0	0,173 (0,163-0,184)
Michaelson 2012 [8]	1,00	33,6	0,93	1,02	2,8 ⁶ (9,5)	10	13,0	0,128 (0,099-0,156)

² Apskaičiuota darant prielaidą, kad yra tiesinė priklausomybė, taikant nurodytą tėvo amžiaus poveikį ir atsižvelgus į vertinamo genomo dydį. Taikyta $\hat{\mu}_g + \frac{m(30-P)}{2N}$ lygtis, čia $\hat{\mu}_g$ yra tyrime nustatytas mutacijų lygis per bp vienai kartai, \hat{m} – nuokrypio kampas tėvo amžiaus poveikiui, P – vidutinis tėvo amžius tyrime ir N yra vertinamo genomo dydis bazių poromis.

³ CG dinuklotidų (CpG) dalis grįsta mutacijomis autosomose ir binominiu 95 % PI. Kai įmanoma, remtasi patvirtintomis mutacijomis. Vis dėlto kai kuriuose tyriuose buvo patvirtinta tik nedidelė dalis mutacijų, todėl čia vertintos visos galimos mutacijos.

⁴ Tyrimai, kuriuose naudotas *Illumina* sekoskaitos metodas.

⁵ Įtraukti broliai ir seserys.

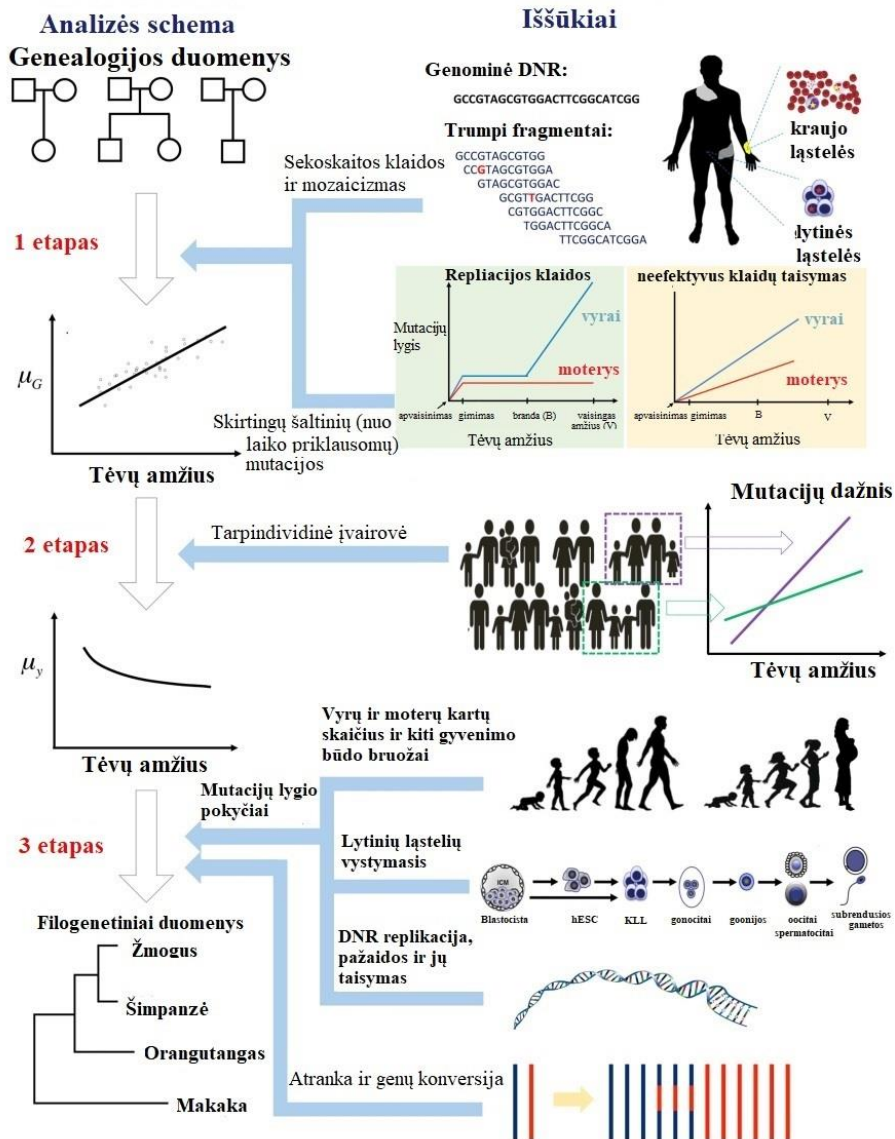
- Nėra duomenų.

⁶ Straipsnyje nepateikta, pagrįsta papildomais autorių komentarais.

Jiang 2013 [131]	-	34,4 ¹⁰	-	1,50	-	32	50,0	0,162 (0,146-0,177)
Francioli 2015 [32]	-	29,4	-	1,20 ⁷	2,1 (31,1)	250	13,0	0,165 (0,158-0,172)
Besenbacher 2015 [5]	1,27 (1,16-1,38)	28,4	1,3 ⁸	2,00	-	10	50,0	0,201 (0,166-0,236)

⁷ Tėvo amžiaus poveikis straipsnyje nėra nurodytas, tačiau apskaičiuotas remiantis Puasono regresijos modeliu pagal autosomose įvykusias *dnM*.

⁸ Pagrįsta vizuali nuokrypio kampo įvertinimu, pateiktu straipsnyje.



10 pav. *De novo* mutacijų dažnio nustatymo etapai naudojant skirtingus genetinės informacijos šaltinius. 1 etapo metu analizuojamas kiekvieno tiriamojo mutacijų spektras ir dažnis individualiai. 2 etapo metu analizuojamos ir lyginamos *dnM* tarp kartų. 3 etapo metu *dnM* analizuojamos evoliucijos kontekste. μ_G – mutacijų dažnis kartai, μ_y – metinis mutacijų dažnis, hESC – žmogaus embrionų kamieninės ląstelės, KLL – pirminės lytinės ląstelės. Adaptuota iš [128]

2.9. *De novo* mutacijos – neurovystymosi ligų priežastis

Medicininis *de novo* mutacijų išaiškinimo aktualumas buvo įvertintas tik ištobulinius NKS metodus tiek, kad tapo įmanoma šias *dnM* analizuoti patikimai, išsamiai ir už prieinamą kainą [132]. Iki tol žmogaus genetikos srityje dėmesys buvo sutelktas tik į paveldimas ligas. Analizuojant paveldimų ligų etiopatogenezę, buvo taikomi tokie tradiciniai ligos genų identifikavimo būdai, pavyzdžiui, pozicinis klonavimas, kurio metu buvo analizuojamos didelės šeimos (bent trys kartos) su daug sergančių narių, žinant jų plačią genealogiją. Vėliau buvo *Sanger* sekoskaita, kuria buvo identifikuojamos ligą lemiančios mutacijos genuose kandidatuose. Dabar taikomi NKS metodai, pavyzdžiui, VES ar VGS, suteikia galimybę identifikuoti ir apibūdinti jei ne visas, tai tikrai daugelį *dnM*, lemiančias genetines ligas [4, 132–134].

Neurovystymosi ligos – tai daugialypės būklės, kurioms būdingi pažinimo, bendravimo bei elgesio ir (arba) motorinių įgūdžių sutrikimai, atsirandantys dėl sutrikusio smegenų vystymosi. Neurovystymosi ligų grupei priskiriama intelektinė negalia (IN), autizmo spektro ligos, bendravimo sutrikimai, dėmesio trūkumo / hiperaktyvumo sutrikimai ir šizofrenija [135–137]. Dabar neurovystymosi ligų diagnostikai ar tam tikrų ligų grupių diferenciacijai biožymenų dar nėra. Šių ligų etiopatogenezės tyrimus sunkina tai, kad ligų grupes apibūdinantys fenotipai dažnai sutampa, pavyzdžiui, tiek šizofrenijai, tiek ir autizmo spektro ligoms būdingas sutrikęs socialinis pažinimas [138–141], o psichozė – šizofrenijai, bipoliniam sutrikimui ir depresiniams sutrikimams [141, 142], todėl kliniškai juos atskirti yra labai sudėtinga.

DnM, kaip neurovystymosi ligų priežastis, pirmiausia buvo priskirta dideliems *de novo* genomo pokyčiams – kopijų skaičiaus pokyčiams (KSP). Yra žinoma, kad didesni nei 100 kb *de novo* KSP bendrojoje žmonių populiacijoje nėra dažni, pasitaiko maždaug 1 iš 50 asmenų [143], tačiau tarp sergančiųjų tokiomis ligomis kaip intelektinė negalia, autizmas ar šizofrenija jie nustatomi maždaug 10 % visų pacientų [144–148]. Taigi pritaikius didelės raiškos genominių lustų technologijas dar praėjusiame dešimtmetyje atrasta daug naujų mikrodelecijų sindromų (pavyzdžiui, 1q21.1, 3q29, 15q13.3, 15q24, 17q12 ir 17q21.31), kurių priežastis yra *de novo* KSP. Daugelis nustatytų ir nustatomų *de novo* KSP yra reti arba išvis pirmą kartą nustatyti asmenims, sergantiems neurovystymosi ligomis, tai iš esmės palengvina jų patogeniškumo vertinimą nustatant diagnozes [145]. Vis dėlto verta

pažymėti, kad *de novo* KSP gali būti tiek pat neutralūs ar patogeniniai kaip ir paveldėti KSP, todėl kaip ir kiekvienas kitas *de novo* įvykis genome, *de novo* KSP negali būti vertinamas vienareikšmiškai [148]. Dėl šios priežasties neurovystymosi ligų etiopatogenzei aiškinti buvo pasiūlyta remtis dviejų KSP mutacijų modeliu [149]. Pagal šį modelį, asmenims, sergantiems neurovystymosi ligomis, turintiems pasikartojančių, dažniausiai paveldėtų KSP, antro, bent dalinio, KSP egzistavimas bet kur genome yra asocijuotas su padidėjusia ligos penetracija ir raiška [149].

De novo mutacijų interpretavimą neurovystymosi ligų kontekste taip pat sunkina genetinis šių ligų bruožų heterogeniškumas. Dėl jo komplikuojasi tiek *de novo* mutacijų aptikimas, tiek ir funkcinis jų interpretavimas. Iki tol, kol VES tapo prieinama pacientams su IN ir kitomis neurovystymosi ligomis, buvo galima identifikuoti tik tuos *de novo* VNP, kurie yra genuose, koduojančiuose baltymus, svarbius sinapsių fiziologijai. Tokiu būdu nesindrominėms IN formoms nustatytos *dnM*, būdingos sinapsių RAS GTPazę aktyvinančiame baltyme 1 (SYNGAP1) [150], o pacientams, sergantiems šizofrenija, *SHANK3* (*SH3* ir dauginiai ankirino pasikartojantys domenai 3) gene [151].

Dabar patogiai naudoti yra sukurta visų identifikuotų patogeninių, naudingų ir neaiškios klinikinės reikšmės *dnM* duomenų bazė *denovo-db* (<http://denovodb.gs.washington.edu>) [152]), kuria laisvai gali naudotis ir ją pildyti visi tyrėjai.

2.9.1. *De novo* mutacijos intelektinės negalios asmenų grupėje

Intelektinė negalia (IN) – heterogeninė neurovystymosi ligų grupė, kuriai būdinga:

- pažinimo sutrikimai;
- prisitaikymo prie aplinkos sutrikimai;
- pasireiškimas raidos metu [153].

IN vertinti naudojamas intelekto koeficientas (IQ). Pagal jį skiriama lengva (IQ 50–69), vidutinė (IQ 35–49), sunki (IQ 20–34) ir gili (IQ < 20) IN. Skaičiuojama, kad genetiniai veiksniai lemia apie 50 % sunkios ir apie 15 % lengvos IN atvejų [154]. Šis sutrikimas gali būti izoliuotas (izoliuota IN)

arba pasireikšti kartu su dismorfiniais požymiais, raidos anomalijomis ar neurologiniais simptomais (sindrominė IN). Genetinės priežastys tiek sindrominės, tiek ir nesindrominės IN atvejais gali būti įvairios: chromosomų anomalijos, mutacijos viename gene (paveldėjimas pagal Mendelio dėsnis arba sporadiniai atvejai) arba mutacijos keliuose genuose kartu su aplinkos poveikiu (daugiaveiksnis paveldėjimas) [155].

IN dažnis populiacijoje yra 1–3 %. Jos priežastys labai heterogeninės, daugeliu atvejų – genetinės, todėl, nepaisant technologijų pažangos genetikos srityje, IN priežastys 60 % pacientų lieka nežinomos [156]. Didelį IN neišaiškinimo procentą lemia nepakankamas tiriamųjų grupės dydis. Per paskutinius 10 metų IN tyrėjams jungiantis visame pasaulyje ir tiriamųjų grupes jungiant iki 6000 ir daugiau tiriamųjų (> 2000 triadų) identifikuojama vis daugiau IN galimai lemiančių genų kandidatų ar genetinių sričių su *dnM*. 2016 metais 2104 triadų tyrimo metu identifikuota dešimt naujų IN genų kandidatų: *DLG4*, *PPM1D*, *RAC1*, *SMAD6*, *SON*, *SOX5*, *SYNCRIP*, *TCF20*, *TLK2* ir *TRIP12* [156]. *DnM* juose yra žalingi įvairiems procesams, pavyzdžiui, chromatino modifikavimui (*TRIP12* ir *TLK2*), lūžiosios X intelektualinės negalios baltymo (*FMRP*) taikinio ir sinapsių plastiškumo genų (*DLG4*) ar embrionuose vykstančių genų (*PPM1D* ir *RAC1*) raiškiai. Kitas 4293 triadų VES tyrimas parodė, kad net 25 % sergančiųjų neurovystymosi ligomis turi *dnM* genuose, asocijuotuose su vystymosi ligomis [157], o kiti 17 % sergančiųjų taip pat turi *dnM* genuose, tačiau jų asociacija su IN dar nėra patvirtinta.

IN etiopatogenezeje, be baltymus koduojančių genų, ypač svarbūs ir **reguliaciniai genomo elementai**. Žinoma, kad *dnM* reguliaciniuose genomo regionuose lemia sunkią ar gilią IN. Šie genomo elementai pirmiausia veikia kaip genų raiškos stiprintuvai arba alternatyvaus baltymų sukirpimo reguliaciniai elementai. 2018 metais publikuotas Short et al tyrėjų darbas, kurio metu ištirti 7930 triadų reguliaciniai genomo elementai, atskleidė 88 *dnM* evoliuciškai konservatyviuose genomo elementuose, kurių raiška vyksta vaisiaus smegenyse [158]. Iš jų daugiau nei 4 % *dnM* lemia autosominiu dominantiniu būdu paveldimą IN ir tik 0,15 % *dnM* lemia didelio penetrantiškumo neurovystymosi ligas [159], todėl apibendrintai skaičiuojama, kad patogeninės nekoduojančiuose genomo regionuose esančios *dnM* yra atsakingos už mažiau nei 5 % visų IN atvejų, kai jos priežastys genetinės, tačiau ne dėl mutacijų egzome. Tai tik įrodo, kad neurovystymosi ligų grupėje išsamiai visų klasių reguliacinių nekoduojančių genomo elementų analizei statistiškai patikimiems rezultatams reikalingos

dešimčių tūkstančių, jei ne šimtų tūkstančių, tėvų-probandų triadų viso genomo sekos.

2.9.2. *De novo* mutacijos asmenų, sergančių autizmo spektro ligomis, grupėje

Autizmo spektro ligos (ASL) – tai smegenų vystymosi nukrypimo būklės, apibūdinamos įvairiu nuosekliu socialinių interesų trūkumu ar sulėtėjimu, komunikacijos stoka ir pasikartojančiomis griežtomis tvarkos veiksmy sekomis. Šie požymiai gali pasitaikyti ir nesant autizmo, todėl jis diagnozuojamas tik turintiems akivaizdžius visus tris simptomus [159]. Autizmo sukeliama raidos nukrypimai yra plataus spektro, todėl kiekvienam individui pasireiškia skirtingo sunkumo simptomais. Daugeliu atvejų ASL būklės išryškėja per pirmuosius penkerius gyvenimo metus ir išlieka tiek paauglystės laikotarpiu, tiek ir visą suaugusiojo gyvenimą. Asmenims, kuriems diagnozuojama ASL, dažnai kartu pasireiškia ir epilepsija, depresija, nerimo ir dėmesio trūkumo / hiperaktyvumo liga, miego sutrikimai, autoimuninės ligos, rezorbcijos sutrikimai (sutrikusi plonųjų žarnų veikla), dėl kurių jie nesugeba pasisavinti reikalingų medžiagų. Šių asmenų intelektas labai varijuoja, nuo gilios intelektualinės negalios iki aukštesniojo intelekto lygio.

Pasaulio sveikatos organizacijos duomenimis, vidutiniškai visame pasaulyje ASL serga vienas iš 160 vaikų. Remiantis epidemiologiniais tyrimais, atliktais per pastaruosius 50 metų, ASL paplitimas didėja visame pasaulyje ir dabar siekia apie 1 % bendrosios populiacijos [160]. Tam galimai turi įtakos diagnostinių kriterijų išplėtimas, atliekamų tyrimų skaičius ir naujų technologijų pritaikymas. Autizmo priežastys dar nėra žinomos. Konkretūs autizmą nulemiantys genai nėra nustatyti. Manoma, kad tai yra genetinio paveldėjimo ir aplinkos veiksnių derinys. Paveldėjimo reikšmę įrodo tai, kad šis sutrikimas labiau paplitęs tarp berniukų (2–4 kartus dažnesnis tarp berniukų nei tarp mergaičių [161, 162], bei monozigotinių dvynių, nes požymių konkordiskumas siekė 70–90 % (dizigotinių dvynių jis tesiekė 10 %) [163, 164]. Naujausiais tyrimų duomenimis, genetinis paveldėjimas lemia 83 % ASL variacijos [165]. Be to, ASL rizika didėja vaikams, kurie jau turi ASL sergantį sibsą, tada rizika sirgti ASL padidėja iki 6,9–18 % [103, 166, 167].

Didelė ASL paveldėjimo dalis aiškinama VNP įvairove. Deja, pirmieji viso genomo asociacijos tyrimai dėl per mažų imčių neaptiko reikšmingai stiprių asociacijos su ASL signalų, tačiau vėlesniuose tyrimuose, kur tiriamųjų imtys buvo dešimt ir daugiau kartų didesnės, asociacijos tyrimų

buvo identifikuota 12 naujų genetinių sričių, asocijuotų su ASL [168], be to, nustatyti 93 ASL viso genomo žymenys, iš kurių net 53 nepriklausomose grupėse buvo patvirtinti pakartotinai [169].

Nepaisant įrodymų, kad ASL etiologijai yra svarbūs dažni genomo variantai, reti (RAD<1 %, angl. *MAF*) genomo variantai lemia dar didesnę individualią riziką sirgti ASL (**4 lent.**). Kaip jau minėta, reti genomo variantai gali būti tiek paveldėti, tiek ir *de novo*. Tai gali būti iškritos, intarpai, KSP ar VNP [170]. *De novo* mutacijų svarba ASL etiologijoje yra glaudžiai susijusi su gamtinės atrankos vaidmeniu ir alelių dažniu. Gamtinė atranka įprastai stengiasi pašalinti žalingus variantus, o vis dėlto išlikę tokie variantai yra teigiamos atrankos požymis [171]. *De novo* variantai susidaro naujai, todėl gamtinės atrankos poveikio iš pradžių nepatiria ir turi stiprų poveikį genams, kurių raiška vyksta smegenyse [172]. Tai leidžia identifikuoti ASL rizikos genus.

4 lent. ASL genetinė architektūra. Adaptuota iš [173]

Mutacijų klasė	Polinkis sirgti ASL, %	Mutacijų klasė	Skirtingos mutacijų klasės, kurias turi ASL pacientai, %
Dažni genomo variantai	49,4	<i>De novo</i> KSP	4-7
<i>De novo</i> genomo variantai	3	<i>De novo</i> VNP	7
Reti paveldėti variantai	3	Reti autosominiu recesyviu būdu paveldimi variantai	3
Viso	55	Su X chromosoma susiję genomo variantai	2
		Iš viso	16–19

Didelių imčių VES genetiniai tyrimai atskleidė, kad maždaug 30 % visų tirtų ASL pacientų, kurių beveik pusė (45 %) yra moteriškos lyties, ASL lėmė *de novo* mutacijos [174]. *De novo* mutacijų indėlis sirgti ASL yra vertinamas

ir prioretizuojamas pagal du veiksnius: 1) aminorūgšties pokyčio poveikį baltymo funkcijai, todėl santykinė rizika ASL visada bus didesnė tiems asmenims, kurie turi baltymo funkcijos praradimo *de novo* mutacijų, nei tiems, kurie turi nesinoniminės pakaitos tipo mutacijų; ir 2) geno tolerantiškumą / netolerantiškumą mutacijoms. Yra genų, kuriuose nustatoma daugiau nei tikėtasi funkciinių mutacijų (tolerantiški genai), ir yra tokių, netolerantiškų, kuriuose mutacijų nėra arba jos yra pavienės, todėl mažai tikėtina, kad tolerantiški genai turi įtakos ASL vystymuisi.

Sudarius genų kandidatų, sietinų su ASL etiologija, sąrašą, toliau yra nustatomi nervinio signalo perdavimo keliai (sinapsių funkcijos, signalo perdavimo neuronais, kanalų aktyvumo ir chromatino modifikacijos), kuriuose genų produktai sąveikauja, vertinami genų-genų, genų-baltymų, baltymų-baltymų sąveikos ir kuriuose šie baltymai veikia [131, 175, 176]. ASL patogenezės etiologijai aiškinti yra nustatytos *de novo* mutacijos sinapses koduojančiuose genuose, pavyzdžiui, *NLGN3* ir *NLGN4X* [176], *SHANK3* [177], *NRXN1* [178] ir *CNTNAP2* [179], dėl kurių sutrinka sinapsių kontaktų vystymasis ir palaikymas. O chromatino reguliavimas yra labai svarbus neuronų vystymuisi ir netinkamas jo struktūros organizavimas gali lemti kognityvius (pažintinių funkcijų) sutrikimus. Chromatino struktūros modifikavimas kontroliuoja ląstelės likimą ir funkciją [180–182], o jį patį modifikuoja daugybė veiksnių, kurie gali būti susiję su ASL etiologija ir kitomis neurologinėmis ligomis, pavyzdžiui, *Coffin-Siris* sindromu [183], *Nicolaides-Baraitser* sindromu [184], *CHARGE* sindromu [185] ar *Rubinstein-Taybi* sindromu [186]. Apibendrintą genų, siejamų su ASL patogenezė, kuriuose nustatytos *de novo* mutacijos, ir biologinių procesų sąrašą galima rasti Neale et al [37].

DARBO METODOLOGIJA

3.1. Tiriamieji

Siekiant įgyvendinti mokslinio darbo užduotis ir tikslą, tyrimams buvo panaudoti dviejų mokslinių tyrimų (LITGEN ir UNIGENE) duomenys.

LITGEN (2011–2015 m.) „Lietuvos populiacijos genetinė įvairovė ir sandaros kitimai, susiję su evoliucija ir dažniausiai paplitusiomis ligomis“ biomedicininiam tyrimui atlikti 2011 m. gegužės 3 d. buvo gautas Vilniaus regioninio biomedicininio tyrimų etikos komiteto leidimas Nr. 158200-05-329-79. Projektas finansuotas Europos socialinio fondo lėšomis; projekto vadovas LMA tikrasis narys akad. prof. habil. dr. Vaidutis Kučinskas.

LITGEN tiriamųjų grupė. LITGEN projektui buvo surinkta etninė sveikų lietuvių populiacijos asmenų grupė. Ją sudarė 1000 tiriamųjų asmenų (279 triados). Projekto metu surinktas jų veninis kraujas, kraujo plazma ir užpildyti anketiniai duomenys. LITGEN atrankos kriterijai:

- sveiki asmenys;
- trijų kartų genealogijoje visi asmenys lietuvių tautybės;
- asmenys, gyvenantys toje pačioje Lietuvos etnolingvistinėje teritorijoje (Vakarų, Šiaurės, Pietų Žemaitija ir Vakarų, Rytų, Pietų Aukštaitija).

Darbui atlikti panaudoti LITGEN projekte dalyvavusių 147 asmenų (49 triadų) viso egzomo sekoskaitos duomenys, reprezentuojantys Lietuvos populiacijos tiriamųjų grupę (toliau pirma grupė).

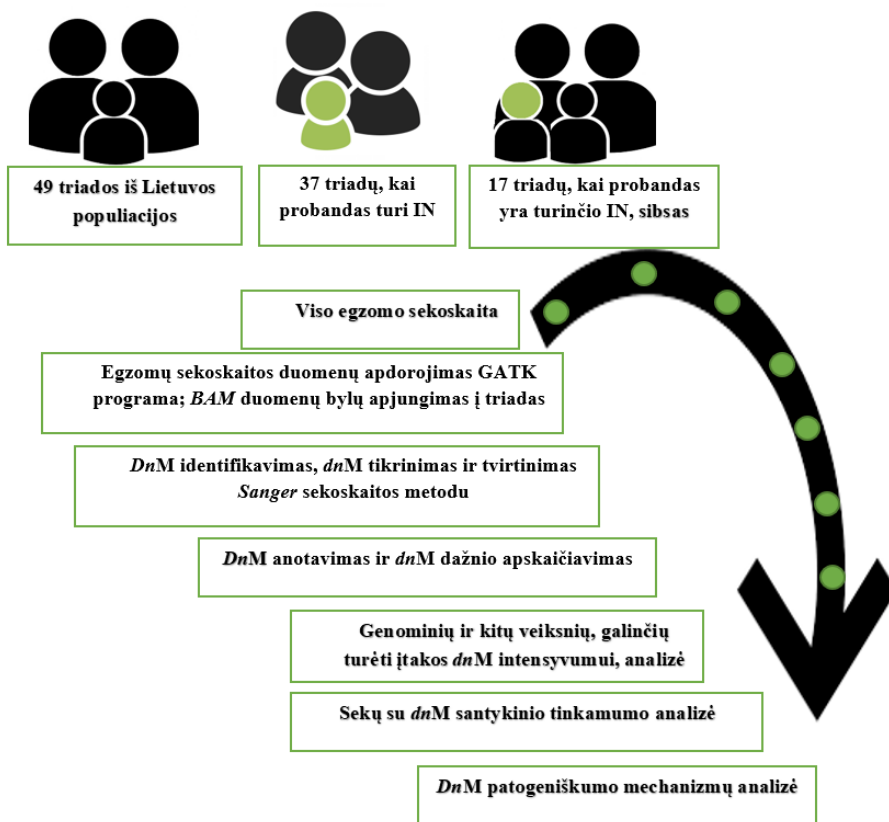
UNIGENE (2012–2016 m.) „Unikalūs genomo persitvarkymai esant įgimtiems nervų sistemos raidos sutrikimams: kilmė, genominiai mechanizmai, funkcinės ir klinikinės pasekmės“ biomedicininiam tyrimui atlikti 2012 m. lapkričio 29 d. buvo gautas Vilniaus regioninio biomedicininio tyrimų etikos komiteto leidimas Nr. 158200-12-563-164. Projektas finansuotas Europos socialinio fondo ir Lietuvos ir Šveicarijos Konfederacijos bendradarbiavimo programos „Moksliniai tyrimai ir plėtra“ fondo lėšomis; projekto vadovas LMA tikrasis narys akad. prof. habil. dr. Vaidutis Kučinskas.

Šiame darbe *de novo* mutacijoms tirti intelektinė negalia buvo pasirinkta kaip genetinės ligos modelis. Tai ypač heterogeninė genetinė liga, kurios priežastys ir pasekmės kiekvienu analizuojamu atveju yra skirtingos, o žmonių, kuriems nustatyta intelektinė negalia, skaičius nuolat didėja [187]. Asmenų su IN sibsų grupė – tai tarpinė tiriamųjų grupė tarp „sveikųjų“ Lietuvos ir asmenų su IN grupių.

UNIGENE tiriamųjų grupės. UNIGENE projektui buvo surinkta 128 tiriamųjų grupė. Ją sudarė tiriamieji, atitinkantys pagrindinį atrankos kriterijų, turintys intelekto negalia, kai intelekto koeficientas pagal *Wecksler* skalę yra < 70 arba yra matomas raidos atsilikimas (vaikams iki 6 metų) tais atvejais, kai yra:

- Įgimtos centrinės nervų sistemos anomalijos,
- Dauginiai raidos defektai / sindrominė intelekto negalia, neaprašyta anksčiau,
- Nežinomos priežasties neurologiniai sindromai (distonija / ataksija / cerebrinis paralyžius),
- Sunki idiopatinė epilepsija,
- Autizmas,
- Izoliuota intelekto negalia, sporadiniai atvejai,
- Subalansuoti chromosominiai persitvarkymai,
- Įtariamasis žinomas sindromas / būklė, kurios genomine etiologija nėra žinoma / pasireiškę nauji klinikiniai požymiai.

Darbo metu analizuoti tik tie tiriamieji, kuriems atlikta viso egzomo sekoskaita. Darbui atlikti UNIGENE dalyvavę tiriamieji suskirstyti į dvi grupes (antrą ir trečią grupes). Antrą tiriamųjų grupę sudarė UNIGENE projekte dalyvavę 111 asmenų (iš viso 37 triados, kai probandas turi intelektinę negalia ar autizmo spektro ligą, o jų tėvai intelektinės negalios neturi). Trečią tiriamųjų grupę sudarė 51 asmuo (iš viso 17 triadų, kai probandas yra asmens, turinčio intelektinę negalia, sibsas, tačiau nei pats, nei jo tėvai intelektinės negalios neturi) (**11 pav.**).



11 pav. *DnM* intensyvumo skirtinguose geno regionuose tyrimo schema

3.2. Tyrimo eiga ir metodai

3.2.1. DNR išgryninimas

Genominė DNR išgryninta iš veninio kraujo leukocitų naudojant standartinę fenolio chloroformo metodą arba automatizuotą branduolio rūgščių išgryninimo sistemą, naudojant magnetines daleles (*TECAN Freedom EVO® 200*, *Tecan Schweiz AG*, Šveicarija) pagal gamintojo rekomendacijas. Išgrynintos genominės DNR koncentracijos ir švarumo matavimas atliktas spektrofotometru pagal VŠĮ VUL SK MGC patvirtintą tyrimo procedūrą.

3.2.2. Viso egzomo sekoskaita

LITGEN tiriamųjų grupės asmenų (triadų), atitinkančių įtraukimo į LITGEN projektą kriterijus, sutikusių dalyvauti tyrime, pasirašiusių Informuoto asmens sutikimo formą, viso egzomo sekoskaita buvo atlikta Vilniaus universiteto Medicinos fakulteto Biomedicinos mokslų instituto Žmogaus ir medicininės genetikos katedros mokslininkų 2011–2013 metais. Naujos kartos sekoskaita atlikta genetiniu analizatoriumi *5500 SOLiD™ Sequencer (Applied Biosystems, JAV)* pagal gamintojo protokolus.

UNIGENE tiriamųjų grupės 7 triadų, kai probandas turi IN ir atitinka įtraukimo į UNIGENE projektą kriterijus, sutikusių dalyvauti tyrime, pasirašiusių Informuoto asmens sutikimo formą, viso egzomo sekoskaita buvo atlikta Vilniaus universiteto Medicinos fakulteto Biomedicinos mokslų instituto Žmogaus ir medicininės genetikos katedros mokslininkų 2013–2016 metais genetiniu analizatoriumi *5500 SOLiD™ Sequencer (Applied Biosystems, JAV)* pagal gamintojo protokolus. Kitų UNIGENE projekte dalyvavusių 30 triadų, kai probandas turi intelektualinę negalią, ir 17 triadų, kai probandas yra asmens, turinčio intelektualinę negalią, sibsas, viso egzomo sekoskaita buvo atlikta Loanos universiteto Biologijos ir medicinos fakulteto Integruotos genomikos centre, Šveicarijoje, dr. Norine Voisin, dr. Lucie Gueneau ir dr. Katrin Mannik. Viso egzomo DNR bibliotekų, paruoštų rinkiniu *SureSelect Human All Exome V5 (Agilent, JAV)*, naujos kartos sekoskaita atlikta genetiniu analizatoriumi *Illumina HiSeq 2500 (Illumina, Inc., JAV)* pagal gamintojo protokolus.

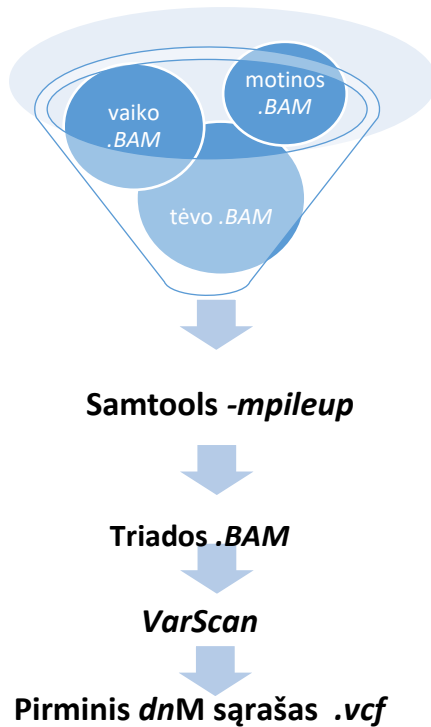
3.2.3. NKS duomenų apdorojimas

LITGEN ir daliai UNIGENE tiriamųjų egzomų, nuskaitytų *5500 SOLiD™ Sequencer* genetiniu analizatoriumi, pirminiai egzomų, padalintų į 75 bp ilgio fragmentus, duomenys buvo generuojami binukleotidiniu spalviniu kodu ir pateikti *.XSQ* formatu. *.XSQ* duomenų bylos *LifeScope™ Genomic Analysis Software 2.5.1* programa sulygiuotos su referentiniu hg19 genomu, pateiktu UCSC duomenų bazėje. Tokiu būdu gautos išvestinės sulygiavimo duomenų bylos *.BAM* formatu (**1 priedas**). Vidutinis horizontalusis padengimas fragmentais buvo 38,5.

UNIGENE tiriamųjų egzomų, nuskaitytų *Illumina HiSeq 2500* genetiniu analizatoriumi, pirminiai egzomų, padalintų į 150–200 bp ilgio fragmentus,

duomenys buvo generuoti ir pateikti *.FASTQ* formato duomenų bylomis. Jose yra informacija apie fragmentus, fragmentų sekos ir kiekvieno fragmento sekoje identifikuotų heterociklinių bazių kokybės įverčiai Q . Kiekvienam egzomui generuotos *.FASTQ* duomenų bylos sujungtos C++ programos *cat* funkcija. *Burrows-Wheeler Aligner* (BWA) [188] bioinformaciniu įrankiu *.FASTQ* bylos sulygiuotos su referentiniu hg38 genomu, pateiktu UCSC duomenų bazėje. Tokiu būdu gautos išvestinės sulygiavimo duomenų bylos *.BAM* formatu. Vidutinis horizontalusis padengimas fragmentais buvo 100.

Visos duomenų bylos *.BAM* formatu toliau apdorotos *The Genome Analysis Toolkit v.3.5* (GATK) įrankiu pagal geros praktikos gaires (2 priedas). Apdorotos *.BAM* duomenų bylos GATK programa, *SAMtools v.1.1* (*Sequence Alignment/Map*) programos [17] *mpileup* funkcija sujungtos pagal triadas po tris bylas į vieną (12 pav.).



12 pav. *DnM* identifikavimo schema.

3.2.4. *De novo* mutacijų identifikavimas ir tikrinimas

Pirminis *de novo* mutacijų identifikavimas atliktas atviros prieigos *VarScan* (*Variant detection in massively parallel Sequencing data*) v.2.4.2 programa [189] (**12 pav.**) pagal protokolą, pateiktą **3 priede**, ir alternatyvia komercine *VarSeq*TM programa. Potenciali *de novo* mutacija buvo identifikuojama, jei vaikas turi genomo variantą, kai nė vienas iš tėvų toje pačioje genomo pozicijoje jo neturi. Rezultatai pateikiami generuotoje *.vcf* formato duomenų byloje. Klaidingai teigiamų *VarScan* programos rezultatų pirminiam atmetimui panaudota *SnpSift v.4* programa. Taikyti konservatyvūs filtravimo kriterijai pateikti **5 lent.** *DnM* sąrašui, gautam *VarSeq*TM programa, pritaikyti tie patys filtravimo kriterijai *Trio Workflow* dalyje. Siekiant atvesti somatines *dnM* (*dnM* dalį, esančią tik nuskaitytose kraujo ląstelėse) arba sekoskaitos artefaktus, *dnM* su mažu (< 0,3) alelių balansu, filtruoti buvo pritaikytas alelių balanso (sekoskaitos fragmentų dalies su alternatyviu aleliu) [0,3; 0,7] intervalas.

5 lent. Duomenų filtravimo kriterijai *.vcf* duomenų byloms

Kriterijus	Pritaikyta reikšmė
Horizontalusis regiono padengimas	>30
Genotipo kokybės įvertis (GQ)	≥ 50*
Referentiniam aleliui alternatyvaus alelio padengimas (DP)	> 20*
Alelių balansas (AB)	[0,3;0,7]*

* Remtasi [1, 5, 32]

Antriniam *dnM* tikrinimui panaudota *Integrative Genomics Viewer (IGV)* programa [190], o galutinai *dnM* patvirtinti atlikta sekų, kuriose nustatytos *dnM*, sekoskaita *Sanger* metodu. *DnM* tikrinti *Sanger* sekoskaitos metodu atsitiktiniu būdu atrinkta 50 *dnM* iš LITGEN projekto duomenų ir 50 *dnM* iš UNIGENE projekto duomenų. Koduojančios geno sekos polimerazės grandininė reakcija atlikta naudojant specifinius oligonukleotidinius pradmenis, sukurtus *Primer Blast* (NCBI) įrankiu (**4 priedas**). Abiejų kryptių sekoskaita atlikta naudojant rinkinį *BigDye® Terminator v3.1 Cycle Sequencing Kit* (Applied Biosystems, JAV). Kapiliarinė elektroforezė atlikta genetiniu analizatoriumi *ABI 3130xL Genetic Analyser* (Applied Biosystems,

JAV). Fluorescencijos signalai analizuoti *DNA Sequence Analysis v5.1* programa (Applied Biosystems, JAV). Nuskaitytos sekos lygintos su referentinėmis genų sekomis.

3.2.5. Biologinės tėvystės tikrinimas

Tais atvejais, kai *dnM* buvo nustatyta reikšmingai dažniau nei visose kitose triadose, buvo atliktas biologinės tėvystės tikrinimas. Biologinės tėvystės tikrinimas buvo atliktas *AmpFLSTR™ Identifiler™* rinkiniu (Applied Biosystems™, JAV) pagal gamintojo rekomendacijas.

3.2.6. Tyrimo duomenų anotavimas

Patvirtintų *dnM* anotacija atlikta *ANNOVAR* (v.20170601 hg19 referentiniu genomui ir v.20170912 hg38 referentiniu genomui) programa [191]. Naudotas plačiausiai taikomų anotacinių įrankių paketas, sudarytas iš *SIFT*, *PolyPhen*, *LRT*, *MutationTaster*, *MutationAssessor*, *FATHMM* (angl. *Functional Analysis through Hidden Markov Models*), *PROVEAN* (angl. *Protein Variation Effect Analyzer*), *CADD* (angl. *Combined Annotation Dependent Depletion*), *GERP++* (angl. *Genomic Evolutionary Rate Profiling*), *PhyloP*, *SiPhy*, *COSMIC* (angl. *Catalogue Of Somatic Mutations In Cancer*) algoritmų anotacinių įrankių (**6 lent.**). Taip pat prijungtos duomenų bazės, suteikiančios informaciją apie geną, kuriame identifikuota *dnM*, *dbSNP* duomenų bazė, tokiu būdu kiekvienam geno variantui priskiriant *rs* kodą, *ClinVar* vertinant geno varianto patogeniškumą, 1000 genomų projekto ir *ExAC* duomenų bazės genomų variantų dažniui skirtingose populiacijose vertinti.

6 lent. Anotavimo įrankių kategorinis vertinimas

Anotavimo įrankis	Kategorinis įvertis
<i>SIFT</i>	D: patogeninis; T: toleruojamas
<i>PolyPhen</i>	D: galimai patogeninis P: tikėtina patogeninis B: neutralus
<i>LRT</i>	D: patogeninis; N: neutralus; U: nėra duomenų
<i>MutationTaster</i>	A: tiesiogiai ligą lemiantis; D: ligą lemiantis; N: polimorfizmas; P: polimorfizmas

MutationAssessor	H: labai patogeniškas; M: vidutiniškai patogeniškas; L: mažai patogeniškas; N: neutralus. H/M funkcinis, L/N nefunkcinis variantas
FATHMM	D: patogeninis; T: toleruojamas
PROVEAN	D: ligą lemiantis; N: polimorfizmas
CADD	Didesni nei >12 įverčiai rodo varianto patogeniškumą
GERP++	Didesni nei >2 įverčiai rodo varianto patogeniškumą
PhyloP	Teigiamos vertės – sekos konservatyvumas, kai sekos evoliucija yra lėtesnė nei tikėtina. Neigiamos vertės – seka evoliucionuoja greičiau nei tikėtasi.
SiPhy	Teigiamos vertės – sekos konservatyvumas, kai sekos evoliucija yra lėtesnė nei tikėtina. Neigiamos vertės – seka evoliucionuoja greičiau nei tikėtasi.
COSMIC	Nurodo žinomus duomenis apie mutacijas somatinėse vėžinėse lastelėse, siejamus su tiriamu genomo variantu.

3.2.7. *De novo* mutacijų dažnio apskaičiavimas

De novo mutacijų dažnis į poziciją vienai kartai (PVK) apskaičiuotas pagal formulę:

$$De\ novo\ dažnis_{PVK} = \frac{\sum_{i=1}^f n_i}{2 \sum_{i=1}^f \sum_{j=1}^N P_j^i(de\ novo)}$$

čia f – triadų skaičius, N – pozicijų, kurias galima vertinti, skaičius ir n_i – identifikuotų dnM skaičius triadai i .

Tikimybė $P_j^i(dnM)$ vertinamai pozicijai j šeimoje i mutuoti, atsižvelgiant į horizontalųjį sekos padengimą, kai dnM yra vieno nukleotido dydžio, skaičiuojama:

$$P_j^i(dnM) = P_j^i(V_{Hetero} | M_{HomR}, T_{HomR}) + P_j^i(V_{Hetero} | M_{HomA}, T_{HomA})$$

Tikimybė P_j^i (*de novo* iškrita – intarpas) vertinamai pozicijai j šeimoje i mutuoti, atsižvelgiant į horizontalųjį sekos padengimą, kai dnM yra iškrita-intarpas, skaičiuojama:

$$P_j^i(\textit{de novo iškrita – intarpas}) = P_j^i(V_{\text{HomR}} | M_{\text{HomA}}, T_{\text{HomA}})$$

čia V , M ir T atitinka vaiką, motiną ir tėvą, o *Hetero*, *HomR* ir *HomA* atitinka heterozigotinį, homozigotinį pagal referentinį ir homozigotinį pagal alternatyvų alelių genotipus.

Pasikliautinieji intervalai (PI 95 %) dnM dažniui buvo skaičiuojami taikant binominį tikslųjį kriterijų. dnM dažnių ir kitiems skaičiavimams buvo naudotas atviros prieigos R paketas, versija 3.4.3 ir R studio, versija 1.2.1335 [192].

3.2.8. Tėvų amžiaus, genetinių ir epigenetinių veiksnių, galinčių turėti įtakos dnM intensyvumui, analizė

Siekiant patikrinti dnM intensyvumo priklausomybės skirtinguose geno regionuose nuo skirtingų geno ypatybių ir tėvų amžiaus hipotezę, atlikta tiesinės regresijos analizė. Jai atlikti buvo vykdomas kiekvienos dnM „antrinis“ anotavimas naudojantis ENCODE [193] projekto duomenimis ir LITGEN bei UNIGENE projektuose dalyvavusių asmenų anketomis.

Remiantis anksčiau atliktu Besenbacher et al. 2015 tyrimu, siekiant surinkti duomenis apie visas identifikuotas dnM ir geno regionų, kuriuose jų yra, ypatybes, pasirinkta limfoblastoidinių ląstelių linija (LCL ir GM12878) [193]. „Antrinio“ anotavimo parametrai ir vertinimo kriterijai pateikti 7 lent. Buvo surinkti šie duomenys: 1) raiškos lygis skirtinguose audiniuose (eQTL). Pagal raiškos lygį dnM buvo suskirstytos į regionus, kai jų raiška specifinė ir nespecifinė; 2) deoksiribonukleazei1 (*Dnazei1*) ypač jautrių geno regionų (angl. *DNase1 hypersensitivity sites*, *DHS*) įverčiai. *DHS* „vertinama“ 1, jei dnM yra *DNazei1* ypač jautriame regione, arba 0, jei nėra 3) CpG dinukleotidų konteksto įverčiai. Jei dnM yra CpG dinukleotidų gausiame regione, „vertinama“ 1, jei ne – 0; 4) trys histonų žymenys (H3K27ac, H3K4me1 ir H3K4me3) iš ENCODE projekto duomenų. Geno regionai buvo vertinami su kiekvienu iš žymenų atskirai. Jei dnM buvo šia žyme žymėtame regione, „vertinama“ 1, jei ne – 0; 5) GERPP++ DNR sekos konservatyvumo įverčiai buvo surinkti pasinaudojus ANNOVAR anotavimo

įrankiu. Esant didesniai nei >2 įverčiui, regionas su *dnM* vertinamas kaip konservatyvus genomo regionas ir žymimas 1, jei įvertis < 2 – nekonservatyvus, 0; 6) pagal LITGEN tiriamųjų apklausos anketas (ar UNIGENE projekte dalyvavusių asmenų duomenis) surinkti duomenys apie tėvų amžių.

7 lent. „Antrinio“ anotavimo parametrai ir vertinimo kriterijai

Anotavimo parametras	Vertinimo kriterijai	Anotavimo šaltinis
Raiškos lygis skirtinguose audiniuose, eQTL	Pagal jį, <i>dnM</i> buvo suskirstytos į genomo pozicijas, kuriose raiška yra audiniui specifinė arba ne	[193–195]
Deoksiribonukleazei1 (<i>DNazei1</i>) ypač jautrūs genomo regionai (angl. <i>DNase1 hypersensitivity sites, DHS</i>)	DHS „vertinama“ 1, jei <i>dnM</i> yra <i>DNszei1</i> ypač jautriame regione, 0, jei nėra	[194, 195]
CpG dinukleotidų kontekstas	Jei <i>dnM</i> yra CpG dinukleotidų gausiame regione „vertinama“ 1, jei ne – 0	[194, 195]
Histonų žymė H3K27ac	Jei <i>dnM</i> yra regione, žymėtame šia žyme, „vertinama“ 1, jei ne – 0	[194]
Histonų žymė H3K4me1	Jei <i>dnM</i> yra regione, žymėtame šia žyme, „vertinama“ 1, jei ne – 0	[194]
Histonų žymė H3K4me3	Jei <i>dnM</i> yra regione, žymėtame šia žyme, „vertinama“ 1, jei ne – 0	[194]
GERPP++ DNR sekos konservatyvumo įvertis	Kai įvertis > 2 , regionas su <i>dnM</i> vertinamas kaip konservatyvus genomo regionas ir žymimas 1, jei	ANNOVAR anotavimo įrankis

	įvertis < 2 – nekonservatyvus, 0	
Tėvų amžius	Amžius metais	LITGEN arba UNIGENE anketos

Surinkus visus „antrinio“ anotavimo duomenis kiekvienai triadai, pagal „0“ arba „1“ vertinimo principą kiekvienai *dnM* suskaičiuota, kiek kiekvieno anotavimo parametru iš viso nustatyta. Tada R paketu, versija 3.4.3, atlikta kiekvieno parametro ir apskaičiuoto *dnM* dažnio koreliacijos analizė. Tų parametru, kurių koreliacija su *dnM* dažniu buvo reikšminga, toliau atlikta tiesinės regresijos analizė.

3.2.9. *De novo* mutacijų spektro sudėties ir šablonų Lietuvos, asmenų su IN ir asmenų su IN sibsų grupėse nustatymas ir vertinimas

Mutacijų spektro sudėties ir šablonų analizei referentiniai mutacijų šablonai naudoti iš <https://cancer.sanger.ac.uk/cosmic/signatures/SBS/> duomenų bazės. Tyrimo metu kiekvienas tripletas, kurio vidurinėje pozicijoje buvo darbo metu nustatyta unikali *dnM*, o 3' ir 5' galuose – nukleotidai pagal referentinę geno seką, buvo priskirtas atitinkamai vienam tripletui iš 96 galimų. Pritaikius Spirmano koreliacijos koeficientą, atlikta Lietuvos populiacijos, asmenų su IN ir asmenų su IN sibsų grupėse nustatytų tripletų dažnio ir <https://cancer.sanger.ac.uk/cosmic/signatures/SBS/> duomenų bazėje esančių mutacijų šablonų, sudarytų iš skirtingų tripletų dažnių kombinacijų, koreliacijos analizė. Remiantis koreliacijos analizės rezultatais tiriamosioms grupėms priskirti mutacijų šablonai. Nustatyti šablonai su *de novo* mutacijomis, būdingomis tiriamosioms grupėms, toliau buvo lyginami su 6570 *de novo* mutacijų pagrindu sudarytu šablonų katalogu, jį sudarė *Rahbari et al.* analizuodami Jutos gyventojų, turinčių Šiaurės ir Vakarų Europos protėvių (CEU), Ibadano Jorubos (Nigerija) (YRI), islandų, škotų ir pietų Dakotos (JAV) populiacijas [31].

3.2.10. Sekų su *de novo* mutacijomis santykinio tinkamumo vertinimas

Aminorūgščių sekų su *dnM* santykinio tinkamumo vertinimas buvo atliktas siekiant išanalizuoti, kaip *dnM* veikia atskirų aminorūgščių tipų santykinį tinkamumą, ir, jei veikia, kokie santykinio tinkamumo šablonai apibūdina Lietuvos populiaciją.

Aminorūgščių sekų, kurias koduojančiuose genuose susidarė *dnM*, atrankos koeficientų arba kitaip santykinio tinkamumo efekto (angl. *fitness effect*) pasiskirstymo skaičiavimai atlikti pagal *TdG12 site-wise mutation-selection* skaičiavimo modelį (swMutSel0) [209]. Iš tikimybės modelių, paremtų genomo pakaitų apibūdinimu kodonais [210–212], šis modelis buvo pasirinktas dėl didžiausio specifiškumo mutabilumo procesui vertinti, kai kiekvienos tiriamos aminorūgšties evoliucija apibūdinama pagal individualią tinkamumo vertę ir nukleotidų lygmeniu esantį mutacijų modelį. Šiame modelyje buvo daroma prielaida, kad genų sekos susidarė nepriklausomai viena nuo kitos, viena kitos neveikia, gamtinės atrankos poveikis labai mažas, todėl nereikšmingas. swMutSel0 modeliu evoliucijos procesas vertintas per ilgą laiką (milijonus metų, nes taikant šį modelį žmogaus DNR seka lyginama su artimiausiais žmogui primatais), o laikas nuo naujų sekų su *dnM* susidarymo iki jų fiksavimo yra momentinis.

Taikant šį modelį, analizuotos tik baltymus koduojančios sekos, kuriose buvo nustatytos *dnM*. Pirmiausia pagal kiekvienos nustatytos *dnM* koordinatas genome *Ensembl* duomenų bazėje buvo surasta atitinkama genomo seka, iš viso apimanti penkis kodonus, išsidėsčiusius iš eilės. Penkių aminorūgščių aprėptis pasirinkta todėl, kad vertinant aminorūgšties su *dnM* santykinį tinkamumą būtų atsižvelgta į jos genominių kontekstą tiek 3', tiek ir 5' galuose. Viduryje esantis kodonas – tas kodonas, kuriame įvyko *dnM*.

Toliau kiekvienai atrinktai sekai *Ensembl* duomenų bazėje lyginamosios genomikos bioinformacinių įrankių grupės dauginio (angl. *multiple*) genomų sulgiavimo įrankiu analizuota žmogaus genomo seka sulgiuota su evoliuciškai artimiausių primatų genomais: šimpanzės, mažosios šimpanzės (angl. *Bonabo*), gorilos, orangutano ir žaliajo paviano (angl. *Olive baboon*). Lygiuoti su evoliuciškai artimiausiais primatų genomais reikia tam, kad taikomu swMutSel0 modeliu būtų galima įvertinti kodono su *de novo* mutacija santykinį tinkamumą evoliucijos eigoje. Toliau kiekvienas genominių sekų sulgiavimas <https://www.ebi.ac.uk> *Multiple Sequence Alignment*>*Clustal Omega* bioinformaciniu įrankiu konvertuotas į .phy (PHYLIP) formato duomenų bylas. RAXML v8.2 (angl. *Randomized*

Axelerated Maximum Likelihood) programa [213] pagal *.phy* duomenų bylas sukurti filogenetiniai lyginamų organizmų medžiai NEWICK formatu (5 priedas). Filogenetinių medžių šakoms apdoroti pagal swMutSel0 modelio kūrėjų rekomendacijas panaudota PAML v14.8 programa [214]. Atlikus filogenetinio medžio analizę PAML funkcija *codeml*, kiekvienam filogenetiniam medžiui surinkta informacija apie κ (tranzicijų ir transversijų santykį), π (nukleotidų sudėtį dalimis), μ (pokyčių dažnį). κ ir π vertės buvo pateiktos *.mlc* duomenų byloje, o μ apskaičiuotas pagal formulę $\mu = 3 \times T_{ds}/T$, čia T_{ds} yra filogenetinio medžio ilgis *ds* (sinoniminių pokyčių) vienetais, T – visas filogenetinio medžio ilgis. Šie parametrai taip pat nurodyti *.mlc* duomenų byloje. Pateikus anksčiau aprašytus parametrus programai TdG12 buvo įvertintas aminorūgščių sekų santykinis tinkamumas prieš mutacijas ir po jų.

3.2.11. *De novo* mutacijų patogeniškumo mechanizmų nustatymas

Patogeninių genomų variantų su identifikuotomis *dnM* galimi patogeniškumo mechanizmai buvo tiriami naudojant tokius bioinformacinius įrankius ir duomenų bazes:

- SWISS-MODEL baltymo struktūros atviros prieigos modeliavimo programą [215];
- STRING programą [216] baltymų sąveikos tyrimui;
- OMIM (angl. *Online Mendelian Inheritance in Man*) interaktyvų žmogaus genų ir genetinių ligų katalogą;
- GeneCards® – žmogaus genų duomenų bazę;
- Ensembl – žmogaus genomo naršyklę;
- ClinVar – patogeninių genomo variantų duomenų bazę;
- Human Splicing Finder – sukirpimui svarbių genomo variantų vertinimo duomenų bazę;
- denovo-db – patogeninių, naudingų ir neaiškios klinikinės reikšmės *dnM* duomenų bazę;
- IDGenetics – genų ir genų variantų, susijusių su IN, duomenų bazę (<http://www.ccgenomics.cn/IDGenetics/>);
- SFARI gene – su autizmu siejamų genų duomenų bazę (<https://gene.sfari.org/database/human-gene/>);

- AutDB – su autizmu siejamų genų duomenų bazė
(<http://autism.mindspec.org/autdb/Welcome.do>)

Vienu atveju, kai buvo identifikuotas tikėtinais patogeninis *ARID1B* geno variantas, jo patogeniškumo mechanizmui nustatyti visa probando RNR buvo išskirta iš veninio kraujo naudojant *Tempus™ Blood RNA Tube* ir *Tempus™ Spin RNA Isolation* rinkinius (Thermo Fisher Scientific, USA) pagal gamintojo rekomendacijas. Kopijinė DNR iš RNR buvo susintetinta naudojant *High-Capacity RNA-to-cDNA* rinkinį (Thermo Fisher Scientific, USA) pagal gamintojo rekomendacijas. kDNR sukirpimo sekos PGR reakcija buvo atlikta naudojant specifinius oligonukleotidinius pradmenis, sukurtus *Primer Blast* (NCBI) įrankiu. PGR produktai nuskaityti naudojant standartinį *Sanger* sekos nustatymo metodą, kaip aprašyta 3.2.4 skyrelyje⁹. Sukirpimo vietos pokyčio patogeniškumui vertinti buvo panaudotos *Mutation Taster* (<http://www.mutationtaster.org/>) ir *Human Splicing Finder* (<http://www.umd.be/HSF3/>) duomenų bazės. Galimi genų variantų su *dnM* efektai vertinti naudojantis *ExPASy* bioinformatikos (<https://www.expasy.org/>) ir Pfam 32.0 (<https://pfam.xfam.org/>) duomenų bazėmis.

3.2.12. Duomenų analizė ir statistiniai metodai

Remiantis LITGEN ir UNIGENE tyrimų duomenims atlikta aprašomoji ir analitinė statistinė analizė. Genomo ypatybės, būdingos genomo regionams su *dnM* (yra / nėra), analizuotos kaip kategoriniai kintamieji. Kiekybiniai kintamieji: *dnM* skaičius ir dažnis genome, tėvų amžius, genomo variantų su *dnM* patogeniškumo vertinimo vertės, aminorūgščių santykinio tinkamumo vertės.

Atliekant genotipo ir fenotipo koreliacijos analizę, genetinių tyrimų rezultatai buvo laikomi aiškinamaisiais kintamaisiais. Fenotipo raiškos rodikliai (IN laipsnis ir kiti fenotipiniai požymiai) buvo atsako kintamieji.

Imčių pasiskirstymui pagal normalųjį skirstinį tikrinti taikytas *Shapiro-Wilk* testas. Koreliacijai vertinti taikyti Pirsono arba Spirmeno koreliacijos

⁹ RNR išsskyrimą, *ARID1B* kDNR PGR reakciją ir sekoskaitą *Sanger* būdu atliko dokt. Evelina Siavrienė.

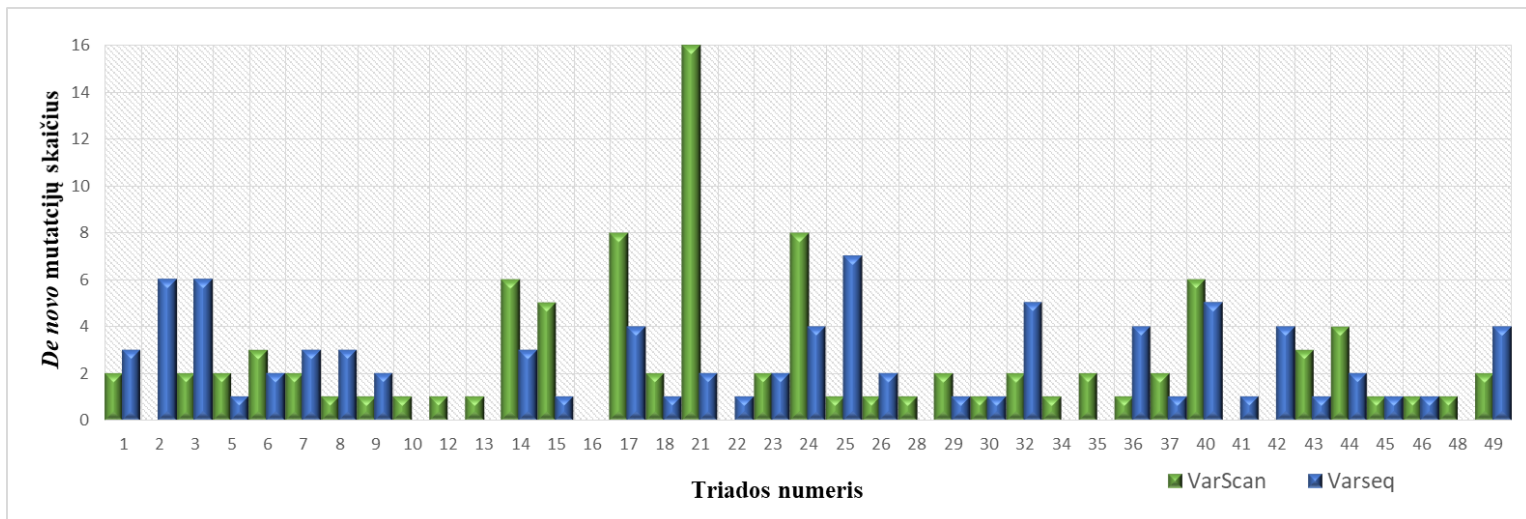
koeficientai. Tiesinės regresijos analizė, *Mann-Whitney U* kriterijaus ir tiksliojo Fišerio kriterijaus, *Shapir-Wilk* testas, koreliacijos koeficientų skaičiavimai atlikti naudojant atviros prieigos R paketą, versija 3.4.3 ir R studio, versija 1.2.1335 [205]. Taikant tiesinę regresiją, koreliacijos analizės metu mažesnė nei 0,05 *p* reikšmė buvo laikoma statistiškai reikšminga. Visi kiti skaičiavimai atlikti *Microsoft Office Professional Plus 2013* „Excel“ programa.

TYRIMO REZULTATAI

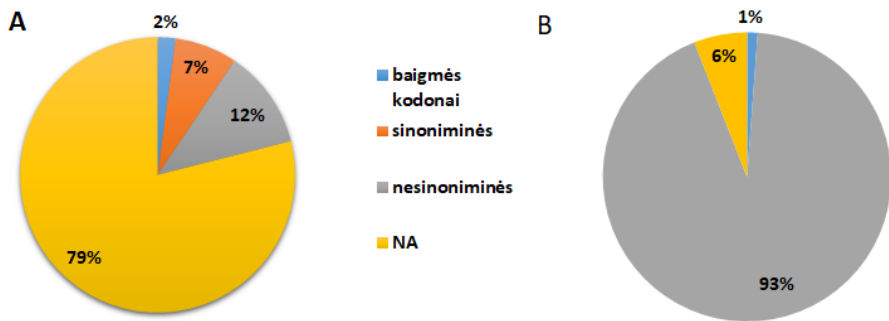
4.1. *De novo* mutacijos, jų skaičius ir dažnis Lietuvos populiacijos genome

Siekiant atrinkti optimalią programą taškinių ir nedidelių (20–30 bp) iškritų-intarpų *dnM* identifikuoti, Lietuvos populiacijos grupės duomenų analizei buvo panaudotos dvi programos – *VarScan* ir *VarSeq*TM. Po pirminio *dnM* identifikavimo *VarScan* programa buvo nustatytos 1752 *dnM* ir *VarSeq*TM programa – 4756 *dnM*. Pritaikius visus filtravimo parametrus ir *dnM* sąrašą patikrinus *IGV* programa ir *Sanger* sekoskaitos metodu, *VarScan* programa 34 triadoms iš viso nustatytos 95 *dnM*, *VarSeq*TM programa – 31 triadai 84 *dnM* (**13 pav., 6 priedas**). Buvo nustatytas išskirtinai didelis *dnM* skaičius triadoms nr. 4 ir nr. 21: 113 ir 16 *dnM* atitinkamai. *VarSeq*TM programa triadai nr. 4 nustatytos 123 *dnM*, o triadai nr. 21 – 2 *dnM*. Šie radiniai paskatino atlikti biologinės tėvystės patikrinimą. Biologinė tėvystė patvirtinta triadai nr. 21, tačiau atmesta triadai nr. 4, todėl triada nr. 4 iš tolesnės duomenų analizės buvo pašalinta. Galutinę Lietuvos populiacijos grupę sudarė 48 triados. 15-kai triadų *VarScan* programa ir 18-kai triadų *VarSeq*TM programa nenustatyta nė vienos *dnM*. Palyginus *dnM*, identifikuotas abiem programomis, tik 5,37 % visų nustatytų *dnM* sutapo (3 *dnM* *MEIS2*, *PGK1* ir *MT1B* genuose). Kiekvienas vaikas vidutiniškai turi po 1,9 (*VarScan*) ir 1,7 *de novo* vieną nukleotido variantą egzome (*VarSeq*TM).

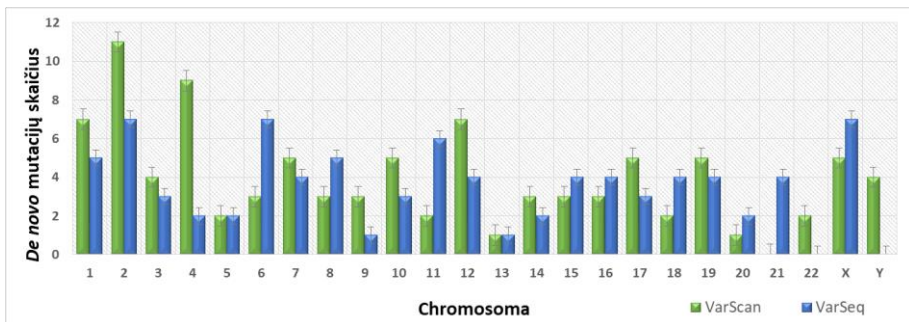
95 *dnM*, nustatytų *VarScan* programa, analizė atskleidė, kad iš jų 20 *dnM* yra egzonuose, tarp kurių yra dvi baigmės kodono *dnM*, septynios sinoniminės ir 11 nesinoniminių vieno nukleotido *dnM*. Tarp 84 *dnM*, nustatytų *VarSeq*TM programa, 80 *dnM* yra egzonuose, iš kurių viena *dnM* yra baigmės kodono ir 78 nesinoniminės vieno nukleotido *dnM* (**14 pav.**). Didžioji *dnM*, nustatytų *VarScan*, dalis yra 1, 2, 4 ir 5 chromosomose, o *VarSeq*TM – 2, 6, 7 ir 11 chromosomose. Nustatytų *dnM* skaičius nekoreliuoja nei su genų tankiu chromosomose ($R_p = 0,09$, p reikšmė = 0,69 *VarScan*; $R_p = 0,07$, p reikšmė = 0,75 *VarSeq*TM), nei su chromosomų dydžiu (**15 pav.**). Pagal abiejų programų duomenis, tranzicijų ir transversijų santykis yra labai panašus – atitinkamai 1,47 ir 1,54 (**16 pav.**), taigi daugiausiai *dnM*, nustatytų abiem *VarScan* ir *VarSeq*TM programomis, sudėtį apibūdina A/G ir C/T pokyčių gausa.



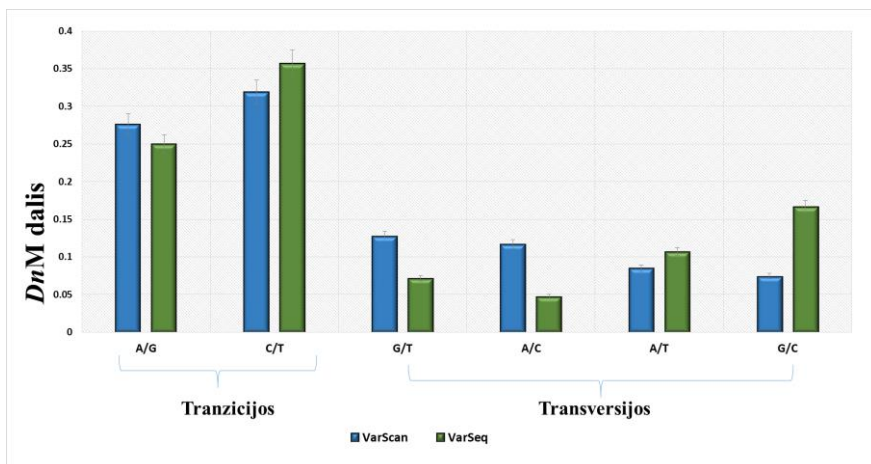
13 pav. DnM, nustatytos VarScan (žalia spalva) ir VarSeq™ (mėlyna spalva) programa, Lietuvos populiacijos grupėje



14 pav. *DnM*, nustatytų *VarScan* (A) ir *VarSeq*TM (B), pasiskirstymas pagal funkcinius pokyčių tipus. NA – nėra duomenų



15 pav. *DnM* skaičiaus pasiskirstymas pagal chromosomą. *DnM*, nustatytos *VarScan* programa, pažymėtos žalia spalva, mėlyna – *VarSeq*TM. Klaidų juostos atitinka kiekvienoje chromosomoje nustatytų *dnM* skaičiaus vidurkio standartinę paklaidą



16 pav. *De novo* tranzicijos ir transversijos, identifikuotos *VarScan* (mėlyna spalva) ir *VarSeq*TM (žalia spalva) programa. Egzome tranzicijos yra ~1,5 karto dažnesnės nei transversijos. Klaidų juostos atitinka *dnM* skaičiaus vidurkio standartinę paklaidą

Pagal *VarScan* programos duomenis, apskaičiuotas *de novo* vieno nukleotido mutacijų dažnis yra $2,74 \times 10^{-8}$ į poziciją vienai kartai (95 % pasikliautinis intervalas (PI): $2,24 \times 10^{-8} - 3,35 \times 10^{-8}$ į poziciją vienai kartai). Pagal *VarSeq*TM duomenis – $2,4 \times 10^{-8}$ į poziciją vienai kartai (95 % PI: $1,96 \times 10^{-8} - 2,99 \times 10^{-8}$). *VarScan* programos duomenimis, *de novo* iškritų-intarpų dažnis šioje grupėje buvo $1,77 \times 10^{-8}$ (95 % PI: $6,03 \times 10^{-9} - 5,21 \times 10^{-8}$) į poziciją vienai kartai, *VarSeq*TM programa jų nenustatyta. Pritaikius *Mann-Whitney U* kriterijų buvo palygintas *de novo* vieno nukleotido mutacijų dažnis Lietuvos populiacijoje su islandų, danų, olandų, kanadiečių ir amerikiečių bendrųjų populiacijų tyrimuose nustatytu *de novo* vieno nukleotido mutacijų dažnio vidurkiu (žr. 3 lent.) Analizės metu tarp Lietuvos ir lygintų populiacijų nustatytas reikšmingas (p reikšmė = $4,49 \times 10^{-9}$) *dnM* dažnio skirtumas.

Tiesinės regresijos modelis parodė, kad *DNaze1* ypač jautrių regionų, regionų, gausių CpG dinukleotidų, skaičius, GERPP++ genomo regionų konservatyvumo vertės ir genomo regionų raiškos lygiai paaiškina ~68–94 % viso *dnM* dažnio (8 lent.). Nei vieno iš epigenetinių žymenų pozicijų skaičius (H3K4Me3 žymė: $R_S = 0,13$, p reikšmė = 0,46; H3K27ac žymė: $R_S = 0,06$, p reikšmė = 0,75; H3K4Me1 žymė nenustatyta nei vienai *dnM* pozicijai), nei

tėvų amžius reikšmingai (tėvo amžius $R_S = 0,21$, p reikšmė = 0,15; motinos amžius $R_S = 0,03$, p reikšmė = 0,86) su dnM dažniu nekoreliavo (**8 priedas**). Tiesinės regresijos modeliai buvo sudaryti tik iš dnM , nustatytų *VarScan* programa, dažnių duomenų, nes *VarSeq*TM programa nustatytų dnM dažnio ir geno savybių (**žr. 7 lent.**), koreliaciją nenustatyta.

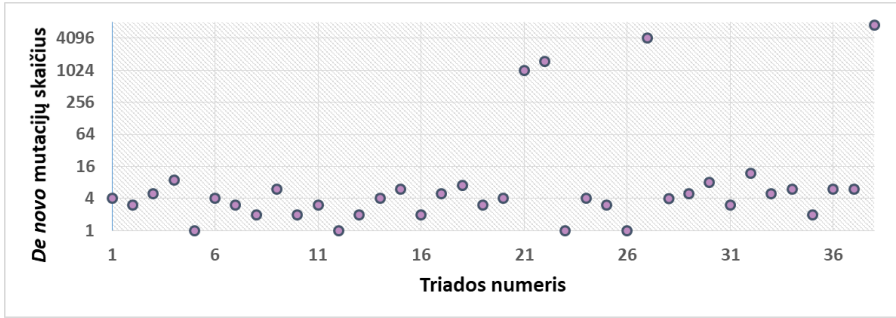
8 lent. DNazei 1 ypač jautrių regionų, regionų, gausių CpG dinukleotidų, skaičiaus, konservatyvumo regionų skaičiaus ir geno regionų su skirtinga raiška skaičiaus įtaka dnM dažnio intensyvumui Lietuvos tiriamųjų grupėje žodžius nukelti prie pavadinimo

Atsako kintamasis	Nepriklausomas kintamasis	$\beta^* \pm SP$	p reikšmė	R^2
dnM dažnis	DNAazei I jautrūs regionai	$2,83 \times 10^{-6} \pm 1,96 \times 10^{-7}$	$2,31 \times 10^{-13}$	0,89
	Konservatyvios pozicijos	$5,28 \times 10^{-6} \pm 5,67 \times 10^{-7}$	$1,49 \times 10^{-6}$	0,88
	Nekonservatyvios pozicijos	$2,56 \times 10^{-6} \pm 2,67 \times 10^{-7}$	$2,45 \times 10^{-9}$	0,8
	Regionai, turtingi CpG dinukleotidų	$2,76 \times 10^{-6} \pm 4,94 \times 10^{-7}$	$8,93 \times 10^{-6}$	0,68
	Regionai su nespecifine raiška	$2,05 \times 10^{-6} \pm 1,144 \times 10^{-7}$	$8,34 \times 10^{-14}$	0,94

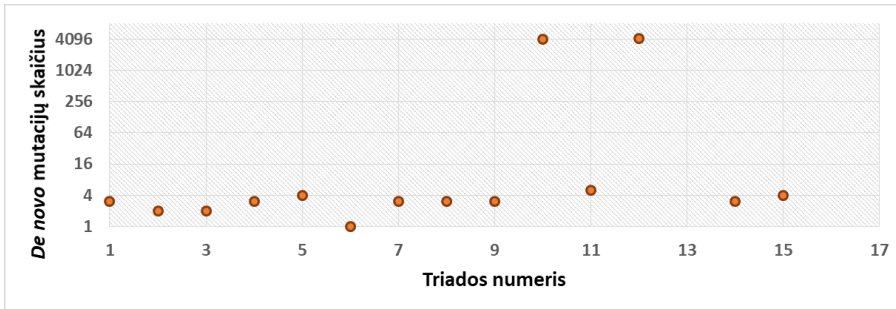
β – regresijos koeficientas, parodantis, kiek didėja dnM dažnis nepriklausomam kintamajam padidėjus vienetu (atitinkamai viena pozicija ar regionu), SP – standartinė paklaida, p reikšmė – tiesinės regresijos modelio p reikšmė, R^2 – determinacijos koeficientas.

4.2. *De novo* mutacijos, jų skaičius ir dažnis asmenų su IN ir jų sibsų genuose

DnM skaičiaus ir dažnio analizė asmenų su IN (toliau antroje grupėje) ir asmenų su IN sibsų grupėse (toliau trečioje grupėje) dėl didesnio jautrumo ir atviros prieigos galimybės buvo atlikta taikant tik *VarScan* algoritimą. Antroje grupėje 37 tirtoms triadoms iš viso nustatyta 14 103 dnM , iš kurių 1016, 1513, 4123 ir 7309 dnM nustatytos keturioms triadoms (atitinkamai nr. 21, 22, 27, 37), o likusios 142 dnM nustatytos 33 triadoms (**17 pav., 6 priedas**). Šie radiniai paskatino atlikti biologinės tėvystės patikrinimą. Biologinė tėvystė patvirtinta visų tirtų triadų, todėl asmenys su reikšmingai didesniu dnM skaičiumi toliau buvo vertinami kaip išskirtys. Vidutiniškai kiekvienai iš 33 triadų nustatyta po 4,2 *de novo* VNP ir 1,3 *de novo* iškritos-intarpo (**18 pav.**).



17 pav. *DnM*, nustatytos asmenų, turinčių intelektualinę negalią, grupėje. 21, 22, 27 ir 37 triados pasižymėjo *dnM* gausa (išskirtys). 21 triadai nustatytos 1016 *dnM*, 22 triadai – 1513, 27 triadai – 4123 *dnM*, 37 triadai – 7309 *dnM*

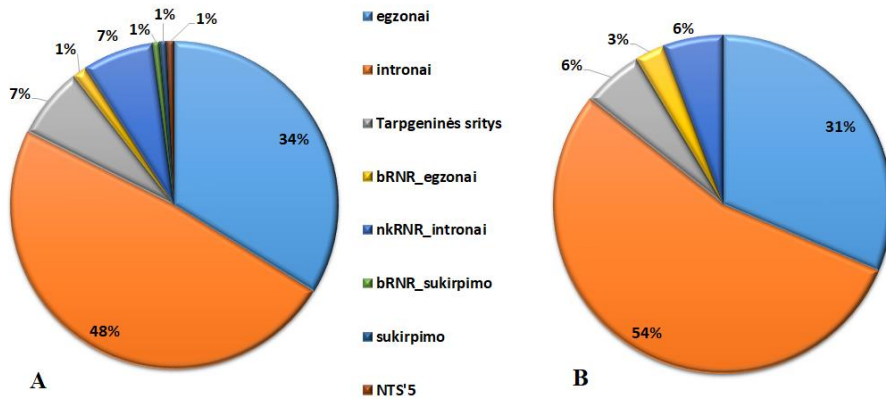


18 pav. *DnM*, nustatytos asmenų, turinčių intelektualinę negalią, sibsų grupėje. 10 ir 12 triados pasižymėjo *dnM* gausa (išskirtys). 10 triadai nustatytos 4130 *dnM*, 12 triadai – 4246 *dnM*

Trečioje grupėje 17 tirtų triadų iš viso nustatyta 8412 *dnM*, iš kurių 4130 ir 4246 *dnM* nustatytos dviem triadoms (atitinkamai nr. 10 ir 12), o likusios 36 *dnM* nustatytos 15-kai triadų. Vidutiniškai kiekvienai iš 15 triadų nustatyta po 2,4 *de novo* VNP ir 0,92 *de novo* iškritos-intarpo (**18 pav., 6 priedas**).

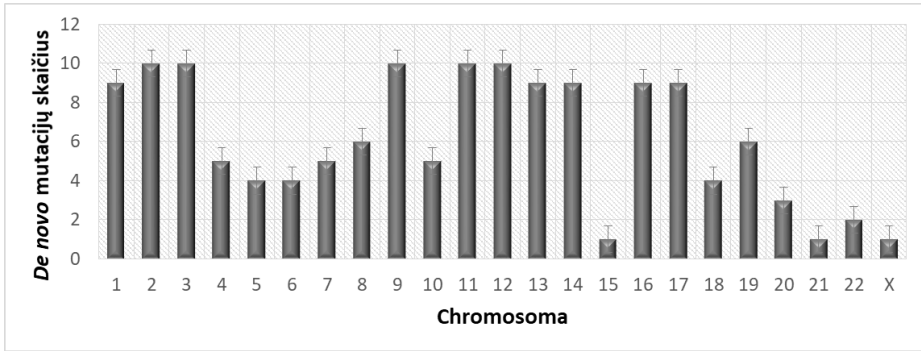
DnM analizė pagal tai, kokiuose funkcinuose genomo regionuose jos nustatytos, atskleidė, kad antra ir trečia grupės yra panašios pagal *dnM*, nustatytų intronuose ir koduojančiuose regionuose (egzonuose), skaičių – maždaug pusė *dnM* visų nustatytų *dnM* yra intronuose (48 %, **19 pav. A** ir 54 % **19 pav. B**) ir apie trečdali, atitinkamai 34 % ir 31 %, *dnM* buvo egzonuose. Vis dėlto antros grupės tiriamųjų egzonuose nustatyta 53 *dnM*, tarp kurių yra dvi baigmės kodono *dnM*, penkios rėmelio poslinkio *dnM*, 13 sinoniminių VNP, 28 nesinoniminių VNP *dnM* ir penkios nuo vieno iki 30 nukleotidų

dydžio *de novo* išskritos, o trečioje grupėje – vienas *de novo* intarpas, septyni nesinoniminiai ir penki sinoniminiai *de novo* VNP. Kitos pavienės *dnM* išsidėsčiusios tarpgeninėse srityse, sukirpimui svarbiuose, netransliuojamuose branduolio RNR regionuose.

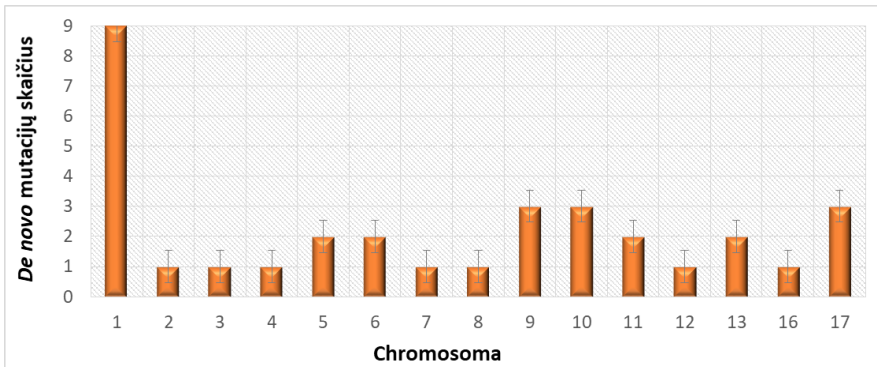


19 pav. *dnM*, nustatytų asmenų su intelektine negalia (**A**) ir asmenų su intelektine negalia sibsams (**B**), pasiskirstymas pagal funkcinius geno regionus. Čia bRNR yra branduolio ribonukleorūgštys, nkRNR – nekoduojanti ribonukleorūgštis, NTS – netransliuojamos geno sritys

Antroje grupėje nustatytos *dnM* yra išsibarčiusios visose chromosomose, išskyrus Y chromosomą (**20 pav.**), trečioje grupėje jų nenustatyta 14, 15, 18–22, X ir Y chromosomose (**21 pav.**). Nustatytų *dnM* skaičius nekoreliuoja nei su genų tankiu chromosomose ($R_S = 0,2873$, p reikšmė = 0,173 antrai grupei ir $R_S = 0,1863$, p reikšmė = 0,3855 trečiai grupei), nei su chromosomų dydžiu (**20, 21 pav.**). O tranzicijų ir transversijų santykis antroje ir trečioje grupėse skiriasi daugiau nei du kartus. Antroje grupėje šis santykis siekė 1,65, o trečioje grupėje – 3,5 (**22 ir 23 pav.**). Taip pat nustatytų *dnM* identifikuoti transversijų struktūros skirtumai. Tokį tranzicijų ir transversijų santykio ir struktūros skirtumą lemia grupių ir nustatytų *dnM* imtis. Jai didėjant tranzicijų / transversijų santykis ir struktūra trečioje grupėje turėtų būti artima pirmoje grupėje nustatytiems. ~ 78 % *dnM*, nustatytų trečiai grupei, spektrą apibūdina C>T ir A>G pokyčiai.



20 pav. *dnM* skaičiaus pasiskirstymas pagal chromosomą asmenų, turinčių intelektualinę negalią, grupėje. Klaidų juostos atitinka kiekvienoje chromosomoje nustatytų *dnM* skaičiaus vidurkio standartinę paklaidą. Y chromosomoje *dnM* nenustatyta



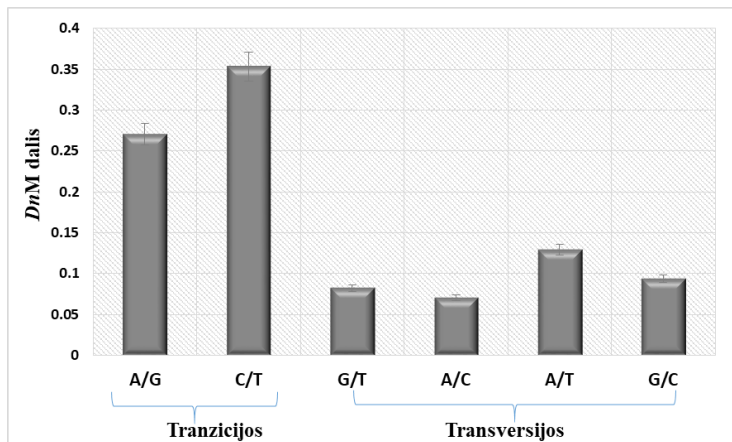
21 pav. *dnM* skaičiaus pasiskirstymas pagal chromosomą asmenų, turinčių intelektualinę negalią, sibsų grupėje. Klaidų juostos atitinka kiekvienoje chromosomoje nustatytų *dnM* skaičiaus vidurkio standartinę paklaidą

Antroje grupėje buvo nustatytos 42 iškritos-intarpai, kurių dydis nuo vieno iki 30 nukleotidų dydžio 24 triadose, o trečioje grupėje 9 triadoms iš viso nustatyta 12 iškritų-intarpų nuo trijų iki 31 nukleotido dydžio. Trečioje grupėje tarp nustatytų *de novo* iškritų-intarpų tik viena *dnM* yra koduojančioje sekoje (8 %), o antroje grupėje – dešimt (24 %), iš kurių net penkios *dnM* yra rėmelio poslinkio tipo (**9 lent.**). Likusios *dnM* išsibarsčiusios kituose nekoduojančiuose geno regionuose.

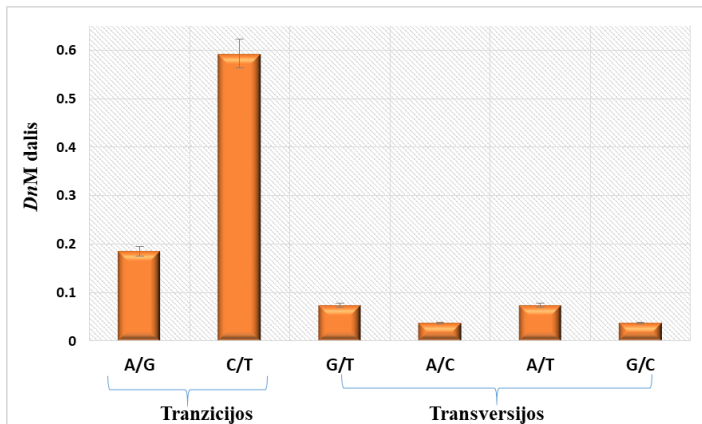
9 lent. *De novo* iškritos-intarpai antroje ir trečioje grupėje

<i>dnM</i> skaičius	<i>dnM</i> dydis, nt	Rėmelio poslinkio <i>dnM</i> skaičius	Ne rėmelio poslinkio <i>dnM</i> skaičius	Nesinoniminių <i>dnM</i> skaičius	<i>dnM</i> intronuose skaičius	<i>dnM</i> kituose* genomo regionuose skaičius
Asmenų su IN grupė (antra grupė)						
42	1-30	5	5	0	18	14
Asmenų su IN sibsų grupė (trečia grupė)						
12	3-31	0	1	0	8	3

* kiti regionai – tai netransliuojamas 5' regionas, branduolio RNR introniniai regionai, tarpgerinės sritys arba nėra duomenų.



22 pav. *De novo* tranzicijos ir transversijos, identifikuotos asmenų, turinčių intelektualinę negalią, grupėje. Egzome tranzicijos yra ~1,65 karto dažnesnės nei transversijos. Klaidų juostos atitinka *dnM* skaičiaus vidurkio standartinę paklaidą



23 pav. *De novo* tranzicijos ir transversijos, identifikuotos asmenų, turinčių intelektualinę negalią, sibsų grupėje. Egzome tranzicijos yra ~1,5 karto dažnesnės nei transversijos. Klaidų juostos atitinka *dnM* skaičiaus vidurkio standartinę paklaidą

Pagal *VarScan* programos duomenis, apskaičiuotas *de novo* vieno nukleotido mutacijų dažnis antroje grupėje yra $3,05 \times 10^{-8}$ į poziciją vienai kartai (95 % pasikliautinis intervalas (PI): $2,52 \times 10^{-8} - 3,69 \times 10^{-8}$ į poziciją vienai kartai). Trečioje grupėje – $7,19 \times 10^{-8}$ į poziciją vienai kartai (95 % PI: $4,91 \times 10^{-8} - 1,05 \times 10^{-7}$). Kaip ir pirmoje grupėje, pritaikius *Mann-Whitney U* kriterijų buvo palygintas asmenų su IN ir asmenų su IN sibsų grupių *de novo* vieno nukleotido mutacijų dažnis su islandų, danų, olandų, kanadiečių ir amerikiečių bendrųjų populiacijų ir autizmo spektro liga sergančiųjų grupės tyrimuose [1, 5, 8, 31, 32, 37, 130, 131] nustatyto *de novo* vieno nukleotido mutacijų dažnio vidurkiu (žr. **3 lent.**) Analizės metu tarp antros ir trečios grupių ir lygintų populiacijų nustatytas reikšmingas (atitinkamai $5,15 \times 10^{-6}$, trečiai grupei p reikšmė = $4,8 \times 10^{-6}$) *dnM* dažnio skirtumas.

De novo iškritų-intarpų dažnis antros grupės genome yra $1,74 \times 10^{-7}$ (95 % PI: $1,30 \times 10^{-7} - 2,30 \times 10^{-7}$) į poziciją vienai kartai. Trečioje grupėje – $1,08 \times 10^{-6}$ į poziciją vienai kartai (95 % PI: $9,41 \times 10^{-7} - 1,23 \times 10^{-6}$). Vidutiniai visų *dnM*, nusekvenuotų *Illumina HiSeq 2500* genetiniu analizatoriumi, kokybės parametrai buvo: GQ = 50, vidutinis horizontalusis padengimas – 100, o *dnM*, nusekvenuotų *5500 SOLiD™ Sequencer* genetiniu analizatoriumi – GQ = 50, vidutinis horizontalusis padengimas – 37,5.

Genetinių ir epigenetinių veiksnių, minėtų 3.2.8 skyrelyje, koreliacijos su *dnM* dažniu analizė atskleidė, kad antroje grupėje su *dnM* dažniu silpnai ($R_S = 0,56$) koreliuoja nespacificinės raiškos regionai. Remiantis tiesinės regresijos

modeliu, nespecifinės raiškos regionai paaiškina tik ~13 % viso *dnM* dažnio (**10 lent.**). Kiti veiksniai reikšmingai *dnM* dažnio neveikė. Trečioje grupėje genetinių ir epigenetinių veiksnių koreliacija su *dnM* dažniu nenustatyta.

10 lent. Genetinių ir epigenetinių veiksnių įtaka *dnM* dažnio intensyvumui antroje grupėje – šie du žodžiai turi būti prie pavadinimo viršuje

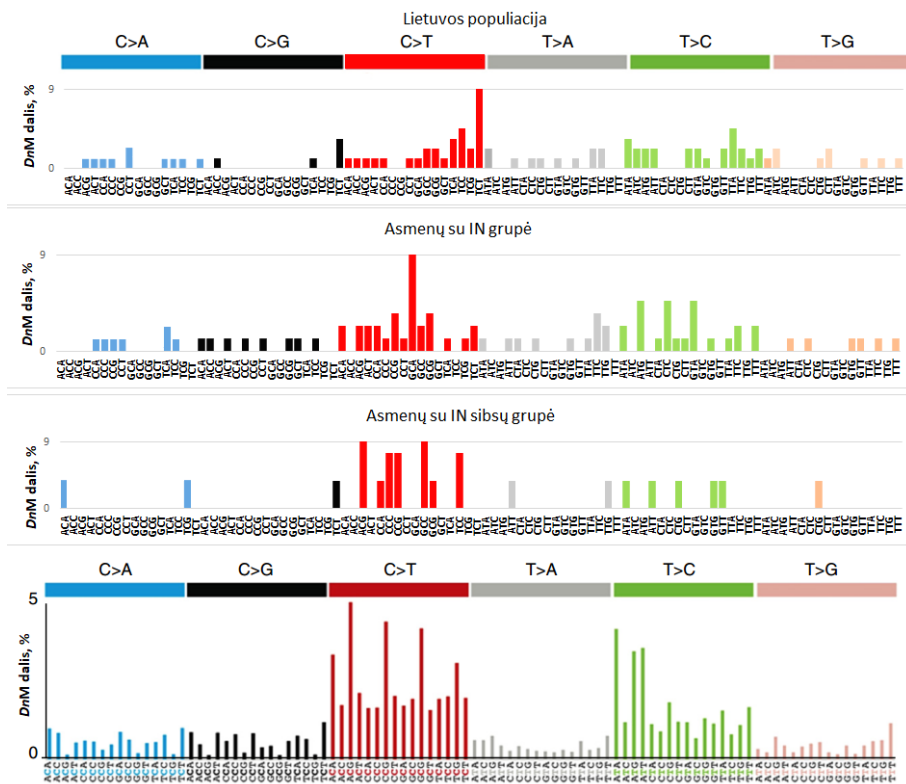
Atsako kintamasis	Nepriklausomas kintamasis	$\beta^* \pm SP$	<i>p</i> reikšmė	R^2
<i>dnM</i> dažnis	Nespecifinės raiškos regionai	$8,56 \times 10^{-7} \pm 3,52 \times 10^{-7}$	0,02	0,13
<p>β – regresijos koeficientas, parodantis, kiek didėja <i>dnM</i> dažnis nepriklausomam kintamajam padidėjus vienetu (atitinkamai viena pozicija ar regionu), SP – standartinė paklaida, <i>p reikšmė</i> – tiesinės regresijos modelio <i>p</i> reikšmė, R^2 – determinacijos koeficientas.</p>				

4.3. *De novo* mutacijų spektro sudėties ir šablonų Lietuvos, asmenų su IN ir asmenų su IN sibsų grupėse analizės rezultatai

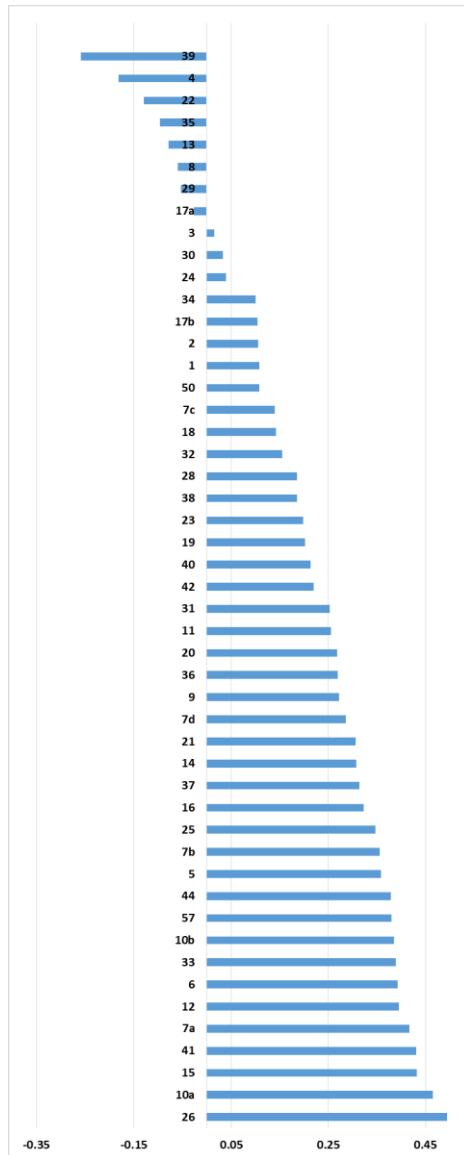
Remiantis *de novo* mutacijų spektro sudėties analizės duomenimis Lietuvos populiacija, kaip ir asmenų su IN, ir asmenų su IN sibsų grupės *dnM* spektrų sudėtis artima Jutos gyventojų, turinčių Šiaurės ir Vakarų Europos protėvių (CEU), Ibadano Jorubos (Nigerija) (YRI), islandų, škotų ir pietų Dakotos (JAV) populiacijose nustatytai [31] (**24 pav.**). Visose grupėse daugiausiai nustatyta C>T (pirmoje grupėje 33 %, antroje – 35 % ir trečioje grupėje 59 %) ir T>C (pirmoje grupėje 28 %, antroje – 27 % ir trečioje grupėje 18,5 %) *dnM* ir statistškai reikšmingų skirtumų tarp grupių *de novo* mutacijų spektrų nenustatyta (*p reikšmė* = 0,575, χ^2 kriterijus).

Vis dėlto, atlikus Lietuvos populiacijoje nustatytų tripletų su *dnM* dažnio koreliacijos analizę su <https://cancer.sanger.ac.uk/cosmic/signatures/SBS/> duomenų bazėje esančiais 49 egzomo šablonais, nustatyta, kad Lietuvos populiaciją daugiausiai apibūdina 26 ($R_S = 0,49$, *p reikšmė* = 3×10^{-7}), 10a ($R_S = 0,46$, *p reikšmė* = $1,9 \times 10^{-6}$), 15 ($R_S = 0,43$, *p reikšmė* = $1,1 \times 10^{-5}$), 41 ($R_S = 0,43$, *p reikšmė* = $1,2 \times 10^{-5}$) ir 7a ($R_S = 0,42$, *p reikšmė* = $2,4 \times 10^{-5}$) mutacijų šablonai – (**25 pav.**); asmenų su IN grupė – 44 ($R_S = 0,41$, *p reikšmė* = $2,9 \times 10^{-5}$), 6 ($R_S = 0,41$, *p reikšmė* = $3,2 \times 10^{-5}$), 5 ($R_S = 0,39$, *p reikšmė* = $7,4 \times 10^{-5}$), 20 ($R_S = 0,39$, *p reikšmė* = $8,4 \times 10^{-5}$) ir 7b ($R_S = 0,39$, *p reikšmė* = $9,8 \times 10^{-5}$) mutacijų šablonai (**26 pav. A**); ir asmenų su IN sibsų grupė – 5

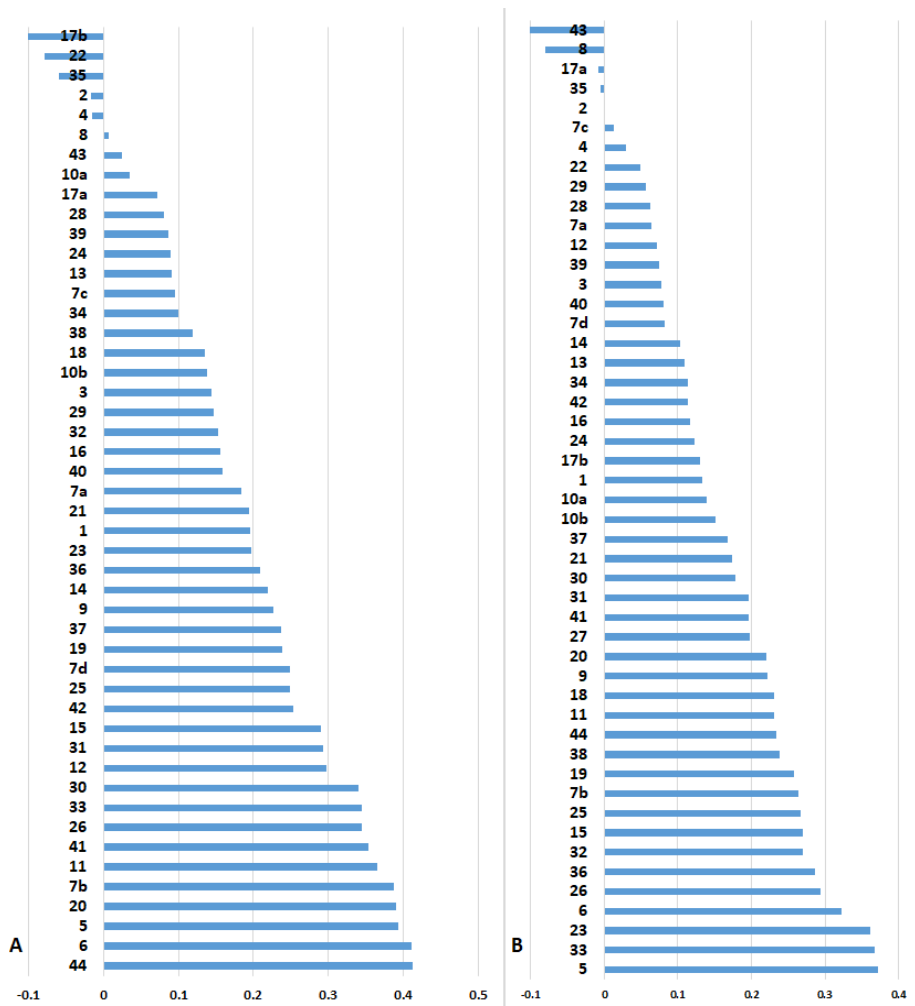
($R_S = 0,37$, p reikšmė = $1,9 \times 10^{-4}$), 33 ($R_S = 0,37$, p reikšmė = $2,3 \times 10^{-4}$), 23 ($R_S = 0,36$, p reikšmė = 3×10^{-4}) mutacijų šablonai (26 pav. B).



24 pav. dnM spektro sudėtis Lietuvos, asmenų su IN ir asmenų su IN sibsų bei Rahbari et al. apibendrintame CEU, YRI, islandų, škotų ir pietų Dakotos (JAV) populiacijų tyrime (apačioje) grupėse. Spalvos atitinka *de novo* mutacijos tipą. Vertikalioje ašyje – dnM dalis procentais, horizontalioje – 96 galimi tripletai, kurių vidurinėje pozicijoje yra atitinkama dnM , o 3' ir 5' galuose – tai nukleotidai pagal referentinę genomo seką.



25 pav. Lietuvos populiacijoje nustatytų tripletų su *dnM* dažnio koreliacija su https://cancer.sanger.ac.uk/cosmic/signatures/SBS/_duomenų bazėje esančiais 49 egzomo šablonais. Horizontalioje ašyje – Spirmano koreliacijos koeficientas, vertikaloje – nurodyti mutacijų šablonų numeriai.



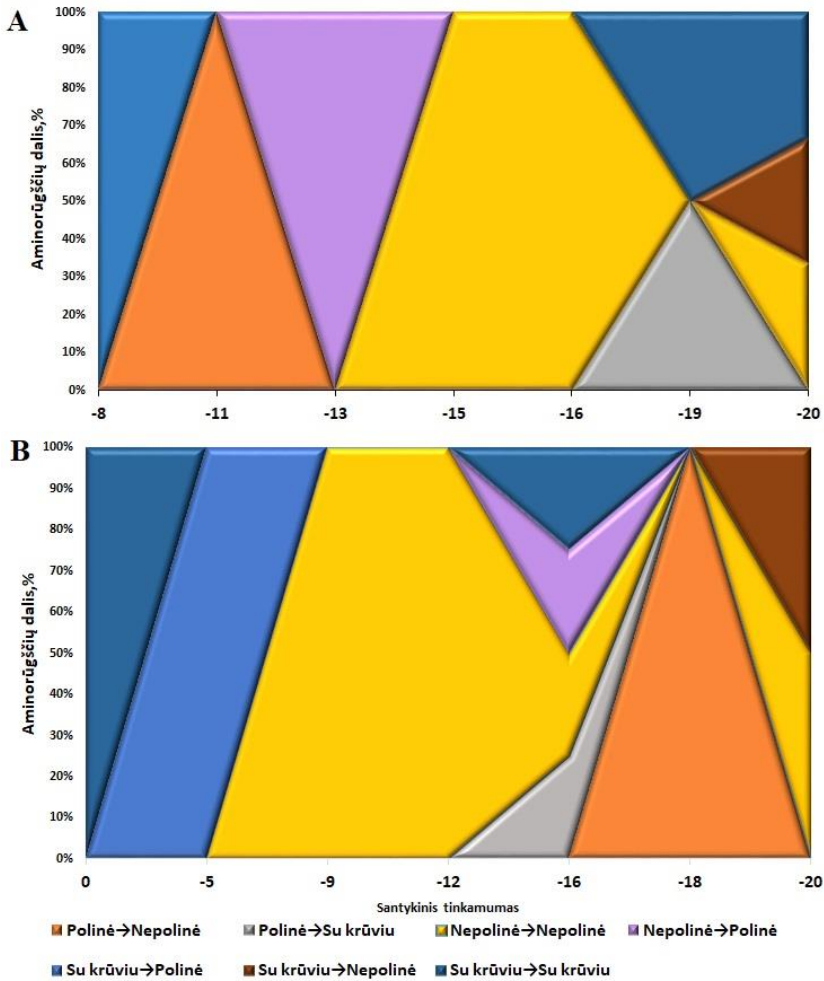
26 pav. Asmenų su IN (A) ir asmenų su IN sibsų (B) grupėse nustatytų tripletų su *dnM* dažnio koreliacija su https://cancer.sanger.ac.uk/cosmic/signatures/SBS/___ duomenų bazėje esančiais 49 egzomo šablonais. Horizontalioje ašyje – Spirmano koreliacijos koeficientas, vertikaloje – nurodyti mutacijų šablonų numeriai

4.4. Baltymų sekų, kuriose nustatytos *de novo* mutacijos, santykinio tinkamumo analizės rezultatai

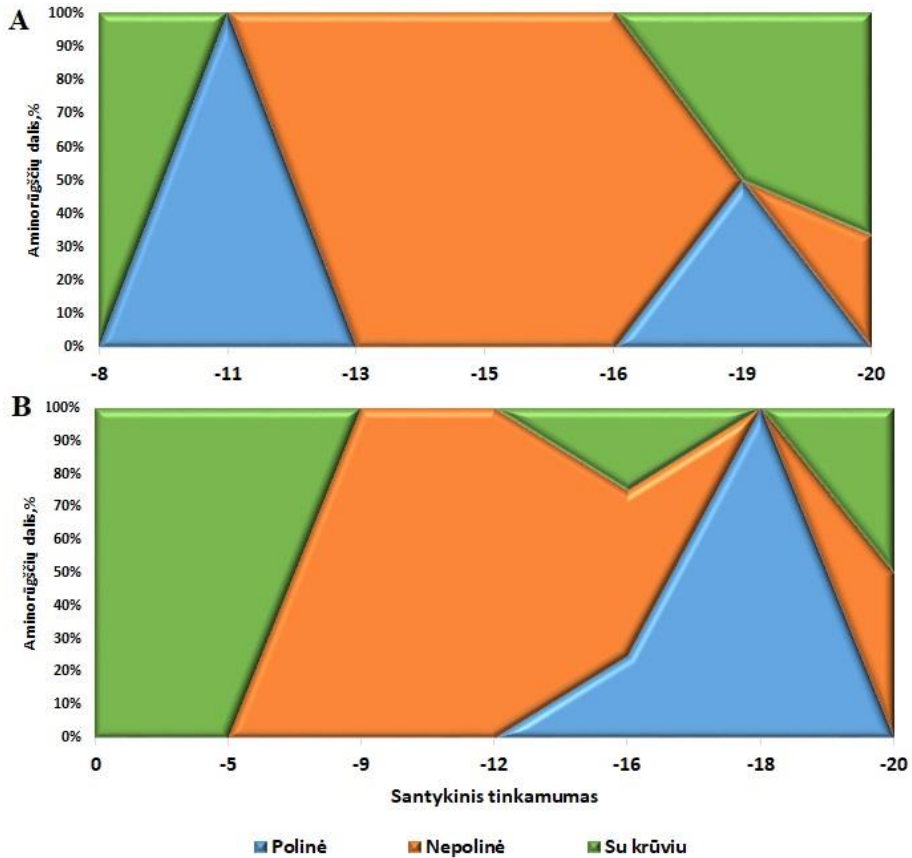
Baltymus koduojančių aminorūgščių sekų, kuriose įvyko nustatytos *dnM* (pirmoje grupėje tokių *dnM* buvo 11, antroje grupėje – 40 ir trečioje grupėje – aštuonios *dnM*) atrankos koeficientų arba kitaip santykinio tinkamumo efekto (angl. *fitness effect*) pasiskirstymo skaičiavimų analizės rezultatai vaizduojami **27–32 paveiksluose**. Čia horizontaliose ašyse pateiktų *swMutSel0* skaičiavimo modeliu apskaičiuotos santykinio tinkamumo verčių reikšmės yra nuo nulio iki -21, o tai reiškia, kad šių aminorūgščių tinkamumo efektas svyruoja nuo neutralaus (reikšmė nuo 0 iki -2) ir vidutiniškai žalingo (reikšmės tarp -2 ir -10) aminorūgščių baltymo funkcijai iki labai žalingo (> -10, kuo didesnė neigiama reikšmė, tuo žalingesnis yra pokytis baltymo funkcijai ir (ar) struktūrai). Teigiamų santykinio tinkamumo reikšmių šios analizės metu nenustatyta, taigi nė viena nustatyta *dnM* koduojančiose aminorūgštis sekose adaptavyvaus efekto nesukėlė. Vertikaliajoje ašyje pateikta analizuojamų aminorūgščių dalis procentais.

Detaliame pirmos grupės santykinio tinkamumo įverčių paveiksle (**27 pav.**) pavaizduota, kad polinės aminorūgštys prieš įvykstant *de novo* mutacijoms turėjo santykinį tinkamumą nuo -11 iki -19, o įvykus *de novo* mutacijoms jos virto nepolinėmis arba krūvį turinčiomis, tačiau abiem atvejais jų santykinis tinkamumas dar labiau sumažėjo – krito nuo -16 iki -18.

Santykinio tinkamumo įverčio reikšmes apibendrinus ir padalijus tik į polines, nepolines ir su krūviu aminorūgštis, nesvarbu, kokio tipo aminorūgštimi įvykus *de novo* mutacijai ji tapo ir kokia aminorūgštimi prieš įvykstant *de novo* mutacijai ji buvo (**28 pav.**), matomi tokie pokyčiai: 1) aminorūgščių su krūviu (kaip ir visų kitų tipų) skaičiaus pasiskirstyme prieš ir po *de novo* mutacijų matomas „veidrodžio atspindžio efektas“, t. y. 67 % didelį neigiamą (nuo -16 iki -20) santykinį tinkamumą turinčios aminorūgštys įvykus *dnM* virto neutraliomis (nulio) ar mažą (iki minus penkių) santykinį tinkamumą turinčiomis aminorūgštimis. Dėl to galima numanyti, kad *dnM* reikšmingo poveikio aminorūgščių ir iš jų sudarytų baltymų struktūrai ir funkcijai neturi; 2) visos nepolinės aminorūgštys įvykus *dnM* įgijo mažiau neigiamas santykinio tinkamumo vertes nuo -13 iki -20, nuo -9 iki -18; 3) polinių aminorūgščių santykinis tinkamumas nedaug sumažėjo ir kito nuo -11 ir -19 iki -16 ir -18.

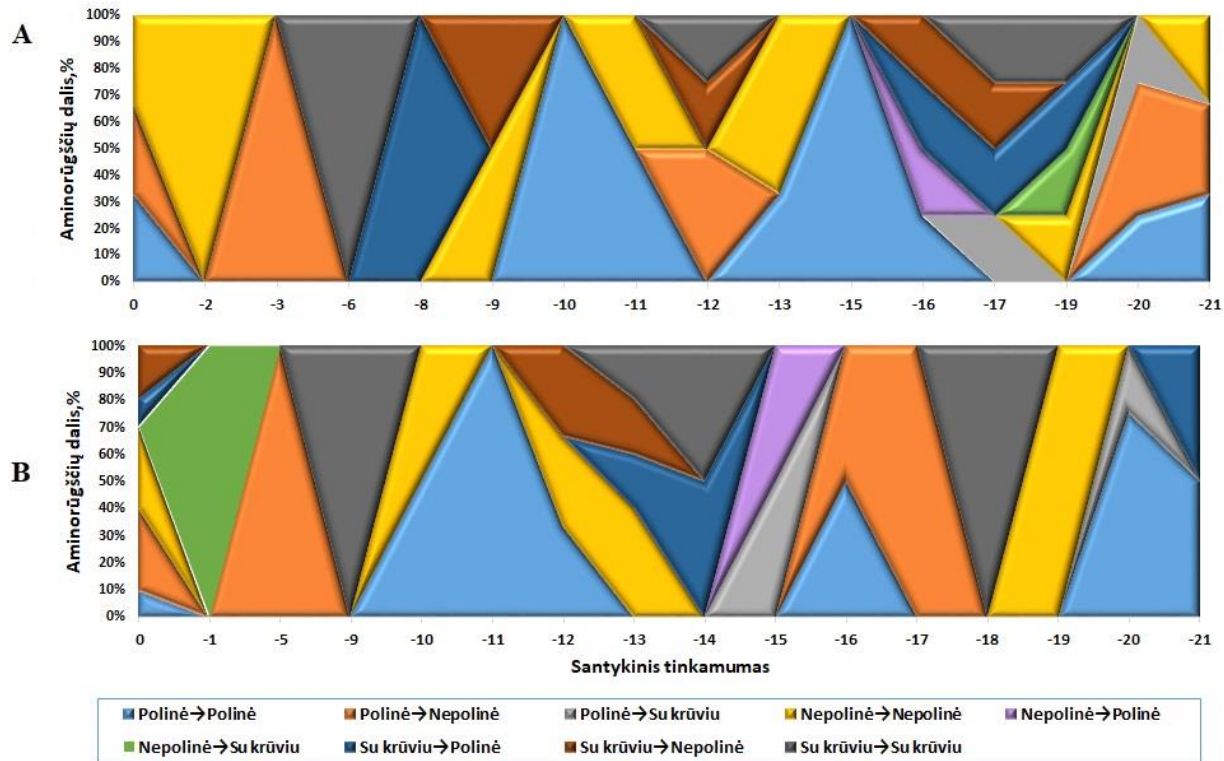


27 pav. Detalūs santykinio tinkamumo įverčiai Lietuvos populiacijos grupėje. **A)** aminorūgščių santykinio tinkamumo įverčiai prieš mutacijas. Spalvos atitinka aminorūgštis, nurodytas legendoje prieš rodyklės; **B)** įverčiai, įvykus *de novo* mutacijoms. Spalvos atitinka aminorūgštis, nurodytas legendoje už rodyklės.

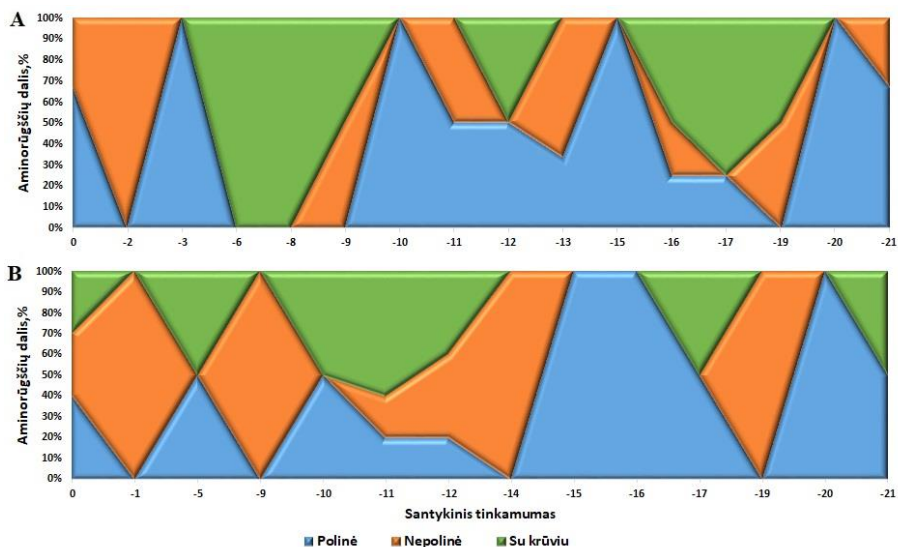


28 pav. Apibendrinti santykinio tinkamumo įverčiai Lietuvos populiacijos grupėje. **A** aminorūgščių santykinio tinkamumo įverčiai prieš mutacijas, **B** – įvykus *de novo* mutacijoms.

29 paveiksle pateikti detalūs antros grupės santykinio tinkamumo įverčiai. Dėl analizuotos aminorūgščių gausos analizė ir išvados apie santykinio tinkamumo efekto pokyčius buvo padarytos remiantis apibendrintais santykinio tinkamumo įverčiais, pateiktais **30 pav.**

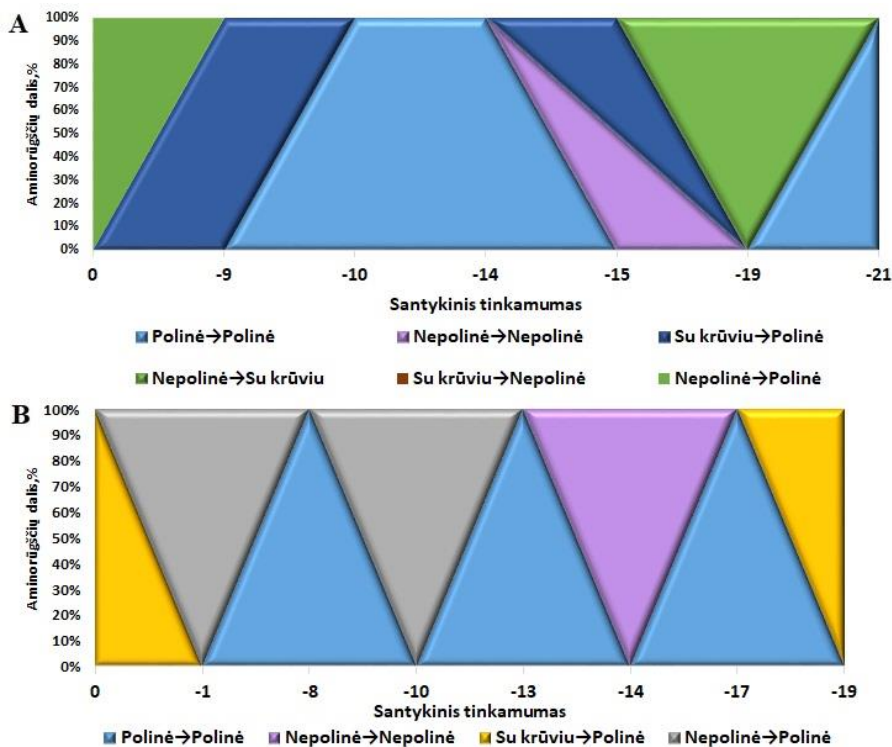


29 pav. Detalūs santykinio tinkamumo įverčiai asmenų, turinčių intelektualinę negalią, grupėje. **A** paveiksle aminorūgščių santykinio tinkamumo įverčiai prieš mutacijas, spalvos atitinka aminorūgštis, nurodytas legendoje prieš rodyklės; **B** – įvykus *de novo* mutacijoms, spalvos atitinka aminorūgštis, nurodytas legendoje už rodyklės



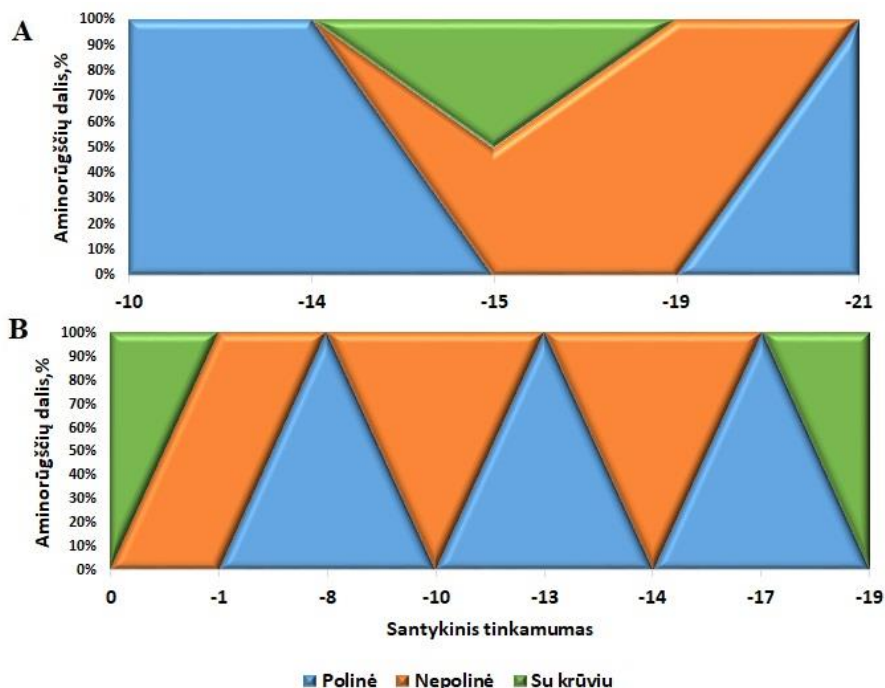
30 pav. Apibendrinti santykinio tinkamumo įverčiai asmenų, turinčių intelektinę negalią, grupėje. **A** paveiksle aminorūgščių santykinio tinkamumo įverčiai prieš mutacijas, **B** – įvykus *de novo* mutacijoms

Antros grupės aminorūgščių su *dnM* santykinio tinkamumo analizė atskleidė, kad polinių aminorūgščių skaičius įvykus *de novo* mutacijoms beveik nekito (prieš *dnM* jų buvo 18, po 14), napolinių – padaugėjo 1,9 karto (prieš *dnM* jų buvo 10, po 19), o krūvį turinčių aminorūgščių, atvirkščiai, – sumažėjo 1,7 karto (prieš *dnM* jų buvo 12, po 7). Taip pat tik trijų polinių aminorūgščių santykinis tinkamumas padidėjo nuo -3, -10 ir -21 iki 0 ir -16, susidariusių napolinių aminorūgščių santykinis tinkamumas didėjo šešioms iš 11 (54,5 %) aminorūgščių, o iš krūvį turinčių aminorūgščių santykinis tinkamumas didėjo keturioms (33 %) ir mažėjo penkioms (42 %) iš 12 aminorūgščių su krūviu (**30 pav.**). Taigi apibendrinus tiek pagal aminorūgščių skaičių, tiek ir santykinio jų tinkamumo pokyčius, galima numatyti, kad baltymų, kuriuos koduoja tirtos aminorūgštys, struktūra įvykus *dnM* turėtų reikšmingai kisti, dėl to galimi baltymų funkcijos pokyčiai, susiję su fenotipiniais požymiais.



31 pav. Detalūs santykinio tinkamumo įverčiai asmenų, turinčių intelektinę negalią, sibsų grupėje. **A** paveiksle aminorūgščių santykinio tinkamumo įverčiai prieš mutacijas, spalvos atitinka aminorūgštis, nurodytas legendoje prieš rodykles; **B** – įvykus *de novo* mutacijoms, spalvos atitinka aminorūgštis, nurodytas legendoje už rodyklės

Trečioje grupėje aminorūgščių skaičius pagal jų tipus reikšmingai nekito, o tiek detaliame **31 pav.**, tiek ir apibendrintame **32 pav.** nepolinių ir aminorūgščių su krūviu santykinis tinkamumas po *dnM* didėja nuo -14 iki -19, nuo -1 iki -14. Polinių aminorūgščių santykinio tinkamumo įverčiai padidėjo nuo -10 ir -21 ir išsibarstė nuo -8 iki -17, o iš krūvių turinčių aminorūgščių vienos aminorūgšties santykinis tinkamumas padidėjo nuo -9 iki 0, o kitos sumažėjo nuo -15 iki -19. Tai rodo, kad visų tipų aminorūgštys įvykus *dnM* tapo labiau neutralesnės nei prieš tai, todėl galima numanyti, kad analizuotos *dnM* reikšmingo poveikio aminorūgščių, taip pat baltymo struktūros ar funkcijų pokyčiams neturi.



32 pav. Apibendrinti santykinio tinkamumo įverčiai asmenų, turinčių intelektinę negalią, sibsų grupėje. **A** paveiksle aminorūgščių santykinio tinkamumo įverčiai prieš mutacijas, **B** – įvykus *de novo* mutacijoms.

Remiantis aminorūgščių santykinio tinkamumo duomenimis buvo patikrinta „žalingų“ ($S < -2$) ir „neutralių“ ($-2 < S < 0$) aminorūgščių dažnių suderinamumo tarp trijų grupių hipotezė. Pagal R paketu apskaičiuotą tiksliąjį Fišerio kriterijų aminorūgščių, turinčių skirtingą santykinį tinkamumą, dažnių pasiskirstymas tarp visų trijų grupių (p reikšmė = 0,63), taip pat tarp pirmos ir antros (p reikšmė = 0,54) bei antros ir trečios (p reikšmė = 0,66) statistiškai reikšmingai nesiskyrė.

4.5. Funkcinė *de novo* mutacijų analizė *in silico*

Be santykinio tinkamumo įverčio analizės, kuri padeda *a priori* vertinti, kaip nustatytos *dnM* gali veikti baltymų funkcijas, buvo atlikta išsami funkcinė *dnM* analizė *in silico*. Siekiant nustatyti, kurios *de novo* mutacijos yra patogeninės ir keičia paveikto baltymo struktūrą ir funkciją tiek koduojančiose aminorūgštis, tiek ir jų nekoduojančiose sekose, buvo atlikta numatomų kategorinių *dnM* patogeniškumo įverčių analizė. Buvo vertinta 11

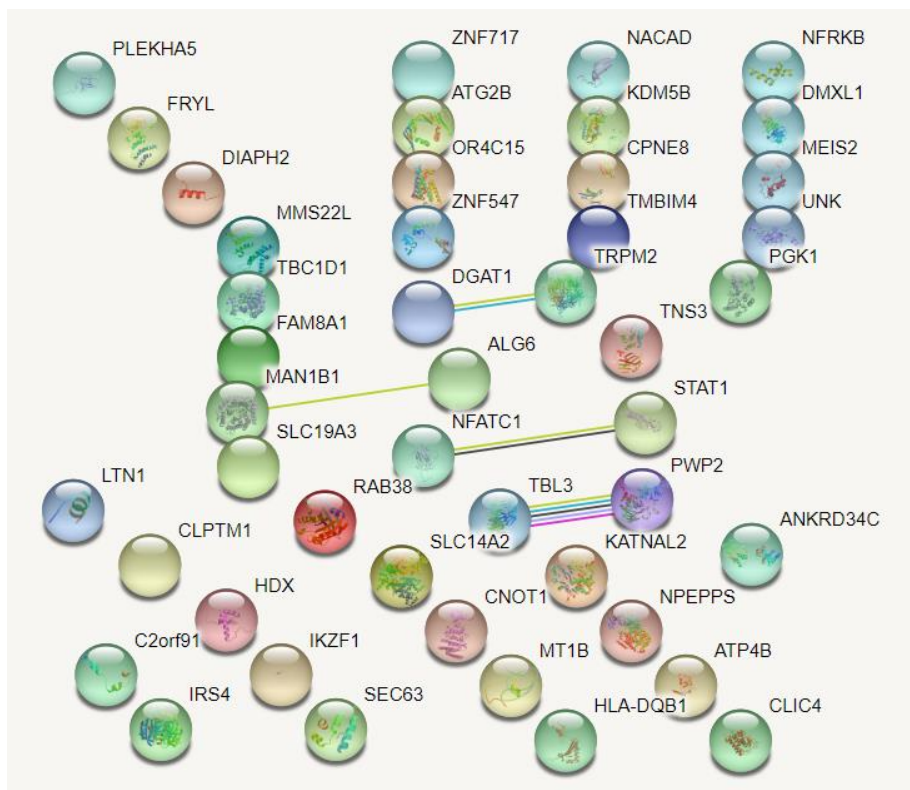
patogeniškumo įverčių: *polyphen HDIV* ir *HVAR, LRT, PROVEAN, CADD, FATHMM, MutationTester, MutationAssessor, SIFT*, koduojančių sekų *Fathmm-MKL* ir *GERP++* (**7 priedas**). Visi šie genomo variantų vertinimo įverčiai plačiai naudojami medicininėje genetikoje, norint atskirti žalingus ir toleruotinus genomo variantus. Remiantis patogeniškumo įverčiais Lietuvos populiacijos grupėje tarp iš viso vertintų 12 *dnM* (*VarScan* duomenys) nustatytos keturios galimai patogeninės mutacijos, surinkusios šešis arba daugiau įverčių kaip patogeninės ar galimai patogeninės (**11 lent.**). Asmenų, turinčių intelektualinę negalią, grupėje (antroje grupėje) patogeniškumas tais pačiais patogeniškumo algoritmais įvertintas 32 *dnM*, trečioje grupėje – devynioms *dnM*. Pirmoje ir trečioje grupėse daugiausiai nustatyta *dnM* (60–67 %), kurių patogeniškumas varijuoja tarp nulio ir penkių įverčių iš 11, o antra grupė išsiskyrė 21 *dnM* (65,6 %), kurios buvo vertintos kaip galimai patogeninės ir surinko 6–11 ar visus patogeniškumo įverčius; tik 34,4 % *dnM* antroje grupėje buvo įvertintos kaip galimai patogeninės pagal 0–5 įverčius. Remiantis *dnM* patogeniškumo įverčių analizės duomenimis buvo patikrinta skirtingo patogeniškumo *de novo* mutacijų dažnio suderinamumo tarp trijų grupių hipotezė. Remiantis tiksliau Fišerio kriterijumi skirtingo patogeniškumo *de novo* mutacijų dažnių pasiskirstymas tiek tarp visų trijų (*p* reikšmė = 0,02), tiek tarp pirmos ir antros (*p* reikšmė = 0,0125) grupių statistiškai reikšmingai skiriasi. O tarp antros ir trečios grupių skirtingo patogeniškumo *de novo* mutacijų dažnių statistiškai reikšmingo skirtumo nenustatyta (*p* reikšmė = 0,21).

11 lent. Apibendrintas *dnM* patogeniškumo vertinimas pagal 11 baltymo funkcijos patogeniškumo vertinimo įverčių

Tiriamoji grupė	Patogeninė pagal 0-3 įverčius	Patogeninė pagal 4-5 įverčius	Patogeninė pagal 6-11 įverčius	Pagal visus įvertintus įverčius patogeninė, kai <i>dnM</i> vertinta ne visais įverčiais
Bendroji Lietuvos populiacija (pirma grupė)	8	0	2	2
Asmenų su IN grupė (antra grupė)	6	5	18	3
Asmenų su IN sibsų grupė (trečia grupė)	4	2	3	0

Įvertinus kiekvienos *dnM* galimą patogeniškumą pagal patogeniškumo vertinimo įverčius, programa STRING buvo atlikta *in silico* baltymų funkcijos ir sąveikos su kitais baltymais ar molekulėmis funkcinė analizė siekiant nustatyti *dnM* funkcinio patogeniškumo mechanizmus.

Pirmoje grupėje galimai patogeninės *dnM* buvo baigmės kodono *dnM* *MEIS2* ir *ULK4* genuose ir nesinoniminės *dnM* – *MT1B* ir *PGK1* genuose (7 priedas). Šių genų koduojami baltymai yra svarbūs neuronų augimui, endocitozės procesui ir apsaugai nuo neigiamo sunkiųjų metalų poveikio. Jie dalyvauja atpalaiduojant vėžio kraujagyslių slopiklį angiostatina ir daugelyje kitų signalo perdavimo kelių. Funkcinio šių baltymų ryšio nenustatyta.



33 pav. Baltymų-baltymų sąveikos. Čia pavaizduoti baltymai, koduojami genų, kuriuose *VarSeq*TM programa nustatytos *dnM*. Spalvotos linijos žymi baltymų ryšį [196]

Lietuvos populiacijoje *dnM*, identifikuotos *VarSeq*TM programa, buvo nustatytos galimai patogeninės 48 taškinės *dnM* (7 priedas). Remiantis

baltymų, kuriuos koduojančiuose genuose šio darbo metu buvo nustatytos *dnM*, sąveikos su STRING duomenų bazėje esančiais baltymais analize, nustatyta, kad tai yra baltymai, svarbūs DNR klaidoms taisyti, chromatinui remodeliuoti, ribosomų biogenezei, imuniniam atsakui, lipidų biosintezei, potransliaciniam baltymų pažymėjimui endoplazminiame tinkle, citoplazmoje esantiems signaliniams keliams, molekulių per membranas transportui, ląstelės citoskeletui reguliuoti, ląstelių augimui ir išgyvenamumui, neuronų atsako, lemiančio kvapo suvokimą, iniciacijai. Baltymų, kuriuos koduojančiuose genuose nustatytos *dnM*, tarpusavio ryšiai nustatyti tik tarp keturių DGAT1 ir TRPM2, MAN1B1 ir ALG6, NFATC1 ir STAT1, TBL3 ir PWP2 baltymų porų (**33 pav.**).

Asmenų su IN grupėje tarp galimai patogeninių mutacijų buvo nustatytos ir patvirtintos *Sanger* sekoskaitos metodu penkios patogeninės *dnM* *ARID1B*, *PACSI*, *TCF4*, *CHD7* ir *MECP2* genuose. Visos nustatytos *dnM* buvo heterozigotinės būklės. Tyrimo metu nustatyta, kad *dnM* *ARID1B* gene lemia autosominiu dominantiniu būdu paveldimą *Coffin-Siris* sindromą (MIM #135900), *dnM* *CHD7* gene – *CHARGE* sindromą (MIM #214800), *dnM* *PACSI* gene – autosominiu dominantiniu būdu paveldimą intelektinę negalią su ligai būdingu fenotipu (MIM #615009), *TCF4* gene – *Pitt-Hopkins* sindromą (MIM #610954) ir *MECP2* gene – *Rett* sindromą (MIM #312750). Likusių 13 nustatytų galimai patogeninių mutacijų, esančių genus koduojančiose sekose, poveikis asmenų fenotipui buvo vertinamas pagal šių genų koduojamų baltymų funkcijas, pateiktas **12 lentelėje**. Dėl patogeninės pagal visus vertinimo įverčius neaprašytos *de novo* mutacijos *CREBBP* gene paciento fenotipas bus pakartotinai vertinamas siekiant nustatyti, ar nustatyta mutacija lemia *Rubinšteino-Taybi* sindromui būdingą fenotipą. *DnM* *SMO*, *RAPGEF1*, *NFATC3*, *CTR9* ir *SYT1* genuose yra sietinos su IN pagal koduojamų baltymų funkcijas, tačiau patogenezės mechanizmams aiškinti ir patvirtinti yra būtini funkciniai tyrimai modeliniuose organizmuose, jie šio darbo metu nebuvo atliekami. Likusios *dnM*, nustatytos genuose, kurie tiesiogiai nedalyvauja smegenų struktūrų vystymesi ar perduodant signalą, todėl su IN etiopatogeneze nebuvo siejamos.

Asmenų su IN sibsų grupėje nustatytos trys galimai patogeninės mutacijos:

- *DNAH8*:NM_001206927:c.C7030G:p.L2344V
- *TLN1*:NM_006289:c.C4120T:p.R1374W (rs556004234)
- *POLD1*:NM_002691:c.T563G:p.L188R (rs768494581)

DNAH8 svarbus spermatozoidų ir kvėpavimo takuose esančių cilijų judrumo procesams, *TLN1* yra ląstelių citoskeleto baltymas, o *POLD1* koduoja δ *DNR* polimerazę ir yra svarbus *DNR* replikacijos ir taisymo procesuose.

12 lent. Genų, kuriuose nustatytos galimai patogeninės *dnM* mutacijos, funkcijos

Genas	Pozicija geno ir baltymo lygmeniu	Koduojamo baltymo funkcija ¹⁰
<i>PPL</i>	NM_002705:c.C135G;p.I45M	PPL yra dezmosomų ir epidermio kornizuoto apvalkalo keratinocitų komponentas.
<i>AARS</i>	NM_001605:c.C1043T;p.T348M (rs762241422)	tRNR sintetazė, II klasės itin konservatyvus fermentas.
<i>SMO</i>	NM_005631:c.A389G;p.Y130C	Iš G baltymų sudarytas receptorius, dalyvaujantis perduodant baltymų aktyvinimo signalą.
<i>SATBI</i>	NM_002971: c.A1259G;p.Q420R	Ląstelės matricos baltymas, besijungiantis su branduolio matricos ir su DNR asocijuotomis molekulėmis.
<i>ESPN</i>	NM_031475: c.G1438A;p.V480M	Daugiafunkcinis su aktinu besijungiantis baltymas
<i>ETNPPL</i>	NM_031279: c.C926T;p.T309M	Dalyvauja medžiagų apykaitoje ir glicerolfosfolipidų sintezėje.
<i>RAPGEF 1</i>	NM_198679:c.T1727C;p.L576P	Baltymas dalyvauja Gtpazių signalo perdavimo keliuose.
<i>NFATC3</i>	NM_173165:c.G1666A;p.D556N	Baltymas yra aktyvintų T ląstelių branduolio transkripcijos veiksnys, prijungiantis DNR.
<i>ESRRG</i>	NM_206594:c.T770A;p.I257N	Hormonų receptorius, prijungiantis DNR.

Duomenys surinkti iš GeneCards duomenų bazės

<i>FASTKD1</i>	NM_024622: c.A1754G:p.Y585C	Baltymų kinazė.
<i>CREBBP</i>	NM_004380:c.A4429C:p.S1477R	CREBBP raiška vyksta daugelyje audinių. Jis dalyvauja koaktyvinant daugelio skirtingų transkripcijos veiksnių transkripciją. Svarbus embriogenezei, augimo ir homeostazės procesuose. Baltymas veikia kaip acetiltransferazė. Acetilo grupę prijungia prie histonų ir nehistoninių baltymų. Stabilizuoja papildomas baltymų sąveikas su transkripcijos kompleksu. Šio geno mutacijos sukelia <i>Rubinšteino-Taybi</i> sindromą (RTS) [186]
<i>GIGYF2</i>	NM_015575: c.G216T:p.Q72H	Tirozino kinazės receptoriaus signalo perdavimo kelio reguliatorius. Siejamas su Parkinsono liga.
<i>SYT1</i>	NM_001135805:c.C1098G:p.D366E	Sinapsinių pūslelių membraninis baltymas, yra Ca ⁽²⁺⁾ jutiklis pūslelių pernašoje ir egzocitozės procese.
<i>CTR9</i>	NM_014633: c.C74G:p.P25R	Baltymas dalyvauja chromatinio reguliavimo / acetilinimo ir ŽIV-1 transkriptų ilginimo procese.
<i>MYOM1</i>	NM_003803:c.G3883A:p.D1295N	Baltymo titino domenai, sarkomerų struktūros dalis.

TYRIMO REZULTATŲ APTARIMAS

Beveik visi dažni genomo variantai jau yra žinomi. Tūkstančiai genetinių variantų ir (ar) genų yra susieti su atitinkamais žmonių bruožais ir ligomis. O retų genetinių variantų nustatymas vis dar sparčiai progresuoja. Šių tyrimų poreikį grindžia tai, kad retų variantų įvairovė tiesiogiai priklauso nuo regionų, iš kurių žmonės yra kilę, taigi kuo šių regionų daugiau, tuo daugiau ir retų variantų. Dėl šios priežasties retų variantų, kurių daugumą sudaro unikalūs *de novo* variantai, analizę būtina atlikti įvairiose populiacijose, o tam reikia didelių pastangų. Šio darbo metu buvo atlikta išsami *de novo* mutacijų pasiskirstymo ir dažnio analizė žmogaus viso egzomo regionuose Lietuvos populiacijos (toliau pirma grupė), triadų, kai probandas turi intelektualinę negalią (toliau antra grupė), ir triadų, kai probandas yra asmens, turinčio intelektualinę negalią, sibsas (toliau trečia grupė), grupėse (iš viso ištirti 275 asmenys).

5.1. Algoritmo *de novo* mutacijoms identifikuoti atranka

Šio darbo metu buvo analizuoti asmenų egzomai, nuskaityti skirtingomis NKS technologijomis – 5500 SOLiD™ Sequencer ir Illumina HiSeq 2500 genetiniu analizatoriumi, bei apdoroti skirtingomis pirminių duomenų apdorojimo programomis – atitinkamai LifeScope™ Genomic Analysis Software 2.5.1 ir GATK programa. Skirtingos duomenų apdorojimo procedūros ir programinė nuskaitymo įranga gali paveikti identifikuojamų bazių, priskirtų referentiniam genomui, ir nustatomų genomo variantų skaičių. Dėl to kiekvienai tiriamai triadai yra identifikuojamas skirtingas bazių ir *de novo* mutacijų skaičius.

Tiriant pirmą grupę buvo siekiama atrinkti optimalią *dnM* nustatymo programą tolesniam darbui, nes tik šioje tiriamųjų grupėje, remiantis žiniomis apie kitų bendrųjų populiacijų vidutinį *dnM* skaičių, buvo galima korektiškai įvertinti *dnM* nustatymo rezultatus. Pažymėtina, kad *dnM* nustatyti buvo pasirinkta VarScan programa, nes Lietuvos populiacijos grupės egzomų sekoskaita atlikta 5500 SOLiD™ Sequencer genetiniu analizatoriumi, ir pirminės sekoskaitos bylos .XSQ formatu buvo iškart paverstos LifeScope™ Genomic Analysis Software 2.5.1 programa į .BAM formato bylas. Šiose bylose dėl naudotos programos nėra informacijos arba ji suformatuota kitaip nei .BAM bylose, gautose egzomus nusekvenavus dabar plačiausiai naudojamu Illumina genetiniu analizatoriumi ir apdorojus GATK programa pagal geros praktikos gaires. Dėl trūkstamos ar LifeScope™ programos

netinkamai suformatuotos informacijos apie nuskaitytus fragmentus *.BAM* bylose buvo analogiškai sudarytos netinkamos *.vcf* bylos plačiai naudojamam, jautriam ir specifiškam *dnM* identifikavimo algoritmui *PhaseByTransmission* [46]. O *VarScan* programa tinkama analizuoti *.BAM* bylas, gautas atlikus sekoskaitą tiek *5500 SOLiD™ Sequencer*, tiek ir *Illumina HiSeq 2500* genetiniu analizatoriumi, nesvarbu, kokia naudota programa *.BAM* byloms generuoti. Remiantis Warden CD et al. (2014) [197] duomenimis, šis algoritmas pateikia rezultatus, kurie > 97 % sutampa su aukštos kokybės duomenimis, gautais *GATK UnifiedGenotyper* ir *HaplotypeCaller* programomis. *VarSeq™* programa buvo pasirinkta kaip atviros prieigos *VarScan* programos analogas, tačiau komercinis, vizualiai patrauklesnis ir patogesnis *Windows* sistemos naudotojams, naudojamas tiek moksliniams, tiek ir klinikiams tyrimams. *VarScan* programa dirbama *Linux* sistemoje.

Pagal tyrimo duomenis, iš viso pirmoje grupėje *VarScan* programa 34 triadoms buvo nustatytos 95 *dnM*, o *VarSeq™* – 31 triadai 84 *dnM*. Nepaisant to, kad abu naudoti algoritmai sukurti taip, kad identifikuotų *dnM* probando egzome, kai tų variantų tėvų egzomuose nėra, tik 5,37 % visų nustatytų *dnM* abiem programomis sutapo. Po *dnM* filtravimo etapo nustatyta, kad *VarScan* algoritmas turi didesnę jautrumą (5,42 %) nei *VarSeq™* algoritmas (1,76 %), o mutacijų nustatymo efektyvumas, jas patikrinus *Sanger* sekoskaitos metodu, skyrėsi tik 4 %. Dėl to galima daryti išvadą, kad nė vienam iš įrankių tikrų *dnM* nepavyko teisingai nustatyti dėl ypač didelio (> 99 %) specifiškumo ir mažo jautrumo. Taigi būtina pabrėžti, kad, nustatant *dnM*, būtina naudoti papildomus griežtus duomenų filtravimo metodus (žr. **5 lent.**) [197]. Juos pritaikius, identifikuotų tikrų *dnM* skaičius kito nuo 1752 iki 95 (*VarScan* duomenims) ir nuo 4756 iki 84 (*VarSeq™* duomenims). Kadangi žinoma, kad esant didesniam (≥ 100) horizontaliam sekų padengimui naudotų programų jautrumas turėtų būti taip pat didesnis, dėl didesnio jautrumo ir atviros prieigos galimybės toliau *dnM* identifiukuoti kitose tiriamųjų grupėse buvo pasirinktas *VarScan* algoritmas. Jį pritaikius antroje ir trečioje grupėje iš nustatytų ir patikrintų *dnM*, *Sanger* sekoskaitos metodu patvirtinta 78 % mutacijų.

5.2. *De novo* mutacijų spektras ir dažnis Lietuvos populiacijoje, asmenų su IN ir asmenų su IN sibsų grupėse

Tyrimų apžvalgoje minėta, kad lytinėse ląstelėse susidariusios mutacijos reprezentuoja fundamentalią evoliucijos jėgą, kuri iškreipia fenotipinę įvairovę, todėl tikslus *de novo* mutacijų dažnio nustatymas itin reikalingas ir gali būti pritaikomas labai plačiai ir įvairiai: mutacijų, lemiančių ligas, interpretacijoms [198–200], gamtinės atrankos tyrimams [200, 201], demografinių įvykių datavimams remiantis genetinė analize [202–205] ar žmogaus mutagenezės tyrimams [124]. Taip pat *de novo* mutacijų dažnis yra pagrindinis ir centrinis biologinis parametras, kuris padeda suprasti evoliucinius fenomenus. Vieni iš svarbiausių – tai „molekulinis laikrodis“: nuolatinis procesas, kurio metu evoliucijos eigoje kaupiasi genomo pokyčiai, ir neutralumo teorija, kuri aiškina neutralių genetinių pokyčių, pavyzdžiui, pokyčių be poveikio baltymų tinkamumui, efekto genomui priežastis.

DnM dažniui nustatyti yra sukurta daug įvairių metodų, grįstų populiacijų genetikos modeliais ar fosilijų datavimo duomenimis. Jiems naudoti reikia daug papildomų netiesioginių įverčių. Daliai metodų reikalingi demografiniai populiacijų modeliai (kuriais remiantis skaičiuojamas laikas iki pirmojo bendro tiriamųjų protėvio), taip pat genomo, suskirstyto į trumpas atkarpas, rekombinacijos dažnis ar, siekiant įvertinti metinį *dnM* dažnį, apytikslė vienos kartos trukmė metais. Naudojantis papildomais duomenimis, *dnM* dažnį galima nustatyti pagal autozigotinius genomo regionus, kilusius iš artimiausio bendro protėvio, kurį galima patikimai identifikuoti. Šis būdas geras tuo, kad, vertinant *dnM* dažnį, galimas somatinių mutacijų indėlis yra minimalus, tačiau prastas tuo, kad nustatomas tik abiejų lyčių *dnM* dažnio vidurkis ir iškart kelioms kartoms.

Kitas būdas – vidutinio *dnM* dažnio nustatymas skirtingais evoliucijos laikotarpiais pagal fosilijų DNR datavimo duomenis. Šiuo būdu analizuojamo objekto, pavyzdžiui, išlikusio žmogaus DNR ir išorinio objekto, pavyzdžiui, šimpanzės DNR, divergencijos laikas yra lyginamas su fosilijų genomais, taip, pagal jų skirtumus, įvertinant *dnM* dažnį [118, 119].

Šiame darbe *dnM* dažnis apskaičiuotas ir vertintas vienu iš tiksliausių metodų pagal tiesiogiai nustatytas *dnM* viso egzomo sekoskaitos metu. Remiantis gautais duomenimis, apskaičiuotas *de novo* vieno nukleotido mutacijų dažnis Lietuvos populiacijoje yra $2,74 \times 10^{-8}$, o *de novo* iškritų-intarpų – $1,77 \times 10^{-8}$ į poziciją vienai kartai (PVK). Antroje grupėje *de novo* vieno nukleotido mutacijų dažnis yra $3,05 \times 10^{-8}$ PVK, trečioje grupėje –

$7,19 \times 10^{-8}$ PVK. *De novo* iškritų-intarpų mutacijų dažnis antros grupės genome yra $1,74 \times 10^{-7}$ PVK, trečioje grupėje – $1,08 \times 10^{-6}$ PVK. Visose trijose grupėse apskaičiuotas *dnM* dažnis yra reikšmingai didesnis (atitinkamai pirmai grupei *p reikšmė* = $4,49 \times 10^{-9}$, $5,15 \times 10^{-6}$, trečiai grupei *p reikšmė* = $4,8 \times 10^{-6}$) nei nustatytas anksčiau publikuotuose tvariuose islandų, danų, olandų, kanadiečių ir amerikiečių bendrųjų populiacijų ir autizmo spektro liga sergančiųjų grupės tyrimuose [1, 5, 8, 31, 32, 37, 130, 131], čia *de novo* vieno nukleotido mutacijų dažnis yra tarp $0,96 \times 10^{-8}$ ir $1,28 \times 10^{-8}$ PVK, o *de novo* iškritų-intarpų – $1,1 \times 10^{-9}$ PVK (išsamią informaciją žr. **3 lent.**). Šiame tyrime nustatytą didesnę *dnM* dažnį galima grįsti faktu, kad tyrimas atliktas naudojant tik egzomo sekoskaitos duomenis. Verta pažymėti, kad egzomai pasižymi reikšmingai (apie 30 %) didesniu mutacijų dažniu nei visas genomai. Tai aiškinama tuo, kad viso genomo heterociklinių bazių struktūra skiriasi nuo egzomo: egzome regionai su guanino ir citozino dinukleotidų pasikartojimais sudaro maždaug 50 % viso egzomo dydžio, o visame genome – apie 40 % [37]. Žinoma, kad žmonių genomuose metilinti citozino ir guanino dinukleotidų gausūs regionai yra labai mutabilūs dėl citozino bazių spontaninio deaminavimo [37]. Remiantis lyginamosios genomikos tyrimų duomenimis, didesnis mutacijų dažnis CpG dinukleotidų gausiuose regionuose, manoma, atsirado maždaug žinduolių taksoninės įvairovės didėjimo (dar vadinamos radiacijos, angl. *mammalian radiation*) metu [32]. Ši hipotezė buvo iškelta, remiantis žiniomis, kad, rūšims diverguojant, CpG dinukleotidų gausiuose koduojančiuose regionuose mutacijų dažnis buvo didesnis nei nekoduojančiuose DNR regionuose. Evoliucijos eigoje, vykstant mutacijoms, koduojantys regionai virto nekoduojančiais regionais, todėl CpG dinukleotidinio genomo konteksto efektas, kaip ir vidutinis mutacijų dažnis, bėgant laikui mažėjo, kol pasiekė mutacijų dažnį, nustatomą aplink koduojančias sekas esančioje nekoduojančioje DNR [206]. Vis dėlto, nors genomo regionuose, kur skirtingas CpG dinukleotidų kontekstas, neutraliems genomo regionams susikurti ir taip pasiekti pusiausvyrą buvo pakankamai laiko, išgryninančioji gamtinė atranka funkcinuose genomo regionuose išlaikė hipermutabilių CpG dinukleotidų dalį [32, 206–208]. Šiame darbe nustatytas didesnis *dnM* dažnis nei kitų tyrėjų genomo tyrimuose, manoma, yra dėl specifinio CpG egzomo sekų konteksto ir (ar) gamtinės atrankos poveikio egzomui. Remiantis tiesinės regresijos modeliais, buvo nustatyta, kad ~68–94 % viso *dnM* dažnio Lietuvos populiacijos genome paaiškina tiems genomo regionams būdingas DNAzei 1 hiperjautrumas, CpG kontekstas, regionų konservatyvumas (remiantis GERPP++ konservatyvumo

įverčiais) ir regionų raiškos lygiai. Asmenų su IN grupėje *dnM* dažnį taip pat, kad ir nedaug (13 %, *p* reikšmė = 0,02), aiškino egzomo regionų raiškos lygiai. Remiantis šiais duomenimis padaryta išvada, kad *dnM* egzome susidaro į promotorius panašiuose, transkripciškai aktyviuose genomo regionuose nepriklausomai nuo DNR sekos konservatyvumo, nors *dnM* dažnis buvo didesnis genuose, kurių raiška yra nespecifinė. Tai sutampa su genų raiškos, transkripcijos veiksnių prisijungimo ir histonų modifikacijų vieta daugelyje organizmų rūšių ir jų audinių žemėlapių duomenimis [193, 195, 209]. Remiantis šiais duomenimis skirtingų rūšių organizmuose tapačių ar panašių audinių genų raiška yra konservatyvi [210], tačiau stipriklių (angl. *enhancers*) ir promotorių sekoms būdinga greita DNR struktūros kaita [211–213]. Pavyzdžiui, neseniai atliktas hepatocitų promotorių ir stipriklių tyrimas 20-čiai žinduolių rūšių atskleidė, kad 25 % stipriklių ir 10 % promotorių visoms rūšims buvo unikalūs net tada, kai sekos buvo ypač konservatyvios [211, 214]. Panašių rezultatų gauta ištyrus ir žmogaus, makakos ir pelės galūnių audinius [215].

Priešingai nei kitų tyrėjų darbuose [1, 3, 32, 43], šiame darbe nustatyta, kad tėvų amžius su *dnM* dažniu Lietuvos populiacijoje nekoreliuoja (tėvams $R_S = 0,2$ *p* reikšmė = 0,09, motinoms $R_S = 0,08$ *p* reikšmė = 0,98). Tai galima paaiškinti faktu, kad tyrimo metu analizuotų triadų tėvų amžius yra pernelyg panašus, todėl nėra pakankama amžiaus ir *dnM* dažnio analizės galia. Taip pat tyrime buvo analizuota tik nedidelė (~1,5 %) viso genomo dalis – egzomas, todėl vidutiniškai kiekviename analizuotame egzome nustatyta po 1,9 (*VarScan*) ir 1,7 (*VarSeq*TM) *de novo* vieno nukleotido mutacijos, o visame genome jų identifikuojama maždaug nuo 40 iki 82 [5,41]. Asmenų su IN grupėje *de novo* VNP nustatyta daugiau – vidutiniškai po 4,2 *de novo* vieną nukleotidą ir 1,3 *de novo* iškritos-intarpo, kurio dydis svyruoja nuo vieno iki 30 nukleotidų. Asmenų su IN sibsų grupėje vidutiniškai kiekvienas tirtas probandas turėjo po 2,4 *de novo* VNP ir 0,92 *de novo* iškritos-intarpo nuo trijų iki 31 nukleotido dydžio. Trečioje grupėje tarp nustatytų *de novo* iškritų-intarpų nustatyta tik viena *dnM* koduojančioje sekoje (8 %), antroje grupėje – dešimt (24 %), iš kurių net penkios *dnM* yra rėmelio poslinkio tipo. Taigi dėl didesnio *dnM*, nustatytų antroje grupėje, skaičiaus buvo padaryta prielaida, kad šios grupės atveju *dnM* galėjo būti IN priežastimi.

Lietuvos populiacijos ir asmenų su IN sibsų grupės buvo panašios pagal funkcinis regionus, kuriuose nustatytos *dnM*. Jų didžiąją dalį (54–64 %) sudarė intronuose esančios *dnM* ir tik 21–24 % jų buvo egzonuose. O asmenų

su IN grupėje *dnM* intronuose buvo 48 % ir 34 % egzonuose. Analogiški grupių panašumai ir skirtumai matyti išanalizavus grupes pagal *dnM* tipus: asmenų su IN grupė išsiskyrė mutacijų, lemiančių pakitusią baltymų sudėtį dėl pasikeitusių aminorūgščių, gausa: dviem baigmės kodono *dnM*, penkiomis rėmelio poslinkio *dnM*, 13 sinoniminėmis, 28 nesinoniminėmis *dnM* ir penkiomis nuo vieno iki 30 nukleotidų dydžio *de novo* delecijomis, o Lietuvos populiacijos ir asmenų su IN sibsų grupėse nustatytų *dnM* pagrindą sudarė 13 sinoniminių ir 18 nesinoniminių *dnM*.

Remiantis *de novo* mutacijų spektro sudėties analizės duomenimis, Lietuvos populiacija, asmenų su IN ir asmenų su IN sibsų grupės *dnM* spektrų sudėtimi nesiskiria (p reikšmė = 0,575, χ^2 kriterijus) ir yra artimos Jutos gyventojų, kilusių iš Šiaurės ir Vakarų Europos (CEU), Ibadano Jorubos (Nigerija) (YRI), islandų, škotų ir pietų Dakotos (JAV) populiacijose nustatytai *dnM* spektrų sudėčiai [31]. Visose grupėse daugiausiai nustatyta C>T (pirmoje grupėje 33 %, antroje – 35 % ir trečioje grupėje 59 %) ir T>C (pirmoje grupėje 28 %, antroje – 27 % ir trečioje grupėje 18,5 %) *dnM*.

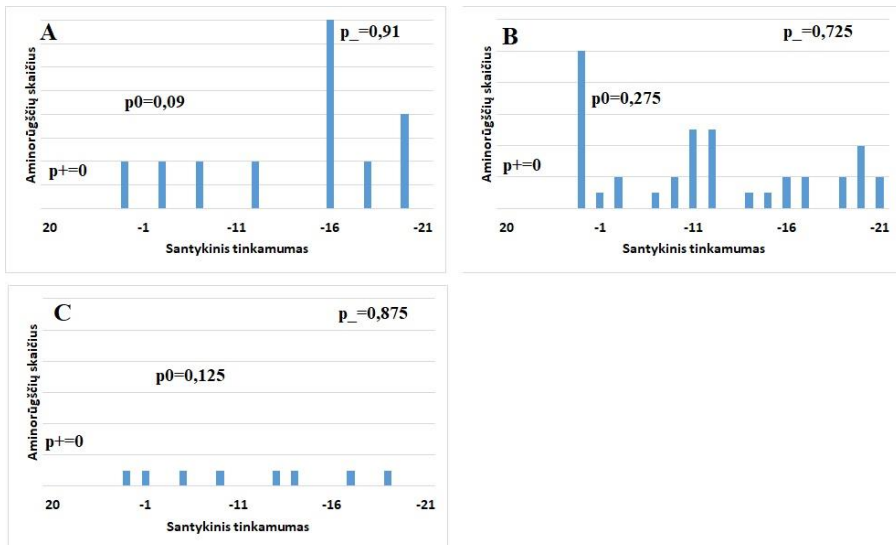
Vis dėlto Lietuvos populiacijoje nustatytų tripletų su *dnM* dažnio koreliacijos su <https://cancer.sanger.ac.uk/cosmic/signatures/SBS/> duomenų bazėje esančiais 49 egzomo šablonais analizė atskleidė, kad visos šio darbo metu analizuotos grupės šablonų sudėtimi skiriasi nuo Rahbari et al. 6570 *de novo* mutacijų pagrindu sudaryto šablonų katalogo ir yra unikalios: Lietuvos populiaciją beveik lygiomis dalimis (42–49 %) apibūdina šeši mutacijų šablonai: 26, 10a, 15, 41 ir 7a. Asmenų su IN grupę apibūdina 44, šeštas, penktas, 20 ir septintas *b* mutacijų šablonai, o asmenų su IN sibsų grupę apibūdina penktas, 33 ir 23 mutacijų šablonai. Rahbari et al. kataloge įtrauktose CEU, YRI, islandų, škotų ir pietų Dakotos populiacijose ~ 75 % visų *dnM* apibūdinamos penktu ir ~ 25 % – pirmu šablonu.

Lietuvos populiacijoje nustatyti *dnM* šablonai, manoma, susidaro dėl DNR grandinių nesutapimo taisymo mechanizmo klaidų (26, 15 šablonas), ϵ polimerazės egzonukleazės domeno mutacijų (10a šablonas), ultravioletinės spinduliuotės metu susidariusių ciklobutano pirimidino dimerų (7a šablonas). 41 mutacijų šablono etiologija nėra žinoma. Asmenų su IN grupėje nustatyti mutacijų šablonai, manoma, susidaro dėl DNR grandinių nesutapimų taisymo mechanizmo klaidų (44, šeštas ir 20 šablonas), mutacijų *POLD1* gene (20 šablonas) ir ultravioletinės spinduliuotės (7b šablonas). Penkto mutacijų šablono etiologija nėra žinoma, tačiau nustatyta, kad šio šablono raiška pasižymi šlapimo pūslės vėžinės ląstelės su *ERCC2* geno mutacijomis.

Asmenų su IN sibsų grupėje nėra vieno nustatyto mutacijų šablono etiologija nėra žinoma.

5.3. Baltymų sekų, kuriose nustatytos *de novo* mutacijos, santykinio tinkamumo analizės rezultatų aptarimas

Populiacijų genetikos tyrimuose asmenų, turinčių įvairių tipų *de novo* mutacijų ir su jomis asocijuotų atrankos koeficientų dažnio pasiskirstymas yra esminis populiacijų genetikos tyrimų objektas [237, 238]. Mutacijų dažnis ir santykinis tinkamumas daro įtaką daugybei evoliucinių ir biologinių reiškinių, pavyzdžiui, genetinei variacijai palaikyti, kiekybinių bruožų kitimui, rekombinacijai, lyties evoliucijai, senėjimo evoliucijai, genų kopijoms išlikti ir t. t. [18]. Kai genome susidaro *de novo* mutacija, ji gali turėti vieną iš trijų poveikių asmens fenotipo tinkamumui (toliau $w = s + 1$, čia w yra santykinis tinkamumas, o s – atrankos koeficientas): mutacija gali būti žalinga ($s < 0$), kai sumažėja asmenų išgyvenimas arba vaisingumas, neutrali ($s \approx 0$), kai mutacija turi tokį mažą poveikį asmenų tinkamumui, kad jos dažnis populiacijoje tiesiogiai priklauso nuo genų dreifo, arba naudinga ($s > 0$), kai asmenų su tokia mutacija išgyvenimo trukmė ilgėja arba jų reprodukcinė sėkmė yra didesnė nei kitų asmenų. Galutinis dnM likimas, ji bus fiksuojama ar pašalinta iš populiacijos genomo, priklauso nuo gamtinės atrankos poveikio stiprio ir genų dreifo efektyviam populiacijos dydžiui, todėl ir tinkamumas yra aprašomas tinkamumo koeficientu S , kuris yra lygus efektyvaus populiacijos dydžio N_e ir gamtinės atrankos s sandaugai ($S = 2N_e \times s$). Šiame darbe buvo analizuotas kiekvienos aminorūgšties su dnM tinkamumas koeficientu S iki gamtinės atrankos poveikio, nes dnM buvo vertintos momentiška. Pritaikius $swMutSel0$ modelį, buvo nustatyta, kad Lietuvos populiacijoje neutralios ($-2 < S < 0$) mutacijos sudarė tik ~9 %, o žalingos ($S < -2$) net 91 %, iš kurių stipriai žalingos ($S < -10$) ~73 % (**34 pav.**). Asmenų su IN grupėje neutralios mutacijos sudarė apie 27,5 %, žalingos ~72,5 %, iš kurių stipriai žalingos sudarė ~60 %, o asmenų su IN sibsų grupėje neutralių mutacijų nustatyta panašiai kaip ir pirmoje grupėje – 12,5 %, žalingų – 87,5 %. Neutralių ir patogeninių mutacijų dalis asmenų su IN grupėje nuo Eyre-Walker A ir Fay JC nustatytos [216, 217] reikšmingai nesiskyrė (p reikšmė = 0,6, χ^2 kriterijus) – atitinkamai 20 % ir 80 %. Adaptyvių mutacijų nėra vienoje iš grupių nenustatyta, nors, naujausiais duomenimis, adaptyvių mutacijų dažnis genome \hat{a} siekia net 13,5 % [218].



34 pav. Santykinio tinkamumo koeficiento S pasiskirstymas genuose, kuriems nustatytos dnM . p_- atitinka žalingų ($S < -10$) dnM dalį, p_+ naudingų ($S > 0$) dnM dalį ir p_0 – neutralių ($-2 < S < 0$) dnM dalį. A) Lietuvos populiacijos B) asmenų su IN grupės C) asmenų su IN sibsų grupės analizės duomenys

Paradoksalu tai, kad, nors šiame darbe, remiantis Lelieveld SH et al. 2016 m. tyrimu [156], buvo kelta hipotezė, kad Lietuvoje asmenų su IN grupėje žalingų mutacijų turėtų būti reikšmingai daugiau nei kitose tirtose grupėse, remiantis tiksluoju Fišerio kriterijumi aminorūgščių su skirtingu santykinu tinkamumu dažnių pasiskirstymas tarp visų trijų grupių statistškai reikšmingai nesiskyrė (p reikšmė = 0,63). Darant prielaidą, kad tokį rezultatą galimai lėmė tai, kad pirmoje ir trečioje grupėse iš visų identifikuotų dnM santykinis tinkamumas buvo vertintas tik atitinkamai 11 % ir 17 % nustatytų dnM , esančių koduojančiose genomo srityse, žalingų dnM dalis kiekvienai grupei buvo vertinta pakartotinai, į analizę įtraukus tas dnM , kurių nebuvo galimybės įtraukti į santykinio tinkamumo analizę dėl jų pozicijų nekoduojančiose genomo sekose. Manant, kad papildomai įtrauktos dnM yra galimai neutralios, nes neturi reikšmingos įtakos tirtų asmenų sveikatai (anketinės apklausos metu tiriamieji įvertino save kaip „sveikus“, nors susirgti ateityje galėjo), tai anksčiau minėta hipotezė buvo patvirtinta, nes

tokiu atveju asmenų su IN grupėje žalingų *dnM* dalis siekia ~20 %, o pirmoje ir trečioje grupėse – atitinkamai ~10 % ir ~13 %. Taip pat santykinio tinkamumo analizės duomenys atskleidžia, kad iš *dnM* skaičiaus (kartu ir intensyvumą) genome prieš gamtinės atrankos poveikį galima tiesiogiai spręsti apie galimą *dnM* mutacijų poveikį genomui ir asmens fenotipui, nes esant reikšmingai didesniai *dnM* skaičiui apie du kartus didėja tikimybė, jog unikalios mutacijos (ypač koduojančiose geno srityse) gali turėti žalingą poveikį asmens fenotipui, taigi šiuo atveju gali būti IN priežastis [241]. O gamtinei atrankai veikiant, žalingų mutacijų poveikis ir *dnM* dažnis genome, kurį asmenys gali toleruoti, priklauso tik nuo to, ar ji veikia absoliutų ar santykinį asmens su *dnM* fenotipo tinkamumą [242].

Neutralumo teoriją įrodo pirmoje grupėje buvęs „veidrodžio atspindžio efektas“, kai dauguma didelį neigiamą santykinį tinkamumą ($S < -16$) turinčių aminorūgščių įvykus *dnM* virto neutraliomis (0) ar mažą (iki -5) santykinį tinkamumą turinčiomis aminorūgštimis. Visos nepolinės aminorūgštys įvykus *dnM* įgijo mažiau neigiamas santykinio tinkamumo vertes, o polinių aminorūgščių įvykus *dnM* santykinis tinkamumas kito nedaug. Aminorūgščių pagal tipus skaičius nekito, o *dnM* reikšmingo poveikio aminorūgščių ir iš jų sudarytų baltymų struktūrai ir funkcijai neturėjo. Antroje grupėje, asmenų su IN grupėje, įvykus *dnM*, reikšmingai kito aminorūgščių sudėtis – nepolinių aminorūgščių – padaugėjo 1,9 karto, o krūvį turinčių aminorūgščių, atvirkščiai, – sumažėjo 1,7 karto. Taip pat tik trijų polinių aminorūgščių santykinis tinkamumas padidėjo nuo -3, -10 ir -21 iki 0 ir -16, susidariusių nepolinių aminorūgščių santykinis tinkamumas didėjo šešioms iš 11 (54,5 %) aminorūgščių, o iš krūvį turinčių aminorūgščių santykinis tinkamumas didėjo keturioms (33 %) ir mažėjo penkioms (42 %) iš 12 aminorūgščių su krūviu. Taigi, apibendrinus pagal tiek aminorūgščių skaičių, tiek ir santykinio jų tinkamumo pokyčius, galima numatyti, kad baltymų, kuriuos koduoja tirtos aminorūgštys, struktūra įvykus *dnM* turėtų reikšmingai kisti, todėl galimi baltymų funkcijos pokyčiai, susiję su fenotipiniais požymiais.

Trečios grupės santykinio tinkamumo duomenys atkartoja pirmos grupės duomenis, nes visų tipų (išskyrus vienos krūvį turinčios aminorūgšties) aminorūgščių santykinio tinkamumo efekto vertės įvykus *dnM* mažėjo, o jų skaičius pagal tipus nekito. Tai rodo, kad visų tipų aminorūgštys įvykus *dnM* tapo labiau neutralesnės nei prieš tai, todėl galima numatyti, kad analizuotos *dnM* reikšmingo poveikio aminorūgščių, o kartu ir baltymo struktūros ar funkcijų pokyčiams neturi.

5.4. Funkcinės *in silico* analizės rezultatų aptarimas

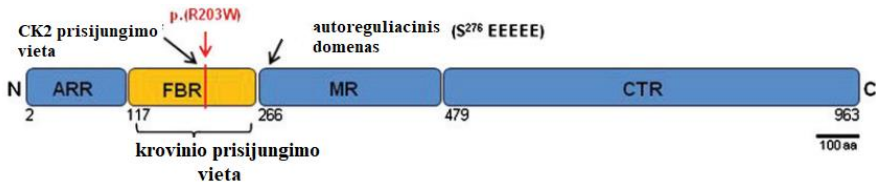
Išplėstinės funkcinės *in silico* analizės rezultatai parodė, kad Lietuvos populiacijos grupėje tarp visų identifikuotų *dnM*, 4 *dnM*, identifikuotos *VarScan* programa, ir 35 *VarSeq*TM programa yra galimai patogeninės. Didelį patogeninių *dnM* skaičiaus skirtumą galima paaiškinti tuo, kad priklausomai nuo *dnM* identifikuoti skirto algoritmo, identifikuotų *dnM* skaičius koduojančiose sekose reikšmingai skyrėsi. Pavyzdžiui, *VarScan* algoritmu nustatytų *dnM* buvo tik 21,05 %, o *VarSeq*TM algoritmu net 95,24 %. Remiantis žiniomis apie baltymų struktūrą ir funkcijas, algoritmų mutacijų poveikiui koduojančioje genomo dalyje vertinti yra gausu, o apie reguliacinius genomo regionus žinių, kartu ir jų struktūros pokyčių vertinimo algoritmų nėra daug, todėl mutacijos, esančios koduojančiose sekose, dažniau įvertinamos kaip patogeninės. Verta pažymėti, kad nors reguliaciniai genomo regionai baltymų struktūrą ir funkcijas veikia netiesiogiai, jie sudaro didelę nekoduojančio genomo dalį – vienam baltymui yra priskiriama ~11 % viso genomo dydžio reguliacinių elementų. Jų išsibartymas genome užtikrina minimalų patogeninių genomo variantų poveikį [243]. Pirmoje grupėje identifikuotos patogeninės *dnM* buvo genuose, kurie koduoja baltymus, svarbius chromatinui remodeluoti, ląstelių citoskeletui reguliuoti, ląstelių augimui ir gyvybingumui moduluoti, citoplazmos signaliniams keliams funkcionuoti ir neuronų atsako iniciacijai. Vis dėlto, nepaisant šių *dnM* įvertinimo patogeniškumo, LITGEN projekto apklausos metu visi tiriamieji save įvertino kaip „sveikus“. Tai rodo, kad, nepaisant galimo *dnM* patogeniškumo, genomai šiuos pokyčius akivaizdžiai toleruoja, todėl fenotipinis *dnM* poveikis nepasireiškia. Tai pasufleruoja, kad šiame tyrime identifikuotos patogeninės *dnM* nėra pakankamai žalingos, kad dėl jų sumažėtų bendras koduojamų baltymų tinkamumas, todėl šios *dnM* geba išlikti daugelį kartų ir išvengti gamtinės atrankos poveikio. Remiantis *Szamecz* tyrimu, kuo dažniau *dnM* atsiranda konservatyviuose genomo regionuose, tuo stipresnis gamtinės atrankos poveikis per genomo apsaugos kompensacinius mechanizmus genetiniams pokyčiams tenka [219]. Žalingas pokyčių poveikis gali būti sušvelnintas keturiais būdais. Dalis genų gali toleruoti sutrumpėjusius baltymų variantus, nes jų funkcinis poveikis gali būti maskuojamas (1) nepilna genų raiška, (2) kompensaciniais baltymų variantais arba (3) dėl mažos baltymo sutrumpėjimo funkcinės reikšmės [220]. Tais atvejais, kai genų pokyčiai, asocijuoti su nesinoniminėmis *dnM*,

jie gali būti kompensuojami (4) genome susikaupus reikalingam skaičiui naudingų mutacijų [219].

Intelektinės negalios asmenų grupėje (antros grupės) funkcinės *dnM* analizės rezultatai leido 13,5 % tiriamųjų indentifikuoti IN genetinę priežastį ir paaiškinti ligos etiopatogenezę. Asmenų, turinčių IN, grupėje tarp galimai patogeninių mutacijų buvo nustatytos ir patvirtintos penkios patogeninės *dnM ARID1B*, *PACS1*, *TCF4*, *CHD7* ir *MECP2* genuose, iš kurių trys buvo iki šiol literatūroje neaprašytos. Remiantis OMIM duomenų baze dabar žinoma ~800 genų, lemiančių IN, o tai sudaro ~19 % žinomų genetines ligas lemiančių genų, taigi daugeliu atvejų, kaip ir šio tyrimo metu, IN priežastys lieka neaiškios.

Visos nustatytos *dnM* buvo heterozigotinės būklės. Atlikus DNR ir RNR tyrimus nustatyta, kad NC_000006.12 (NM_020732.3): c.5025+2T>C *dnM ARID1B* gene lemia autosominiu dominantiniu būdu paveldimą *Coffin-Siris* sindromą (MIM #135900). Probando kDNR sekoskaita *Sanger* sekoskaitos metodu atskleidė, kad *dnM* suardo 19 introno 5' gale esantį sukirpimo centrą, todėl 19 egzonus yra praleidžiamas. Vieno egzono praleidimas sukirpimo metu yra dažnas patogeninės pre-mRNR rezultatas, dėl kurio susidaro atviro DNR skaitymo rėmelio poslinkiai, formuojasi pirmalaikiai baigmės kodonai arba vyksta trumpesnių baltymų sintezė [221]. Dėl ląstelėse esančių tiek laukinio tipo, tiek ir mutantinių transkriptų, mutantiniai transkriptai visiškai nesuyra, o vietoj to sintetinamas sutrumpėjęs baltymas NP_065783.3:p.(Thr1633Valfs*11). Sutrumpėjęs ARID1B lemia BAF250 domeno, kuris yra nuo ATP priklausomo į SWI/SNF panašaus chromatino modifikavimo komplekso dalis, reguliuojanti genų raišką, praradimą. BAF250 gali veikti panašiai kaip E3 ubikvito ligazė, kurios taikinis yra H2B histonai [222]. Pelių embrioninių ląstelių su neveikliu BAF250b kompleksu tyrimai parodė, kad BAF250 yra ypač svarbus ankstyvajai embriogenezei, nes modifikuotų ląstelių atsinaujinimas sutrinka, o diferenciacija didėja [223–225]. Su ARID1B asocijuoti BAF kompleksai dalyvauja neuronų diferenciacijoje ir žinduolių smegenų vystymesi. Ypač jų raiška svarbi piramidinių neuronų vystymuisi, kai formuojasi kompleksinė dendridinė architektūra [222, 223]. Taip pat žinoma, kad ARID1B haplonepakankamumas lemia GABAerginių interneuronų sumažėjimą smegenų žievėje. Taip smegenų vystymosi metu paveikiami eksitaciniai ir neurotransmisiją slopinantys procesai. Negana to, dėl BAF komplekso sąveikos su histonais ARID1B haplonepakankamumas gali veikti ir per epigenetinių reguliavimą.

NM_018026.4:c.607C>T (Arg203Trp) *dnM* *PACS1* gene lemia autosominiu dominantiniu būdu paveldimą intelektinę negalią su ligai būdingu fenotipu (MIM #615009). Šią mutaciją 2015 metais nustatė ir IN etiopatogenezę paaiškino Gadzicki D et al. [226]. *PACS1* yra Goldžio aparato judėjimo per membranas reguliatorius, kuris nukreipia baltymus reikalinga kryptimi [227]. Embriogenezės metu didelė jo raiška vyksta smegenyse, taip pat reguliuojama ir postnataliniame periode [228]. Tiriant šį pokytį, buvo iškelta hipotezė, kad dėl c.607C>T *dnM* šalia proksimalinio CK2 prisijungimo vietos motyvo kinta *PACS1* prisijungimo savybės (35 pav.) [226]. Tyrimai su zebražuvėmis ir priešprasmėmis (angl. *antisense*) DNR sekomis leido šią hipotezę patvirtinti. Jų metu nustatyta, kad šis geno variantas daro dominantinį neigiamą poveikį, dėl kurio kinta zebražuvių neuronų su SOX10 baltymu išgyvenimas, nes dėl jo formavosi veido dismorfija [229].



35 pav. *PACS1* baltymas sudarytas iš 962 aminorūgščių ir keturių domenų: su atropinu 1 susijusio regiono (ARR), furino krovinio prisijungimo regiono (FBR), viduriniojo regiono (MR) ir C galinio regiono (CTR). *DnM* vieta pažymėta raudonai, CK2 – kazeino kizazė 2, aa – aminorūgštys. Adaptuota iš [251]

CHD7 gene nustatyta *de novo* 18 nukleotidų duplikacija (c.6341_6358dup, p.Asp2119_Pro2120ins6); NM_017780; NP_060250; MIM# 214800). Nustatyta, kad mutacija yra evoliuciškai konservatyviame regione, o baltymo modeliavimas parodė, kad DNR sekos, kurioje susidarė *dnM*, konformacija iš linijinės grandinės pakito į trumpą α spiralę. Žinoma, kad *CHD7* koduoja chromodomeno DNR prijungiančią helikazę, kuri yra svarbi ankstyvoje embriogenezėje ir kontroliuoja genų raišką per chromatino modifikavimą ląstelės ciklo metu [230], todėl net trimis iš keturių atvejų būtent *CHD7* gene identifikuojamos mutacijos asmenims su *CHARGE* (MIM #214800) sindromui būdingais fenotipiniais požymiais. Tirtam asmeniui taip pat patvirtintas *CHARGE* sindromas, o mutacija buvo ne tik *de novo*, bet ir unikali, iki šiol neprašyta. *CHARGE* sindromu sergantiems asmenims

būdingas kolobomos, širdies defekto, sutrikusio augimo ar vystymosi, genitalijų ir ausų anomalijos, kurtumo ir chuanos atrezijos (užakusių nosies landų) požymių derinys.

TCF4 gene nustatyta NM_001243234:c.1259G>A:p.420R>Q (rs121909121) dnM, lemianti autosominiu dominantiniu būdu paveldimą *Pitt-Hopkins* sindromą (MIM #610954). *TCF4* genas koduoja spiralės-kilpos-spiralės struktūros transkripcijos veiksnį (bHLH), dar žinomą kaip E baltymų šeimos narį. Jis yra evoliuciškai konservatyvus ir ypač svarbus neuronų determinacijai ir diferenciacijai, nes geba prisijungti DNR kaip homodimerus arba heterodimerus [231, 232]. Kai dėl įvykusios mutacijos *TCF4* negali tinkamai sąveikauti su *ASCL1*, audiniams specifiniu HLH baltymu, *PHOX-RET* kelyje sutrinka atskirų smegenų struktūrų vystymasis [233]. Be to, pelių tyrimais buvo parodyta, kad embrionų vystymasis tiesiogiai priklauso nuo *TCF4* baltymo kopijų skaičiaus, o kai jų skaičius nesiekia trijų – pelių embrionai neišgyvena [234]. Šio baltymo raiška didelė žmogaus smegenyse, plaučiuose, širdyje ir kituose raumenyse. *Pitt-Hopkins* sindromą turintiems asmenims būdinga intelektinė negalia, plati burna, išskirtiniai veido bruožai, hiperventiliacija su pertrūkiais, po kurios būna apnėja.

DnM NM_001316337:c.194C>T :p.65T>M (rs28934906) *MECP2* gene lemia *Rett* sindromą (MIM #312750). *MECP2* yra daugiafunkcinis baltymas, kuris gali keisti genų raišką ir medžiagų apykaitą daugeliu būdų: gali aktyvinti AKT/mTOR signalo perdavimo kelią, dalyvauja vykstant alternatyviam genų sukirpimui, dalyvauja mikroRNR ir ilgų nekoduojančių RNR raiškoje, taip pat epigenetiniame chromatinio modifikavimo procese [235–239]. Pagrindinė jo funkcija – DNR grandinėje atpažinti ir prisijungti specifiskai metilintas ir turingas A/T bazių citozino liekanas (5MeCyt). Mutacijos *MECP2* gene lemia specifiskumo prijungti 5MeCyt praradimą ir kartu yra *Rett* sindromo priežastis. *Rett* sindromo fenotipą lemia *MECP2* raiška, nuo kurios priklauso genų, kurie keičia neurobiologinį aktyvumą, tinklų formavimą ir funkciją, epigenetinis reguliavimas. *Rett* sindromą turintiems asmenims (daugiausiai juo serga moterys, nes tai su X chromosoma dominantiniu būdu paveldima liga) būdingas sutrikęs vystymasis nuo šešių iki 18 mėnesių amžiaus, įgytų įgūdžių regresija, kalbos praradimas, stereotipiniai (paprastai rankų) judesiai, mikrocefalija, traukuliai ir, žinoma, intelektinė negalia. *Rett* sindromas klasifikuojamas kaip

neurologinė ir autizmo spektro liga [235, 237, 238]. Dėl *MECP2* geno dozės kompensavimo skirtumų X chromosomoje gali skirtis fenotipinė sindromo raiška.

Visų penkių patvirtintų patogeninių *dnM* atvejais tirtų asmenų fenotipai buvo itin panašūs į publikacijose viešintų asmenų su tapačiais intelektinę negalią lemiančiais sindromais fenotipus.

Apibendrinant reikia pasakyti, kad šiame darbe atlikta *dnM* spektro, pasiskirstymo ir jų genetinio bei epigenetinio konteksto analizė suteikė žinių ir įžvalgų apie Lietuvos genomo genetinę įvairovę ir galimas jos priežastis. Šie duomenys ir rezultatai didina tyrėjų gebėjimus ir kartu palengvina tikrų patogeninių *dnM* atskyrimą nuo toleruotino *dnM* fono ir padeda identifikuoti galimas intelektinės negalios priežastines *dnM*.

IŠVADOS

1. *De novo* taškinėms ir nedidelėms iškritoms-intarpams (iki 20 bp) nustatyti genome, kai jis nuskaitomas skirtingomis technologijomis, tinkamesnis yra atviros prieigos algoritmas *VarScan*, nes jis pasižymėjo didesniu jautrumu, yra tinkamas dirbti su taikininėmis sekomis, egzomu ar visu genomu, analizuoja įvairių genetinių analizatorių (*Illumina*, *SOLID*, *Life/PGM*) duomenis, pagrįstas genomo variantų, kurie atitinka norimus horizontaliojo regionų padengimo, bazių kokybės, varianto alelio dažnio ir statistinio reikšmingumo įverčius, euristiniu / statistiniu nustatymo metodu.

2. a) Lietuvos populiacija, asmenų su IN ir asmenų su IN sibsų grupės *dnM* spektro sudėtimi tarpusavyje nesiskiria (p reikšmė = 0,575) ir yra artimos Jutos gyventojų, kilusių iš Šiaurės ir Vakarų Europos (CEU), Ibadano Jorubos (Nigerija) (YRI), islandų, škotų ir pietų Dakotos (JAV) populiacijose nustatytai *dnM* spektrų sudėčiai: visose grupėse daugiausiai nustatyta C>T (pirmoje grupėje 33 %, antroje – 35 % ir trečioje grupėje 59 %) ir T>C (pirmoje grupėje 28 %, antroje – 27 % ir trečioje grupėje 18,5 %) tipo *dnM*.

b) *DnM* spektro susiformavimo mechanizmų pagrindu visos tirtos grupės yra unikalios:

- Lietuvos populiacijoje nustatytos *dnM*, manoma, susidaro dėl DNR grandinių nesutapimo taisymo mechanizmo klaidų, ϵ polimerazės egzozonukleazės domeno mutacijų ir ultravioletinės spinduliuotės metu susidariusių ciklobutano pirimidino dimerų.

- Asmenų su IN grupėje nustatytos *dnM*, manoma, susidaro dėl DNR grandinių nesutapimo taisymo mechanizmo klaidų, mutacijų *POLD1* gene ir ultravioletinės spinduliuotės.

- Asmenų su IN sibsų grupėje nėra nustatytos *dnM* etiologija nėra žinoma.

3. Visų tirtų grupių apskaičiuotas *de novo* mutacijų dažnis yra reikšmingai didesnis (pirmos grupės p reikšmė = $4,49 \times 10^{-9}$, antros grupės p reikšmė = $5,15 \times 10^{-6}$, trečios grupės p reikšmė = $4,8 \times 10^{-6}$) nei islandų, danų, olandų, kanadiečių ir amerikiečių bendrųjų populiacijų ir autizmo spektro liga sergančiųjų grupės tyrimuose nustatytas. Tai nulėmė analizuotos genomo dalies – egzomo genetinė struktūra.

- *De novo* vieno nukleotido ir *de novo* iškritų-intarpų dažnis Lietuvos populiacijoje atitinkamai yra $2,74 \times 10^{-8}$ ir $1,77 \times 10^{-8}$ į poziciją vienai kartai; asmenų su IN grupėje atitinkamai $3,05 \times 10^{-8}$ ir $1,74 \times 10^{-7}$ į poziciją vienai kartai ir asmenų su IN sibsų – atitinkamai $7,19 \times 10^{-8}$ ir $1,08 \times 10^{-6}$ į poziciją vienai kartai.

4. *De novo* mutacijos egzome susidaro į promotorius panašiuose, transkripciškai aktyviuose genomo regionuose neatsižvelgiant į DNR sekos konservatyvumą.

- Lietuvos populiacijos grupės tiriamųjų egzomuose *de novo* mutacijų intensyvumui turi įtakos (~68–94 %) DNRazei 1 ypač jautrūs regionai, CpG kontekstas, regionų raiškos lygiai.
- Intelektinės negalios asmenų grupėje *de novo* mutacijų intensyvumą nežymiai (13 %) įtakojo egzomo regionų raiškos lygiai.

5. Intelektinės negalios asmenų egzomuose patogeninių mutacijų yra reikšmingai daugiau (*p* reikšmė = 0,02) nei Lietuvos populiacijos ir intelektinės negalios asmenų sibsų grupėse, todėl tikėtina, kad unikalios *de novo* mutacijos (ypač koduojančiose genomo srityse) gali turėti žalingą poveikį asmens fenotipui ir būti intelektinės negalios priežastis.

- 13,5 % tiriamųjų, turinčių intelektinę negalią, buvo nustatytos penkios patogeninės *de novo* mutacijos *ARID1B*, *PAC1*, *TCF4*, *CHD7* ir *MECP2* genuose, iš kurių trys (*ARID1B*, *CHD7* ir *PAC1* genuose) buvo iki šiol literatūroje neaprašytos, atitinkamai lemiančios *Coffin-Siris*, *CHARGE* ir autosominiu dominantiniu būdu paveldimą intelektinę negalią.

6. Lietuvos populiacijai būdingas „veidrodžio atspindžio efektas“, kai didelį neigiamą santykinį tinkamumą ($S < -16$) turinčios aminorūgštys įvykus *dnM* virsta neutraliomis (0) ar mažą (iki -5) santykinį tinkamumą turinčiomis aminorūgštimis, patvirtinantis neutralumo teoriją.

LITERATŪROS SARAŠAS

1. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, et al. Rate of de novo mutations and the importance of father-s age to disease risk. *Nature*. 2012.
2. Jónsson H, Sulem P, Kehr B, Kristmundsdottir S, Zink F, Hjartarson E, et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature*. 2017.
3. Wong WSW, Solomon BD, Bodian DL, Kothiyal P, Eley G, Huddleston KC, et al. New observations on maternal age effect on germline de novo mutations. *Nat Commun*. 2016.
4. Acuna-Hidalgo R, Veltman JA, Hoischen A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biology*. 2016.
5. Besenbacher S, Liu S, Izarzugaza JMG, Grove J, Belling K, Bork-Jensen J, et al. Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. *Nat Commun*. 2015.
6. MW Nachman. Haldane and the first estimates of the human mutation rate. *Journal of Genetics*. 2008.
7. Kondrashov AS. Direct estimates of human per nucleotide mutation rates at 20 loci causing mendelian diseases. *Hum Mutat*. 2003.
8. Michaelson JJ, Jian M, Liu G, Greer D, Bhandari A, Wu W, et al. Whole-Genome Sequencing in Autism Identifies Hot Spots for De Novo Germline Mutation. *Cell*. 2012.
9. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001.
10. Altshuler DM, Gibbs RA, Peltonen L, Schaffner SF, Yu F, Dermitzakis E, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010.
11. Abecasis GR, Auton A, Brooks L, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012.
12. Gibson G. Rare and common variants: Twenty arguments. *Nature Reviews Genetics*. 2012.
13. Sherman SL, Petersen MB, Freeman SB, Hersey J, Pettay D, Taft L, et al. Non-disjunction of chromosome 21 in maternal meiosis I: Evidence for a maternal age-dependent mechanism involving reduced recombination. *Hum Mol Genet*. 1994.
14. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009.
15. Marco-Sola S, Sammeth M, Guigó R, Ribeca P. The GEM mapper: Fast, accurate and versatile alignment by filtration. *Nat Methods*. 2012.

16. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 2008.
17. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* 2011.
18. Katju V, Bergthorsson U. Old trade, new tricks: Insights into the spontaneous mutation process from the partnering of classical mutation accumulation experiments with high-throughput genomic approaches. *Genome Biol Evol.* 2019.
19. Halligan DL, Keightley PD. Spontaneous Mutation Accumulation Studies in Evolutionary Genetics. *Annu Rev Ecol Evol Syst.* 2009.
20. Uktverytė I. Lietuvos etnolingvistinių grupių genetinės struktūros analizė remiantis informatyviais genomo žymenimis: daktaro disertacija. Vilnius: Vilniaus universitetas. Prieiga per eLABa – nacionalinė Lietuvos akademinė elektroninė biblioteka; 2014.
21. Rančelis T. Patogeninių genomo variantų ir jų genų, lemiančių autosominės recesyvias ligas, įvairovės analizė, panaudojant viso egzomo sekoskaitą: daktaro disertacija. Vilnius: Vilniaus universitetas. Prieiga per eLABa – nacionalinė Lietuvos akademinė elektroninė biblioteka; 2016.
22. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, Wilkie AOM, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet.* 2014.
23. Boomsma DI, Wijmenga C, Slagboom EP, Swertz MA, Karssen LC, Abdellaoui A, et al. The Genome of the Netherlands: Design, and project goals. *Eur J Hum Genet.* 2014.
24. Clarke L, Zheng-Bradley X, Smith R, Kulesha E, Xiao C, Toneva I, et al. The 1000 Genomes Pproject: Data management and community access. *Nature Methods.* 2012.
25. Ségurel L, Wyman MJ, Przeworski M. Determinants of Mutation Rate Variation in the Human Germline. *Annu Rev Genomics Hum Genet.* 2014.
26. Korona DA, Lecompte KG, Pursell ZF. The high fidelity and unique error signature of human DNA polymerase ϵ . *Nucleic Acids Res.* 2011.
27. Schmitt MW, Matsumoto Y, Loeb LA. High fidelity and lesion bypass capability of human DNA polymerase δ . *Biochimie.* 2009.
28. Kunkel TA, Erie DA. Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu Rev Genet.* 2015.
29. Haldane JBS. The rate of spontaneous mutation of a human gene. *J Genet.* 1935.
30. Roach JC, Glusman G, Smit AFA, Huff CD, Hubley R, Shannon PT, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* (80-). 2010.
31. Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S, et al. Timing, rates and spectra of human germline mutation. *Nat*

Genet. 2016.

32. Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, Renkens I, et al. Genome-wide patterns and properties of de novo mutations in humans. *Nat Genet.* 2015.

33. Goldmann JM, Wong WSW, Pinelli M, Farrah T, Bodian D, Stittrich AB, et al. Parent-of-origin-specific signatures of de novo mutations. *Nat Genet.* 2016.

34. Gilissen C, Hehir-Kwa JY, Thung DT, Van De Vorst M, Van Bon BWM, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature.* 2014.

35. B.J. O, L. V, S. G, E. K, N. K, B.P. C, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature.* 2012.

36. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature.* 2012.

37. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature.* 2012.

38. Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jørgensen ME, et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science (80-).* 2015.

39. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science (80-).* 2010.

40. Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, et al. Genetic Signatures of Strong Recent Positive Selection at the Lactase Gene. *Am J Hum Genet.* 2004. 41. Fu W, Akey JM. Selection and Adaptation in the Human Genome. *Annu Rev Genomics Hum Genet.* 2013.

42. Hurst LD. Fundamental concepts in genetics: Genetics and the understanding of selection. *Nature Reviews Genetics.* 2009.

43. Francioli LC. Inherited and de novo variation in human genomes. 2015.

44. Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, et al. A large genome center's improvements to the Illumina sequencing system. *Nat Methods.* 2008.

45. Smith DR, Quinlan AR, Peckham HE, Makowsky K, Tao W, Woolf B, et al. Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res.* 2008.

46. Depristo MA, Banks E, Poplin R, Garimella K V., Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011.

47. H L, B H, A W, T F, J R, N H, et al. The Sequence Alignment/Map

format and SAMtools. *Bioinformatics*. 2009.

48. Maki H. Origins of Spontaneous Mutations: Specificity and Directionality of Base-Substitution, Frameshift, and Sequence-Substitution Mutageneses. *Annu Rev Genet*. 2002.

49. Lindahl T, Wood RD. Quality control by DNA repair. *Science*. 1999.

50. Shendure J, Akey JM. The origins, determinants, and consequences of human mutations. *Science*. 2015.

51. Russell J. *iGenetics A molecular approach*. 3rd edition. Pearson Education, Inc; 2010.

52. Crow JF. The origins, patterns and implications of human spontaneous mutation. *Nature Reviews Genetics*. 2000.

53. Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov G V., Mirkin SM, Sunyaev SR. Human mutation rate associated with DNA replication timing. *Nat Genet*. 2009.

54. Koren A, Polak P, Nemesh J, Michaelson JJ, Sebat J, Sunyaev SR, et al. Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am J Hum Genet*. 2012.

55. Green P, Ewing B, Miller W, Thomas PJ, Thomas J, Touchman J, et al. Transcription-associated mutational asymmetry in mammalian evolution. *Nat Genet*. 2003.

56. Chen CL, Rappailles A, Duquenne L, Huvet M, Guilbaud G, Farinelli L, et al. Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res*. 2010.

57. Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, Hess JM, et al. Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. *Cell*. 2016.

58. Schuster-Böckler B, Lehner B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature*. 2012.

59. Fryxell KJ, Zuckerkandl E. Cytosine deamination plays a primary role in the evolution of mammalian isochores. *Mol Biol Evol*. 2000.

60. Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet*. 2012.

61. Petljak M, Alexandrov LB. Understanding mutagenesis through delineation of mutational signatures in human cancer. *Carcinogenesis*. 2016.

62. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin A V, et al. Signatures of mutational processes in human cancer. *Nature*. 2013.

63. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. *Nat Genet*. 2015.

64. Chan K, Resnick MA, Gordenin DA. The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis. *DNA*

Repair (Amst). 2013.

65. Roberts SA, Sterling J, Thompson C, Harris S, Mav D, Shah R, et al. Clustered Mutations in Yeast and in Human Cancers Can Arise from Damaged Long Single-Strand DNA Regions. *Mol Cell*. 2012.

66. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*. 2012.

67. Sakofsky CJ, Roberts SA, Malc E, Mieczkowski PA, Resnick MA, Gordenin DA, et al. Break-induced replication is a source of mutation clusters underlying kataegis. *Cell Rep*. 2014.

68. Carvalho CMB, Pehlivan D, Ramocki MB, Fang P, Alleva B, Franco LM, et al. Replicative mechanisms for CNV formation are error prone. *Nat Genet*. 2013.

69. Stults DM, Killen MW, Pierce HH, Pierce AJ. Genomic architecture and inheritance of human ribosomal RNA gene clusters. *Genome Res*. 2008.

70. Neumann R, Lawson VE, Jeffreys AJ. Dynamics and processes of copy number instability in human α -globin genes. *Proc Natl Acad Sci*. 2010.

71. Weber JL, Wong C. Mutation of human short tandem repeats. *Hum Mol Genet*. 1993.

72. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte R V, Schwartz S, et al. Suppl to: Recent segmental duplications in the human genome. *Science* (80-). 2002.

73. Sun JX, Helgason A, Masson G, Ebenesersdóttir SS, Li H, Mallick S, et al. A direct characterization of human mutation based on microsatellites. *Nat Genet*. 2012.

74. Paul C, Robaire B. Ageing of the male germ line. *Nature Reviews Urology*. 2013.

75. Goriely A. Decoding germline de novo point mutations. *Nature Genetics*. 2016.

76. Goriely A, Wilkie AOM. Paternal age effect mutations and selfish spermatogonial selection: Causes and consequences for human disease. *American Journal of Human Genetics*. 2012.

77. Uchimura A, Higuchi M, Minakuchi Y, Ohno M, Toyoda A, Fujiyama A, et al. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Res*. 2015.

78. Goriely A, Wilkie AOM. Missing heritability: paternal age effect mutations and selfish spermatogonia. *Nat Rev Genet*. 2010.

79. Lappalainen T, Sammeth M, Friedländer MR, 'T Hoen PAC, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013.

80. Giannoulatou E, McVean G, Taylor IB, McGowan SJ, Maher GJ, Iqbal Z, et al. Contributions of intrinsic mutation rate and selfish selection to

levels of de novo HRAS mutations in the paternal germline. *Proc Natl Acad Sci*. 2013.

81. Arnheim N, Calabrese P. Germline Stem Cell Competition, Mutation Hot Spots, Genetic Disorders, and Older Fathers. *Annu Rev Genomics Hum Genet*. 2016.

82. Giannoulatou E, McVean G, Taylor IB, McGowan SJ, Maher GJ, Iqbal Z, et al. Contributions of intrinsic mutation rate and selfish selection to levels of de novo HRAS mutations in the paternal germline. *Proc Natl Acad Sci U S A*. 2013.

83. Maher GJ, McGowan SJ, Giannoulatou E, Verrill C, Goriely A, Wilkie AOM. Visualizing the origins of selfish de novo mutations in individual seminiferous tubules of human testes. *Proc Natl Acad Sci*. 2016.

84. Yoon SR, Choi SK, Eboeime J, Gelb BD, Calabrese P, Arnheim N. Age-dependent germline mosaicism of the most common noonan syndrome mutation shows the signature of germline selection. *Am J Hum Genet*. 2013.

85. Goriely A, McVean GAT, Røjmyr M, Ingemarsson B, Wilkie AOM. Evidence for selective advantage of pathogenic FGFR2 mutations in the male germ line. *Science* (80-). 2003.

86. Goriely A, Hansen RMS, Taylor IB, Olesen IA, Jacobsen GK, McGowan SJ, et al. Activating mutations in FGFR3 and HRAS reveal a shared genetic origin for congenital disorders and testicular tumors. *Nat Genet*. 2009.

87. Choi SK, Yoon SR, Calabrese P, Arnheim N. Positive selection for new disease mutations in the human germline: Evidence from the heritable cancer syndrome multiple endocrine neoplasia type 2B. *PLoS Genet*. 2012.

88. Goriely A, McGrath JJ, Hultman CM, Wilkie AOM, Malaspina D. “Selfish spermatogonial selection”: A novel mechanism for the association between advanced paternal age and neurodevelopmental disorders. *American Journal of Psychiatry*. 2013.

89. Fulton N, Martins Da Silva SJ, Bayne RAL, Anderson RA. Germ cell proliferation and apoptosis in the developing human ovary. *J Clin Endocrinol Metab*. 2005.

90. Acuna-Hidalgo R, Bo T, Kwint MP, Van De Vorst M, Pinelli M, Veltman JA, et al. Post-zygotic Point Mutations Are an Underrecognized Source of de Novo Genomic Variation. *Am J Hum Genet*. 2015.

91. Ju YS, Martincorena I, Gerstung M, Petljak M, Alexandrov LB, Rahbari R, et al. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature*. 2017.

92. Smith TB, Dun MD, Smith ND, Curry BJ, Connaughton HS, Aitken RJ. The presence of a truncated base excision repair pathway in human spermatozoa that is mediated by OGG1. *J Cell Sci*. 2013.

93. Dobson AT, Raja R, Abeyta MJ, Taylor T, Shen S, Haqq C, et al. The unique transcriptome through day 3 of human preimplantation development.

Hum Mol Genet. 2004.

94. Zhang P, Zuchelli M, Bruce S, Hambiliki F, Stavreus-Evers A, Levkov L, et al. Transcriptome profiling of human pre-implantation development. *PLoS One*. 2009.

95. Titus S, Li F, Stobezki R, Akula K, Unsal E, Jeong K, et al. Impairment of BRCA1-related DNA double-strand break repair leads to ovarian aging in mice and humans. *Sci Transl Med*. 2013.

96. Wei H, Liu X, Yuan J, Li L, Zhang D, Guo X, et al. Age-Specific Gene Expression Profiles of Rhesus Monkey Ovaries Detected by Microarray Analysis. *Biomed Res Int*. 2015.

97. Hassold T, Hunt P. To err (meiotically) is human: The genesis of human aneuploidy. *Nature Reviews Genetics*. 2001.

98. Gao Z, Moorjani P, Sasani TA, Pedersen BS, Quinlan AR, Jorde LB, et al. Overlooked roles of DNA damage and maternal age in generating human germline mutations. *Proc Natl Acad Sci*. 2019.

99. Huang AY, Xu X, Ye AY, Wu Q, Yan L, Zhao B, et al. Postzygotic single-nucleotide mosaicism in whole-genome sequences of clinically unremarkable individuals. *Cell Res*. 2014.

100. Youssoufian H, Pyeritz RE. Human genetics and disease: Mechanisms and consequences of somatic mosaicism in humans. *Nat Rev Genet*. 2002.

101. Tomasetti C, Vogelstein B, Parmigiani G. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc Natl Acad Sci*. 2013.

102. Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* (80-). 2015.

103. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhom SF, et al. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *N Engl J Med*. 2014.

104. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman P V, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*. 2014.

105. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med*. 2014.

106. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* (80-). 2015.

107. Laurie CCC a, Rice K, Doheny KKF, Zelnick LR, McHugh CP, Ling H, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet*. 2012.

108. Narasimhan VM, Rahbari R, Scally A, Wuster A, Mason D, Trust

W. A direct multi-generational estimate of the human mutation rate from autozygous segments seen in thousands of parentally related individuals. *bioRxiv*. 2016.

109. Lipson M, Loh PR, Sankararaman S, Patterson N, Berger B, Reich D. Calibrating the Human Mutation Rate via Ancestral Recombination Density in Diploid Genomes. *PLoS Genet*. 2015.

110. Palamara PF, Francioli LC, Wilton PR, Genovese G, Gusev A, Finucane HK, et al. Leveraging Distant Relatedness to Quantify Human Mutation and Gene-Conversion Rates. *Am J Hum Genet*. 2015.

111. Scally A, Durbin R. Revising the human mutation rate: Implications for understanding human evolution. *Nat Rev Genet*. 2012.

112. Moorjani P, Amorim CEG, Arndt PF, Przeworski M. Variation in the molecular clock of primates. *Proc Natl Acad Sci U S A*. 2016.

113. CROW JF. Age and Sex Effects on Human Mutation Rates: An Old Problem with New Complexities. *J Radiat Res*. 2006.

114. Scally A. The mutation rate in human evolution and demographic inference. *Current Opinion in Genetics and Development*. 2016.

115. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet*. 2014.

116. Begun DR. Fossil record of miocene hominoids. In: *Handbook of Paleoanthropology, Second Edition*. 2015.

117. Hublin JJ, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, et al. New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. *Nature*. 2017.

118. Ciochon RL, White JL. Walter C. Hartwig (ed.). 2002. *The Primate Fossil Record*. Cambridge University Press, Cambridge, 503 p., cloth, ISBN 0-521-66315-6. *J Paleontol*. 2004.

119. Steiper ME, Young NM. Timing primate evolution: Lessons from the discordance between molecular and paleontological estimates. *Evol Anthropol*. 2008.

120. Jensen-Seaman MI H-BK. *Molecular Clocks: Determining the Age of the Human– Chimpanzee Divergence*. 2013.

121. Mailund T, Munch K, Schierup MH. *Lineage Sorting in Apes*. *Annu Rev Genet*. 2014.

122. Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, et al. Insights into hominid evolution from the gorilla genome sequence. *Nature*. 2012.

123. Wall JD. Estimating ancestral population sizes and divergence times. *Genetics*. 2003.

124. Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. *Genetics*. 2000.

125. Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. Genetic evidence for complex speciation of humans and chimpanzees. *Nature*. 2006.

126. Perelman P, Johnson WE, Roos C, Seuánez HN, Horvath JE, Moreira MAM, et al. A molecular phylogeny of living primates. *PLoS Genet.* 2011.
127. Li WH, Tanimura M. The molecular clock runs more slowly in man than in apes and monkeys. *Nature.* 1987.
128. Moorjani P, Gao Z, Przeworski M. Human Germline Mutation and the Erratic Evolutionary Clock. *PLoS Biol.* 2016.
129. Conrad DF, Keebler JEM, Depristo MA, Lindsay SJ, Zhang Y, Casals F, et al. Variation in genome-wide mutation rates within and between human families. In: *Nature Genetics.* 2011.
130. Campbell CD, Chong JX, Malig M, Ko A, Dumont BL, Han L, et al. Estimating the human mutation rate using autozygosity in a founder population. *Nat Genet.* 2012.
131. Jiang YH, Yuen RKC, Jin X, Wang M, Chen N, Wu X, et al. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am J Hum Genet.* 2013.
132. Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, et al. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *American Journal of Human Genetics.* 2015.
133. Veltman JA, Brunner HG. De novo mutations in human genetic disease. *Nature Reviews Genetics.* 2012.
134. American Psychiatric Association. DSM-5 Development. *J Okla Dent Assoc.* 2013.
135. Reiss AL. Childhood developmental disorders: An academic and clinical convergence point for psychiatry, neurology, psychology and pediatrics. *J Child Psychol Psychiatry Allied Discip.* 2009.
136. Rapoport JL, Giedd JN, Gogtay N. Neurodevelopmental model of schizophrenia: Update 2012. *Molecular Psychiatry.* 2012.
137. Korkmaz B. Theory of mind and neurodevelopmental disorders of childhood. *Pediatr Res.* 2011.
138. Sugranyes G, Kyriakopoulos M, Corrigall R, Taylor E, Frangou S. Autism spectrum disorders and schizophrenia: Meta-analysis of the neural correlates of social cognition. *PLoS ONE.* 2011.
139. King BH, Lord C. Is schizophrenia on the autism spectrum? *Brain Research.* 2011.
140. Lugnégård T, Unenge Hallerbäck M, Hjärthag F, Gillberg C. Social cognition impairments in Asperger syndrome and schizophrenia. *Schizophr Res.* 2013.
141. Owoeye O, Kingston T, Scully PJ, Baldwin P, Browne D, Kinsella A, et al. Epidemiological and clinical characterization following a first psychotic episode in major depressive disorder: Comparisons with schizophrenia and bipolar i disorder in the cavan-monaghan first episode psychosis study (CAMFEPS). *Schizophr Bull.* 2013.
142. Domschke K. Clinical and molecular genetics of psychotic

depression. *Schizophr Bull.* 2013.

143. Vissers LELM, De Vries BBA, Veltman JA. Genomic microarrays in mental retardation: From copy number variation to gene, from research to diagnosis. *Journal of Medical Genetics.* 2010.

144. Koolen DA, Pfundt R, De Leeuw N, Hehir-Kwa JY, Nillesen WM, Neefs I, et al. Genomic microarrays in mental retardation: A practical workflow for diagnostic applications. *Human Mutation.* 2009.

145. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of de novo copy number mutations with autism. *Science* (80-). 2007.

146. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al. Structural Variation of Chromosomes in Autism Spectrum Disorder. *Am J Hum Genet.* 2008.

147. Xu B, Roos JL, Levy S, Van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet.* 2008.

148. Vermeesch JR, Balikova I, Schrandt-Stumpel C, Fryns J-P, Devriendt K. The causality of de novo copy number variants is overestimated. *Eur J Hum Genet.* 2011.

149. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A, et al. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet.* 2010.

150. Hamdan FF, Gauthier J, Spiegelman D, Noreau A, Yang Y, Pellerin S, et al. Mutations in SYNGAP1 in Autosomal Nonsyndromic Mental Retardation . *N Engl J Med.* 2009.

151. Gauthier J, Champagne N, Lafrenière RG, Xiong L, Spiegelman D, Brustein E, et al. De novo mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia . *Proc Natl Acad Sci.* 2010.

152. Turner TN, Yi Q, Krumm N, Huddleston J, Hoekzema K, Stessman HAF, et al. NAR Breakthrough Article denovo-db: A compendium of human de novo variants. *Nucleic Acids Res.* 2017.

153. Marrus N, Hall L. Intellectual Disability and Language Disorder. *Child and Adolescent Psychiatric Clinics of North America.* 2017.

154. Hagberg B, Kyllerman M. Epidemiology of mental retardation—A Swedish survey. *Brain and Development.* 1983.

155. Preikšaitienė E, Kasnauskienė J, Utkus A K V. No Title. *Neurol Semin.* 2012;16:283–288.

156. Lelieveld SH, Reijnders MRF, Pfundt R, Yntema HG, Kamsteeg E-J, de Vries P, et al. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat Neurosci.* 2016.

157. Deciphering Developmental Disorders, McRae JF, Clayton S, Fitzgerald TW, Kaplanis J, Prigmore E, et al. Prevalence and architecture of

de novo mutations in developmental disorders. *Nature*. 2017.

158. Short PJ, McRae JF, Gallone G, Sifrim A, Won H, Geschwind DH, et al. De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature*. 2018.

159. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM). CoDAS. 2013.

160. Fombonne E. Epidemiology of pervasive developmental disorders. *Pediatric Research*. 2009.

161. Loomes R, Hull L, Mandy WPL. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2017.

162. Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg IC, Jakobsson G, et al. A Twin Study of Autism in Denmark, Finland, Iceland, Norway and Sweden. *J Child Psychol Psychiatry*. 1989.

163. Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, et al. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med*. 1995.

164. Sandin S, Lichtenstein, P, Kuja-Halkola R, Hultman C, Larsson H RA. The heritability of autism spectrum disorder analysis method. *JAMA Pediatr*. 2017;318:1182–4.

165. Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L, et al. Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics*. 2011.

166. Risch N, Hoffmann TJ, Anderson M, Croen LA, Grether JK, Windham GC. Familial recurrence of autism spectrum disorder: Evaluating genetic and environmental contributions. *Am J Psychiatry*. 2014.

167. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism*. 2017.

168. Grove J, Ripke S, Als TD, Mattheisen M, Walters R, Won H, et al. Common risk variants identified in autism spectrum disorder. *bioRxiv*. 2017.

169. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*. 2014.

170. Polimanti R, Gelernter J. Widespread signatures of positive selection in common risk alleles associated to autism spectrum disorder. *PLoS Genet*. 2017.

171. Uddin M, Tammimies K, Pellicchia G, Alipanahi B, Hu P, Wang Z, et al. Brain-expressed exons under purifying selection are enriched for de novo mutations in autism spectrum disorder. *Nat Genet*. 2014.

172. J K. Discovery and characterization of 102 genes associated with autism from exome sequencing of 37,269 individuals. 2018.

173. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et

al. Most genetic risk for autism resides with common variation. *Nat Genet.* 2014.

174. Gilman SR, Chang J, Xu B, Bawa TS, Gogos JA, Karayiorgou M, et al. Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat Neurosci.* 2012.

175. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet.* 2014.

176. Jamain S, Quach H, Betancur C, Råstam M, Colineaux C, Gillberg C, et al. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet.* 2003.

177. Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, et al. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet.* 2007.

178. Consortium AGP, Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, et al. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet.* 2007.

179. Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, Ikeda M, et al. A Common Genetic Variant in the Neurexin Superfamily Member CNTNAP2 Increases Familial Risk of Autism. *Am J Hum Genet.* 2008.

180. van Bokhoven H. Genetic and Epigenetic Networks in Intellectual Disabilities. *Annu Rev Genet.* 2011.

181. Jakovcevski, M; Akbarian S. Epigenetic mechanisms in neurodevelopmental and neurodegenerative disease. *Nat Med.* 2013.

182. Ronan JL, Wu W, Crabtree GR. From neural development to cognition: Unexpected roles for chromatin. *Nature Reviews Genetics.* 2013.

183. Tsurusaki I, Okamoto, N., Ohashi H, Kosho T, Imai Y H-KY al. Mutations affecting components of the SWI/SNF complex cause coffin–siris syndrome. *Nat Genet.* 2012;44:376–8.

184. Van Houdt JKJ, Nowakowska BA, Sousa SB, Van Schaik BDC, Seuntjens E, Avonce N, et al. Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. *Nat Genet.* 2012.

185. Vissers LELM, Van Ravenswaaij CMA, Admiraal R, Hurst JA, De Vries BBA, Janssen IM, et al. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet.* 2004.

186. Roelfsema JH, White SJ, Ariyürek Y, Bartholdi D, Niedrist D, Papadia F, et al. Genetic Heterogeneity in Rubinstein-Taybi Syndrome: Mutations in Both the CBP and EP300 Genes Cause Disease. *Am J Hum Genet.* 2005.

187. Oliver C, Woodcock K, Adams D. The importance of aetiology of intellectual disability. In: *Learning disability: a life cycle approach to valuing people.* 2010.

188. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010.
189. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012.
190. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform*. 2013.
191. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010.
192. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2016.
193. Bernstein B, Birney E, Dunham I, Green E, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012.
194. GTEx Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017.
195. The GTEx Consortium, Ardlie KG, Deluca DS, Segre A V., Sullivan TJ, Young TR, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* (80-). 2015.
196. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017.
197. Warden CD, Adamson AW, Neuhausen SL, Wu X. Detailed comparison of two popular variant calling packages for exome and targeted exon studies. *PeerJ*. 2014.
198. Zuckerkandl E, Pauling L. Molecular Disease, Evolution and genic heterogeneity. In: *Horizons in biochemistry*. 1962.
199. Kumar S. Molecular clocks: Four decades of evolution. *Nature Reviews Genetics*. 2005.
200. Bromham L, Penny D. The modern molecular clock. *Nature Reviews Genetics*. 2003.
201. Takahata N, Kimura M. The Neutral Theory of Molecular Evolution. *Princ Med Biol*. 1994.
202. JUKES TH, CANTOR CR. Evolution of Protein Molecules. In: *Mammalian Protein Metabolism*. 2013.
203. Birky CW, Walsh JB. Effects of linkage on rates of molecular evolution. *Proc Natl Acad Sci*. 1988.
204. Kumar S, Subramanian S. Mutation rates in mammalian genomes.

Proc Natl Acad Sci. 2002.

205. Kondrashov FA, Kondrashov AS. Measurements of spontaneous rates of mutations in the recent past and the near future. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2010.

206. Subramanian S, Kumar S. Neutral substitutions occur at a faster rate in exons than in noncoding DNA in primate genomes. *Genome Res*. 2003.

207. Schmidt S, Gerasimova A, Kondrashov FA, Adzuhbei IA, Kondrashov AS, Sunyaev S. Hypermutable non-synonymous sites are under stronger negative selection. *PLoS Genet*. 2008.

208. Branciamore S, Chen Z-X, Riggs AD, Rodin SN. CpG island clusters and pro-epigenetic selection for CpGs in protein-coding exons of HOX and other transcription factors. *Proc Natl Acad Sci*. 2010.

209. Sakabe NJ, Savic D, Nobrega MA. Transcriptional enhancers in development and disease. *Genome Biology*. 2012.

210. Weirauch MT, Hughes TR. Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. *Trends in Genetics*. 2010.

211. Villar D, Flicek P, Odom DT. Evolution of transcription factor binding in metazoans-mechanisms and functional implications. *Nature Reviews Genetics*. 2014.

212. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature*. 2014.

213. Necsulea A, Kaessmann H. Evolutionary dynamics of coding and non-coding transcriptomes. *Nature Reviews Genetics*. 2014.

214. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, et al. Enhancer evolution across 20 mammalian species. *Cell*. 2015.

215. Cotney J, Leng J, Yin J, Reilly SK, Demare LE, Emera D, et al. XThe evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell*. 2013.

216. Eyre-Walker A, Keightley PD, Smith NGC, Gaffney D. Quantifying the slightly deleterious mutation model of molecular evolution. *Mol Biol Evol*. 2002.

217. Fay JC, Wyckoff GJ, Wu CI. Positive and negative selection on the human genome. *Genetics*. 2001.

218. Uricchio LH, Petrov DA, Enard D. Exploiting selection at linked sites to infer the rate and strength of adaptation. *Nat Ecol Evol*. 2019.

219. Szamecz B, Boross G, Kalapis D, Kovács K, Fekete G, Farkas Z, et al. The Genomic Landscape of Compensatory Evolution. *PLoS Biol*. 2014.

220. Bartha I, Rausell A, McLaren PJ, Mohammadi P, Tardaguila M, Chaturvedi N, et al. The Characteristics of Heterozygous Protein Truncating Variants in the Human Genome. *PLoS Comput Biol*. 2015.

221. Anna A, Monika G. Splicing mutations in human genetic disorders:

- examples, detection, and confirmation. *Journal of Applied Genetics*. 2018.
222. Li XS, Trojer P, Matsumura T, Treisman JE, Tanese N. Mammalian SWI/SNF-A Subunit BAF250/ARID1 Is an E3 Ubiquitin Ligase That Targets Histone H2B. *Mol Cell Biol*. 2010.
223. Yan Z, Wang Z, Sharova L, Sharov AA, Ling C, Piao Y, et al. BAF250B-Associated SWI/SNF Chromatin-Remodeling Complex Is Required to Maintain Undifferentiated Mouse Embryonic Stem Cells. *Stem Cells*. 2008.
224. Boyer LA, Tong IL, Cole MF, Johnstone SE, Levine SS, Zucker JP, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*. 2005.
225. Flores-Alcantar A, Gonzalez-Sandoval A, Escalante-Alcalde D, Lomelí H. Dynamics of expression of ARID1A and ARID1B subunits in mouse embryos and in cells during the cell cycle. *Cell Tissue Res*. 2011.
226. Gadzicki D, Döcker B, Schubach M, Menzel M, Schmorl B SF et al. Expanding the phenotype of a recurrent de novo variant in PACS1 causing intellectual disability. *Clin Genet*. 2015;88:300–2.
227. Wan L, Molloy SS, Thomas L, Liu G, Xiang Y, Rybak SL, et al. PACS-1 defines a novel gene family of cytosolic sorting proteins required for trans-Golgi network localization. *Cell*. 1998.
228. Youker RT, Shinde U, Day R, Thomas G. At the crossroads of homeostasis and disease: roles of the PACS proteins in membrane traffic and apoptosis. *Biochem J*. 2009.
229. Schuurs-Hoeijmakers JHM, Oh EC, Vissers LELM, Swinkels MEM, Gilissen C, Willemsen MA, et al. Recurrent de novo mutations in PACS1 cause defective cranial-neural-crest migration and define a recognizable intellectual-disability syndrome. *Am J Hum Genet*. 2012.
230. Sanlaville D, Etchevers HC, Gonzales M, Martinovic J, Clément-Ziza M, Delezoide AL, et al. Phenotypic spectrum of CHARGE syndrome in fetuses with CHD7 truncating mutations correlates with expression during human development. *J Med Genet*. 2006.
231. Kageyama R, Nakanishi S. Helix-loop-helix factors in growth and differentiation of the vertebrate nervous system. *Curr Opin Genet Dev*. 1997.
232. Soosaar A, Chiaramello A, Zuber MX, Neuman T. Expression of basic-helix-loop-helix transcription factor ME2 during brain development and in the regions of neuronal plasticity in the adult brain. *Mol Brain Res*. 1994.
233. Brockschmidt A, Todt U, Ryu S, Hoischen A, Landwehr C, Birnbaum S, et al. Severe mental retardation with breathing abnormalities (Pitt - Hopkins syndrome) is caused by haploinsufficiency of the neuronal bHLH transcription factor TCF4. *Hum Mol Genet*. 2007.
234. Zhuang Y, Cheng P, Weintraub H. B-lymphocyte development is regulated by the combined dosage of three basic helix-loop-helix genes, E2A,

E2-2, and HEB. *Mol Cell Biol*. 1996.

235. Schwartzman JS, Velloso R de L, D'Antino MEF, Santos S. The eye-tracking of social stimuli in patients with Rett syndrome and autism spectrum disorders: a pilot study. *Arq Neuropsiquiatr*. 2017.

236. Ricciardi S, Boggio EM, Grosso S, Lonetti G, Forlani G, Stefanelli G, et al. Reduced AKT/mTOR signaling and protein synthesis dysregulation in a Rett syndrome animal model. *Hum Mol Genet*. 2011.

237. Gonzales ML, Adams S, Dunaway KW, LaSalle JM. Phosphorylation of Distinct Sites in MeCP2 Modifies Cofactor Associations and the Dynamics of Transcriptional Regulation. *Mol Cell Biol*. 2012.

238. Long SW, Ooi JYY, Yau PM, Jones PL. A brain-derived MeCP2 complex supports a role for MeCP2 in RNA processing. *Biosci Rep*. 2010.

239. Young JI, Hong EP, Castle JC, Crespo-Barreto J, Bowman AB, Rose MF, et al. Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc Natl Acad Sci U S A*. 2005.

PRIEDAI

1 PRIEDAS

Lifescop programos parametrai

Parametrai	Įverčiai
<i>SAET</i>	
<i>Update Quality Values</i>	TRUE
<i>Trusted Quality Values</i>	25
<i>Support Votes</i>	3
<i>Maximum Correction per Read</i>	0
<i>K-mer size</i>	0
<i>Genome Length</i>	30 000 000
<i>Position of Error Inflation Point</i>	0
<i>Disable Random Sampling for Large Data</i>	FALSE
<i>Trusted Frequency</i>	0
<i>On Target Ratio</i>	0,5
<i>Number of Recursive Runs</i>	1
<i>Maximum Correction per Read</i>	0
<i>K-mer size</i>	0
<i>Genome Length</i>	30 000 000
<i>Position of Error Inflation Point</i>	0
<i>Disable Random Sampling for Large Data</i>	FALSE
<i>Trusted Frequency</i>	0
<i>On Target Ratio</i>	0,5
<i>Number of Recursive Runs</i>	1
<i>Fragment Mapping</i>	
<i>Mapping QV Threshold</i>	0
<i>Create Unmapped BAM Files</i>	FALSE
<i>Reference Weight</i>	8
<i>Second Map Gapped Algorithm</i>	GLOBAL
<i>Base Quality Filter Threshold</i>	10
<i>Map in Base Space</i>	FALSE
<i>Add Color Sequence</i>	TRUE
<i>BAM soft clip</i>	FALSE
<i>BAM Stats</i>	
<i>Whether to combine Data from Both the Strands for Coverage in WIG format</i>	1
<i>Bin Size for Coverage in WIG file format</i>	100
<i>Maximum insert size</i>	100 000
<i>Insert size bin</i>	100

<i>Primary Alignments only for Coverage in WIG file format</i>	1
<i>Maximum Coverage</i>	500 000
<i>Small Indel</i>	
<i>Detail Level</i>	3
<i>Number Alignment per Pileup</i>	1 000
<i>Random Seed</i>	94 404
<i>Min Num Evid</i>	2
<i>Max Num Evid</i>	-1
<i>ConsGroup</i>	1
<i>Max Reported Alignments</i>	-1
<i>Min Mapping Quality (MAPQ)</i>	8
<i>Min Best Mapping Quality</i>	10
<i>Min Anchor Mapping Quality</i>	-1
<i>SNP Finding</i>	
<i>Call Stringency</i>	Medium
<i>Skip High Coverage Positions (Het)</i>	FALSE
<i>Minimum Mapping QV</i>	8
<i>Detect Adjacent SNPs</i>	FALSE
<i>Polymorphism rate</i>	0,001
<i>Include Reads with unmapped Mate</i>	FALSE
<i>Exclude Reads with Indels</i>	TRUE
<i>Require Only Uniquely Mapped Reads</i>	FALSE
<i>Minimum Ratio of the Filtered Reads and Raw reads</i>	0
<i>Require alleles to be present in both strands</i>	FALSE
<i>Minimum Base QV a Read for a Position</i>	28
<i>Minimum Color QV of a Read for a Position</i>	7
<i>Min Base QV of the non-reference allele of the position</i>	28
<i>Minimum Unique Start Positions of Less Common Allele</i>	0
<i>The Less Common Allele on Both Strands</i>	FALSE
<i>Minimum Allele Ratio (Het)</i>	0,15
<i>Minimum Coverage (Het)</i>	2
<i>Minimum Unique Start Positions (Het)</i>	2
<i>Minimum non-Reference Color QV (Het)</i>	7
<i>Minimum non-Reference Base QV (Het)</i>	28
<i>Minimum Ratio of Valid Reads (Het)</i>	0,65
<i>Minimum Valid Tricolor Count (Het)</i>	2
<i>Minimum Coverage (Hom)</i>	1

2 PRIEDAS

.BAM duomenų bylų apdorojimas *The Genome Analysis Toolkit v.3.5* (GATK) įrankiu pagal Geros praktikos gaires

1. Fragmentų prilygiavimas prie referentinio geno:
<code>bwa mem -M -t 10 /kelias_į_referentinį_genomą/hg38.fa /kelias_į_bylą/1_fastq_fragmentų_grupė.fq /kelias_į_bylą/2_fastq_fragmentų_grupė.fq samtools view -h -b -@ 10 -o /kelias_į_bylą/1.bam</code>
2. Informacijos, reikalingos dirbant GATK programa, apie fragmentų grupę priskyrimas:
<code>java -jar /kelias_į_bylą/picard.jar AddOrReplaceReadGroups I=/kelias_į_bylą/1.bam O=/kelias_į_bylą/2.bam RGID=tiriamasis RGLB=lib1 RGPL=Illumina RGPU=1_1 RGSM=tiriamasis RGDT=data</code>
3. Fragmentų išrūšiuojimas pagal referentinio geno koordinatas:
<code>java -jar /kelias_į_bylą/picard.jar ReorderSam I=/kelias_į_bylą/2.bam O=/kelias_į_bylą/3.bam R=/kelias_į_referentinį_genomą/hg38.fa</code>
4. BAM duomenų bylos išrūšiuojimas pagal chromosomų koordinatas:
<code>java -jar /kelias_į_bylą/picard.jar SortSam I=/kelias_į_bylą/3.bam O=/kelias_į_bylą/4.bam SORT_ORDER=coordinate</code>
5. Duplikuotų DNR fragmentų pažymėjimas, kad analizės metu būtų vertinamas tik vienas iš jų:
<code>java -jar /kelias_į_bylą/picard.jar MarkDuplicates I=/kelias_į_bylą/4.bam O=/kelias_į_bylą/5.bam METRICS_FILE=/kelias_į_bylą/5.txt TMP_DIR=/kelias_į_bylą/tmp</code>
6. BAM duomenų bylos indeksų sukūrimas genominiams pozicijoms:
<code>java -jar /kelias_į_bylą/picard.jar BuildBamIndex I=/kelias_į_bylą/5.bam O=/kelias_į_bylą/5.bai</code>
7. Pirmas iškritų-intarpų regionų perlygiavimo prie referentinio geno etapas. DNR fragmentams perlygiuoti reikalingo regionų sąrašo sudarymas:
<code>java -jar /kelias_į_bylą/gatk.jar -T RealignerTargetCreator -R /kelias_į_referentinį_genomą/hg38.fa -S STRICT -I /kelias_į_bylą/5.bam -o /kelias_į_bylą/5.intervals -allowPotentiallyMisencodedQuals</code>
8. Antras iškritų-intarpų regionų perlygiavimo prie referentinio geno etapas. DNR fragmentų perlygiavimas:
<code>java -jar /kelias_į_bylą/gatk.jar -T IndelRealigner -R /kelias_į_referentinį_genomą/hg38.fa -S STRICT -I /kelias_į_bylą/5.bam -targetIntervals /kelias_į_bylą/5.intervals -o /kelias_į_bylą/6.bam -known /kelias_į_bylą/Mills_and_1000G_gold_standard.indels.hg38.vcf.gz -allowPotentiallyMisencodedQuals</code>
9. Fragmentų išrūšiuojimas pagal chromosomų koordinatas:
<code>java -jar /kelias_į_bylą/picard.jar SortSam I=/kelias_į_bylą/6.bam O=/kelias_į_bylą/7.bam SORT_ORDER=coordinate</code>
10. BAM duomenų bylos indeksų sukūrimas genominiams pozicijoms:

java -jar kelias_i_byla_picard.jar BuildBamIndex I=/kelias_i_byla/7.bam O=/kelias_i_byla/7.bai
11. Nukleotidų kokybės įverčių kalibravimas:
java -jar /kelias_i_byla_gatk.jar -R /kelias_i_referentini_genoma/hg38.fa -T BaseRecalibrator -I /kelias_i_byla/7.bam -S STRICT -knownSites /kelias_i_byla/dbsnp_146.hg38.vcf.gz -o /kelias_i_byla/7.table --covariate QualityScoreCovariate --covariate ReadGroupCovariate --covariate ContextCovariate -- covariate CycleCovariate -allowPotentiallyMisencodedQuals
12. Nukleotidų kokybės įverčių kalibravimo duomenų pritaikymas sekoskaitos duomenims:
java -jar /kelias_i_byla_gatk.jar -R /kelias_i_referentini_genoma/hg38.fa -T PrintReads -I /kelias_i_byla/7.bam -S STRICT -o /kelias_i_byla/8.bam -BQSR /kelias_i_byla/7.table - allowPotentiallyMisencodedQuals

3 PRIEDAS

***DnM* identifikavimo *VarScan* (*Variant detection in massively parallel Sequencing data*) v.2.4.2 programa komandinės eilutės ir parametrai**

1. Tėvo, motinos ir vaiko .BAM duomenų bylų sujungimas:
samtools mpileup -B -q 1 -f ref.fasta tėvo.bam motinos.bam vaiko.bam >triada.mpileup
2. <i>dnM</i> nustatymas:
java -jar /kelias_i_byla/VarScan.v2.4.2.jar trio /kelias_i_byla/triada.mpileup /kelias_i_byla/triada --min-covage 20 --min-var-freq 0.30 --p-value 0.05 -adj-varfreq 0.05 -adj-p-value 0.05

Parametras	Reikšmė
Mažiausias horizontalusis padengimas (angl. <i>min covage</i>), kad variantas būtų nustatytas	20
Mažiausias varianto dažnis (angl. <i>min varfreq</i>)	0.30
<i>P</i> reikšmė (angl. <i>p-value</i>)	0.05
Greta esančios bazės mažiausias dažnis (angl. <i>adj varfreq</i>)	0.05
Greta esančios bazės <i>p</i> reikšmė (angl. <i>adj p-value</i>)	0.05

Pradmenų sekos *dnM* tikrinti *Sanger* sekoskaitos metodu

Bendroji Lietuvos populiacija		
Genas	Tiesioginis pradmuo	Atvirkštinis pradmuo
<i>MEIS2</i>	cgacggagattaccagtca	ttgagtgggattctggggac
<i>ULK4</i>	actcactgatgggattgtctt	tggacttctagactccacgc
<i>OR2T34</i>	acctctccattacceactgc	cggcagaattcatctgtgg
<i>PGK1</i>	gccatcctcccatgactaa	agcagatcagagagtgggtt
<i>DNAH6</i>	gggggtggtggaaggaaatt	aagcatggggatagggtgtg
<i>KIF13A</i>	tgaccaattgtttcatcgctt	gcaccaatttctgctgcttc
<i>SEC31A</i>	cccggccataatctgttcaag	tgattcaacacagccatgcc
<i>TDG</i>	caaaacaaccagtggaaacca	acctcatgaagctgacacca
<i>SPOCD1</i>	ctggagcaaggactcgct	atgttcatgaagcggccatc
<i>ACTR6</i>	gcattacactggaaatgtaacc	actcaaccttacttcaacatgg
<i>MT1B</i>	atgtaccatcagaacacagct	agaatgtagcaaaccggtcag
<i>GTF2H2C</i>	ctcacaacctttctttgcattgt	cgcaagtgtgaaggctgcta
<i>PCDHAC1</i>	gtaccaagttgcaccagagc	cattgtactccccacgcaac
<i>STK4</i>	ctgtatcggcctccaagaa	tcttctgcttcttctgggg
<i>PPAT</i>	attggctgggcagagaagag	caacaacgaagggtgacaa
<i>REXO4</i>	ttactgcaaaaggtggcgg	caacaagtcagccaccagtg
<i>ADHFE1</i>	ggcaacagaatgagactccg	ccgacagcaacataggcatc
<i>NBEA</i>	tgaacaccagtcacattgcc	tcccagaaccttcaaaagcttg
<i>ZFPM2</i>	tgcatgtgaaaaggtgaaagga	cactggaagttgctgtcgac
<i>UBR4</i>	tggcaagatccactgcttca	caagcttgcttggccagaag
<i>SLC27A6</i>	gcccagtgaccaagcttaa	ggccagaagacagctctgat
<i>PDZRN4</i>	cagtgagggcaagaaggaga	accgaaactggtagcatcct
<i>VNIR4</i>	tcccatgtatgtgactggca	tcaccaggatgaggatgctc
<i>CEP89</i>	ttctccaggacacactcagc	ttatttcttccctgcctgc
<i>SYNM</i>	gatgaagctggcctggacta	tccactttgaatgggacgga
<i>SLFN13</i>	gcaagttcaacatggaggca	catccctgttggccatttcc
<i>TBCK</i>	tggcactaggttaatttgagacg	gcactctccttccatcct
<i>KIF13A</i>	ggctgcccttaaacctetca	gcaccaatttctgctgcttc
<i>CEACAM5</i>	tgacggacgattcagcatct	ggagcatgaagaccagggaa
<i>TAC1</i>	agaagcttgctttgttctgt	aacccaaagattcaggttgg

<i>MAP9</i>	agctctttgacatgctggtg	gcagtagtcacatggctgc
<i>RASA1</i>	ggtcagcagccttgtttaca	acacctggatatggtacctgt
<i>STK38</i>	ctceccgtctcctcccttt	gcaattaggatgtggggctg
<i>PMPCB</i>	agattgctggcagctttct	tgtccgtccaagtgcagtat
<i>MAD1L1</i>	tgatccgctactgaggatt	aaccacctcgtagtctctg
<i>CENPP</i>	tggtttacgttctccagagt	cctggctcatgaactcctct
<i>ATP2A2</i>	ggtgacaaagtctctgctga	tggagcgcagactaaagcact
<i>SOS2</i>	ggagcgcagtcagaagacct	acacactaacacagacctgc
<i>UNK</i>	acttcattctctgctccgg	acttgggtgcagtagacgtca
<i>ZNF256</i>	ggagaagtcacaccggagaa	tttcccactgtgaattcgc
<i>LONP1</i>	atcctctgcttctatggccc	cccggattgctctaccagat
<i>AXL</i>	ccctactggtgggtgaatgt	atagttgggtgggagtgtgg
<i>SMARCA1</i>	cccactcctaattctggtgga	gcccgtttctcttttctgc
<i>ARSD</i>	gggtcccttgggtactgt	ggctccatggtaacctgact
<i>CCDC181</i>	ttgaaaaggacctggagtgg	acatttaagcaatctacgcctca
<i>ELP3</i>	agagaagctgtcaggggttg	ctaggatccgagccaccaat
<i>FAM122C</i>	tgcagccagtgactcagtag	ccttcccagcttcttctgatgg
<i>PARP4</i>	tgccaattttcccacagtt	ctgttcctgaatgcgcaat
<i>VWA8</i>	ttgatgagcaaccagtggt	aagcttctgaaggatgggct
Asmenų su IN ir asmenų su IN sibsų grupėms		
Genas	Tiesioginis pradmuo	Atvirkštinis pradmuo
<i>HOXA1</i>	cgacgaccgcttcttagt	gggtaccaccacttacgtc
<i>MAPK3</i>	cacctcctatccacaggca	tttatcccacaccaccctc
<i>ARID1B</i>	gagagtacgtgggctttgga	gacggctcagtgtcatcaac
<i>PAC1</i>	taaccacgctcactctgtt	gttctctcacctctgccaat
<i>AKAP8L</i>	aggaggagaaggagcaggag	ctcctcgtcgtcctccac
<i>BECN1</i>	tggggccagacagatgtg	agagetccatttgctgagt
<i>ZCWPW1</i>	ccaagcccactgtcattc	acctgcagactccctctcta
<i>NCOA3</i>	tcgagtgtctagtctccca	tctttacctgttctggcca
<i>ABC5</i>	tgcccttgataaagccagga	tgaaggtcagatggtgtgga
<i>AHDC1</i>	cctgcctgactccttgag	gattccggcttcaactctgg
<i>COL5A3</i>	ggcgctccctttgtcttg	ctecatccccatccaacct
<i>KMT2A</i>	aggatcctgccccaaagaaa	tttctggcggtcctgtagt
<i>PPL</i>	ccgcagcatcttaacaagg	tcaccgattccaaacagatcc
<i>AARS</i>	aggtgtgctttgggtagga	agcacagaagggtgaagcaga

<i>TCF4</i>	catccctttgtcccacagc	aggatcaggagcttggtctg
<i>TRANK1</i>	tagaacaccaccagaggcag	cctcttgatgtgctcctgga
<i>BRPF1</i>	aaaatcaccaaccgctgac	ctagttcccagagcccacag
<i>SMO</i>	gtgggtcagagtgaggagg	ttcagggaagcgggtcagg
<i>IRF5</i>	ccacacgcactctctgtaga	ccaggacctcagagagaagc
<i>SATB1</i>	cctttcatttctgcagggct	tctttcccttcgtcctgg
<i>ESPN</i>	acaggcagctgacatctaca	ccctatgccactagtgagg
<i>ETNPPL</i>	acatcgtcacaatgggaaaac	ccggcaatactgagctttattct
<i>RAPGEF1</i>	gctgctgtgggaagaatgg	gctggtagatgtgctcgttc
<i>AMPD2</i>	ctttgacctggcatcactg	atacctcggcgatctccttg
<i>NFATC3</i>	gcaggtatttggaaactccgc	cacttacagcactcaacggg
<i>GORASP1</i>	caagcacagaggcctagat	ccaccaaaagcagcaaggaat
<i>MSH3</i>	agcctgaaatccactcctc	ctcctccagccctatcatcg
<i>MECP2</i>	tcgaaaaggttaggcgacaca	cttttacctgcacaccctc
<i>ESRRG</i>	cacatttgttggtggctgaac	gtacctggaatatgcttcgcc
<i>TAB2</i>	caaatcagcaaggccaccag	agaggaagtgtggaggttc
<i>QRICH2</i>	acctcgtgcagatcatcagc	ccataccttgctgacctattcc
<i>MYADML2</i>	ctcaagatcgtccaggcctt	tgaggtagaggagcacagc
<i>C6orf201</i>	accccaaagccgaacagaat	tggggcccatgtgcttttat
<i>CREBBP</i>	cgatttactggcacgttca	gtgacaaaagccaccacctt
<i>GIGYF2</i>	tgtgtttggtcctacagatacc	aaaagaataagagtttgcgtgggg
<i>SYT1</i>	ttcagaaagtcaggtggtg	catggcatcaacttctctct
<i>CTR9</i>	tcacgaaaagcagcaaatga	tgcagttgtgtgttctctg
<i>PRAME</i>	gtgctcgtagacctgttctt	ctggcccaggtaaggagaaa
<i>MYOM1</i>	catcaagcttgtgtcattcca	acgttccctcatctcatcc
<i>PLB1</i>	ctatgacacgcaggaggact	cgtctggagagctgtgtctt
<i>DNAH8</i>	gtacggcatggcttgatgac	cagtgtccagctgccaaac
<i>IGSF3</i>	ggaagaggaaagggaggagg	gagagcctctggaagcgaa
<i>TLN1</i>	atggcattgaggttgaggga	tccatcctactgccctctct
<i>POLD1</i>	cctctgatcatcctcccac	gatggaagaagctgggacca
<i>TRAPPC12</i>	gccttcatttccgtcagcaa	gtcactccctgcaaactg
<i>SVEP1</i>	ctgaagcctcccgaatgg	tctctgaaccggaccacaaa
<i>HIST1H1E</i>	gcctcaacatgtccgagac	tgatcgggctgtttcttctc
<i>ZNF189</i>	tgtgcagacagtgtggaaaa	ctgcgctgttgactgaaact

WNT3A	acactcaccaccggatcttg	atgccaatcttgatgccctc
CHD7	agagtgggtgggagtgtggac	ttacctgccctctgttttg

5 PRIEDAS

Tinkamumo įverčiams nustatyti naudotos komandinės eilutės

1. Filogenetinio medžio paruošimas su RAxML programa:
raxmlHPC -f a -x 12345 -p 12345 -N 100 -m GTRGAMMA -s /kelias_i_byla/seka.phy -n seka
2. Filogenetinio medžio šakų ilgio sutvarkymas:
bin/codeml codeml.ctl
3. Tinkamumo įverčių apskaičiavimas:
java -cp /kelias_i_byla/tdg12.jar tdg.Analyse -s /kelias_i_byla/seka.phy -t /kelias_i_byla/RAxML_bestTree -gc standard -site 7 -tau 1e-6 -kappa 8.0004 -pi 0.21051,0.19380,0.40010,0.19559 -mu 3.0 > /kelias_i_byla/fitness.out

6 PRIEDAS

De novo genomo variantų katalogas

Chr	Pozicija	Pradžia	Pabaiga	Alt	cytoBand	Func.refGene	Genas	pokytis	snp147
<i>Lietuvos populiacijos grupėje nustatytos dnM</i>									
Y	22089056	22089056	G	A	Yq11.2 22	Tarpgeninė sritis	<i>KDM5D</i> <i>,TTY10</i>	NA	NA
X	135462743	135462743	T	A	Xq26.3	Intronas	<i>ADGRG</i> <i>4</i>	NA	NA
Y	13846451	13846451	A	G	Yq11.2 1	Tarpgeninė sritis	<i>GYG2P</i> <i>1</i>	NA	rs111213896
X	77365375	77365375	A	G	Xq21.1	Egzonas	<i>PGK1</i>	Nesinoniminis VNV	NA
Y	13661326	13661326	G	T	Yq11.2 1	Tarpgeninė sritis	<i>GYG2P</i> <i>1</i>	NA	NA
X	129370527	129370527	C	T	Xq26.1	Egzonas	<i>ZNF280</i> <i>C</i>	Nesinoniminis VNV	NA
Y	59146621	59146621	T	G	Yq12	Tarpgeninė sritis	<i>SPRY3</i> <i>,VAMP7</i>	NA	NA
X	72224741	72224741	T	G	Xq13.2	NTS3'	<i>PABPC</i> <i>1L2B</i>	NA	rs376528143
X	35971646	35971646	C	G	Xp21.1	Intronas	<i>CFAP4</i> <i>7</i>	NA	NA
1	240078659	240078659	G	A	1q43	Tarpgeninė sritis	<i>CHRM3</i> <i>,RPS7P</i> <i>5</i>	NA	NA
1	66083861	66083861	C	A	1p31.3	Intronas	<i>LEPR</i>	NA	NA
1	17185602	17185602	A	G	1p36.13	Upstream	<i>MIR367</i> <i>5</i>	NA	rs79106853

1	169513583	169513583	G	T	1q24.2	Egzonas	<i>F5</i>	Sinonimimis VNV	rs6037
1	196965137	196965137	A	T	1q31.3	Intronas	<i>CFHR5</i>	NA	NA
1	17195034	17195034	G	A	1p36.13	Tarpgeninė sritis	<i>MIR367 5,CROC C</i>	NA	rs112930634
1	212238392	212238392	A	G	1q32.3	Intronas	<i>DTL</i>	NA	NA
2	11359095	11359095	A	T	2p25.1	Egzonas	<i>ROCK2</i>	Nesinonimimis VNV	NA
2	200826683	200826683	C	A	2q33.1	Intronas	<i>C2orf47</i>	NA	NA
2	54870021	54870021	T	A	2p16.2	Intronas	<i>SPTBN1</i>	NA	NA
2	190722170	190722170	G	A	2q32.2	Intronas	<i>PMS1</i>	NA	rs735558
2	11357395	11357395	A	G	2p25.1	Intronas	<i>ROCK2</i>	NA	NA
2	32142951	32142951	A	G	2p22.3	Intronas	<i>MEMO 1</i>	NA	NA
2	197645222	197645222	C	T	2q33.1	NTS3	<i>GTF3C 3</i>	NA	NA
2	223797847	223797847	C	T	2q36.1	Intronas	<i>ACSL3</i>	NA	NA
2	87105886	87105886	T	G	2p11.2	bRNR_Intronas	<i>ANAPC IP1</i>	NA	rs57494924
2	113418840	113418840	C	T	2q13	Intronas	<i>SLC20A 1</i>	NA	rs751111651
2	230312261	230312261	G	A	2q36.3	Intronas	<i>DNER</i>	NA	rs10490182
3	11488058	11488058	A	G	3p25.3	Intronas	<i>ATG7</i>	NA	NA
3	133896978	133896978	T	C	3q22.2	Intronas	<i>RYK</i>	NA	NA

3	41607621	41607621	C	A	3p22.1	Egzonas	<i>ULK4</i>	Baigmės kodonas	NA
3	10115166	10115166	C	T	3p25.3	Intronas	<i>FANCD2</i>	NA	NA
4	49635958	49635958	G	C	4p11	Tarpgeninė sritis	<i>CWH43</i>	NA	NA
4	120981450	120981450	A	C	4q27	Intronas	<i>MAD2L1</i>	NA	NA
4	123195492	123195492	T	C	4q27	Intronas	<i>KIAA1109</i>	NA	NA
4	144447373	144447373	T	C	4q31.21	Intronas	<i>SMARCA5</i>	NA	NA
4	9214042	9214042	T	C	4p16.1	Downstream	<i>USP17L10</i>	NA	NA
4	49247454	49247454	A	G	4p11	Tarpgeninė sritis	<i>CWH43</i>	NA	rs3883610
4	1389765	1389765	G	A	4p16.3	NTS3'	<i>CRIPAK</i>	NA	NA
4	79403747	79403747	C	A	4q21.21	Intronas	<i>FRAS1</i>	NA	rs924149
4	71629209	71629209	T	A	4q13.3	Intronas	<i>RUFY3</i>	NA	NA
5	130825204	130825204	G	A	5q31.1	Intronas	<i>RAPGEF6</i>	NA	NA
5	49450645	49450645	T	C	5q11.1	Tarpgeninė sritis	<i>EMB</i>	NA	NA
6	11201157	11201157	C	T	6p24.2	NTS3'	<i>NEDD9</i>	NA	rs3798734
6	10697763	10697763	T	C	6p24.2	Intronas	<i>PAK1IP1</i>	NA	NA
6	144171374	144171374	T	C	6q24.2	Intronas	<i>LTV1</i>	NA	NA
7	33015903	33015903	G	A	7p14.3	Intronas	<i>FKBP9</i>	NA	NA

7	143560565	143560565	T	G	7q35	Egzonas	<i>TCAF1</i>	Sinoniminis VNV	rs200302861
7	64643549	64643549	G	A	7q11.21	Tarpgeninė sritis	<i>CCT6P 3,ZNF9 2</i>	NA	rs2671978
7	33054313	33054313	G	A	7p14.3	NTS3'	<i>NT5C3 A</i>	NA	NA
7	92855683	92855683	A	C	7q21.3	Egzonas	<i>HEPAC AM2</i>	Nesinoniminis VNV	NA
8	48866530	48866530	G	A	8q11.21	Intronas	<i>PRKDC</i>	NA	NA
8	9577866	9577866	C	A	8p23.1	Intronas	<i>TNKS</i>	NA	NA
8	381331	381331	T	G	8p23.3	Intronas	<i>FBXO2 5</i>	NA	rs776602003
9	91940638	91940638	A	T	9q22.2	Intronas	<i>SECISB P2</i>	NA	NA
9	91940633	91940633	C	A	9q22.2	Intronas	<i>SECISB P2</i>	NA	NA
9	75357478	75357478	C	G	9q21.13	Intronas	<i>TMC1</i>	NA	NA
10	125602021	125602021	C	T	10q26.13	Intronas	<i>CPXM2</i>	NA	rs10794571
10	42387123	42387123	T	C	10q11.21	Tarpgeninė sritis	<i>,LOC441 666</i>	NA	NA
10	17032281	17032281	C	T	10p13	Intronas	<i>CUBN</i>	NA	rs2291521
10	47668721	47668721	G	A	10q11.22	Intronas	<i>ANTXRL</i>	NA	rs4570521
10	94369005	94369005	T	C	10q23.33	Intronas	<i>KIF11</i>	NA	NA
11	51436127	51436127	G	T	11p11.12	Tarpgeninė sritis	<i>OR4A5,O R4C46</i>	NA	rs138669462

11	57463500	57463500	T	C	11q12.1	Egzonas	<i>ZDHHC5</i>	Sinonimimis VNV	rs1783978
12	12633044	12633044	A	G	12p13.2	Intronas	<i>DUSP16</i>	NA	NA
12	9162528	9162528	G	A	12p13.31	NTS3	<i>KLRG1</i>	NA	NA
12	47178811	47178811	G	T	12q13.11	Intronas	<i>SLC38A4</i>	NA	NA
12	96421364	96421364	A	C	12q23.1	Intronas	<i>LTA4H</i>	NA	NA
12	22666120	22666120	A	G	12p12.1	Intronas	<i>C2CD5</i>	NA	NA
12	53207583	53207583	C	G	12q13.13	Egzonas	<i>KRT4</i>	Nesinonimimis VNV	rs76773498
12	10571687	10571687	A	G	12p13.2	Egzonas	<i>KLRC3</i>	Nesinonimimis VNV	NA
13	21173710	21173710	T	G	13q12.11	Intronas	<i>IFT88</i>	NA	NA
14	34029438	34029438	T	A	14q13.1	Intronas	<i>NPAS3</i>	NA	NA
14	36231843	36231843	G	C	14q13.2	bRNR_Intronas	<i>RALGAP A1P</i>	NA	NA
14	75283980	75283980	G	A	14q24.3	Egzonas	<i>YLPM1</i>	Sinonimimis VNV	NA
15	81601117	81601117	C	T	15q25.1	Egzonas	<i>IL16</i>	Nesinonimimis VNV	NA
15	37188859	37188859	G	A	15q14	Egzonas	<i>MEIS2</i>	Baigmēs kodonas	NA
15	34542753	34542753	C	T	15q14	Intronas	<i>SLC12A6</i>	NA	NA
16	56686929	56686929	G	T	16q12.2	Egzonas	<i>MT1B</i>	Nesinonimimis VNV	NA
16	68842480	68842480	G	C	16q22.1	Intronas	<i>CDH1</i>	NA	rs33963999

16	33677598	33677598	C	A	16p11.2	Tarpgeninė sritis	<i>ENPP7P 13,LINC 00273</i>	NA	NA
17	45673872	45673872	T	C	17q21.32	Intronas	<i>NPEPPS</i>	NA	NA
17	33772544	33772544	A	G	17q12	Egzonas	<i>SLFN13</i>	Sinoniminis VNV	NA
17	39084464	39084464	T	A	17q21.2	Intronas	<i>KRT23</i>	NA	NA
17	37657710	37657710	T	A	17q12	Intronas	<i>CDK12</i>	NA	NA
17	40963623	40963623	C	T	17q21.31	Intronas	<i>BECN1</i>	NA	NA
18	9775362	9775362	C	T	18p11.22	Intronas	<i>RAB31</i>	NA	rs149779563
18	22806355	22806355	A	G	18q11.2	Egzonas	<i>ZNF521</i>	Sinoniminis VNV	NA
19	40196627	40196627	G	A	19q13.2	Egzonas	<i>LGALS1 4</i>	Nesinoniminis VNV	rs72480733
19	30503146	30503146	T	C	19q12	Intronas	<i>UR11</i>	NA	NA
19	58152020	58152020	T	C	19q13.43	Intronas	<i>ZNF211</i>	NA	rs3746221
19	14736288	14736288	C	G	19p13.12	Intronas	<i>ADGRE3</i>	NA	rs34381727
19	55351066	55351066	C	T	19q13.42	Egzonas	<i>KIR2DS4</i>	Nesinoniminis VNV	rs1130503
20	45174026	45174026	A	C	20q13.12	Egzonas	<i>OCSTAM P</i>	Sinoniminis VNV	rs202396
22	29451139	29451139	T	G	22q12.1	NTS3'	<i>ZNRF3</i>	NA	rs200424365
22	24325783	24325783	G	A	22q11.23	Egzonas	<i>GSTT2,G STT2B</i>	Nesinoniminis VNV	rs200087397
Asmenų su IN grupėje nustatytos dnM									

Chr	Pradžia	Pabaiga	Ref	Alt	cytoBand	Regionas	Genas	EgzonasFunc.refGene	snp147
X	154031355	154031355	G	A	Xq28	Egzonas	MECP2	Nesinonimimis VNV	rs28934906
1	149950131	149950131	G	C	1q21.2	Egzonas	OTUD7B	Sinonimimis VNV	NA
1	97699506	97699506	C	T	1p21.3	Egzonas	DPYD	Sinonimimis VNV	rs6670886
1	27550994	27550994	G	-	1p36.11	Egzonas	AHDC1	Rėmelio poslinkio delecija	rs749294057
1	3872647	3872647	A	G	1p36.32	Intronas	DFFB	NA	rs75402433
1	3872648	3872648	T	C	1p36.32	Intronas	DFFB	NA	rs79548119
1	6445909	6445909	G	A	1p36.31	Egzonas	ESPN	Nesinonimimis VNV	NA
1	3872614	3872614	-	GGCCCTGTCCCT GCCGC	1p36.32	Intronas	DFFB	NA	rs202183040
1	109625330	109625330	C	T	1p13.3	Egzonas	AMPD2	Nesinonimimis VNV	rs748496694
1	216564242	216564242	A	T	1q41	Egzonas	ESRRG	Nesinonimimis VNV	NA
1	26889836	26889836	G	A	1p36.11	Egzonas	GPN2	Sinonimimis VNV	NA
1	169773475	169773475	A	T	1q24.2	Intronas	C1orf112	NA	NA
2	29073917	29073917	C	T	2p23.2	Egzonas	C2orf71	Sinonimimis VNV	rs549344507
2	74514879	74514879	C	T	2p13.1	Egzonas	TLX2	Sinonimimis VNV	NA

2	233832151	233832151	-	AAAA	2q37.1	Intronas	<i>MROH2 A</i>	NA	rs76276350
2	127637301	127637301	C	A	2q14.3	Intronas	<i>MYO7B</i>	NA	rs2288656
2	218824425	218824425	T	C	2q35	Intronas	<i>PRKAG3</i>	NA	NA
2	170401293	170401293	T	C	2q31.1	Egzonas	<i>FASTKD 1</i>	Nesinonimimis VNV	NA
2	28032985	28032985	T	G	2p23.2	Intronas	<i>RBKS</i>	NA	NA
2	28032986	28032986	C	A	2p23.2	Intronas	<i>RBKS</i>	NA	NA
2	28032987	28032987	A	C	2p23.2	Intronas	<i>RBKS</i>	NA	NA
2	233613741	233613741	G	T	2q37.1	Egzonas	<i>GIGYF2</i>	Nesinonimimis VNV	NA
3	175471628	175471628	T	A	3q26.31	Intronas	<i>NAALAD L2</i>	NA	rs201261597
3	175471629	175471629	G	-	3q26.31	Intronas	<i>NAALAD L2</i>	NA	rs67167200
3	36831070	36831070	G	A	3p22.2	Egzonas	<i>TRANK1</i>	Nesinonimimis VNV	NA
3	9740827	9740839	CATGCAGAG GCTG	-	3p25.3	Egzonas	<i>BRPF1</i>	Rėmelio poslinkio delecija	NA
3	18386559	18386559	T	C	3p24.3	Egzonas	<i>SATB1</i>	Nesinonimimis VNV	NA
3	39098275	39098275	C	G	3p22.2	Egzonas	<i>GORASP 1</i>	Nesinonimimis VNV	NA
3	16358081	16358081	G	T	3p24.3	Intronas	<i>RFTN1</i>	NA	rs774103191
3	100064534	100064534	A	G	3q12.2	Intronas	<i>NIT2</i>	NA	rs372927365
3	99885147	99885147	T	-	3q12.1	Intronas	<i>CMSS1</i>	NA	NA

3	137963914	137963914	C	T	3q22.3	Intronas	<i>ARMC8</i>	NA	rs567697082
4	49147291	49147291	C	G	4p11	Tarpgeninė sritis	<i>CWH43</i>	NA	NA
4	49147291	49147291	C	G	4p11	Tarpgeninė sritis	<i>CWH43</i>	NA	NA
4	69410865	69410865	A	T	4q13.3	Tarpgeninė sritis	<i>UGT2B2</i> 8; <i>UGT2B</i> 4	NA	NA
4	108749239	108749239	G	A	4q25	Egzonas	<i>ETNPPL</i>	Nesinoniminis VNV	rs753029468
4	87615711	87615728	AATAGTAGT GACAGCAGC	-	4q22.1	Egzonas	<i>DSPP</i>	Ne rėmelio poslinkio delecija	NA
4	110972455	110972455	T	G	4q25	NTS3	<i>ELOVL6</i>	NA	NA
5	4095534	4095534	-	CTTTCC TTTTCT TTT	5p15.33	Tarpgeninė sritis	<i>IRX1;LO</i> <i>C101929</i> <i>153</i>	NA	rs70963758
5	4095539	4095539	T	C	5p15.33	Tarpgeninė sritis	<i>IRX1;LO</i> <i>C101929</i> <i>153</i>	NA	rs1689760
5	710908	710908	T	A	5p15.33	Tarpgeninė sritis	<i>TPPP;Z</i> <i>DHHC11</i>	NA	rs368146265
5	80654917	80654917	C	G	5q14.1	Egzonas	<i>MSH3</i>	Nesinoniminis VNV	rs2405877
6	157203998	157203998	T	C	6q25.3	Sukirpimas	<i>ARID1B</i>	NA	NA
6	92888027	92888027	G	C	6q16.1	Tarpgeninė sritis	<i>CASC6;E</i> <i>PHA7</i>	NA	NA
6	149378814	149378814	C	A	6q25.1	Egzonas	<i>TAB2</i>	Baigmės kodonas	NA

6	4099363	4099363	A	C	6p25.2	Egzonas	<i>C6orf201</i>	Nesinoniminis VNV	NA
7	27095698	27095700	TGG	-	7p15.2	Egzonas	<i>HOXA1</i>	Ne rėmelio poslinkio delecija	rs753969638
7	100402540	100402540	G	A	7q22.1	Egzonas	<i>ZCWPW1</i>	Baigmės kodonas	rs1046667014
7	20755558	20755560	TCA	-	7p21.1	Egzonas	<i>ABCB5</i>	Ne rėmelio poslinkio delecija	NA
7	129203441	129203441	A	G	7q32.1	Egzonas	<i>SMO</i>	Nesinoniminis VNV	NA
7	128947298	128947327	ACTCTGCAG CCGCCCACT CTGCGGCCG CCT	-	7q32.1	Egzonas	<i>IRF5</i>	ne rėmelio poslinkio delecija	rs199508964
8	133021896	133021896	G	T	8q24.22	Intronas	<i>TG</i>	NA	NA
8	37751607	37751607	C	A	8p11.23	Intronas	<i>ERLIN2</i>	NA	NA
8	144427027	144427027	C	T	8q24.3	Intronas	<i>VPS28</i>	NA	rs61498915
8	35750476	35750476	-	A	8p12	Intronas	<i>UNC5D</i>	NA	rs111565377
8	144427025	144427025	-	GT	8q24.3	Intronas	<i>VPS28</i>	NA	rs139513541
8	26906286	26906286	T	G	8p21.2	Tarpgeninė sritis	<i>ADRA1A ;STMN4</i>	NA	NA
9	134796954	134796954	A	G	9q34.3	Intronas	<i>COL5A1</i>	NA	rs3827852
9	76175233	76175237	AATGA	-	9q21.13	Intronas	<i>PCSK5</i>	NA	rs774970963

9	134848781	134848804	GATCACACC TCCCAGGAA GCACAG	-	9q34.3	bRNR_Intronas	<i>LOC101 448202</i>	NA	NA
9	76175208	76175222	AATGGAATG GAATGA	-	9q21.13	Intronas	<i>PCSK5</i>	NA	rs750861488
9	135714735	135714735	-	GG	9q34.3	Intronas	<i>KCNT1</i>	NA	rs141291153
9	134848803	134848803	A	G	9q34.3	bRNR_Intronas	<i>LOC101 448202</i>	NA	rs34435434
9	131621977	131621977	A	G	9q34.13	Egzonas	<i>RAPGEF 1</i>	Nesinoniminis VNV	NA
9	33036799	33036799	C	T	9p21.1	Intronas	<i>DNAJA1</i>	NA	rs112603003
10	104055446	104055446	A	T	10q25.1	Intronas	<i>COL17A 1</i>	NA	rs1747676
10	7824305	7824305	C	-	10p14	Intronas	<i>TAF3</i>	NA	rs199762284
10	7824301	7824301	T	-	10p14	Intronas	<i>TAF3</i>	NA	rs141447291
10	1124278	1124278	C	T	10p15.3	Egzonas	<i>WDR37</i>	Sinonimimis VNV	NA
10	7824301	7824301	T	-	10p14	Intronas	<i>TAF3</i>	NA	rs141447291
11	1191646	1191646	T	C	11p15.5	Egzonas	<i>MUC5A C</i>	Nesinoniminis VNV	rs75826663
11	48163428	48163431	CTGG	-	11p11.2	Intronas	<i>PTPRJ</i>	NA	rs75225930
11	63914846	63914846	G	-	11q13.1	Intronas	<i>RCOR2</i>	NA	rs113350066
11	406488	406488	C	T	11p15.5	Egzonas	<i>SIGIRR</i>	Sinonimimis VNV	NA
11	66211206	66211206	C	T	11q13.2	Egzonas	<i>PACSI</i>	Nesinoniminis VNV	rs398123009

11	118481886	118481887	AA	-	11q23.3	Egzonas	<i>KMT2A</i>	Rémelio poslinkio delecija	NA
11	120429672	120429672	A	G	11q23.3	Intronas	<i>ARHGEF12</i>	NA	rs760767574
11	10774247	10774247	C	G	11p15.3	Egzonas	<i>CTR9</i>	Nesinoniminis VNV	NA
11	58877237	58877237	T	C	11q12.1	Intronas	<i>FAM111B</i>	NA	NA
11	67926692	67926692	A	G	11q13.2	Intronas	<i>KMT5B</i>	NA	NA
12	14681492	14681492	A	T	12p12.3	Intronas	<i>GUCY2C</i>	NA	rs75088100
12	50138675	50138675	C	T	12q13.12	Intronas	<i>CERS5</i>	NA	NA
12	9182165	9182165	A	T	12p13.31	Intronas	<i>PZP</i>	NA	rs12814448
12	122191519	122191519	G	C	12q24.31	Egzonas	<i>LRRC43</i>	Sinoniminis VNV	NA
12	14882821	14882821	G	A	12p12.3	Intronas	<i>MGP</i>	NA	NA
12	10215790	10215790	G	A	12p13.2	Egzonas	<i>CLEC9A</i>	Sinoniminis VNV	NA
12	79842733	79842733	C	G	12q21.2	Egzonas	<i>SYT1</i>	Nesinoniminis VNV	NA
12	40643544	40643544	T	G	12q12	Intronas	<i>LRRK2</i>	NA	NA
12	40643549	40643549	G	A	12q12	Intronas	<i>LRRK2</i>	NA	NA
13	87618586	87618588	TGT	-	13q31.2	bRNR_Intronas	<i>MIR4500HG</i>	NA	rs397771872
13	87618586	87618588	TGT	-	13q31.2	bRNR_Intronas	<i>MIR4500HG</i>	NA	rs397771872

13	87618603	87618605	TCA	-	13q31.2	bRNR_Intronas	<i>MIR4500 HG</i>	NA	NA
13	87618586	87618588	TGT	-	13q31.2	bRNR_Intronas	<i>MIR4500 HG</i>	NA	rs397771872
13	87618586	87618588	TGT	-	13q31.2	bRNR_Intronas	<i>MIR4500 HG</i>	NA	rs397771872
13	33096413	33096413	A	C	13q13.1	Intronas	<i>N4BP2L 2</i>	NA	NA
13	33096414	33096414	T	G	13q13.1	Intronas	<i>N4BP2L 2</i>	NA	NA
13	33096415	33096415	A	G	13q13.1	Intronas	<i>N4BP2L 2</i>	NA	NA
13	33096416	33096416	A	G	13q13.1	Intronas	<i>N4BP2L 2</i>	NA	NA
14	73987008	73987008	A	-	14q24.3	Intronas	<i>ENTPD5</i>	NA	rs36015948
14	73987004	73987004	T	-	14q24.3	Intronas	<i>ENTPD5</i>	NA	rs199818408
14	20382493	20382493	C	T	14q11.2	Intronas	<i>TEPI</i>	NA	NA
14	70011458	70011458	A	T	14q24.2	Intronas	<i>SMOC1</i>	NA	rs59684948
14	28792102	28792102	A	C	14q12	bRNR_Egzonas	<i>LINC015 51</i>	NA	rs34203558
14	73874252	73874252	G	-	14q24.3	Intronas	<i>PTGR2</i>	NA	NA
14	23939371	23939371	A	C	14q11.2	Intronas	<i>NGDN</i>	NA	NA
14	23939372	23939372	C	A	14q11.2	Intronas	<i>NGDN</i>	NA	NA
14	60398779	60398779	A	T	14q23.1	bRNR_Intronas	<i>LRRC9</i>	NA	NA
14	60398780	60398780	A	T	14q23.1	bRNR_Intronas	<i>LRRC9</i>	NA	NA
15	56921007	56921007	T	C	15q21.3	Intronas	<i>TCF12</i>	NA	NA

16	30117194	30117194	C	G	16p11.2	Egzonas	<i>MAPK3</i>	Nesinoniminis VNV	NA
16	4910877	4910877	G	C	16p13.3	Egzonas	<i>PPL</i>	Nesinoniminis VNV	NA
16	2060687	2060687	C	T	16p13.3	Egzonas	<i>TSC2</i>	Sinoniminis VNV	rs45517153
16	70268299	70268299	G	A	16q22.1	Egzonas	<i>AARS</i>	Nesinoniminis VNV	rs762241422
16	582231	582231	G	-	16p13.3	Intronas	<i>PIGQ</i>	NA	rs67976359
16	68166907	68166907	G	A	16q22.1	Egzonas	<i>NFATC3</i>	Nesinoniminis VNV	NA
16	70538514	70538514	A	T	16q22.1	Intronas	<i>SF3B3</i>	NA	rs766430910
16	3786782	3786782	T	G	16p13.3	Egzonas	<i>CREBBP</i>	Nesinoniminis VNV	NA
16	48258419	48258419	C	A	16q12.1	Intronas	<i>ABCC11</i>	NA	rs9927238
16	48258424	48258424	T	C	16q12.1	Intronas	<i>ABCC11</i>	NA	rs9933821
16	14312667	14312667	C	G	16p13.12	Intronas	<i>MKL2</i>	NA	rs776362770
17	7107769	7107769	A	G	17p13.1	Intronas	<i>ASGR2</i>	NA	rs62058754
17	61404805	61404805	C	G	17q23.2	Intronas	<i>TBX2</i>	NA	rs56298202
17	61404808	61404808	C	-	17q23.2	Intronas	<i>TBX2</i>	NA	rs201348033
17	42818609	42818610	TG	-	17q21.31	Egzonas	<i>BECN1</i>	Rėmelio poslinkio delecija	NA
17	80962539	80962539	G	A	17q25.3	Egzonas	<i>RPTOR</i>	Sinoniminis VNV	rs17848671
17	61404808	61404808	C	-	17q23.2	Intronas	<i>TBX2</i>	NA	rs201348033

17	76292312	76292341	ACCAGGTTG CACCAAACC ACGCTGAAC TAT	-	17q25.1	Egzonas	<i>QRICH2</i>	Ne r�emelio poslinkio delecija	rs758349471
17	81941177	81941177	C	T	17q25.3	Egzonas	<i>MYADM L2</i>	Nesinoniminis VNV	NA
18	55228987	55228987	C	T	18q21.2	Egzonas	<i>TCF4</i>	Nesinoniminis VNV	rs121909121
18	7035945	7035945	A	C	18p11.31	Intronas	<i>LAMA1</i>	NA	NA
18	3090782	3090782	C	T	18p11.31	Egzonas	<i>MYOM1</i>	Nesinoniminis VNV	NA
19	36151659	36151659	-	C	19q13.12	Intronas	<i>COX7A1</i>	NA	rs112834485
19	22634473	22634474	CG	-	19p12	NTS5	<i>ZNF492</i>	NA	rs148291606
19	15380381	15380381	T	G	19p13.12	Egzonas	<i>AKAP8L</i>	Nesinoniminis VNV	rs767901026
19	42855568	42855568	C	G	19q13.2	bRNR_Egzonas	<i>PSG10P</i>	NA	rs200711372
19	10001778	10001778	G	A	19p13.2	Egzonas	<i>COL5A3</i>	Nesinoniminis VNV	rs150865244
19	4156872	4156872	C	-	19p13.3	Intronas	<i>CREB3L 3</i>	NA	rs11322852
19	12868024	12868024	A	G	19p13.13	Intronas	<i>MAST1</i>	NA	NA
19	55788770	55788770	G	-	19q13.42	Intronas	<i>NLRP11</i>	NA	rs377100270
20	47652547	47652547	-	G	20q13.12	Egzonas	<i>NCOA3</i>	frameshift insertion	NA
20	34741039	34741039	G	A	20q11.22	Egzonas	<i>NCOA6</i>	Sinoniminis VNV	NA
21	14381245	14381245	G	A	21q11.2	Egzonas	<i>HSPA13</i>	Sinoniminis VNV	rs767243273

22	22819430	22819430	A	G	22q11.22	Tarpgeninė sritis	<i>GGTLC2</i> ; <i>MIR650</i>	NA	rs774937594
22	22892482	22892482	G	A	22q11.22	Egzonas	<i>PRAME</i>	Nesinoniminis VNV	rs201064885
Asmenų su IN sibsų grupėje nustatytos dnM									
Chr	Start	End	Ref	Alt	cytoBand	Func.refGene	Genas	EgzonasFunc.refGene	avsnp147
11	47137255	47137255	C	T	11p11.2	Intronas	<i>C11orf49</i>	NA	NA
16	2056182	2056182	G	A	16p13.3	Intronas	<i>TSC2</i>	NA	NA
1	3872614	3872614	-	GGCCCTGTCCCT GCCGC	1p36.32	Intronas	<i>DFFB</i>	NA	rs202183040
2	28604758	28604758	C	T	2p23.2	Egzonas	<i>PLB1</i>	Nesinoniminis VNV	rs34289907
4	87615747	87615747	A	G	4q22.1	Egzonas	<i>DSPP</i>	Nesinoniminis VNV	rs150637282
1	3872614	3872614	-	GGCCCTGTCCCT GCCGC	1p36.32	Intronas	<i>DFFB</i>	NA	rs202183040
10	130107315	130107315	A	G	10q26.3	bRNR_Egzonas	<i>CTAGE7</i> <i>P</i>	NA	NA
18	23572246	23572246	T	C	18q11.2	Intronas	<i>NPC1</i>	NA	NA
6	38872575	38872575	C	G	6p21.2	Egzonas	<i>DNAH8</i>	Nesinoniminis VNV	NA
17	42669400	42669400	C	T	17q21.2	Intronas	<i>PLEKHH</i> <i>3</i>	NA	NA
3	38722419	38722419	G	A	3p22.2	Intronas	<i>SCN10A</i>	NA	NA
1	116579663	116579663	-	TCC	1p13.1	Egzonas	<i>IGSF3</i>	nonframeshift insertion	rs576658823

1	45629686	45629686	A	G	1p34.1	Intronas	<i>GPBPIL 1</i>	NA	NA
11	8100511	8100511	C	T	11p15.4	Egzonas	<i>TUB</i>	Sinonimimis VNV	rs138124865
6	89080770	89080770	C	T	6q15	NTS5'	<i>PNRC1</i>	NA	NA
9	35710880	35710880	G	A	9p13.3	Egzonas	<i>TLN1</i>	Nesinonimimis VNV	rs556004234
19	50402098	50402098	T	G	19q13.33	Egzonas	<i>POLD1</i>	Nesinonimimis VNV	rs768494581
1	222979750	222979750	A	G	1q41	Intronas	<i>DISP1</i>	NA	NA
2	3388400	3388400	C	T	2p25.3	Egzonas	<i>TRAPPC 12</i>	Sinonimimis VNV	NA
5	10986610	10986610	-	G	5p15.2	Intronas	<i>CTNND2</i>	NA	rs36031475
1	168293289	168293289	G	T	1q24.2	Intronas	<i>TBX19</i>	NA	rs12096093
7	38245283	38245283	-	GG	7p14.1	Tarpgeninė sritis	<i>STARD3 NL;TAR P</i>	NA	rs200076851
8	144427025	144427025	-	GT	8q24.3	Intronas	<i>VPS28</i>	NA	rs139513541
19	1234760	1234760	C	T	19p13.3	Intronas	<i>CBARP</i>	NA	rs752082048
9	110513026	110513026	G	A	9q31.3	Egzonas	<i>SVEP1</i>	Sinonimimis VNV	NA
1	168293287	168293287	-	TT	1q24.2	Intronas	<i>TBX19</i>	NA	rs373712384
12	9182165	9182165	A	T	12p13.31	Intronas	<i>PZP</i>	NA	rs12814448
6	26156583	26156583	G	T	6p22.2	Egzonas	<i>HIST1H1 E</i>	Nesinonimimis VNV	NA
9	101409420	101409420	G	A	9q31.1	Egzonas	<i>ZNF189</i>	Nesinonimimis VNV	rs747885994

13	87618586	87618588	TGT	-	13q31.2	bRNR_Intronas	<i>MIR4500 HG</i>	NA	rs397771872
16	582231	582231	G	-	16p13.3	Intronas	<i>PIGQ</i>	NA	rs67976359
5	710908	710908	T	A	5p15.33	Tarpgeninė sritis	<i>TPPP;Z DHHC11</i>	NA	rs368146265
13	87618586	87618588	TGT	-	13q31.2	bRNR_Intronas	<i>MIR4500 HG</i>	NA	rs397771872
19	613554	613554	-	GGGGCTGGGGCC	19p13.3	Intronas	<i>HCN2</i>	NA	rs778777728
1	209628212	209628212	C	T	1q32.2	Intronas	<i>LAMB3</i>	NA	rs767847211
1	228022772	228022772	C	T	1q42.13	Egzonas	<i>WNT3A</i>	Sinoniminis VNV	rs745929272
10	100978934	100978934	G	A	10q24.31	Egzonas	<i>SEMA4G</i>	Sinoniminis VNV	rs781655724
10	7824301	7824301	T	-	10p14	Intronas	<i>TAF3</i>	NA	rs141447291

7 PRIEDAS

DnM patogeniškumo įvertinimas

Lietuvos populiacija (<i>VarScan</i>)																
Chr	Pozicija	Ref	Alt	SIFT	Polyphen2 HDIV	Polyphen2 HVAR	LRT	Mutation Taster	Mutation Assessor	FATHMM	PROVEAN	CADD	fathmm- MKL coding	GERP++	Patogeninių variantų skaičius	
15	81601117	C	T	D	B	B	N	N	L	T	N	8,498	N	3,86	2	
15	37188859	G	A	.	.	.	D	A	.	.	.	40	D	6,16	6	
2	11359095	A	T	T	B	B	N	N	N	D	N	6,42	D	0,364	2	
16	56686929	G	T	D	D	P	U	N	.	T	D	19,07	D	1,83	6	
19	40196627	G	A	T	D	B	.	P	N	T	N	15,67	N	-0,44	3	
22	24325783	G	A	.	B	B	U	N	N	.	.	5,248	N	-0,017	0	
3	41607621	C	A	.	.	.	U	A	.	.	.	46	D	4,54	6	
X	77365375	A	G	D	P	B	D	D	H	D	D	23,6	D	5,88	11	
X	129370527	C	T	T	B	B	.	N	N	T	N	0,002	N	-1,89	0	
7	92855683	A	C	T	B	B	N	N	N	T	N	11,72	D	4,94	2	
12	53207583	C	G	T	B	B	N	N	L	D	N	2,968	N	4,25	2	
12	10571687	A	G	T	B	B	N	N	L	T	N	0,517	N	-3,32	0	

Lietuvos populiacija (VarSeq)														
Chr	Pozicija	Ref	Alt	SIFT	Polyphen2 HDIV	Polyphen2 HVAR	LRT	Mutation Taster	Mutation Assessor	FATHMM	PROVEAN	CADD	fathmm- MKL coding	Patogeninių variantų skaičius
2	228552272	G	C	D	D	D	D	D	M	D	D	26,3	D	9
18	44595633	C	G	T	B	B	D	D	N	D	N	24,4	D	5
16	2027281	T	C	D	P	P	D	D	M	T	D	25	D	8
12	66539620	C	T	D	.	.	.	23,9	D	3
11	87847209	G	T	T	P	B	D	D	N	T	N	19,64	D	5
14	96811025	T	C	T	D	D	U	D	N	T	N	23,4	D	5
19	57889422	G	C	T	D	P	.	N	L	T	D	20,8	N	4
17	73805968	C	T	.	D	D	D	D	M	.	.	35	D	6
1	63867927	A	G	D	D	D	D	D	H	D	D	28,2	D	10
3	75790880	A	G	D	D	D	.	P	H	T	D	22,6	N	8
6	17605166	T	C	T	D	D	D	D	L	.	D	24,6	D	7
13	114309167	G	A	D	D	D	D	D	M	T	D	25,6	D	8
X	83591810	A	G	D	D	D	D	D	L	T	N	25,9	D	7
X	77365375	A	G	D	P	B	D	D	H	D	D	23,6	D	9
1	202704613	C	G	T	P	P	D	D	L	D	N	25,3	D	7
6	97729204	A	T	D	D	P	D	D	L	T	N	24,7	D	7

9	139994174	G	A	D	D	D	U	D	M	D	D	24,7	D	8
18	77171237	C	A	T	D	D	D	D	M	T	N	25,5	D	6
8	145541635	T	C	.	D	D	D	D	L	.	.	25,8	D	6
17	21318782	G	A	D	B	B	D	D	L	T	D	22,4	D	6
X	107977333	G	A	D	D	P	D	D	M	D	D	22,8	D	9
X	96200528	T	C	T	P	P	N	N	L	T	N	19,7	D	4
1	25153595	G	C	T	P	B	D	D	N	D	N	19,71	D	6
6	108246042	C	G	T	P	P	D	D	L	T	N	27,3	D	6
19	45488503	T	C	T	P	P	D	D	L	.	D	29,5	D	7
15	37188859	G	A	.	.	.	D	A	.	.	.	40	D	4
2	42165754	C	T	P	.	.	.	15,32	N	2
4	38019979	G	A	D	D	D	D	D	M	T	D	29,7	D	8
11	129743683	A	T	D	D	D	D	D	L	.	D	29,4	D	8
18	43223957	G	A	T	P	B	D	D	L	T	N	22,9	D	5
21	30359058	T	A	D	D	P	N	D	M	T	D	26,7	D	7
12	19408017	G	A	D	D	D	N	D	L	T	D	25,6	D	7
6	17605166	T	C	T	D	D	D	D	L	.	D	24,6	D	7
7	45122324	C	A	D	D	P	.	P	L	T	D	24	N	6
17	45669359	T	G	D	D	D	D	D	H	T	D	29,9	D	9
6	32632782	G	A	T	P	B	U	P	L	T	N	0,001	N	2
7	47342713	G	C	D	D	D	N	D	L	D	N	27,5	D	7

15	79586767	G	C	D	P	B	.	N	N	T	N	15,55	D	4
16	56686929	G	T	D	D	P	U	N	.	T	D	19,07	D	6
16	58577327	A	C	D	B	B	.	N	.	T	N	0,049	N	1
21	45820184	A	G	D	D	D	D	D	M	T	D	23,8	D	8
11	55322081	C	T	D	D	P	.	N	L	T	D	17,25	N	5
5	118513069	A	T	T	P	B	N	D	L	T	D	16,92	D	5
7	50444255	A	G	T	P	B	D	D	.	T	N	11,28	D	4
21	45533648	C	T	D	D	D	D	D	M	T	D	28,1	D	8
2	191844593	C	T	D	.	.	.	27,6	D	3
4	48592700	T	C	D	D	D	U	D	M	T	D	25,9	D	7
12	39170101	C	T	T	P	B	D	D	M	T	D	25	D	6

Asmenų su intelektine negalia grupė

Chr	Pozicija	Ref	Alt	SIFT	Polyphen2 HDIV	Polyphen2 HVAR	LRT	Mutation Taster	Mutation Assessor	FATHMM	PROVEAN	CADD	fathmm- MKL coding	GERP++	Patogeninių variantų skaičius
11	1191646	T	C	2,585	N	-6,66	0
6	157203998	T	C	D	.	.	.	25,6	D	5,63	4
11	66211206	C	T	D	D	D	D	D	M	T	D	35	D	4,79	9
19	15380381	T	G	T	B	B	N	D	N	T	N	22,7	D	3,54	4
19	42855568	C	G	T	.	0,578	N	-1,21	0
7	100402540	G	A	.	.	.	N	A	.	.	.	35	N	2,54	4
19	10001778	G	A	T	D	P	U	N	L	D	N	12,52	N	3,9	4
16	4910877	G	C	D	D	P	D	N	M	T	N	24,7	D	0,778	6
16	70268299	G	A	D	D	D	D	D	M	T	D	33	D	5,62	9
18	55228987	C	T	D	D	D	D	A	M	D	D	35	D	5,89	10
3	36831070	G	A	T	B	B	N	N	N	T	N	2,798	N	-5,76	0
7	129203441	A	G	D	D	D	D	D	M	T	D	26,5	D	5,5	9
3	18386559	T	C	D	D	D	D	D	L	T	D	24,4	D	6,05	9
1	6445909	G	A	D	D	D	D	D	L	D	N	26,2	D	4,37	9
4	108749239	G	A	D	D	D	D	D	H	T	D	33	D	5,32	10
9	131621977	A	G	D	D	D	D	D	L	T	D	30	D	5,57	9

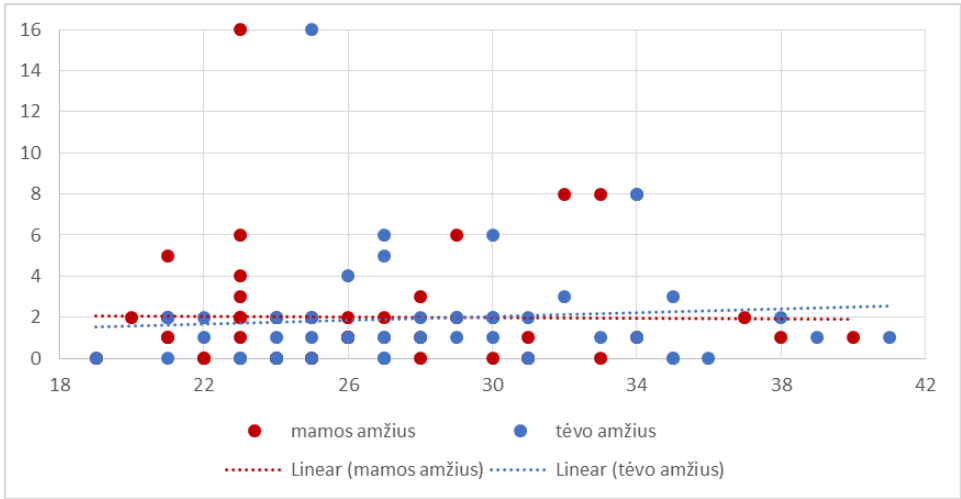
1	109625330	C	T	T	B	B	N	D	N	D	N	24,7	D	4,13	5
16	68166907	G	A	D	D	D	D	D	M	D	D	32	D	5,73	9
3	39098275	C	G	T	P	B	N	N	N	T	N	7,11	N	1,23	1
5	80654917	C	G	T	B	B	.	D	N	D	N	13,43	D	3,14	4
X	154031355	G	A	D	D	D	D	A	L	D	D	26,9	D	5,48	10
1	216564242	A	T	D	D	D	D	D	H	D	D	30	D	4,33	11
6	149378814	C	A	.	.	.	N	A	.	.	.	36	D	6,16	5
17	81941177	C	T	T	B	B	N	N	N	T	N	8,345	N	1,56	0
2	170401293	T	C	D	D	D	D	D	M	T	D	28	D	4,8	9
6	4099363	A	C	T	B	B	N	N	N	T	D	0,382	N	1,09	1
16	3786782	T	G	D	D	P	D	D	M	D	D	24,1	D	5,25	10
2	233613741	G	T	T	D	D	D	D	L	T	D	23,1	D	1,93	7
12	79842733	C	G	D	D	D	D	D	M	T	D	27,6	D	6,16	9
11	10774247	C	G	D	D	D	D	D	M	T	D	31	D	5,82	9
22	22892482	G	A	D	P	P	N	N	L	T	D	22,5	N	2,54	5
18	3090782	C	T	D	P	B	D	D	L	T	D	24,4	D	5,7	9

Asmenų su intelektine negalia sibsų grupė

Chr	Pozicija	Ref	Alt	SIFT	Polyphen2 HDIV	Polyphen2 HV AR	LRT	Mutation Taster	Mutation Assessor	FATHMM	PROVEAN	CADD	fathmm- MKL coding	GERP++	Patogeninių variantų skaičius
16	2056182	G	A	T	.	.	.	N	.	D	N	5,619	N	-2,47	1
2	28604758	C	T	T	D	P	N	P	M	T	N	21,7	N	-5,37	4
4	87615747	A	G	.	B	B	.	P	N	D	N	0,226	N	-1,69	1
6	38872575	C	G	T	D	D	D	D	N	T	D	24,2	D	5,62	8
9	35710880	G	A	D	D	P	N	D	M	T	D	27,6	D	5,76	8
19	50402098	T	G	T	B	P	N	D	L	T	D	20,9	D	3,25	6
19	1234760	C	T	T	.	.	U	N	.	.	N	5,787	N	0,124	0
6	26156583	G	T	T	B	P	D	D	M	T	N	18,33	N	5,11	5
9	101409420	G	A	T	P	B	N	N	N	T	N	22,8	N	4,43	3

8 PRIEDAS

DnM skaičiaus ir tėvų amžiaus koreliacijos grafikas



PADĖKA

Nuoširdžiai dėkoju:

asmenims, dalyvusiems LITGEN ir UNIGENE projektuose;
moksliniam vadovui **akad. prof. habil dr. Vaidučiui Kučinskui** už įkvėpimą mokslui;

prof. dr. Audronei Jakaitienei už pagalbą skaičiuojant *de novo* mutacijų dažnius ir patarimus atliekant statistinius skaičiavimus;

disertacijos Gynimo tarybos pirmininkei **prof. habil. dr. Zitai Aušrelei Kučinskienei** bei nariams **prof. dr. Česlovui Venclovui, prof. habil. dr. Limui Kupčinskui, prof. dr. Loretai Cimbalistienei, prof. dr. Alexandrai Zhernakovai** už įdomius diskusijos klausimus;

recenzentėms **doc. dr. Almai Molytei, dr. Ingridai Domarkienei** ir **dr. Živilei Maldžienei** už pastabas, padėjusias patobulinti šį darbą;

akad. prof. habil. dr. Vaidučiui Kučinskui ir **VU MF BMI ŽMGK mokslininkams**, vykdžiusiems LITGEN ir UNIGENE projektus, už galimybę analizuoti lietuvių populiacijos sveikų asmenų egzomo, asmenų, turinčių intelektualinę negalią, jų sveikų tėvų ir sibsų egzomų sekoskaitos rezultatus;

VšĮ VUL SK MGC kolegoms **doc. dr. Laimai Ambrozaitytei, Daivai Kazlauskaitėi, Dalytei Pliaugo, Vilijai Deveikienei** už bendrą darbą laboratorijoje;

Dokt. Evelinai Siavrienei už eksperimentinę *ARID1B* geno funkcinę analizę, draugišką palaikymą, nuoširdumą ir supratimą;

Doc. dr. Eglei Preikšaitienei už profesinius patarimus, draugišką palaikymą, nuoširdumą ir supratimą;

Dr. Aušrai Morkūnienei už visada gerą nuotaiką;

Už kantrybę, supratimą ir visokeriopą paramą noriu ypač padėkoti savo artimiausiems žmonėms – sūnums **Vytautui Žaibui** ir **Ryčiui**, vyrui **Vytautui**.

Už skatinimą nesustoti – **mamai**.

UŽRAŠAMS

Vilniaus universiteto leidykla
Saulėtekio al. 9, LT-10222 Vilnius
El. p. info@leidykla.vu.lt,
www.leidykla.vu.lt
Tiražas 25 egz.