



VILNIAUS UNIVERSITETO
MEDICINOS FAKULTETO
ŽMOGAUS IR MEDICININĖS GENETIKOS KATEDRA

MAGISTRO DARBAS

LIETUVOS POPULIACIJOS GENETINĖ ANALIZĖ PAGAL DNR KOPIJŲ SKAIČIAUS
POKYČIUS

Magistrantė

KAROLINA MARGYTĖ

(parašas)

Darbo vadovas

Prof. habil. dr. VAIDUTIS KUČINSKAS

(parašas)

Konsultantė

Doktorantė ALINA URNIKYTĖ

(parašas)

VU MF Žmogaus ir medicininės
genetikos katedros vedėjas

prof. (HP) dr. ALGIRDAS UTKUS leidžiama ginti

(parašas)

Darbo įteikimo data _____

Registracijos Nr. _____

2017

**Vilniaus universiteto studijuojančiojo,
teikiančio baigiamąjį darbą,
GARANTIJA**

**WARRANTY
of Vilnius University Student Thesis**

Vardas, pavardė: Karolina Margytė
Padalinys: VU MF Žmogaus ir medicininės
genetikos katedra
Studijų programa: Medicinos genetika
Darbo pavadinimas: Lietuvos populiacijos
genetinė analizė pagal DNR kopijų skaičiaus
pokyčius
Darbo tipas: Magistro darbas

Name and Surname: Karolina Margytė
Faculty: VU MF Department of Human and
Medical Genetics
Study programme: Medical genetics
Thesis topic: Lithuanian population genetic
analysis of copy number variation
Thesis type: Master thesis

Garantuojau, kad mano baigiamasis darbas yra
parengtas sąžiningai ir savarankiškai, kitų asmenų
indėlio į parengtą darbą nėra. Jokių įstatymų
nenumatytų mokėjimų už šį darbą niekam nesu
mokėjęs.
Šiame darbe tiesiogiai ar netiesiogiai panaudotos
kitų šaltinių citatos yra pažymėtos literatūros
nuorodose.

I guarantee that my thesis is prepared in good faith
and independently, there is no contribution to this
work from other individuals. I have not made any
illegal payments related to this work.
Quotes from other sources used in this thesis,
directly or indirectly, are indicated in literature
references.

Aš, Karolina Margytė, patvirtinu (pažymėti)
I, Karolina Margytė, confirm (check)

Patvirtinu, kad baigiamasis darbas yra pateiktas į Vilniaus universiteto studijų informacinę sistemą.
I declare that this thesis is submitted to the Vilnius University Study Information System.

(vardas, pavardė / name,
surname)

(parašas / signature)

(data / date)

Embargo laikotarpis / Embargo

Prašau nustatyti šiam baigiamajam darbui toliau nurodytos trukmės embargo laikotarpį:
I am requesting an embargo of this thesis for the period indicated below:

embargo laikotarpis nereikalingas / *no embargo requested;*

_____ mėnesių / *months [embargo laikotarpis negali viršyti 60 mėn. /an embargo period shall not exceed 60 months].*

Embargo laikotarpio nustatymo priežastis / *reason for embargo:*

.....

(vardas, pavardė / name,
surname)

(parašas / signature)

(data / date)

Darbo vadovo patvirtinimas:

Patvirtinu, kad baigiamasis darbas atitinka reikalavimus ir leidžiu darbą ginti:

I confirm that this thesis has been prepared in accordance with the requirements of students' theses of Vilnius University and can be defended.

(vardas, pavardė / name,
surname)

(parašas / signature)

(data / date)

Katedros (Padalinio) patvirtinimas, kad atspausdintas baigiamasis darbas buvo pateiktas ir užregistruotas:

(vardas, pavardė)

(parašas)

(data)

TURINYS

SANTRUMPŲ SĄRAŠAS.....	5
ĮVADAS	6
1. LITERATŪROS APŽVALGA.....	8
1.1 DNR kopijų skaičiaus pokyčiai.....	8
1.2 DNR Kopijų skaičiaus pokyčių susidarymo principai	10
1.3 DNR kopijų skaičiaus pokyčių nustatymas.....	12
1.3.1 PennCNV programa	12
1.3.2 QuantiSNP programa	14
1.3.3 cnvPartition programa	15
1.4 Genomo kitimų duomenų bazė	17
2. TYRIMO METODAI IR IMTIS	18
2.1 Tiriamoji medžiaga	18
2.2 Tyrimo eiga.....	19
2.3 Tyrimo metodai.....	20
2.3.1 DNR išskyrimas	20
2.3.2 Genotipavimas VNP Illumina lustais	22
2.3.3 Kopijų skaičiaus pokyčių aptikimas PennCNV programa	26
2.3.4 DNR kopijų skaičiaus pokyčių regionų konstravimas	26
2.3.5 Kopijų skaičiaus pokyčių regionų anotacija.....	27
2.3.6 DNR kopijų skaičiaus pokyčių duomenų apdorojimas Excel skaičiuokle.....	28
2.3.7 Kopijų skaičiaus pokyčių regionų atvaizdavimas PhenoGram programine įranga.....	28
3. TYRIMO REZULTATAI.....	29
4. REZULTATŲ APTARIMAS.....	37
5. IŠVADOS	40
SANTRAUKA.....	41
SUMMARY.....	42
PADĖKA	43

LITERATŪROS SAĶAŠAS	44
PRIEDAI.....	48

SANTRUMPŲ SĄRAŠAS

BAF – B alelio dažnis (angl. *B Allele Frequency*, BAF)

BIR – trūkių indukuotos replikacijos (angl. *break-induced replication*)

DNR – deoksiribonukleorūgštis

FISH – fluorescencinė *in situ* hibridizacija

FoSTeS – reparacinės šakutės sustojimas ir šablono perjungimas (angl. *Fork Stalling and Template Switching*)

HMM – paslėptas Markovo modelis (angl. *Hidden Markov Model*, HMM)

Kb – kilobazė

KSP – DNR kopijų skaičiaus pokytis

KSPR – DNR kopijų skaičiaus pokyčių regionai

LRR – R Logaritmo santykis (angl. *log R Ratio*)

LGH – lyginamoji genomo hibridizacija

Mb – megabazė

NGS – nehomologiškas DNR galų sujungimas

PGR – polimerazės grandininė reakcija

VNP – vieno nukleotido polimorfizmas

ĮVADAS

Pagrindiniai žmonių genomo pokyčių šaltiniai yra vieno nukleotido polimorfizmai (VNP) (angl. *Single-Nucleotide Polymorphisms*, SNPs), mažos insercijos ar delecijos ir dideli pokyčiai. Dideli pokyčiai gali būti kopijų skaičiaus pokyčiai (KSP), t.y. chromosomų segmetų delecijos ar duplikacijos arba neutralūs KSP, tokie kaip inversijos arba chromosominės translokacijos. KSP apibrėžiami kaip didesni nei 1 Kb dydžio chromosominiai kopijų skaičiaus pokyčiai, tokie kaip duplikacijos ar delecijos lyginant su referentiniu genomu (1, 2).

Pirmieji KSP tyrimai parodė didesnę genomo kintamumą nei buvo tikėtasi sveikuose asmenyse, nemaža dalis žmonių genomų gali skirtis vienas nuo kito (3). 2006 metais buvo sudarytas pirmasis bendros populiacijos genomo KSP fizinis žemėlapis ir paaiškėjo, kad iki 12 % žmogaus genomo gali skirtis nuo referentinio dėl KSP (2). Tačiau remiantis genomo kitimų duomenų bazės naujausiais duomenimis žmogaus genomas gali skirtis nuo referentinio dėl KSP įvairovės net iki 22 % (4). KSP yra labiausiai paplitusi genomo įvairovės forma, dydžio atžvilgiu, tarp individų.

Neseniai visas dėmesys buvo sutelktas identifikuoti ligas susijusias su VNP iš plataus masto genomo asociacijos tyrimų (angl. *genome-wide association studies*, GWAS) duomenų. Patobulėjus genotipavimo metodams ir KSP nustatymo algoritmams, naujausi tyrimai nustatė daug KSP gemalinėse ląstelėse, kurie apima genetines sritis sukeliančias ligas, tokias kaip infekcijas, vėžį, autoimunines ir neuropsichiatrines ligas (4).

Yra atlikti moksliniai tyrimai, kurie analizavo KSP HapMap mėginiuose ar didelių populiacijų imtyse, tačiau žinios apie KSP charakteristikas tarp unikalių populiacijų ar populiacijų viduje nėra išsamios. Be to, rezultatai apie KSP iš skirtingų tyrimų yra riboti dėl duomenų konsolidavimo. Taip pat skirtingos etninės grupės (unikalios populiacijos) pateikia skirtingus KSP genome, kurie gali prisidėti prie fenotipo variacijos ir skirtingo polinkio ligoms. Pacientų ir bendros žmonių populiacijos referentinis KSP sąrašas gali padėti išaiškinti tikslią klinikinę interpretaciją, pacientams, kuriems buvo nustatyti KSP. Esamose KSP duomenų bazėse nėra patalpinti specifinių populiacijų KSP duomenys (5).

Šio darbo pagrindinis tikslas buvo nustatyti KSP, jų charakteristikas, įvairovę ir pasiskirstymą lietuvių populiacijoje.

Darbo tikslas: nustatyti DNR kopijų skaičiaus pokyčius iš plataus masto genotipavimo duomenų ir įvertinti jų charakteristikas ir pasiskirstymą lietuvių populiacijoje.

Tiksliui įgyvendinti išsikelti uždaviniai:

1. įdiegti ir įsisavinti kopijų skaičiaus pokyčių aptikimo programą *PennCNV*;

2. nustatyti DNR kopijų skaičiaus pokyčius *PennCNV* programa bei sukonstruoti kopijų skaičiaus pokyčių regionus *cnv_tool_v1.4.2.py* programa;
3. anotuoti nustatytus DNR kopijų skaičiaus pokyčių regionus *SG-ADVISED* serveryje;
4. atlikti nustatytų DNR kopijų skaičiaus pokyčių bei jų regionų įvairovės ir pasiskirstymo analizę lietuvių populiacijoje;
5. atlikti lietuvių populiacijoje nustatytų kopijų skaičiaus pokyčių palyginimo analizę su kitomis populiacijomis.

Magistro darbo metu savarankiškai atlikti darbai:

- Žaliavinių duomenų iš 770K *HumanOmniExpress-12 v1.1 Illumina* lustų apdorojimas ir išvesties bylų sukūrimas *GenomeStudio v2011.1* programa;
- DNR kopijų skaičiaus pokyčių nustatymas, *PennCNV v2014May07* programa, 293-ims Lietuvos populiacijos asmenims iš LITGEN projekto;
- DNR kopijų skaičiaus pokyčių regionų konstravimas *cnv_tool_v1.4.2.py* programa;
- DNR kopijų skaičiaus pokyčių regionų atvaizdavimas ideogramoje;
- DNR kopijų skaičiaus pokyčių regionų anotacija *SG-ADVISED* serveryje;
- DNR kopijų skaičiaus pokyčių regionų analizė Genomo kitimų duomenų bazėje;
- Duomenų aprašomoji statistinė analizė ir atvaizdavimas *Excel 2016* skaičiuokle.

Darbo naujumas ir reikšmė

Nuo pirmųjų mokslinių straipsnių apie KSP praėjo 13 metų ir kiekvienais metais susidomėjimas KSP ir jų charakteristikomis, tokiais kaip pasiskirstymu genome ir jų dažniu skirtingose populiacijose, vis augo, kadangi KSP įvairovė nėra galutinai ištirta.

KSP yra pagrindinis genetinis polimorfizmas prisidedantis prie genetinės įvairovės ir žmonių evoliucijos. Klinikinis KSP taikymas diagnostikos tikslams daugiausiai priklauso nuo turimų duomenų apie KSP toje populiacijoje. Tam tikrų populiacijų KSP yra prieinami duomenų bazėse ir tai padeda atskirti neaiškios kilmės KSP nuo gerybinių ar patologinių KSP. Ankstesni tyrimai apie KSP pasiskirstymą keletose viso pasaulio populiacijų parodė, kad didelė dalis KSP yra specifiniai tik tam tikroms populiacijoms (6). Todėl šio darbo pagrindinis tikslas yra nustatyti KSP, jų charakteristikas, įvairovę ir pasiskirstymą lietuvių populiacijoje.

1. LITERATŪROS APŽVALGA

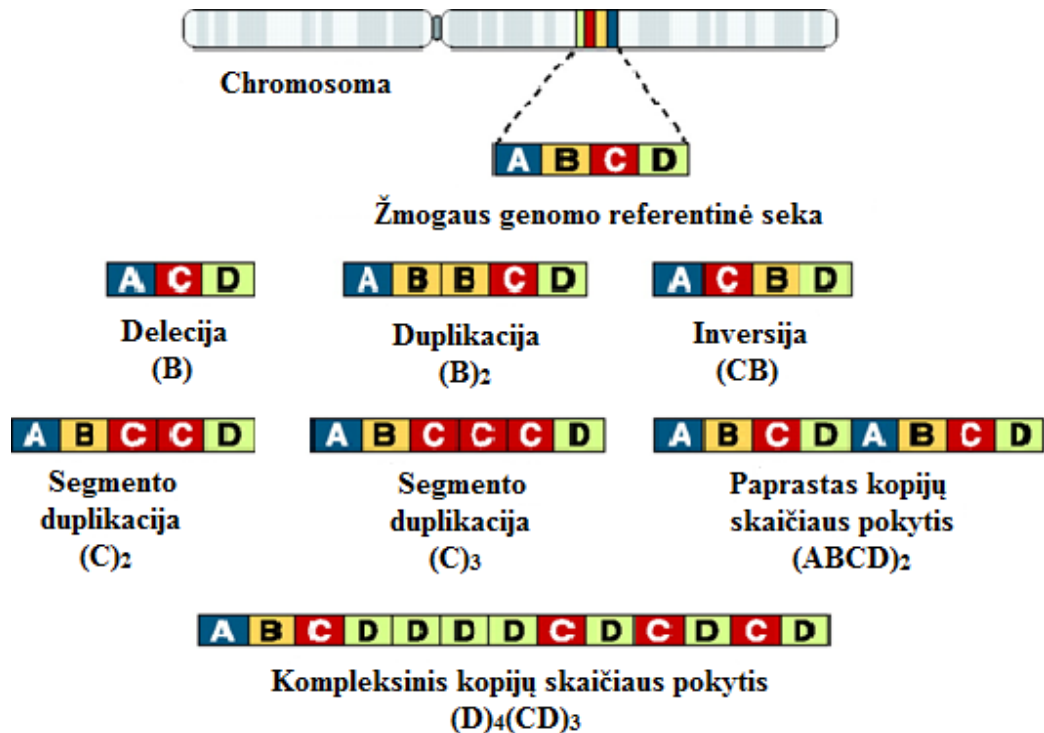
1.1 DNR kopijų skaičiaus pokyčiai

Du moksliniai tyrimai 2004 m. paskelbė, kad DNR kopijų skaičiaus pokyčiai sudaryti iš didelių genominių segmentų egzistuoja tarp sveikų asmenų (3, 7). Skirtingų žmonių genomai skiriasi apie 0,1 % (8). Žmogaus genetinė įvairovė susidaro nuo didelių, mikroskopiškai matomų chromosominių anomalijų iki vieno nukleotido pokyčių. DNR kopijų skaičiaus pokyčiai (KSP) yra chromosomos segmento kopijų skaičiaus pokytis svyruojantis nuo kelių šimtų bazių porų (bp) iki megabazių (Mb) ir skiriasi nuo tikėtosi kopijų skaičiaus, kuris yra referentiniame genome, dėl delecijų, duplikacijų, insercijų ar invertuotų duplikonų (žr. 1.1 pav.) (9). KSP dažniausiai randami žmonių, primatų, graužikų ar nariuotakojų, tokių kaip *Drosophila melanogaster* genomuose (2). Vidutiniškai kiekvieno žmogaus genome yra daugiau nei 1000 KSP, kurie sudaro ~4 milijonus bp (10).

Identifikuota, kad žmonėse, daugiau kaip 6,2 milijonai skirtingų KSP persidengia su ~500 000 genominių regionų (11). Jie reikšmingai prisideda prie genetinių pokyčių, kurie apima daugiau nukleotidų genome, nei vieno nukleotido polimorfizmai (VNP) ir tai gali nulemti apie 0,8 % genominių skirtumų tarp dviejų individų (12). Bet to, KSP pasižymi didesniu mutacijų dažniu nei VNP (13). KSP yra viena dažniausių genomo įvairovės formų žmonių populiacijoje (2). Apie 5-12 % žmogaus genomo, įskaitant tūkstančius genų, gali turėti skirtingą kopijų skaičių ir šie pokyčiai gali būti *de novo* (pasireiškia pirmą kartą tėvų gemalinėse ląstelėse) arba paveldėti iš tėvų (14). Įdomu yra tai, kad ~20-25 % nustatytų KSP šimpanzėse ir makakose persidengia su žmogaus genome nustatytais KSP (15). Kadangi KSP gali persidengti su genais, jie gali keisti genų dozę, gali sutrikdyti koduojamas sekas ar pakeisti genų raišką. Taip pat gali įtakoti evoliucinius pokyčius (16). Tokie apibūdinimai kaip retas ar dažnas KSP, taikomi nurodant varianto dažnį, o ne genetinės srities būseną. Pagal susitarimą, reti KSP yra tie, kurių dažnis yra mažiau nei 1 % populiacijoje, o dažni tie, kurių KSP dažnis populiacijoje yra didesnis nei 1 % (17).

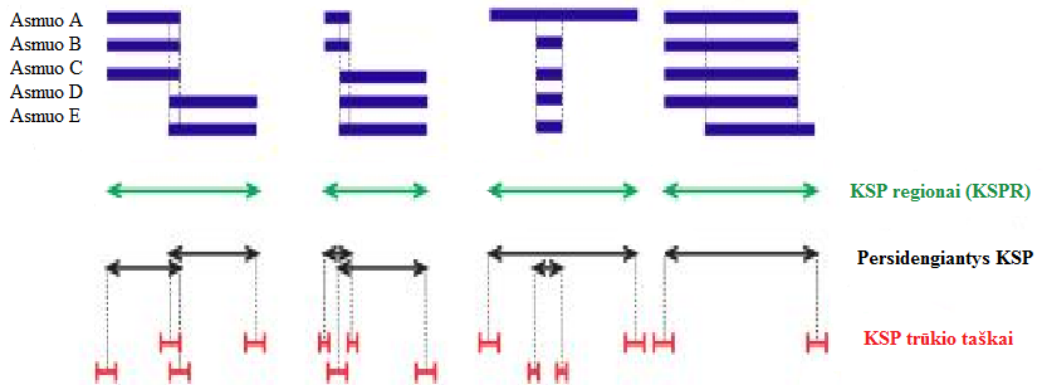
Atsižvelgiant į asociacijas tarp ligos ir KSP buvo pasiūlyti du skirtingi struktūrinių pokyčių modeliai. Pirmasis susijęs su dideliais struktūriniais pokyčiais (paprastai padidėja arba sumažėja keletu šimtų kilobazių porų), kurie yra reti populiacijoje (<1 %), tačiau tai lemia didelę dalį ligų, pvz., neurologinius ir neurokognityvinius sutrikimus. Antrasis apima kelių kopijų genų šeimas (angl. *MultiCopy gene families*), kurios varijuoja kopijų skaičiumi ir prisideda prie ligos jautrumo, kaip buvo pastebėta iš požymių susijusių su imuniteto genų funkcijomis (18). Šiuos KSP poveikius yra sunku suprasti ir nuspėti, tačiau jie būdingi daugeliui ligų ir įtakoja farmakologinį atsaką, pavyzdžiui, *CYP2D6* KSP atveju (19). Be to, dažni ir reti KSP yra siejami su polinkiu į daugelį genetinių ligų įskaitant įvairius vėžinius susirgimus, autoimunines ligas, šizofreniją ir autizmą

(20). Žmonių reti KSP yra susiję su įvairiais sutrikimais tokiais kaip apsigimimai, neurologiniai sutrikimai ir polinkiu į vėžį (21). Dažni KSP prisideda prie tokių ligų, kaip ŽIV, maliarija, lėtinės obstrukcinės plaučių ligos ir Krono liga. Dėl jų poveikio žmonių ligoms KSP gali būti naudojami tiek diagnozuojant tam tikras ligas, tiek jas gydant (22).



1.1 pav. Delecijų ir duplikacijų lyginimas su referentiniu genomu. Nehomologinės rekombinacijos tarp sekų (segmentų duplikacijos, mažas kopijų pasikartojamumas, ar duplikonai), tai yra daugiau kaip 90 % visų delecijų ir duplikacijų priežastis. Invertuoti duplikonai gali sukelti genetinės medžiagos inversijas, tačiau kitokio tipo pokyčiai gali atsirasti priklausomai nuo duplikono sudėtingumo, kurie dažnai turi lygegrečiai orientuotas sekas ir tokiu atveju gali pasireikšti delecijos arba duplikacijos.

Šalia esantys KSP dažnai yra persidengiantys ir jeigu jie persikloja >40 %, tada jie formuoja persidengiančių KSP regioną (KSPR) (angl. *Overlapping copy number variant regions*, CNVRs) (23). KSPR vertinami kaip persidengiantys jeigu toje srityje yra du aiškūs skirtingų ilgių variantai, duplikacija ar delecija, kurie abu persidengia ir yra nustatyti keletui individų (žr. 1.2 pav.). Manoma, kad KSPR gali apimti >13 % (~416 Mb) viso žmogaus genomo (24). Šie regionai gali persidengti su genais, kurie galimai sukelia tam tikras ligas (25).



1.2 pav. Penkių asmenų keturi persidengiantys KSP regionai pavaizduoti mėlyna spalva, o punktyrinės linijos rodo persidengimą. Persidengiantys KSP ir sudarantys KSPR pavaizduoti žalia spalva. Nepriklausomi greta esantys KSP vaizduojami juoda spalva, raudona spalva žymimi KSP trūkio taškai.

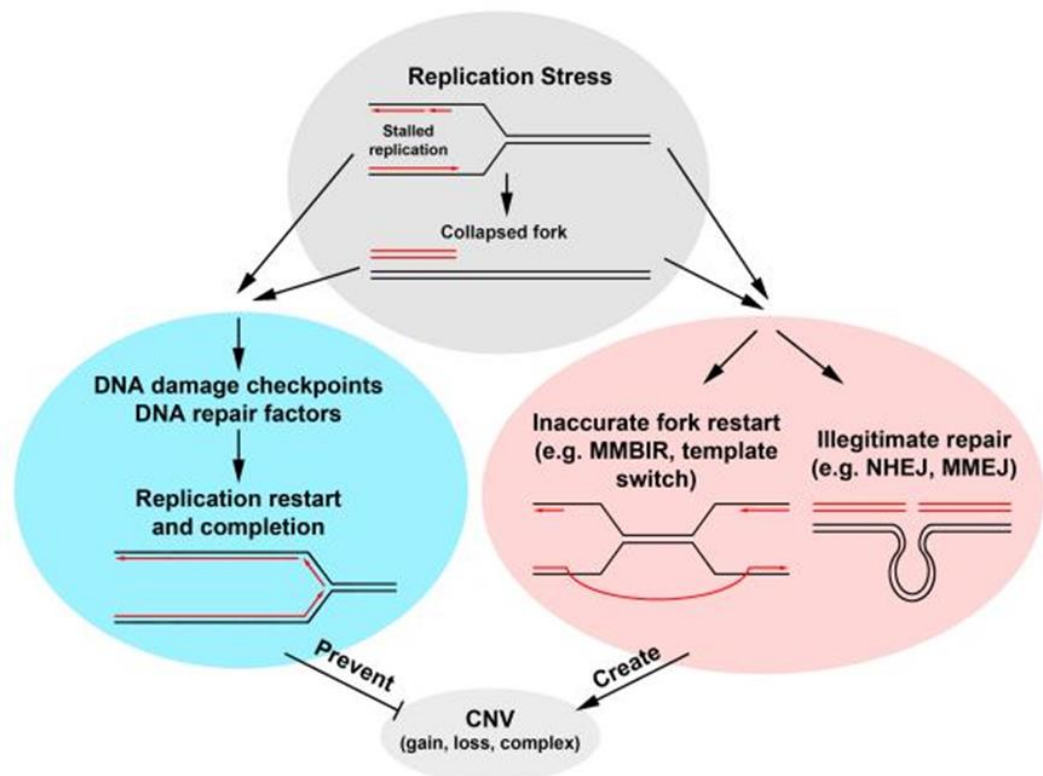
1.2 DNR Kopijų skaičiaus pokyčių susidarymo principai

Pasikartojantys KSP (angl. *Recurrent CNVs*) su tais pačiais trūkio taškais yra delecijos ir reciprokinės duplikacijos, kurios manoma įvyksta nevienodu mejozės ar ne alelinės homologinės rekombinacijos (NAHR) metu, paprastai nesusiligiuojant didelėms gretimoms segmentinėms duplikacijoms ar pasikartojančioms sekoms. Todėl šie KSP atsiranda tokiu pačiu principu kaip ir ankščiau nustatytos žmogaus mikrodelecijos/mikroduplikacijos lemiančios Charcot-Marie Tooth ir Prader-Willi sindromus, pasikartojantys KSP gali išplėsti šių sindromų klasę.

Mažiau žinomi molekuliniai mechanizmai dėl kurių susidaro nepasikartojantys KSP (angl. *non-recurrent CNVs*). Įdomu tai, kad duomenys iš sekų su trūkio taškais ir eksperimentinės sistemos yra iš mitozinių ląstelių, o ne mejozinių. Klaidos DNR replikacijos, o ne mejozės homologinės rekombinacijos metu sukelia pastebimus trūkio taškų susijungimus (žr. 1.3 pav.). Be to, replikacijos inhibitoriai, kuriuos sukuria „replikacinis stresas“ gali sudaryti panašius KSP, kaip buvo aptarta anksčiau.

Paprasčiausias modelis, kuris atitinka abiejų trūkio taškų sekų duomenis ir indukuotas replikacinio streso yra tas, pagal kurį yra pakitęs suirusių replikacinių šakučių atkūrimas. Taigi molekuliniai mechanizmai neabejotinai dalyvauja įvairiuose procesuose, tokiuose kaip ląstelės ciklo patikros taškai, atkuriant normalią replikaciją ir neleidžiant susidaryti KSP, taip pat kaip DNR replikacijos ir reparacijos faktoriai, kurie užkerta kelią KSP. Vėliau buvo pasiūlyta daugiau susidarymo mechanizmų. Šių mechanizmų principai yra dviejų ar daugiau dvigubų grandinių trūkių nehomologinių galų susijungimas (NGS), arba mikrohomologinių galų susijungimas (angl. *microhomology-mediated end-joining*) susijęs su šablono pakeitimu. Svarbiausias iš vėlesnių mechanizmų modelių yra „reparacinės šakutės sustojimas ir šablono perjungimas“ (angl. *Fork Stalling and Template Switching*, FoSTeS), kurį pasiūlė Jennifer A. Lee ir kiti autoriai. Kitas mechanizmas, pasiūlytą Hastings PJ ir kitų autorių 2009 m., susijęs su mikrohomologų

dalyvavimu trūkių sukėltoje replikacijoje (angl. *microhomology-mediated break-induced replication*, MMBIR) (26, 27). Šie modeliai yra grindžiami šablono perjungimo principu pasiūlyti streso sukeltoms amplifikacijoms *E.coli* srityse, kuriose sustojusi replikacija ir mielėse trūkių indukuotos replikacijos (angl. *break-induced replication*, BIR) daugiausia aprašomos mielėse. FoSTeS modelyje replikacijos šakutės susiduria su mažais kopijų skaičiaus pasikartojimais (angl. *low-copy repeats*) ar sritimis kurias sunku replikuoti, tuomet šakutės yra linkę sustoti, o to pasekoje yra pereinama į kitą aktyvią šakutę, apeinant DNR pažaidą ar atnaujinant replikaciją. MMBIR modelis remiasi reparacijos šablono perjungimu vienpusiam dvigrandžiam trūkiui, susiformavusiam suirus replikacinėms šakutėms mikrohomologinėse srityse, o ne ilgesnėse homologinėse srityse, kurios dažniausiai stebimos BIR sukeltuose pakitimuose. Šie molekuliniai mechanizmų modeliai paaiškina KSP susidarymo principus. Vis dėlto, didžioji dalis duomenų su šiais modeliais, gaunami iš trūkio taškų susijungimo sekose, kur KSP randama normaliuose genomuose ar atsiradę pacientams. Yra aiškus poreikis tiksliems visų galimų modelių bandymams, kurie paaiškintų KSP formavimąsi (28).



1.3 pav. Galimi KSP susidarymo mechanizmai.

Sąlygos sukeliančios ląstelių replikacijos stresą pavėlina replikaciją, o tai sukelia replikacinių šakučių suirimą (angl. *fork collapse*), ko pasekoje susidaro dvigrandis trūkis viename DNR sekos gale (viršutinis apskritimas). Kiekviena iš šių struktūrų aktyvuoja DNR pažaidų reparacijos mechanizmus, kurie turėtų atkurti replikaciją ties ta vieta, kurioje buvo pažeista replikacinė šakutė

(apskritimas kairėje). Reparacijos mechanizmai padeda apsaugoti genomo vientisumą, užkirsti kelią KSP ir kitų struktūrinių variantų susidarymui. Vis dėlto, sustojusi replikacija ar replikacinių šakučių suirimas taip pat gali įtakoti naujus arba alternatyvius reparacinius kelius. Sustojusi replikacinė šakutė gali pradėti replikaciją iš naujo, toliau už pažeidimo, naudojant šablono perjungimą arba MMBIR kelią, susidarant KSP. Ilgo nuotolio dviejų DNR trūkių galų susijungimai taip pat gali vesti prie didelių delecijų sekoje, susidarant KSP (apskritimas dešinėje). Tikimasi, kad mutacijos, kurios slopina ląstelių gebėjimą tinkamai reaguoti į replikacinių šakučių sustojimą ar suirimą, gali sukelti padidėjusį KSP dažnį (28).

1.3 DNR kopijų skaičiaus pokyčių nustatymas

Fizinis žemėlapis ir DNR seka gauta Žmogaus Genomo Projekto (angl. Human Genome Project) metu dalinai pakeitė klasikinius citogenetinius metodus suteikiant galimybę nustatyti nesubalansuotus submikroskopinius pokyčius visame genome. Technologiniai pasiekimai, kaip lyginamoji genomo hibridizacija (LGH) pirmą kartą leido nustatyti KSP mokslinių tyrimų metu 1990-aisiais m. (29, 30). Tolimesni pasiekimai buvo genominiai lustai, labiau naudojami klinikinėje diagnostikoje. Šie pasiekimai apėmė genominius klonus (pvz., bakterijų dirbtinės chromosomos, *BACs*), kurių pagalba buvo galima nustatyti vienos kopijos (*angl. single-copy*) delecijas ar duplikacijas. Po to genominiai klonai buvo pakeisti sintetiniais oligonukleotidais tam, kad būtų galima viso genomo mastu tiksliau nustatyti trūkių taškus (31).

VNP lustai buvo naudojami genotipuoti šimtus tūkstančių VNP visame genome. Dabar jie naudojami ir KSP nustatymui remiantis žymens hibridizacijos signalu kiekviename luste. Plačiausiai naudojami VNP lustai yra *Illumina* ir *Affymetrix*. Lustų plėtra paskatino skaičiavimo analizės metodikos pažangą kiekybinės biologijos srityje. Yra sukurta įvairių KSP nustatymo įrankių iš VNP genotipavimo duomenų (32). Dažniausiai naudojami yra *PennCNV*, *QuantiSNP* ir *cnvPartition*.

1.3.1 *PennCNV* programa

PennCNV yra laisvai prieinama programinė įranga KSP nustatymui iš VNP genotipavimo duomenų. *PennCNV* programa galima analizuoti duomenis gautus iš VNP *Illumina* ir *Affymetrix* lustų. Tinkamai paruošus duomenų formatą taip pat galima analizuoti kitų VNP lustų duomenis. *PennCNV* paketas naudoja paslėptą Markovo modelį (angl. *Hidden Markov Model*, HMM), kuriuo remiantis yra teigiama, kad analizuojamų duomenų pasiskirstymas priklauso nuo nestebimų/paslėptų būsenų. Buvo pasitelktos penkios paslėptos būsenos, kurios atspindi galimus KSP: abiejų kopijų delecijas (0 kopijų), vienos kopijos deleciją (1 kopija), normalią būseną (2 kopijos), vienos kopijos duplikaciją (3 kopijos) ir dviejų ar daugiau kopijų duplikacijas (4 ir daugiau kopijų) (žr. 1.1 lentelę). KSP didesni nei keturios kopijos dažniausiai neatskiriami nuo

keturių kopijų signalo intensyvumo, tad tokie KSP yra priskiriami prie 4 kopijų (1). Siekiant tiksliau pademonstruoti signalo intensyvumą duomenyse, šiame pakete taip pat buvo įdiegtas R logaritmo santykis (angl. *log R Ratio*, LRR) ir B alelio dažnis (angl. *B Allele Frequency*, BAF). LRR tai normalizuota signalo intensyvumo reikšmė, kuri yra apskaičiuojama kaip normalizuoto kiekvienos genetinės srities signalo intensyvumo suma su tikėtiniu signalo intensyvumu. BAF yra pagrįstas santykiniu signalu tarp fluorescuojančių alelių žymenų (angl. *probes*), o jo reikšmės gali būti 0 (AA), 0,5 (AB) arba 1 (BB). Be to, *PennCNV* apima populiacijos alelių dažnį kiekvienam VNP žymeniui ir atstumus tarp gretimų VNP. Keletas tyrimų parodė, kad KSP yra paveldimi ir pasiūlė naudoti šeimos arba triadų informaciją (2, 33). Triadų informacijos panaudojimas nustatant KSP leidžia lengviau aptikti naujus pokyčius probandui. Todėl buvo pasitelktas Bajeso metodas, su kuriuo būtų galima naudoti šeimos triadų informaciją galutiniam KSP patvirtinimui ir KSP ryšių nustatymui. Tikslus KSP ribų nustatymas atliekamas identifikuojant trūkio taškus naudojant polimerazės grandininę reakciją (PGR) arba atliekant pakartotinę sekoskaitą (34).

PennCNV turi daugybę analizės galimybių. Viena svarbiausių yra ta, kad galima naudoti šeimų triadų duomenis. *PennCNV* programa taip pat turi didelį skaičių pasirinkimų leidžiančių platesnę gautų rezultatų analizę pvz., palyginti gautus duomenis su žinomų genų bibliotekomis ar pakeisti formatą, kuris būtų tinkamas kitoms programoms, tokioms kaip *BeadStudio's Chromosome Browser* ar UCSC (angl. *University of California Santa Cruz*, <http://www.genome.ucsc.edu/>) (35).

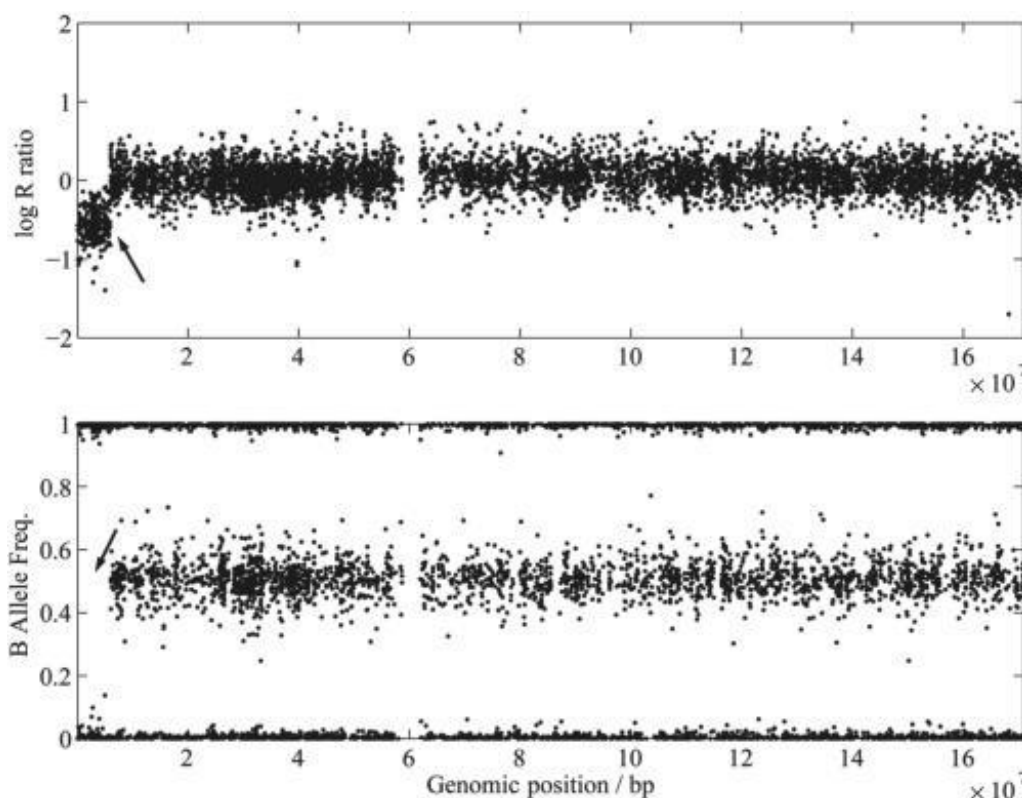
1.1 lentelė. Penkios paslėptos būsenos.

Bendras kopijų skaičius	Genotipas	Apibūdinimas (autosominiai žymenys)
0	Nulinis	Abiejų kopijų delecija
1	A, B	Vienos kopijos delecija
2	AA, AB, BB	Normali būseną
3	AAA, AAB, ABB, BBB	Vienos kopijos duplikacija
4	AAAA, AAAB, AABB, ABBB, BBBB	Dviejų ar daugiau kopijų duplikacija

Šiame darbe *PennCNV* programinė įranga buvo pasirinkta todėl, kad ji sujungia informaciją, įskaitant BAF, LRR, žymens signalo intensyvumą, informaciją identifikuojant KSP panaudojant HMM algoritimą (36). Taip pat lyginant su kitomis programomis, *PennCNV* programa nustato mažiausiai klaidingai teigiamų ir klaidingai neigiamų KSP, pasižymi aukštu jautrumu ir specifiškumu ir nustato mažesnius KSP, nei kitos programos (37).

1.3.2 *QuantiSNP* programa

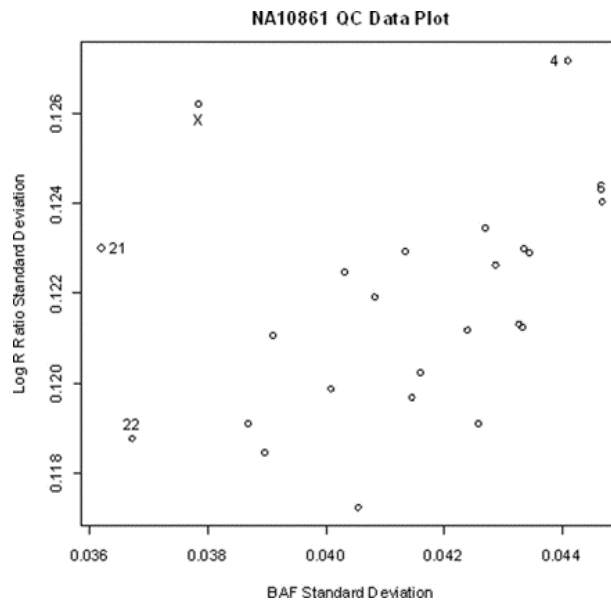
QuantiSNP yra laisvai prieinama programinė įranga KSP nustatymui iš VNP *Illumina* genotipavimo duomenų. Ji taip pat naudoja HMM ir 5 paslėptas būsenas. Tačiau *QuantiSNP* programa LRR ir BAF vertina kiekvieną atskirai, o *PennCNV* programa juos vertina kartu. Paprastai LRR signalo intensyvumo nukrypimai nuo 0, didesni arba mažesni, keičia subalansuotą kopijų skaičių, o BAF nukrypimus nuo 0, 0,5 ar 1 (aleliai AA, AB, BB) taip pat gali sukelti KSP (38). LRR ir BAF pavyzdys pateikiamas 1.4 paveiksle.



1.4 pav. 6 chromosomos duomenys. Viršuje pateikiama LRR, apačioje BAF, šios reikšmės pateikiamos kiekvienam VNP žmeniui, kiekvienoje chromosomoje. Iš LRR matoma delecija, kuri pažymėta rodykle, o iš BAF matomas heterogeniško praradimas ties heterozigotine būseną (0,5), kuri taip pat pažymėta rodykle.

QuantiSNP algoritmo standartinė išvesties byla suteikia sąrašą nustatytų pokyčių, glaustas statistines išvadas apie atliktą kokybės kontrolę bei Bajeso faktorių su prognozėmis. Jos leidžia vartotojui vertinti atvejus ir apskaičiuoti bei nustatyti trūkio vietas. Duomenų kokybės kontrolė yra labai svarbus analizės etapas. Duomenys su dideliu signalo triukšmu dažnai sukelia klaidingai teigiamus rezultatus, todėl patariama atlikti griežtą patikrą šiame etape ir pašalinti klaidingus duomenis. Tam yra naudojamas Bajeso faktorius kiekvienam nustatytam KSP regionui. Tai sustiprina tikimybę, kad KSP tam tikrame regione yra prieš nulinę hipotezę, kuri teigia, kad toje srityje pokyčių nėra. Kuo didesnė Bajeso faktoriaus vertė, tuo stipresnis įrodymas, kad tame regione yra KSP. Patariama, kad Bajeso faktoriaus vertė būtų didesnė nei 10 (39). *QuantiSNP*

išvesties byla leidžia vartotojui atvaizduoti BAF vidurkius ir standartinius nuokrypius chromosomoms ar mėginiams (žr. 1.5 pav.).



1.5 pav. Grafiniai *QuantiSNP* kokybės kontrolės duomenys. Svarbu naudoti kokybės kontrolę iš algoritmo duomenų, kad būtų galima pašalinti klaidingus mėginius, kurie nebūtų aptinkami atliekant standartinę genotipavimo analizę. Atvaizdavimas rodo BAF įverčius kiekvienai chromosomai iš analizuoto mėginio numeriu NA10861. Galima matyti kad 4 ir X chromosomos yra atskirtos.

Vartotojai gali modifikuoti parametrus reikalingus jų analizei ir duomenų rinkiniui, pavyzdžiui, keičiant ilgio parametrus galima tiksliau aptikti įvairaus dydžio pokyčius būtent tam mėginių rinkiniui. Vėlesnė *QuantiSNP* versija yra lankstesnė. Ja galima analizuoti duomenis ne tik gautus iš standartinių *Illumina Infinium* lustų, bet ir *Affymetrix* duomenis, taip pat buvo įrodytas tikslumas analizuojant *Illumina GoldenGate* duomenis, kuriuose VNP padengimas yra tinkamas (35).

1.3.3 *cnvPartition* programa

cnvPartition programos tikslas yra identifikuoti genomo sritis, kuriose yra pakitęs DNR kopijų skaičius, panaudojant VNP *Illumina* lustų išvesties bylas, kuriose yra LRR ir BAF duomenys. Kadangi LRR yra žymens intesyvumas susietas su tikėtiniu intesyvumu, tad bet kokie nukrypimai nuo 0 reikšmės yra KSP įrodymas. BAF yra hibridizuoto mėginio proporcija, kuris turi B alelį kaip nurodyta *Infinium* tyrime. Esant normaliam mėginiui, diskretus BAF 0.0, 0.5 ir 1.0 yra tikėtinas kiekvienai sričiai (prilygsta AA, AB ir BB).

Nukrypimai nuo šių verčių rodo pakitusį kopijų skaičių. Pavyzdžiui, jeigu tam tikros srities BAF yra 0,66, tai gali reikšti, kad yra dvi B alelio kopijos ir viena A alelio kopija. Analizuojant LRR ir BAF kartu yra gaunami tikslesni rezultatai aptinkant KSP.

Vertinant kopijų skaičių labai svarbu atsižvelgti į *cnvPartition* modelio LRR ir BAF reikšmes kiekvienam iš 14 galimų kopijų skaičiaus pokyčių tipų (žr. 1.2 lentelę).

1.2 lentelė. *cnvPartition* programos LRR ir BAF pasiskirstymo parametrai, vidurkis ir standartinis nuokrypis kiekvienam aleliui būdingam kopijų skaičiui (DD-dviguba delecija).

Genotipas	Kopijų skaičius	LRR vidurkis	LRR standartinis nuokrypis	BAF vidurkis	BAF standartinis nuokrypis
DD	0	-5	2	NA	NA
A	1	-0,45	0,18	0	0,03
B	1	-0,45	0,18	1	0,03
AA	2	0	0,18	0	0,03
AB	2	0	0,18	0,5	0,03
BB	2	0	0,18	1	0,03
AAA	3	0,3	0,18	0	0,03
AAB	3	0,3	0,18	1/3	0,03
ABB	3	0,3	0,18	2/3	0,03
BBB	3	0,3	0,18	1	0,03
AAAA	4	0,75	0,18	0	0,03
AAAB	4	0,75	0,18	0,25	0,03
ABBB	4	0,75	0,18	0,75	0,03
BBBB	4	0,75	0,18	1	0,03

Modeliuojant DNR kopijų skaičiaus pokyčius šiuo būdu, galima apskaičiuoti preliminarius KSP lyginant kiekvieną tirtą genetinę sritį su visomis 14-a LRR ir BAF vertėmis. Tai yra, stebint tam tikras LRR ir BAF vertes iš duotų 14-os, galima apskaičiuoti savąją. Pavyzdžiui, norint apskaičiuoti tikėtinas LRR ir BAF vertes duotam AAB (L_{AAB}) genotipui naudojam formulę (AAB parametrai standartiniam normaliajam tankiui yra duoti 1.2 lentelėje):

$$L_{AAB} = \frac{1}{0,18\sqrt{2\pi}} \exp - \frac{(LRR - 0,3)^2}{2(0,18)^2} + \frac{1}{0,03\sqrt{2\pi}} \exp - \frac{(BAF - 1/3)^2}{2(0,03)^2}$$

Tikimybės gali būti paskaičiuojamos ir homozigotinėms delecijoms (L_{DD}). Homozigotinei delecijai būdingos labai mažos LRR reikšmės (< 1), tačiau BAF gali būti bet kuri vertė nuo 0 iki 1. Todėl homozigotinių delecijų tikimybės skaičiuojamos pagal formulę:

$$L_{DD} = \frac{1}{2\sqrt{2\pi}} \exp + \frac{(LRR - (-5))^2}{2(2)^2}$$

Galiausiai visos keturios KSP reikšmės yra susumuojamos:

$$L_0 = L_{DD}$$

$$L_1 = L_A + L_B$$

$$L_2 = L_{AA} + L_{AB} + L_{BB}$$

$$L_3=L_{AAA}+L_{AAB}+L_{ABB}+L_{BBB}$$

$$L_4=L_{AAAA}+L_{AAAB}+L_{ABBB}+L_{BBBB}$$

Kur L_k reiškia sveikąjį kopijų skaičių. Preliminarus kopijų skaičiaus įvertis (X) yra apibrėžiamas kaip penkių modeliuotų KSP vidurkis:

$$x = \frac{L_1 + 2L_2 + 3L_3 + 4L_4}{L_0 + L_1 + L_2 + L_3 + L_4}$$

Kiekvienam žymeniui priskyrus kopijų skaičių yra nustatomi trūkio taškai naudojant atskyrimo (angl. *partition*) metodą. Šis metodas identifikuoja sritis, kuriose numatytas KSP regiono ribose ir už jo yra skirtingas. Tam, kad priskirti KSP nustatytai sričiai pagal trūkio taškus yra sudedamos visos tos srities žymenų bei visų aleliui būdingų kopijų skaičiaus būsenų, duotam kopijų skaičiui, tikimybės bei atrenkamos tos skaičiaus kopijos, kurių suminis rezultatas yra didžiausias

(https://www.illumina.com/documents/products/technotes/technote_cnv_algorithms.pdf).

1.4 Genomo kitimų duomenų bazė

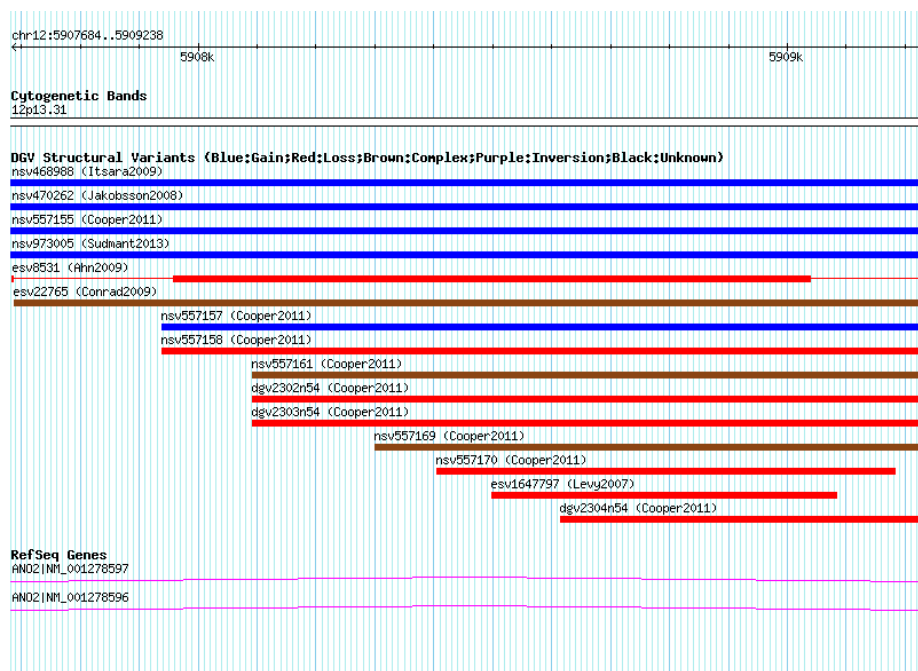
Genomo kitimų duomenų bazės tikslas yra pateikti visapusiškai apibendrintą informaciją apie struktūrinius, subalansuotus/nesubalansuotus, variantus žmogaus genome. Per pastaruosius dešimt metų naujos genomines technologijas su padidėjusia skiriamąją geba atskleidė, kad struktūriniai pokyčiai pasiskirstę visame žmogaus genome bei yra dažnai susiję su ligomis, dėl nesubalansuotų DNR pokyčių, tokių kaip KSP ar mažesnių insercijų/delecijų (40). Subalansuoti DNR pokyčiai, tokie kaip inversijos ir translokacijos yra retesni žmogaus genome, tačiau gali būti svarbūs chromosomų evoliucijai ir ligoms.

Ši duomenų bazė priskiria struktūrinius variantus (pvz. KSP ir inversijas) kaip genominius pokyčius, kurie susiję su DNR struktūriniais pokyčiais didesniais nei 50 bp. Duomenų bazėje yra pateikiami tik kliniškai sveikų asmenų, struktūriniai variantai. Genomo kitimų duomenų bazė pateikia naudingus duomenis susiejant genominius variantus su fenotipu (11). Genomo naršyklė yra grafinė vartotojo sąsaja, kuri naudoja *GMOD/Gbrowse* platformą (41). Ši platforma garantuoja greitą interaktyvų našumą ir yra atviro kodo ir tai leidžia greitą atvaizdavimą, tikrinimą ir duomenų užklausas įvairiais lygmenimis. Taip pat galima nustatyti, kad tuo pačiu metu būtų vykdoma paieška ir kitose duomenų bazėse, tokiose kaip *RefSeq*, *OMIM* (<http://omim.org>), *DECIPHER* ir *ISCA* konsorciumo.

Duomenų bazė yra nuolat atnaujinama recenzuojamų mokslinių tyrimų naujais duomenimis. Šiuo metu genominių kitimų duomenų bazėje yra saugoma 72-iejų skirtingų mokslinių tyrimų

informacija, kurie sudaro 552 586 KSPR. Naujausias genominių kitimų duomenų bazės atnaujinimas 2016 metų gegužės 15 dieną.

Paieška duomenų bazėje atliekama naudojant genominę naršyklę, kurioje reikia įvesti chromosomos numerį, regiono pradžią ir pabaigą (chr12:5907684-5909238). Rezultate gaunama informacija apie toje srityje esančius KSP, genus, ligas lemiančius genus, aprašytus struktūrinius pokyčius (žr. 1.6 pav.). Spragtelėjus ant pokyčio yra nukreipiama į detaliau aprašytą varianto pokytį (11).



1.6 pav. Genominių kitimų duomenų bazės sugeneruoti duomenys pagal užklausą: chr12:5907684-5909238. Mėlynai žymima duplikacija, raudonai – delecija, rudai – kompleksas, juodai – neaiškus variantas. Rožinės spalvos linijos pateikia *RefSeq* genus persidengiančius su šia sritimi. Prie kiekvieno pateikto KSP šalia yra pateikiama apie mokslinį tyrimą, kuriame buvo nustatytas tas KSP.

2. TYRIMO METODAI IR IMTIS

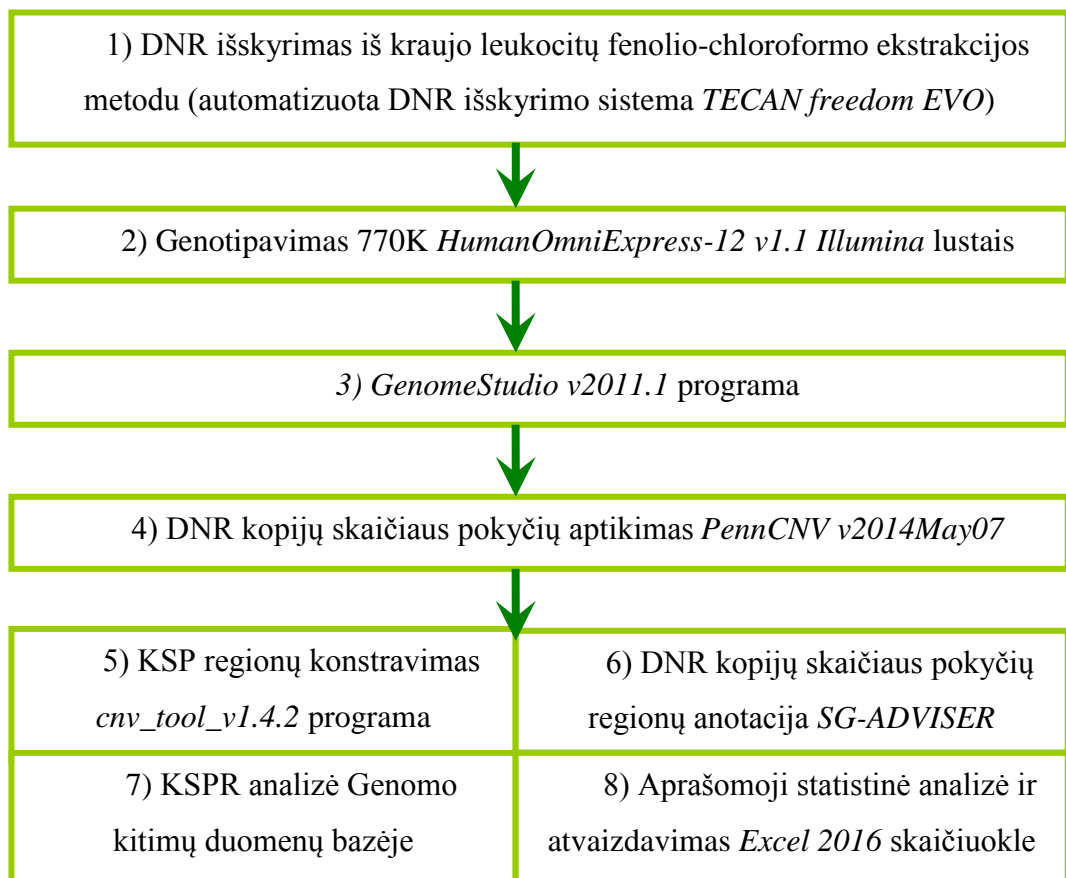
2.1 Tiriamoji medžiaga

DNR kopijų skaičiaus pokyčių analizei buvo naudojama projekto LITGEN (Lietuvos populiacijos genetinė įvairovė ir sandaros kitimai, susiję su evoliucija ir dažniausiai paplitusiomis ligomis) bendros Lietuvos populiacijos 301-o asmens DNR. Visų mėginių genotipavimas buvo atliktas VU MF Žmogaus ir medicininės genetikos katedros darbuotojų 2014 metais. Veninio kraujo ėminiai, DNR išskyrimui, buvo surinkti iš atsitiktinai parinktų asmenų iš šešių Lietuvos etnolingvistinių grupių (Vakarų, Pietų, Šiaurės Žemaičių; Rytų, Vakarų ir Pietų Aukštaičių), kurių tėvai ir seneliai gyveno tame pačiame rajone. Mėginiai surinkti ir DNR išskirta 2011-2013 metais VU MF Žmogaus ir medicininės genetikos katedroje. Tyrimams atliekamiems LITGEN projekte

yra gautas Vilniaus regioninio biomedicininų tyrimų etikos komiteto leidimas (Nr. 158200-05-329-79) ir visų tiriamųjų sutikimai dalyvauti LITGEN projekto tyrimuose.

2.2 Tyrimo eiga

VU MF Žmogaus ir medicininės genetikos katedroje naudojant DNR išskyrimą iš kraujo leukocitų fenolio-chloroformo ekstrakcijos metodu ir automatizuotą DNR išskyrimą *TECAN freedom EVO* sistema buvo išskirta 301-o asmens DNR. Genotipavimas atliktas 301-am asmeniui naudojant VNP *Illumina 770K HumanOmniExpress-12 v1.1* genotipavimo lustus. Gauti duomenys buvo apdorojami *GenomeStudio v2011.1* programa ir sukuriamos išvesties bylos reikalingos sekanciam analizės etapui. Kopijų skaičiaus pokyčiams aptikti buvo naudojama *PennCNV v2014May07* programa (34). Šiek tiek modifikavus *PennCNV* išvesties bylas buvo sukuriamos įvesties bylos KSP regionų konstravimui *cnv_tool_v1.4.2* programa. Toliau iš sukonstruotų KSPR buvo sudaroma įvesties byla KSPR anotacijai *SG-ADVISED* serveryje. Sekančiame etape buvo atliekama KSPR analizė Genomo kitimų duomenų bazėje ir galiausiai aprašomoji statistinė analizė ir atvaizdavimas buvo atliktas naudojant *Excel 2016* skaičiuoklę. Mokslinio darbo eiga pateikiama 2.1 paveiksle.



2.1 pav. Mokslinio darbo eiga.

2.3 Tyrimo metodai

2.3.1 DNR išskyrimas

Tiriamųjų asmenų DNR buvo išskirta iš kraujo leukocitų automatizuota DNR išskyrimo sistema *TECAN freedom EVO* bei fenolio–chloroformo ekstrakcijos metodu VU MF Žmogaus ir medicininės genetikos katedros darbuotojų.

DNR išskyrimo automatizuota *TECAN Freedom EVO* sistema darbo eiga:

I. Reagentų paruošimas:

- Etanolis 70 %;
- Izopropanolis (propanol-2, CH₃-CHOH-CH₃) 99 %;
- *eLysis* buferis;
- *Alcohol wash* tirpalas (etanolis + izopropanolis);
- *Elution* buferis;
- *SV Lysis* buferis, skiedimas 1/5;
- Magnetinės dalelės (jeigu magnetinės dalelės yra laikomos 4 °C arba turi nuosėdų, prieš naudojimą inkubuoti 37 °C 30 minučių, po to, supurtyti 5 sekundes ir nusukti). *eLysis* buferis, *Elution* buferis, *SV Lysis* buferis, *Alcohol wash* tirpalas bei magnetinės dalelės įeina į reagentų rinkinį skirtą DNR išskyrimui *TECAN Freedom EVO* sistema, kurių platintojas įmonė *Promega*.

II. Kasdienės priežiūros atlikimas:

- Patikrinti ar nėra sistemoje nutekėjimų;
- Ištuštinti atliekų indą (angl. *waste bottle*);
- Skysčių sistemos (angl. *Liquid system*) indą pripildyti dejonizuotu vandeniu sumaišytu su nedideliu kiekiu izopropanolio (rekomenduojama įpilti 1 % izopropanolio tam kad išvengti oro burbulų);
- Nuvalyti antgalius ir patikrinti ar nėra vandens lašų;
- Patikrinti, kad nebūtų oro burbulų nutekėjimo vamzdžiuose (oro burbulai veikia pipetavimo tikslumą). Šaltas vanduo gali būti oro burbulų priežastis, todėl rekomenduojama pripildyti skysčių sistemos butelį vieną dieną prieš naudojimą, tam kad vanduo būtų kambario temperatūros.

III. Reagentų bei pagalbinių medžiagų paruošimas darbui su *TECAN*:

- Sudėti tuščius 1,5 ml mėgintuvėlius į tam skirtą stovą;
- Sudėti į stovą atidarytus mėgintuvėlius su krauju (10 mm/5 ml arba 16 mm/10 ml mėginiai);
- Į stovėlį sudėti 50 ml tuščius mėgintuvėlius bei padėti juos ant magnetinio paviršiaus (angl. *magnet position*);
- Sudėti į stovėlį aštuonis 50 ml mėgintuvėlius pripildytus reikiamu kiekiu magnetinių dalelių.

IV. Roboto paleidimas:

- Uždaryti roboto duris;
- Įjungti robotą spaudžiant mygtuką „ON/OFF“, laikyti nuspaudus 0,5 sek. (žalia šviesa yra indikacija, kad instrumentas įjungtas);
- Įjungti kaitinimo bloką (angl. *heater*) ir temperatūros kontrolę;
- Įjungti roboto programą: „*Start Evoware Standard*“;
- Įvesti naudotojo vardą bei slaptažodį;
- Pažymėti „*Edit an existing process*“, tada spausti „*Start your selection*“ (žalia rodyklė);
- Eiti į „*Favorites*“, pasirinkti „*Tecan_gDNA_XL_10mm_tubes_V4_0*“, spausti „*Run*“;
- Plauti vandens srove (angl. *flush*) skysčių sistema: dešiniu pelės klavišu spausti ant „*Flush*“, tada „*Run*“; nustatyti: „*Volume in waste*“ 30 ml; „*Volume in cleaner*“ 30 ml bei pažymėti „*Use fast wash module*“;
- Spausti „*Run*“, tada vėl spausti „*Run*“;
- Ant „*Start recovery mode*“ pažymėti 0 = no
- Įvesti kraujo mėginių skaičių;
- Spausti „*No*“;
- Pradėti procesą su viena eliuacija (jeigu norime antros eliuacijos, reikia naudoti 2 ml mėgintuvėlius);
- Spausti „*START the application*“;
- Pasibaigus procesui spausti „*Exit*“ ir „*OK*“;
- Išjungiant instrumentą spausti mygtuką „ON/Off“ laikant 2 sekundes.

DNR išskyrimo fenolio-chloroformo ekstrakcijos metodu darbo eiga:

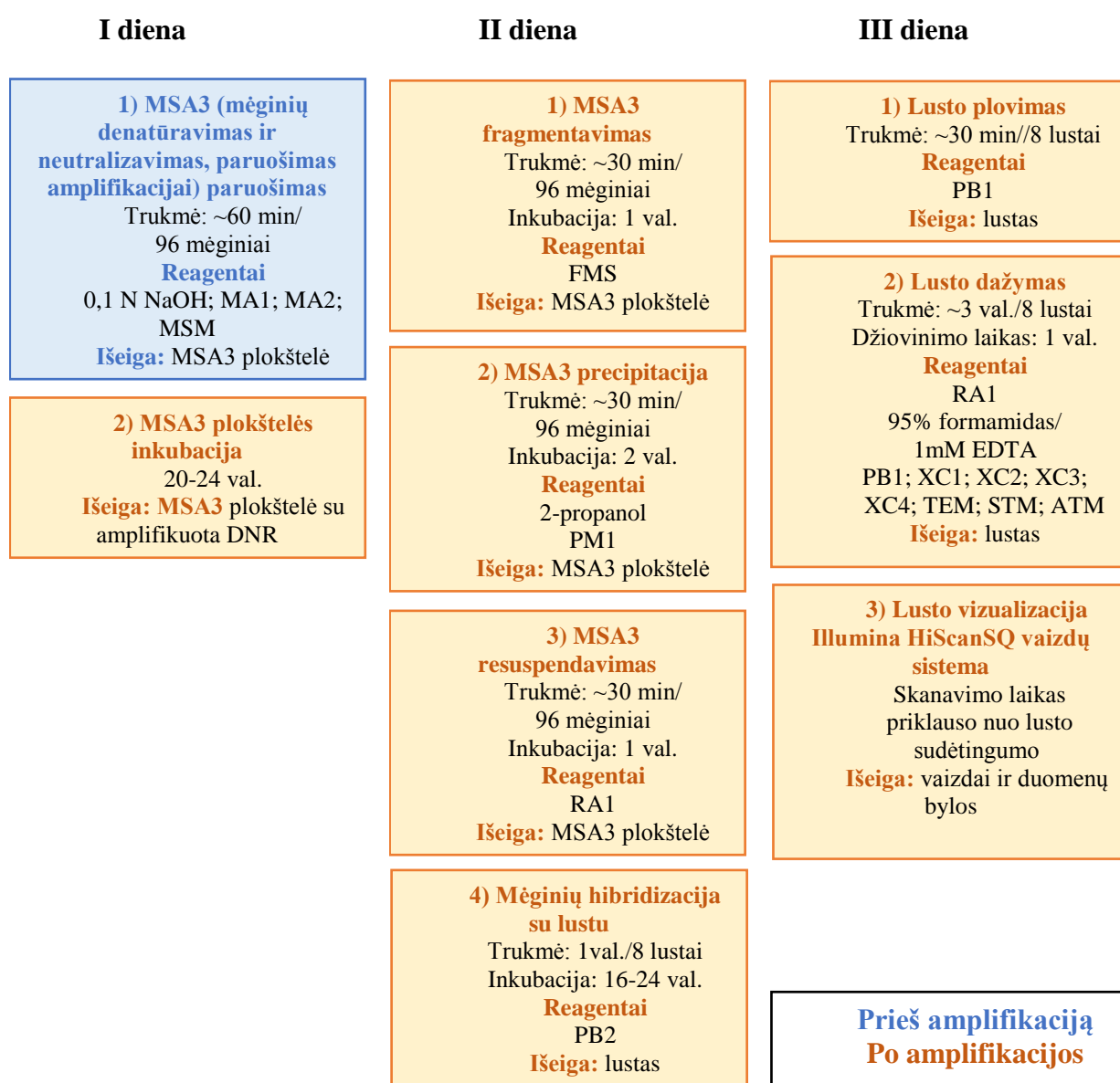
1. Veninis kraujas imamas į 5–10 ml vakuuminį mėgintuvėlį su antikoaguliantu EDTA.
2. Kraujas perpilamas į sterilų 45 ml mėgintuvėlį, iki 35 ml pripilama lizuojančio buferio. Centrifuguojama 10 min, 10 °C temperatūroje (šaldomojoje „*Universal 32R, Hettich*“ centrifugoje), greitis – 3000 aps/min.
3. Viršnuosėdinis skystis nupilamas, nuosėdos sumaišomos pipetuoiant, pripilama lizuojančio buferio iki 10 ml ir centrifuguojama 10 min, 10 °C temperatūroje, greitis – 3000 aps/min.
4. Viršnuosėdinis skystis nupilamas, nuosėdos sumaišomos pipete, pripilama lizuojančio buferio iki 5 ml. Centrifuguojama 10 min, 10 °C temperatūroje, greitis – 3000 aps/min.
5. Viršnuosėdinis skystis nupilamas ir įpilama 400 µl ląstelių branduolių pernešimo buferio (10 mM Tris-HCl, pH 10,5; 1 mM EDTA; 0,15 mM NaCl).
6. Visas turinys sterilia pipete pernešamas į 2 ml sterilų mėgintuvėlį, įpilama 20 µl 10 % natrio dodecilsulfato tirpalo ir 10 µl 2 % proteinazės K tirpalo.
7. Mišinys inkubuojamas 37 °C temperatūroje termostate 16 val.
8. Į mėgintuvėlį įpilama 400 µl fenolio ir mėgintuvėlis vartomas 10 min. Po to centrifuguojama 2 min, greitis – 5000 aps/min.

9. Viršutinis sluoksnis nusiurbiamas ir perkeliamas į 2 ml sterilų mėgintuvėlį, įpilama 200 µl fenolio ir 200 µl chloroformo. Mėgintuvėlis vartomas 5–10 min. Centrifuguojama 2 min, greitis – 5 000 aps/min.
10. Viršutinis sluoksnis nusiurbiamas ir perkeliamas į 2 ml sterilų mėgintuvėlį, įpilama 500 µl chloroformo ir izoamilo alkoholio mišinio (24:1). Mėgintuvėlis vartomas 5 min. Po to, centrifuguojama 2 min, greitis – 5 000 aps/min.
11. Viršutinis sluoksnis nusiurbiamas ir perkeliamas į 2 ml sterilų mėgintuvėlį, įpilama 400 µl chloroformo. Mėgintuvėlis vartomas 5 min. Po to, centrifuguojama 2 min, greitis – 5000 aps/min.
12. Viršutinis sluoksnis nusiurbiamas ir padalijus į 2 dalis perkeliamas į 0,5 ml ir 1,5 ml sterilius mėgintuvėlius. Į kiekvieną mėgintuvėlį įpilama po 20 µl 3 M natrio acetato tirpalo ir po 500 µl 96 % šalto etanolio.
13. Mėgintuvėlius vartant, DNR iškrinta į nuosėdas. Centrifuguojama 5 min, greitis – 10000–13200 aps/min.
14. Etanolis nupilamas ir įpilama 500 µl 70 % etanolio. Centrifuguojama 5 min, greitis – 10000–13200 aps/min.
15. Etanolis nupilamas ir įpilama 500 µl 96 % etanolio. Centrifuguojama 5 min, greitis –10000–13200 aps/min.
16. Vienas (0,5 ml) mėgintuvėlis saugomas -20 °C temperatūroje. Kito mėgintuvėlio etanolis nupilamas ir mėgintuvėlyje likęs turinys džiovinamas 30 °C temperatūroje (koncentratoriuje „Eppendorf“ 5301).
17. Išdžiovinta DNR užpilama 100 µl TE buferinio tirpalo (10 mM Tris-HCl; 1 mM EDTA, pH 8) ir laikoma kambario temperatūroje, kol ištirpsta.
18. Išskirtos DNR koncentracija ir švarumas nustatomas spektrofotometru („NanoDropR ND-1000“). DNR tirpalo švarumas patikrinamas matuojant tirpalo optinį tankį 260 nm ir 280 nm ilgio bangomis. Optinių tankių santykis turi būti 1,7–1,9.
19. Išmatavus DNR tirpalo koncentraciją, jis praskiedžiamas dejonizuotu vandeniu iki 500 mg/l. DNR tirpalas laikomas -20 °C arba -70 °C temperatūroje.
20. Tolesniems tyrimams naudojamas DNR tirpalas praskiedžiamas iki 50 mg/l DNR mėginių koncentracijos.

2.3.2 Genotipavimas VNP *Illumina* lustais

Genotipavimas buvo atliktas VU MF Žmogaus ir medicininės genetikos katedros darbuotojų. Buvo genotipuota bendros lietuvių populiacijos 301-o asmens DNR panaudojant VNP *Illumina HumanOmniExpress-12v1.1* lustus, turinčius apie 700 000 VNP žymenų (*Illumina*, San Diego,

CA, USA). Genotipavimo reakcija remiasi viso genomo amplifikacija pagausinant DNR iki 1000 kartų. Po to DNR yra fragmentuojama, nusodinama, resuspenduojama bei hibridizuojama su oligonukleotidiniais pradmenimis, kurie yra imobilizuoti luste. Toliau vykdoma vieno nukleotido pratęsimo reakcija atliekant pratęsto nukleotido žymėjimą ir žymės sustiprinimą dažais. Fluorochromais žymėti nukleotidai detektuojami *Illumina iScan* arba *HiScanSQ* optine sistema, kuri identifikuoja spalvos ir signalo intensyvumą. Galutiniai *HiScanSQ* duomenys analizuojami *GenomeStudio* programa. Genotipavimo darbo schema pateikta 2.2 paveiksle, pagal *Illumina* protokolą (https://support.illumina.com/content/dam/illumina-support/documents/myillumina/a4877346-aa39-4f9e-b519-eb5cfbf52d6a/infinium_hd_ultra_manual_euc_revb_11328095.pdf).



2.2 pav. Genotipavimo VNP *Illumina* lustais darbo schema.

GenomeStudio v2011.1 programa

GenomeStudio – *Illumina* lustais genotipuotų duomenų pirminės analizės programinė įranga. Atlieka žaliavinių duomenų normalizaciją, klasifikaciją (angl. *clustering*) ir genotipų klasifikavimą. Po genotipavimo, 301-o genotipuoto mėginio DNR žaliaviniams duomenims buvo atlikta kokybės kontrolė. Prieš atliekant kokybės kontrolę turėjo būti sukurtas naujas analizės projektas *GenomeStudio* programoje (pateikta žemiau). Aštuoni, iš 301-o, mėginiai neatitiko kokybės kontrolės pagal KSP klasifikacijų dažnį (angl. *Call Rate*), kuris buvo mažesnis nei 98 (žr. 2.3 pav.). Toliau buvo analizuojami 293, iš 301, mėginiai. Sekantis žingsnis buvo sugeneruoti 293 išvesties bylas (žr. 2.4 pav.), kurios yra reikalingos KSP nustatymui *PennCNV v2014May07* programa.

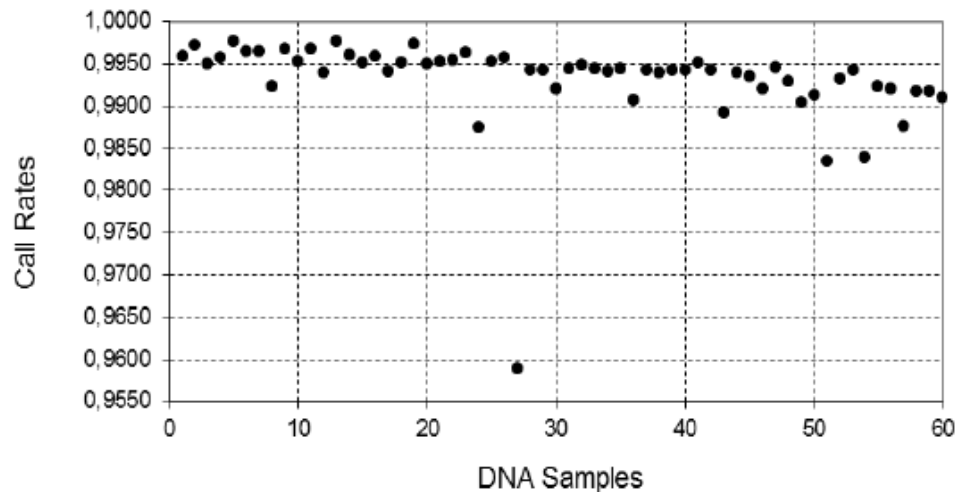
Projekto kūrimo *GenomeStudio* programoje darbo eiga:

1. atidaryti *GenomeStudio* programą;
2. „*New Project*“ lange spausti „*Genotyping*“ arba eiti į „*File | New Project | Genotyping*“;
3. spausti „*Next*“;
4. nurodyti „*Project Repository*“ kur bus išsaugomas projektas;
5. susikurti projekto pavadinimą „*Project Name*“ dėžutėje;
6. spausti „*Next*“;
7. pasirinkti „*Load Sample Intensities by Selecting Directories with Intensity*“;
8. spausti „*Next*“;
9. pasirinkti projektui reikalingas bylas: „*SNP Manifest*“ (.bpm byla), „*Data Repository*“ (VNP intensyvumo bylos);
10. spausti „*Next*“;
11. įkelti „*Cluster file*“ (.egt);
12. spausti „*Finish*“.

Kokybės kontrolės kriterijai

Blogos kokybės DNR mėginių bei klaidingai teigiamų KSP pašalinimui:

- ***Call Rate*** (KSP klasifikacijų dažnis) (>99 kokybė labai gera; >98 kokybė gera; <98 kokybė nepakankama);
- ***Standard Deviation (SD) of LRR*** (R logaritmo santykio standartinis nuokrypis) (<0,3). Tik mėginiai su šaukimų dažniu >98 ir R logaritmo santykio standartiniu nuokrypiu <0,3 buvo atrinkti tolimesniam tyrimui (žr. 2.4 pav.).



2.3 pav. *GenomeStudio* programoje sugeneruoto „*Call Rate*“ su DNR mėginių skaičiumi grafiko pavyzdys, atvaizduojantis 60-ies genotipuotų mėginių kokybę. „*Call Rate*“ yra paskaičiuojamas kiekvienam mėginiui (https://www.researchgate.net/figure/236861869_fig1_Figure-1-Call-rate-for-each-DNA-sample).

Išvesties bylų kūrimas *GenomeStudio* programa

- 1) Eiti į „*Analysis*“ → „*Reports*“ → „*Report Wizard*“;
- 2) Pasirinkti „*Final Report*“;
- 3) Lange „*Displayed Fields*“ pasirinkti laukus pagal eiliškumą: „*SNP Name, Chr, Position, Log R Ratio, B Allele Freq*“ (žr. 2.4 pav.);
- 4) Spausti „*Next*“;
- 5) Pasirinkti bylų išsaugojimo direktoriją;
- 6) Spausti „*Finish*“.

KSP analizei reikalingose bylose turi būti pateikiamas VNP numeris, chromosoma, vieta chromosomoje, R logaritmo santykis ir B alelio dažnis.

KSP aptikimui *PennCNV* programa bylos turi būti modifikuojamos nutrinant nuo [*Header*] iki [*Name*] bylų dalį (žr. 2.4 pav.).

```
[Header]
GSGT Version      1.8.4
Processing Date   5/5/2016 8:39:18 AM
Content           HumanOmniExpress-12v1-1_b.bpm
Num SNPs          726000
Total SNPs        726000
Num Samples       48
Total Samples     48
File              47 of 48
[Data]
SNP Name          Chr      Position      Log R Ratio    B Allele Freq
rs1000000         12      126890980     0.0724         0.4338
rs1000002         3       183635768     0.0248         0.0000
rs10000023        4       95733906      0.0042         0.4384
rs1000003         3       98342907      0.1480         0.0094
rs10000030        4       103374154     -0.7178        0.8583
rs10000037        4       38924330      0.0868         1.0000
rs10000041        4       165621955     0.1142         0.0000
```

2.4 pav. *GenomeStudio* programa sugeneruotos išvesties bylos pavyzdys.

2.3.3 Kopijų skaičiaus pokyčių aptikimas *PennCNV* programa

PennCNV yra laisvai prieinama programinė įranga skirta kopijų skaičiaus pokyčių nustatymui panaudojant plataus masto VNP genotipavimo duomenis. Ši programa analizuoja signalo intensyvumų duomenis, gautus iš VNP *Illumina* genotipavimo lustų. KSP tipo ir vietos nustatymui buvo sukurti tam tikri statistiniai modeliai, kurių pagrindinis yra HMM, kuriuo remiantis yra sudaromas algoritmas (1).

PennCNV v2014May07 programai paleisti reikėjo sukurti ir įvesti komandą su parametrais. Programos paleidimas buvo vykdomas *Perl* programavimo kalba, nes *PennCNV* skriptas parašytas *Perl* kalba. Analizei naudota komandinė eilutė:

```
perl detect_cnv.pl -test -hmm lib/hhall.hmm -pfb lib/OmniExpress_hg19.pfb.new -conf -list LITGEN_FinalReports.txt -log output/ LITGEN_FinalReports.log -out output/ LITGEN_FinalReports.rawcnv -gcmodel omni2.5.hg19.gcmodel
```

kur “./detect_cnv.pl” yra *PennCNV* vykdymo skriptas; .hmm ir .pfb *PennCNV* naudojamos bylos; po jų seka įvesties ir išvesties bylos. Ši komanda skirta visų tiriamųjų, esančių įvesties byloje, KSP analizei. Rezultate gaunama byla su visų tiriamųjų KSP rezultatais. Gautiems rezultatams buvo atlikta kokybės kontrolė remiantis šiais kriterijais: R logaritmo santykio standartinis nuokrypis <0,3; B alelio dažnio poslinkis (angl. *BAF_drift*) < 0,01; kintantis koeficientas (angl. *waviness factor*) tarp -0,04 ir 0,04.

Sugeneruoti rezultatai 293 asmenims *PennCNV* programa buvo sukelti į *Excel* 2016 skaičiuoklę. *PennCNV* programai yra nustatyti tam tikri pasiklivimo įverčių slenksčiai, kada KSP reikia atmesti kaip klaidingai teigiamą, todėl buvo įvykdyta potencialių KSP atranka. *PennCNV* programa sugeneruotiems rezultatams atrinkti pasiklivimo lygmuo (angl. *confidence level*) turi būti ne mažesnis nei 10 ir į KSP turi įeiti ne mažiau nei 7 VNP. Tarp rezultatų nėra lyties chromosomų KSP, kadangi *PennCNV* taiko skirtingą metodą 23 chromosomai analizuoti, todėl rezultatai gaunami tik 22 autosomoms. Be to, KSP nustatyti X chromosomoje turi didelį klaidingai teigiamų rezultatų dažnį, todėl daugelis tyrėjų lyties chromosomų neįtraukia analizuojant KSP.

2.3.4 DNR kopijų skaičiaus pokyčių regionų konstravimas

KSPR buvo konstruojami naudojant *cnv_tool_v1.4.2.py* programą. Programa parašyta *Python* programavimo kalba. KSPR yra sudaromi iš persidengiančių KSP atliekant: persidengiančių KSP atranką (pavienių KSP atmetimas), persidengiančių KSP apjungimą į vieną KSP regioną. Prieš paleidžiant programą reikėjo pasiruošti, *PennCNV* programa nustatytų KSP, įvesties bylą formatu: kodas (*ID_code*), mėginio numeris (*sample_nr*), chromosoma (*chromosome*), pradžia (*start*), pabaiga (*end*), ilgis (*length*), KSP tipas (*cnv_type*) (0,1,3, ar 4), o bylos antraštė nutrinama. Programai paleisti buvo naudojama ši komandinė eilutė:

```
start python cnv_tool_v1.4.2.py -c ksp.txt -r -e ksp_regionai.txt
```

Komanda skirta kopijų skaičiaus pokyčių regionams sudaryti, kur -c – parametras įvesties bylai (ksp.txt); -r – parametras KSPR konstravimui; -e – parametras išvesties bylai (ksp_regionai.txt). Išvesties bylos formatas: KSPR, chromosoma, pradžios ir pabaigos koordinatės, KSPR ilgis bazių poromis, KSP skaičius, delecijos, duplikacijos (žr. 2.5 pav).

CNVRs	chr	start	end	length	n_cnvs	losses	deletions	
		gains	duplications					
CNVR	1	149039031	149201987	162956	9	5	4	0
		0						
CNVR	1	12866590	12880356	1376611	0	8	3	0
CNVR	1	33850889	33877807	269184	0	4	0	0
CNVR	1	150664077	150679358	152814	0	4	0	0
CNVR	1	249170711	249199102	2839122	0	22	0	0
CNVR	1	25889421	25900240	108191	0	1	0	0
CNVR	1	163348444	163359017	105731	0	1	0	0
CNVR	1	239093239	239097408	41691	0	1	0	0
CNVR	1	247559156	247568051	88951	0	1	0	0
CNVR	1	244490369	244497339	69701	0	1	0	0
CNVR	1	151428605	151464911	363062	0	2	0	0
CNVR	1	245636915	245647726	108117	2	5	0	0
CNVR	1	157145225	157161346	161211	0	1	0	0
CNVR	1	46182670	46221440	387702	0	2	0	0
CNVR	1	51423567	51441961	183942	0	2	0	0

2.5 pav. *cnv_tool_v1.4.2.py* programa sukonstruotų KSPR pavyzdys.

2.3.5 Kopijų skaičiaus pokyčių regionų anotacija

Anotacijai buvo naudojamas laisvai prieinamas genomo anotacijų serveris *SG-ADVISER* (angl. *Scripps Genome Annotation and Distributed Variant Interpretation Server*). Tai aukštos kokybės daugiakomponentė skaičiavimo sistema susieta su duomenų bazėmis tokiomis kaip: *DGV*, *ClinVar*, *DECIPHER*, kurių pagalba anotacija vyksta tiksliau ir greičiau. Šio serverio pagrindinis tikslas yra teikti išsamią KSP anotaciją be sudėtingos programinės įrangos ir papildomų skaičiavimų ar paieškų (42).

Anotacijai atlikti buvo sudaryta įvesties byla, kurioje turi būti šie stulpeliai: chromosoma (*Chromosome*), pradžia (*Begin*), pabaiga (*End*), varianto tipas (*VarType*) ir pastabos (*Notes*) (žr. 2.6 pav.). Anotacijos darbo eigos etapai:

1. Eiti į <https://genomics.scripps.edu/ADVISER/Home.jsp> nuorodą;
2. Eiti į „CNVs“ → „Upload File“;
3. Pasirinkti „Unregistered Upload“;
4. Spausti „Choose File“ ir pasirinkti savo pasiruoštą įvesties bylą;
5. Spausti „Upload CNV file“;
6. Priklausomai nuo įkeliamo įvesties bylos dydžio, įkėlimas gali užtrukti nuo 10 ir daugiau minučių. Neuždaryti naršyklės, kol negauta žinutė, kad įvesties byla sėkmingai patalpinta ir nėra gautas išvesties bylos parsisiuntimo kodas.

Chromosome	Begin	End	VarType	Notes
chr18	77900829	77931598	loss	-
chr14	86287347	86310164	loss	-
chr1	25889421	25900240	loss	-
chr1	163348444	163359017	gain	-
chr1	239093239	239097408	gain	-
chr1	247559156	247568051	loss	-
chr1	244490369	244497339	loss	-
chr1	157145225	157161346	loss	-
chr1	169761107	169773317	loss	-
chr1	72818226	72838529	loss	-
chr1	178984086	179022236	loss	-
chr1	78585086	78611730	loss	-
chr1	47496737	47532614	loss	-
chr1	79501129	79519555	loss	-

2.6 pav. Anotacijai, *SG-ADVISED* serveryje, paruoštos įvesties bylos pavyzdys.

Gautoje, su anotuotais KSP, išvesties byloje pirmuose šešiuose stulpeliuose pateikiama informacija apie pokyčio vietą (chromosoma, pradžia, pabaiga, padėtis, atstumas) ir pokyčio tipą, o likę stulpeliai yra *SG-ADVISED* serverio anotuoti KSP: geno pavadinimas, geno tipas, kodavimo poveikis, baltymo pozicija, egzono bazės įeinančios/neįeinančios į KSP, *DGV* persidengimas su delecija ar duplikacija, *Welllderly* persidengimas su delecija ar duplikacija, žinomi sindromai susiję su delecija ar duplikacija, *ClinVar* persidengimas su delecija ar duplikacija, *1000 Genomes* persidengimas su delecija ar duplikacija, miRNR genomai, asociacija su *OmimGene*, baltymų domeno geno ontologija, *HGMD* Genų ligų asociacija, *COSMIC* genai, *MSKCC* vėžio genai, *Atlas* onkologija, *Sanger* vėžio genai, *Sanger* somatinių ar lytinių ląstelių navikas, *Sanger* vėžio sindromai, *Mitelman* duomenų bazė, vaistų bankas, genų ontologija, ligų ontologija.

2.3.6 DNR kopijų skaičiaus pokyčių duomenų apdorojimas *Excel* skaičiuokle

Šiame darbe duomenų atvaizdavimui ir aprašomajai statistinei analizei atlikti buvo naudojama *Excel 2016* skaičiuoklė. Naudojant *Frequency* funkciją buvo apskaičiuoti KSP dažniai taip pat atliktas grafinis KSP ir KSPR įvairovės ir pasiskirstymo atvaizdavimas.

2.3.7 Kopijų skaičiaus pokyčių regionų atvaizdavimas *PhenoGram* programine įranga

Naudojant *PhenoGram* programinę įrangą tyrėjai gali sukurti chromosomines ideogramas, kuriose linijomis ir skirtingomis spalvomis leidžia atvaizduoti chromosomines sritis ir/ar regionus. Naudojant šią programinę įrangą galima atvaizduoti įvairius duomenis, tokius kaip genomo padengimą VNP iš genotipavimo duomenų ir chromosomų padengimą VNP, genomo padengimą KSPR, taip pat jau yra galimybė atvaizduoti ir modelinių organizmų duomenis chromosomose.

Atvaizdavimui *PhenoGram* serveryje buvo sudaryta įvesties byla, kurioje pateikiama informacija: chromosoma (CHR), pradžia (POS), pabaiga (END) ir varianto tipas (POSCOLOR), kuriame skaičiais 1 ir 2 žymimos delecijos ir duplikacijos (žr. 2.7 pav.). KSPR atvaizdavimo *Phenogram* programine įranga eiga:

1. Eiti į <http://visualization.ritchie-lab.psu.edu/phenograms/document> nuorodą;

- Eiti į „PHENOGRAM“ → „PLOT“;
- Spausti „Choose File“ ir pasirinkti savo pasiruoštą įvesties bylą;
- Prie „Title“ įvesti norimą genolapio pavadinimą, prie „Genome**“ pasirinkti „human“, prie „Phenogram spacing“ pasirinkti „Proximity algorithm“, prie „Phenotype color“ pasirinkti „Exhaustive search colors“, prie „Image format“ pasirinkti „PNG“, uždėti varnelę prie „Chromosomes only?“;
- Spragtelėti „Additional options“ pagal eiliškumą pasirinkti: „Larger Font, Annotation, Transparent lines“;
- Spausti „Plot“;

CHR	POS	END	POSCOLOR
7	11740020	11753281	2
17	3111045	3119767	2
1	17597720	17610899	2
1	240127293	240147503	2
1	9321241	9400868	2
1	8989235	9061382	2
1	108316189	108756690	2
2	228243310	228258288	2
2	32437338	32478629	2
2	55680496	55798108	2
3	118730933	118803278	2
3	165204343	165406029	2
3	65358578	65624729	2
4	34900365	34917769	2
4	9371116	9456977	2

2.7 pav. Atlikto atvaizdavimo, *PhenoGram* programine įranga, įvesties bylos pavyzdys.

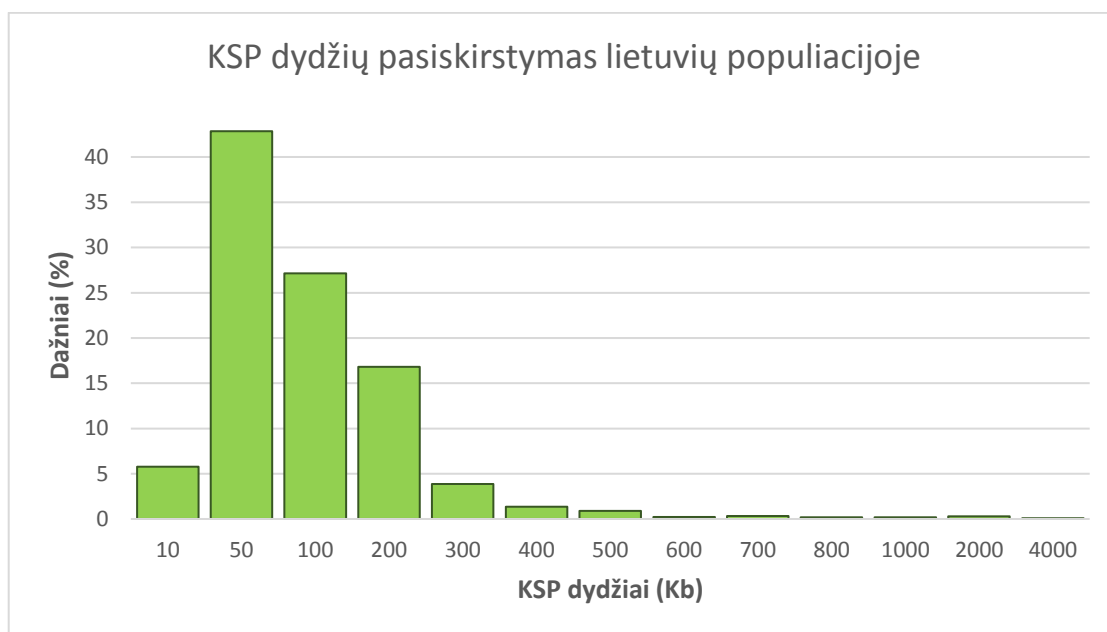
3. TYRIMO REZULTATAI

DNR kopijų skaičiaus pokyčių analizei buvo naudojami LITGEN projekto 293 bendros lietuvių populiacijos asmenų genotipuotų 770K *HumanOmniExpress-12 v1.1 Illumina* lustais duomenys. Kopijų skaičiaus pokyčių nustatymui naudota *PennCNV v2014May07* programa (34). Iš viso *PennCNV* programa nustatyti 8742 KSP 293 asmenims. Po atliktos kokybės kontrolės, kurios metu buvo tikrinami tiriamieji neatitinkantys šių kriterijų: $LRR < 0,3$, $BAF_drift < 0,01$ ir *waviness factor* tarp $-0,04$ ir $0,04$, buvo pašalintas 51 tiriamasis. Toliau buvo vykdomas nustatytų KSP filtravimas remiantis šiais kriterijais: *confidence level* ≥ 10 ir VNP skaičius, kuris turi įeiti į KSP ne mažesnis nei 7 VNP. Po filtravimo liko 3535 potencialūs KSP. KSP nustatymo metu buvo analizuojamas tik autosominis genomus dėl *PennCNV* programos naudojamo kito metodo nustatant KSP lyties chromosomose. Darbo metu nustatyta, kad 82,3 % bendros lietuvių populiacijos asmenų yra KSP nešiotojai. Vienam asmeniui vidutiniškai tenka apie 12,1 KSP, kurių dydžių vidurkis yra 86,67 Kb (nuo 1,5 Kb iki 3,5 Mb), o mediana 51,77 Kb. Nustatytas delecijų ir duplikacijų santykis yra 4,5:1. Nustatyti KSP apima 9,5 % viso žmogaus genomo (žr. 3.1 lentelę).

3.1 lentelė. Kopijų skaičiaus pokyčių charakteristikos lietuvių populiacijoje.

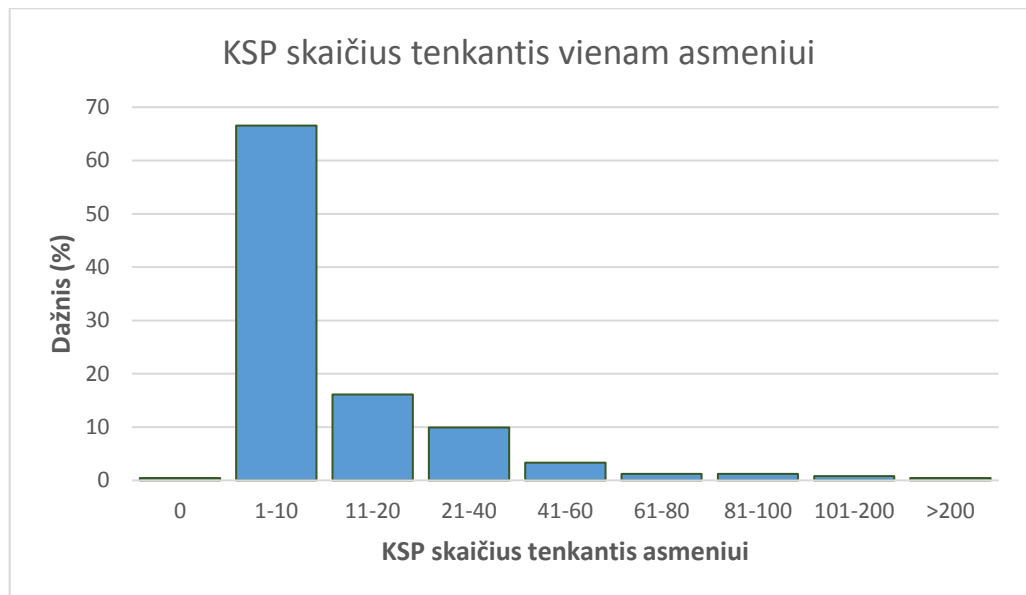
Kopijų skaičiaus pokyčiai (KSP)	Lietuvių populiacija
Mėginių skaičius	293
Nustatytas KSP skaičius	3535
KSP nešiotojų skaičius	241 (82,3 %)
KSP/vienam asmeniui	12,1
KSP duplikacijos	638 (18 %)
KSP delecijos	2897 (82 %)
KSP dydžių vidurkis	86,67 Kb
KSP dydžių mediana	51,8 Kb
Genomo dalis, kurioje nustatyti KSP	306,4Mb (9,5 %)

Atlikus KSP dydžių pasiskirstymo analizę bendroje lietuvių populiacijoje, nustatyta, kad dažniausiai, net 42,8 %, pasitaiko KSP, kurių dydžiai varijuoja nuo 10 Kb iki 50 Kb. Didesni nei 2 Mb KSP nustatyti tik trys ir jie sudaro 0,08 % visų KSP bendroje lietuvių populiacijoje (žr. 3.1 grafiką).



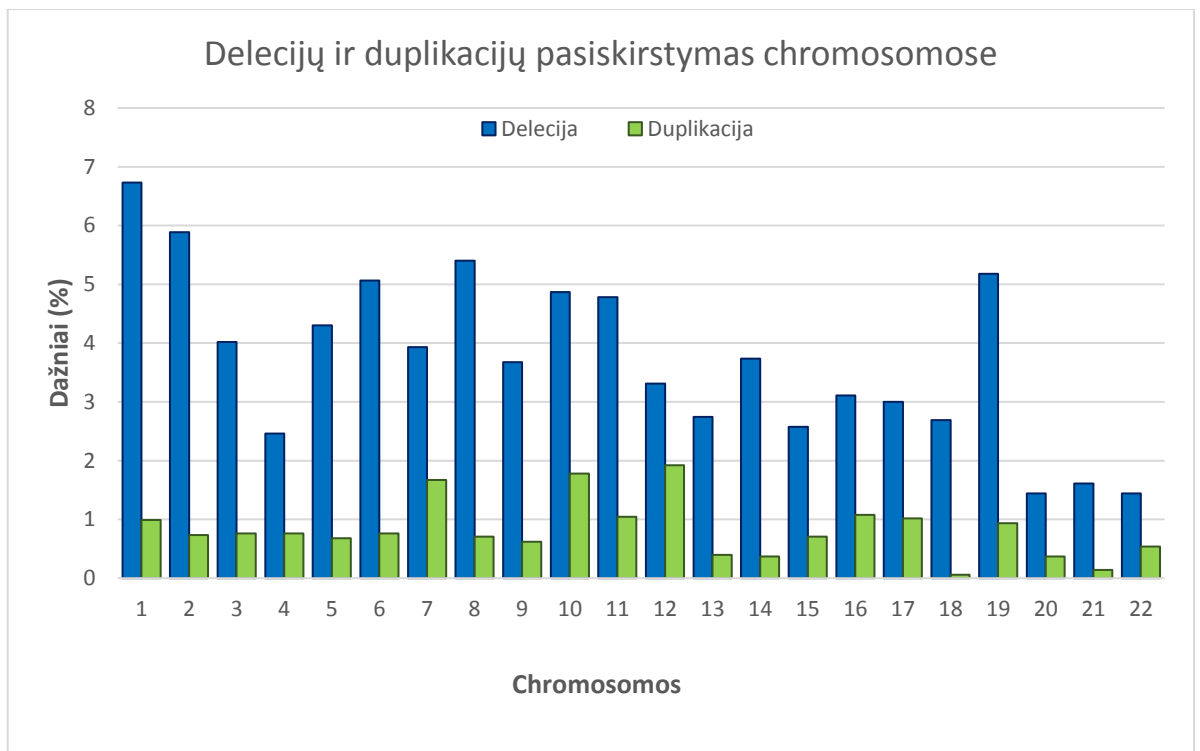
3.1 grafikas. Kopijų skaičiaus pokyčių dydžių pasiskirstymas lietuvių populiacijoje. Kopijų skaičiaus pokyčių dydžiai suskirstyti į grupes: nuo 0-10 Kb priklauso grupei „10”, didesni nei 10 Kb iki 50 Kb priklauso grupei „50” ir t.t.

Sekančiu analizės etapu buvo nustatyta kiek KSP tenka vienam asmeniui. Nustatyta, kad asmenys gali turėti nuo 1 iki 212 KSP, tačiau dažniausiai vienam asmeniui lietuvių populiacijoje tenka nuo 1 iki 10 KSP ir tai apima 66,5 % visų nustatytų KSP. Daugiau nei 200 KSP turėjo tik vienas asmuo, tai apima 0,4 % bendros lietuvių populiacijos (žr. 3.2 grafiką).



3.2 grafikas. Kopijų skaičiaus pokyčiai tenkantys vienam asmeniui lietuvių populiacijoje.

Iš visų nustatytų KSP delecijų buvo 2897 (82 %) , o duplikacijų 638 (18 %). Atlikus kopijų skaičiaus pokyčių pasiskirstymą chromosomose nustatyta, kad delecijos dominavo 1, 2, 8 ir 19 chromosomoje, o duplikacijos 7, 10 ir 12 chromosomoje. Mažiausiai delecijų nustatyta 20 (51 delecija, 1,44 %), 21 (57 delecijos, 1,61 %) ir 22 (51 delecija, 1,44 %) chromosomoje, tuo tarpu mažiausiai duplikacijų nustatyta 18 (2 duplikacijos, 0,06 %) ir 21 (5 duplikacijos, 0,1 %) chromosomoje. KSP pasiskirstė visose chromosomose (žr. 3.3 grafiką).



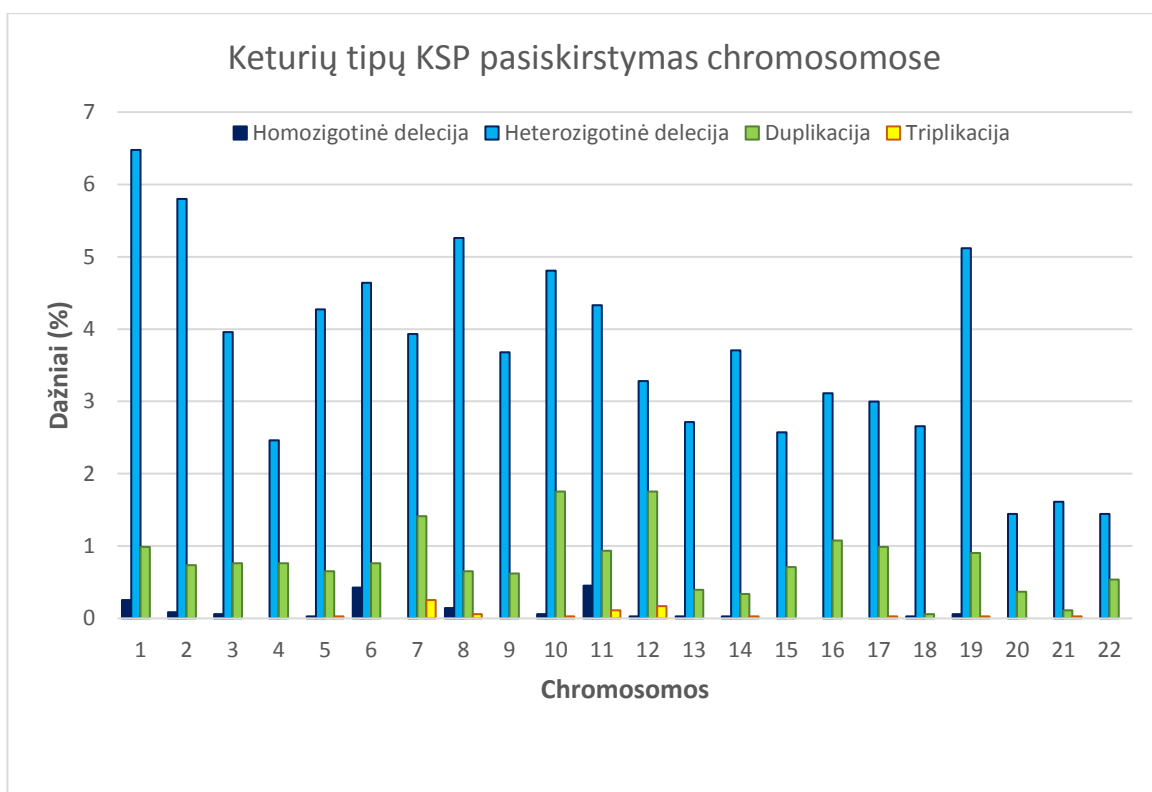
3.3 grafikas. Delecijų ir duplikacijų pasiskirstymas chromosomose lietuvių populiacijoje.

Taip pat buvo atlikta keturių tipų kopijų skaičiaus pokyčių t.y. homozigotinių delecijų, heterozigotinių delecijų, duplikacijų ir triplikacijų, pasiskirstymo analizė. Visose chromosomose daugiausia nustatyta heterozigotinių delecijų, net 2838 (80,3 %) ir duplikacijų 611 (17,2 %). Homozigotinių delecijų ir triplikacijų nustatyta mažiau ir ne visose chromosomose (žr. 3.2 lentelę).

3.2 lentelė. Keturių tipų (homozigotinės ir heterozigotinės delecijos, duplikacijos ir triplikacijos) kopijų skaičiaus pokyčių pasiskirstymas chromosomose lietuvių populiacijoje.

Kopijų skaičiaus pokyčiai (KSP)	Lietuvių populiacija
KSP duplikacijos	611 (17,2 %)
KSP triplikacijos	27 (0,8 %)
KSP homozigotinės delecijos	59 (1,7 %)
KSP heterozigotinės delecijos	2838 (80,3 %)

Homozigotinių delecijų nenustatyta 4, 7, 15, 16, 17, 20, 21 ir 22 chromosomoje, o triplikacijų nenustatyta 1, 2, 3, 4, 6, 9, 13, 15, 16, 18, 20 ir 22 chromosomoje. Daugiausia heterozigotinių delecijų nustatyta 1, 2, 8 ir 19 chromosomoje, homozigotinių delecijų 1, 6 ir 11 chromosomoje, duplikacijų 7, 10 ir 12 chromosomoje, o triplikacijų 7, 11 ir 12 chromosomoje. Tik 5, 8, 10, 11, 12, 14 ir 19 chromosomoje nustatyta visų keturių tipų KSP. Tuo tarpu 4, 9, 15, 16, 20 ir 22 chromosomoje nustatytos tik heterozigotinės delecijos ir duplikacijos (žr. 3.4 grafiką).



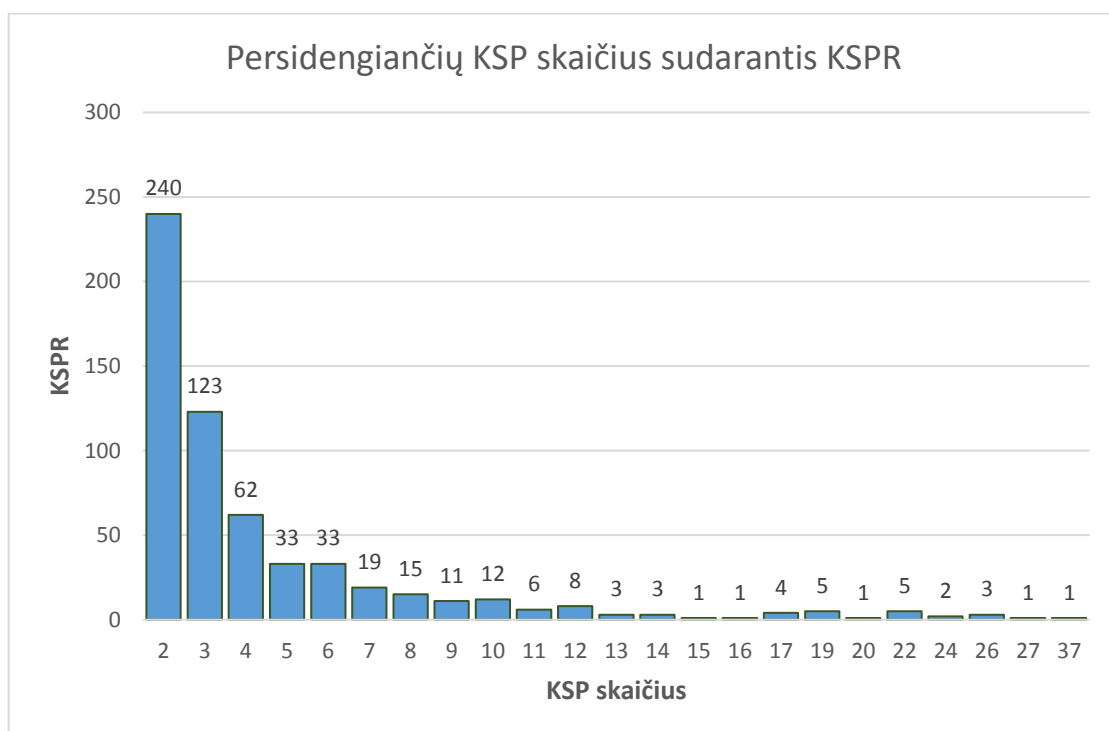
3.4 grafikas. Keturių tipų (homozigotinės ir heterozigotinės delecijos, duplikacijos ir triplikacijos) kopijų skaičiaus pokyčių pasiskirstymas chromosomose lietuvių populiacijoje.

Sekančiame analizės etape buvo sukonstruoti kopijų skaičiaus pokyčių regionai bendrai lietuvių populiacijai remiantis *cnv_tool_v1.4.2.py* programa. Iš viso sukonstruoti 1438 KSPR, kurių dydžių vidurkis yra 66 Kb, o mediana 42,7 Kb. Nustatyti KSPR apima 2,9 % (94,8 Mb) viso lietuvių genomo (žr. 3.3 lentelę).

3.3 lentelė. Kopijų skaičiaus pokyčių regionų charakteristikos.

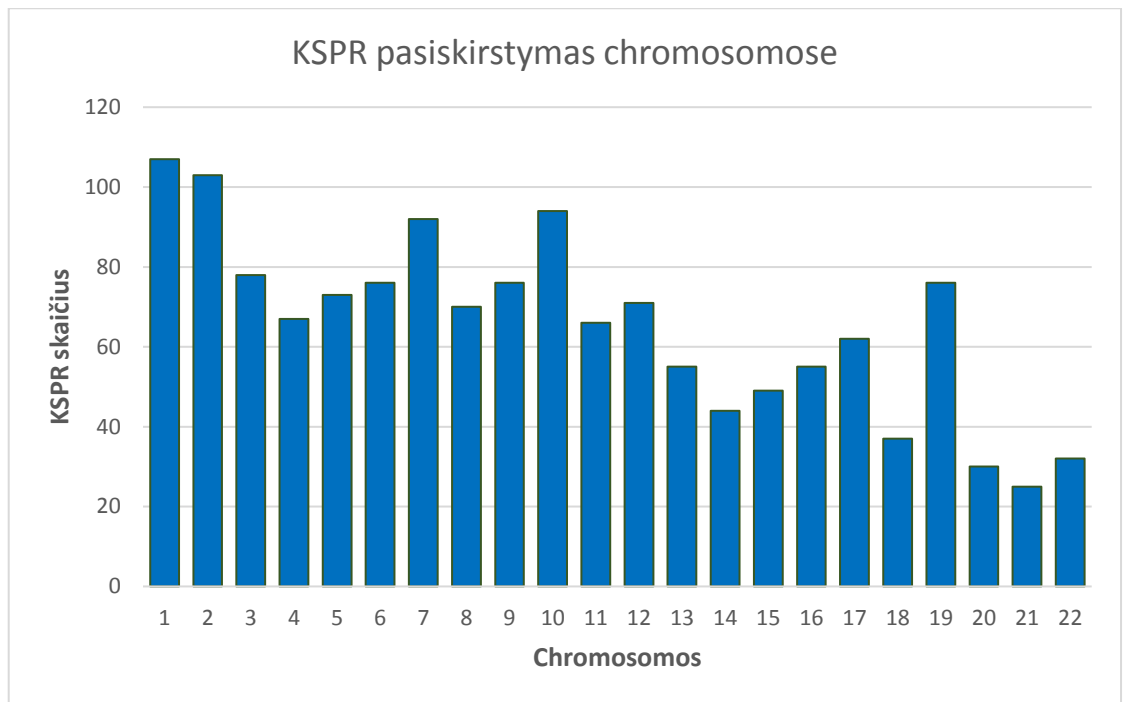
Kopijų skaičiaus pokyčių regionai (KSPR)	Lietuvos populiacija
Iš viso nustatyta KSPR	1438
KSPR dydžių vidurkis	66 Kb
KSPR dydžių mediana	42,7 Kb
Genomo dalis, kurioje nustatyti KSPR	94,8 Mb (2,9 %)

Kopijų skaičiaus pokyčių regionus dažniausiai sudaro 2 arba 3 persidengiantys KSP. Nustatytas vienas KSP regionas su 37 persidengiančiais KSP (žr. 3.5 grafiką).



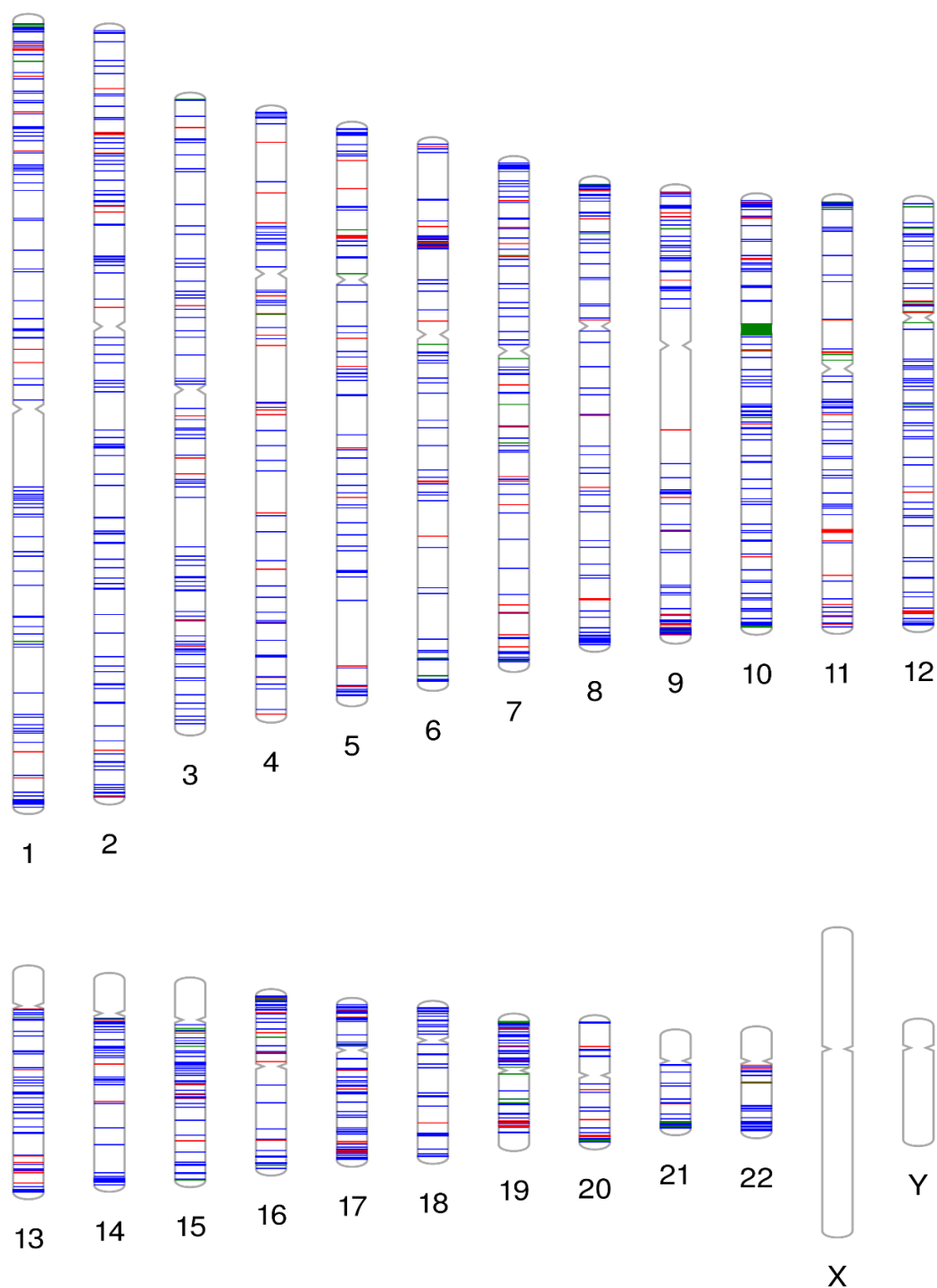
3.5 grafikas. Persidengiantys KSP, kurie sudaro KSPR lietuvių populiacijoje.

Kopijų skaičiaus pokyčių regionų pasiskirstymas chromosomose pateiktas 3.6 grafike. Daugiausiai KSPR nustatyta 1 chromosomoje-107 ir 2 chromosomoje - 103, o mažiausiai 21 chromosomoje -25.



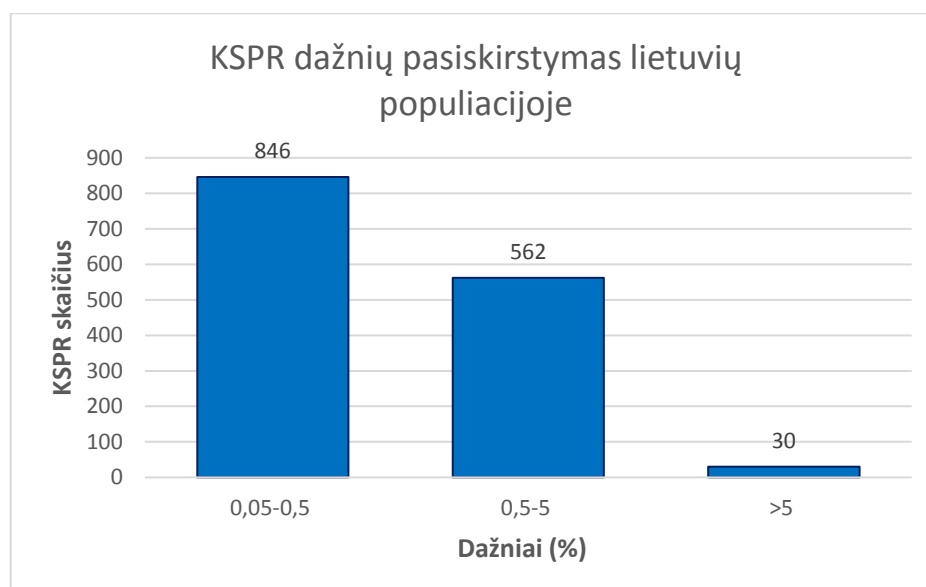
3.6 grafikas. Kopijų skaičiaus pokyčių regionų pasiskirstymas chromosomose.

Taip pat buvo sukonstruota KSP regionų ideograma *PhenoGram* programa, kuri padeda atvaizduoti kopijų skaičiaus pokyčių ar jų regionų pasiskirstymą chromosomose. Programą galima rasti šioje nuorodoje <http://visualization.ritchielab.psu.edu/phenograms/document>. Remiantis sudaryta KSPR ideograma stebimas gana tolygus KSPR per visą genomą. Tačiau yra stebimas didesnis KSPR tankis telomerinėse chromosomų srityse. Didžiausias daugia-alelinis KSP regionas stebimas 10 chromosomoje, net 3,4 Mb dydžio (žalia spalva). Tuo tarpu 19 chromosomoje nustatytas didžiausias KSPR tankis, aptinkama delecijų, duplikacijų ir daugia-alelinių sričių. Šioje chromosomoje 1 Mb susitelkę ~1,28 KSPR (~3,66 KSP), kai 17-oje chromosomoje KSPR tankis atrodo gan didelis, bet 1 Mb tenka tik ~0,76 KSPR, o 10-oje chromosomoje ~0,69 KSPR (žr. 3.1 pav.).



3.1 pav. Kopijų skaičiaus pokyčių regionų ideograma. Mėlyna spalva vaizduoja delecijas, raudona - duplikacijas, žalia - ir delecija ir duplikacija (daugia-alelinė sritis).

Paskaičius KSPR dažnių pasiskirstymą nustatyta, kad lietuvių populiacijoje didžioji dalis KSPR (58,8 %) yra reti, jų dažnis yra 0,05-0,5 %. KSPR, kurių dažnis yra 0,5-5%, sudaro 39,1 % (562) visų KSPR, o dažni KSPR, kurių dažnis >5 %, sudaro 2,1 % (30) visų KSPR (žr. 3.7 grafiką).



3.7 grafikas. Kopijų skaičiaus pokyčių regionų dažnių pasiskirstymas lietuvių populiacijoje.

Paskutiniame analizės etape buvo atlikta KSPR anotacija. Šiam analizės etapui buvo pasitelktas *SG-ADVISER* (angl. *Scripps Genome Annotation and Distributed Variant Interpretation Server*) serveris. Atlikus nustatytų KSPR palyginimą su Genomo kitimų duomenų bazėje (angl. *Database of Genomic Variants*, vėliausias atnaujinimas 2016 m. gegužės mėn.) esamais KSP, nustatyta, kad net 98,2 % identifikuotų KSPR persidengia su sritimis esančiomis DGV, o likę 1,8 %, nebuvo nustatyti DGV ir yra unikalūs lietuvių populiacijai. Net 70 % nustatytų unikalų lietuvių populiacijai KSPR yra reti, su 0,4 % dažniu.

Dauguma nustatytų KSPR (87 %) persidengia su žinomais UCSC (<http://genome.ucsc.edu/>) genais. KSP delecijos persidengiančiose su genais srityse nustatytos dažniau nei duplikacijos (84,7 % delecijų, o duplikacijų 15,3 %). Anotacijos metu KSP buvo susieti su genais, jų sritimis ir funkcijomis. Anotacijos rezultatai parodė, kad nustatyti KSP galimai yra susiję tiek su Mendelinio tipo, tiek su daugiaveiksnėmis ligomis. Išsamesni anotacijos rezultatai pateikiami 3.4 lentelėje.

3.4 lentelė. Su KSP anotuoti genai, jų sritys ir funkcija.

KSPR tipas ir dydis (bp)	Genas	Sritis	Anotacija	Funkcija	Fenotipas
Delecija 61 164	ANKS3	16p13.3	ankyrin repeat and sterile alpha motif domain containing 3	Transkripcijos reguliacija	-
Delecija 27 255	C16orf91	16p13.3	chromosome 16 open reading frame 91	Jonų transportas	Pancitopenija, osteomielitas

Delecija ir duplikacija 161 843	METTL9	16p12.2	methyltransferase like 9	Imuninis atsakas	-
Delecija 76 498	CCDC88 A	2p16.1	coiled-coil domain containing 88A	Baltymų fosforilinimas	Vėžys (karcinoma)
Delecija ir duplikacija 8 198	SIK1	21q22.3	salt inducible kinase 1	Neigiamas transkripcijos reguliavimas	Vėžys
Delecija 44 695	COL18A 1	21q22.3	collagen type XVIII alpha 1 chain	Angiogenezė	Glomerulonefritas, leukemija, šlapimo pūslės vėžys
Duplikacija 16 722	AMY2B	1p21.1	amylase, alpha 2B	mRNR brendimas	-
Delecija 49 063	MIPEP	13q12.1 2	mitochondrial intermediate peptidase	Proteolizė	-
Duplikacija 58 672	BNIP2	15q22.2	BCL2 interacting protein 2	Apoptozė	Neuroblastoma
Delecija 94 915	CKAP5	11p11.2	cytoskeleton associated protein 5	Ląstelės ciklas	-
Delecija 86 776	NRXN1	2p16.3	neurexin 1	Ląstelių adhezija	Autizmas, limfoblastinė leukemija
Delecija ir duplikacija 32 548	ADGRA 1	10q26.3	adhesion G protein-coupled receptor A1	Signalų perdavimas	-
Delecija 17 561	NOM1	7q36.3	nucleolar protein with MIF4G domain 1	Biologiniai procesai	Lėtinis inksto transplantato atmetimas

4. REZULTATŲ APITARIMAS

Panašus KSP tyrimas Lietuvoje buvo atliktas 2014-aisiais metais A. Urnikytės magistrinio darbo metu, kuriame KSP buvo nustatomi Lietuvos populiacijai ir etnolingvistinėms grupėms dviem skirtingom programom (*QaantiSNP 2.0* ir *cnvPartition 3.2.0*). Šio darbo pagrindinis tikslas buvo nustatyti KSP *PennCNV v2014May07* programa bei įvertinti KSP charakteristikas būdingas lietuvių populiacijoje.

Šio tyrimo metu naudojant VNP *Illumina HumanOmniExpress-12v1.1* lustų 293-ųjų asmenų duomenis bei *PennCNV v2014May07* programą, buvo nustatyti 3535 KSP, kurie sudarė 1438 KSPR lietuvių populiacijoje. Nustatyta, kad net 82,3 % lietuvių yra nors vieno KSP nešiotojai (žr.

3.1 lentelę), o vienam asmeniui vidutiniškai tenka ~12,1 KSP, kurių vidutinis dydis yra 86,67 Kb (nuo 1,5 Kb iki 3,5 Mb), o mediana 51,77 Kb. Delecijų ir duplikacijų santykis yra 4,5:1, o nustatyti KSP apima 9,5 % viso žmogaus genomo. KSP charakteristikos kiekviename tyrime gali skirtis dėl galimų genotipavimo klaidų, pasirenkamų metodų KSP nustatymui ir ar KSP nustatymas atliekamas vienu metodu ar keletu (rezultatų patikimumas didėja jeigu tyrimas atliekamas taikant daugiau nei vieną metodą), atliekamos kokybės kontrolės ir KSP filtravimo pasirenkamų kriterijų ir jų verčių.

Tokio tipo tyrimas naudojant dvi KSP nustatymo programas *QuantiSNP* ir *cnvPartition*, buvo atliktas Lietuvoje jau anksčiau A. Urnikytės. Buvo nustatyta, kad didžioji dalis Lietuvos populiacijos, 65,4 % yra nors vieno KSP nešiotojai, o vienam asmeniui vidutiniškai tenka 1,67 KSP, kurių vidutinis dydis yra 141,9 Kb, o mediana 78,2 Kb (5). Panašius tyrimo rezultatus gavo ir Švedijos mokslininkai atlikę švedų populiacijos tyrimą naudojant *PennCNV* ir *Birdsuite* programas. Jie nustatė, kad vienam švedui vidutiniškai tenka 7,1 KSP, kurių dydžių vidurkis yra 109,8 Kb. Nustatytas delecijų ir duplikacijų santykis 2,6:1 (43). Tailandiečių populiacijoje, KSP tyrimo metu, naudojant *PennCNV* ir *CBS* (angl. *Circular Binary Segmentation*), buvo nustatyta, kad vienam tailandiečiui vidutiniškai tenka 8 KSP, kurių dydžių vidurkis yra 59,8 Kb (6). W. Chen atliko trijų Europos šalių KSP tyrimą, t.y., Kroatijos, Škotijos ir Italijos populiacijų naudojant *QuantiSNP* ir *cnvPartition* KSP nustatymo programas. Kroatijos populiacijoje vienam kroatui vidutiniškai tenka apie 1,43 KSP ir jų dydžių vidurkis yra 216 Kb. Škotijos populiacijoje vienam škotui vidutiniškai tenka apie 0,91 KSP, o jų dydžių vidurkis yra 192,6 Kb. Italijos populiacijoje vienam italui vidutiniškai tenka apie 1,77 KSP, o jų dydžių vidurkis yra 201,6 Kb (44). Visa informacija apie KSP nustatymą skirtingose populiacijose pateikta 4.1 lentelėje.

Šio tyrimo metu sudarius KSPR lietuvių populiacijoje nustatyti 1438 KSPR iš kurių 58,83 % KSP (846 iš 1438) yra pavieniai, o like 41,17 % (592 iš 1438) yra aptinkami dažniau nei vieną kartą. Didžiausias daugia-alelinis KSP regionas stebimas 10 chromosomos 39088829..42509794 srityje, net 3,4 Mb dydžio (žr. 3.1 pav.). Tai paaiškina Nguyen su kitais mokslininkais atliktas tyrimas, kurio metu buvo nustatyta, kad centromeriniuose ir telomeriniuose chromosomų regionuose KSP pasireiškia dažniau, galbūt dėl pasikartojančių sekų šiuose genominiuose regionuose (45). Taip pat daug pasikartojančių sekų nustatoma ir akrocentrinių chromosomų trumpuosiuose pečiuose, todėl genotipavimo lustuose žymenų šioms sritims nėra. Tuo tarpu 19 chromosomoje nustatytas didžiausias KSPR tankis aptinkama delecijų, duplikacijų ir daugia-alelinių sričių. Šioje chromosomeoje 1 Mb susitelkę 3,66 KSP. Tai gali paaiškinti, kad santykis tarp chromosomų dydžio ir variantų skaičiaus nėra proporcingas. Kai kurios chromosomos turi gana didelį skaičių KSP, o kitos kur kas mažesni, lyginant su jų dydžiu. Pavyzdžiui, 18 chromosoma turi tik ~ 19,8 % KSP DNR sekoje, kai 16, 17, 19 ir 22 chromosomos turi >41 %

KSP. Tai gali atspindėti interchromosominius skirtumus genomo struktūroje, pavyzdžiui, 19 chromosomoje yra gausu segmentinių duplikacijų ir tandemiškai susikaupusių genų šeimų ir tai gali lemti dažnesnį KSP susidarymą (46).

Taip pat buvo apskaičiuoti KSPR dažniai lietuvių populiacijoje. Buvo nustatyta, kad daugiausia KSPR, 58,8 %, yra reti, kurių dažnis 0,05-0,5 %, o dažni, kurių dažnis >5 %, sudaro 2,1 % visų KSPR. Labai panašūs rezultatai buvo gauti Urnikytės A. tyrimo metu. Buvo nustatyti 57,5 % retų KSPR, o dažnų KSPR 1,1 % (5).

Atlikus nustatytų KSPR palyginimą su Genomo kitimų duomenų bazėje esamais KSP, nustatyta, kad net 98,2 % identifikuotų KSPR persidengia su sritimis esančiomis DGV, o likę 1,8 %, nebuvo nustatyti DGV, yra unikalūs lietuvių populiacijai. Net 70 % nustatytų unikalių lietuvių populiacijai KSPR yra reti, su 0,4 % dažniu. Dauguma nustatytų KSPR (87 %) persidengia su žinomais UCSC genais. KSP delecijos persidengiančiose su genais srityse nustatytos dažniau nei duplikacijos (84,7 % delecijų, o duplikacijų 15,3 %), skirtingai nei ankstesniame Lietuvos populiacijos tyrime, kurio metu didesnioji dalis su genais persidengiančių sričių buvo nustatyta duplikacijos, o ne delecijos. Lyginant šio ir ankščiau atlikto KSP tyrimo Lietuvoje rezultatus yra stebimi skirtumai. Šiuos skirtumus gali įtakoti tai, kad A.Urnikytės tyrime KSP nustatymui buvo taikomos dvi skirtingos KSP nustatymo programos ir atrinkti tik tie KSP, kurie buvo nustatyti abiem programomis, o šiame tyrime buvo taikoma viena programa.

4.1 lentelė. Skirtingų populiacijų KSP charakteristikos.

Taikytas metodas	<i>PennCNV</i>	<i>QuantiSNP</i> ir <i>CNVPartition</i> (Urnikytė A ir kt., 2016)	<i>PennCNV</i> ir <i>Birdsuite</i> (Teo ir kt., 2011)	<i>PennCNV</i> ir <i>CBS*</i> (Suktitipat B, 2014)	<i>QuantiSNP</i> ir <i>cnvPartition</i> (Chen W ir kt., 2011)		
Populiacija	Lietuvos	Lietuvos	Švedų	Tailandiečių	Kroatija	Škotija	Italija
Imtis	293	286	87	3017	965	691	1133
Nustatyti KSP	3535	478	623	23458	1384	630	2002
KSP/vienam asmeniui	12,1	1,67	7,1	8	1,43	0,91	1,77
KSP dydžių vidurkis (Kb)	86,67	141,9	109,8	59,8	216	192,6	201,6
KSP dydžių mediana (Kb)	51,77	78,2	28,6	-	-	-	-

*CBS- Circular Binary Segmentation

5. IŠVADOS

1. Buvo įdiegta ir įsisavinta KSP aptikimo programa *PennCNV v2014May07*;
2. *PennCNV* programa buvo nustatyta 3535 potencialūs KSP ir *cnv_tool_v1.4.2.py* programa buvo sukonstruoti 1438 KSPR. Tam, kad patvirtinti nustatytus KSP reikalingas patvirtinimas kitais metodais;
3. Palyginus KSPR su publikuotais Genomo kitimų duomenų bazėje, buvo nustatyta, kad didžioji dalis identifikuotų KSPR yra aprašyti ir tik 1,8 % yra unikalūs lietuvių populiacijai. Taip pat anotacijos rezultatai parodė, kad dažniausios su genais persidengiančių sričių funkcijos yra transkripcijos reguliacija, jonų transportas, imuninis atsakas, baltymų fosforilinimas, angiogenezė, mRNR brendimas, proteolizė, apoptozė, ląstelės ciklas, ląstelių adhezija, signalo perdavimas;
4. Nustačius 3535 KSP, paaiškėjo, kad 82,3 % lietuvių yra KSP nešiotojai, o vienam asmeniui vidutiniškai tenka 12,1 KSP. Sudarius 1438 KSPR nustatyta, kad lietuvių populiacijoje daugiausia KSPR (58,8 %) yra reti, kurių dažnis 0,05-0,5 %;
5. Atliekant KSP nustatymą *PennCNV* programa yra nustatoma daugiau vienam asmeniui tenkančių KSP, bet mažesnis jų dydžių vidurkis, o *QuantiSNP* programa nustatoma mažiau vienam asmeniui tenkančių KSP, bet didesni jų dydžių vidurkiai.

SANTRAUKA

„Lietuvos populiacijos genetinė analizė pagal DNR kopijų skaičiaus pokyčius“

DNR kopijų skaičiaus pokyčiai (KSP) yra chromosomos segmento kopijų skaičiaus pokytis svyruojantis nuo kelių šimtų bazių porų (bp) iki megabazių (Mb) ir skiriasi nuo tikėtosi kopijų skaičiaus, kuris yra referentiniame genome, dėl delecijų, duplikacijų, insercijų ar invertuotų duplikonų (1, 2). Šio darbo tikslas buvo nustatyti DNR kopijų skaičiaus pokyčius iš 293-ųjų LITGEN projekto dalyvių plataus masto genotipavimo duomenų ir įvertinti jų charakteristikas ir pasiskirstymą lietuvių populiacijoje. Iš viso nustatyta 3535 potencialūs KSP ir 82,3 % bendros lietuvių populiacijos asmenų yra KSP nešiotojai. Taip pat buvo sukonstruoti 1438 KSPR, kurie apima apie 94,8 Mb viso autosominio žmogaus genomo. Paskaičiavus KSPR dažnių pasiskirstymą nustatyta, kad lietuvių populiacijoje didžioji dalis KSPR (58,8 %) yra reti, jų dažnis yra iki 0,5 %. Iš visų nustatytų KSPR net 98,2 % persidengia su sritimis esančiomis Genomo kitimų duomenų bazėje (angl. *Database of Genomic Variants*, DGV), o likę 1,8 %, nebuvo nustatyti DGV ir yra unikalūs lietuvių populiacijai.

Raktiniai žodžiai: DGV, LITGEN projektas, kopijų skaičiaus pokyčiai, KSP, KSPR

SUMMARY

“Lithuanian population genetic analysis of copy number variation”

Copy number variation (CNV) is a chromosomal segment copy number variation which ranges from few hundred base pairs (bp) to few megabases (Mb) in length and varies in comparison with a reference genome, because of deletions, duplications, insertions or inverted duplikons (1, 2). The aim of this study was to identify DNA copy number variations for 293 LITGEN project participants from large-scale genotyping data and evaluate their characteristics and distribution in Lithuanian population. There were 3535 autosomal high-confidence CNVs identified and 82,3 % individuals of the Lithuanian population are CNV carriers. Also overlapping CNVs were merged and constructed 1438 CNVR which covers 94,8 Mb autosomal human genome. More than half identified CNVRs (58,8 %) in Lithuanian population are rare, with frequency $\geq 0,5$ %. From all identified CNVR even 98,2 % overlapped with regions in Database of Genomic Variance (DGV) and the remaining 1,8 % were not found in DGV and are unique for Lithuanian population.

Key words: DGV, LITGEN project, copy number variation, CNV, CNVRs

PADĖKA

Nuoširdžiai dėkoju magistro baigiamojo darbo vadovui, prof.habil.dr. Vaidučiui Kučinskui už suteiktą galimybę baigiamąjį magistro darbą atlikti Žmogaus ir medicininės genetikos katedroje. Taip pat dėkoju magistro baigiamojo darbo konsultantei, doktorantei Alinai Urnikytei, už skirtą laiką, dėmesį, vertingas pastabas ir visapusišką pagalbą rengiant magistro baigiamąjį darbą.

LITERATŪROS SĄRAŠAS

1. Wang K, Chen Z, Tadesse MG, Glessner J, Grant SF, Hakonarson H, et al. Modeling genetic inheritance of copy number variations. *Nucleic Acids Res.* 2008;36(21):e138.
2. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature.* 2006;444(7118):444-54.
3. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. *Science.* 2004;305(5683):525-8.
4. Park RW, Kim TM, Kasif S, Park PJ. Identification of rare germline copy number variations over-represented in five human cancer types. *Mol Cancer.* 2015;14:25.
5. Urnikyte A, Domarkiene I, Stoma S, Ambrozaityte L, Uktveryte I, Meskiene R, et al. CNV analysis in the Lithuanian population. *BMC Genet.* 2016;17(1):64.
6. Suktitipat B, Naktang C, Mhuantong W, Tularak T, Artiwet P, Pasomsap E, et al. Copy number variation in Thai population. *PLoS One.* 2014;9(8):e104355.
7. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. *Nat Genet.* 2004;36(9):949-51.
8. Zhao M, Zhao Z. CNVannotator: a comprehensive annotation server for copy number variation in the human genome. *PLoS One.* 2013;8(11):e80170.
9. Dierssen M, Herault Y, Estivill X. Aneuploidy: from a physiological mechanism of variance to Down syndrome. *Physiol Rev.* 2009;89(3):887-920.
10. Pirooznia M, Goes FS, Zandi PP. Whole-genome CNV analysis: advances in computational approaches. *Front Genet.* 2015;6:138.
11. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* 2014;42(Database issue):D986-92.
12. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. *Nature.* 2010;464(7289):704-12.
13. Gschwind AR, Singh A, Certa U, Reymond A, Heckel T. Diversity and regulatory impact of copy number variation in the primate *Macaca fascicularis*. *BMC Genomics.* 2017;18(1):144.
14. Ostrovnaya I, Nanjangud G, Olshen AB. A classification model for distinguishing copy number variants from cancer-related alterations. *BMC Bioinformatics.* 2010;11:297.
15. Chaignat E, Yahya-Graison EA, Henrichsen CN, Chrast J, Schütz F, Pradervand S, et al. Copy number variation modifies expression time courses. *Genome Res.* 2011;21(1):106-13.
16. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, et al. Copy number variation: new insights in genome diversity. *Genome Res.* 2006;16(8):949-61.

17. Lee C, Scherer SW. The clinical context of copy number variation in the human genome. *Expert Rev Mol Med*. 2010;12:e8.
18. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet*. 2011;12(5):363-76.
19. Johansson I, Ingelman-Sundberg M. Genetic polymorphism and toxicology--with emphasis on cytochrome p450. *Toxicol Sci*. 2011;120(1):1-13.
20. Merikangas AK, Corvin AP, Gallagher L. Copy-number variants in neurodevelopmental disorders: promises and challenges. *Trends Genet*. 2009;25(12):536-44.
21. Ledet EM, Hu X, Sartor O, Rayford W, Li M, Mandal D. Characterization of germline copy number variation in high-risk African American families with prostate cancer. *Prostate*. 2013;73(6):614-23.
22. Rodriguez-Revena L, Mila M, Rosenberg C, Lamb A, Lee C. Structural variation in the human genome: the impact of copy number variants on clinical diagnosis. *Genet Med*. 2007;9(9):600-6.
23. Tsuang DW, Millard SP, Ely B, Chi P, Wang K, Raskind WH, et al. The effect of algorithms on copy number variant detection. *PLoS One*. 2010;5(12):e14456.
24. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med*. 2010;61:437-55.
25. Blackburn A, Göring HH, Dean A, Carless MA, Dyer T, Kumar S, et al. Utilizing extended pedigree information for discovery and confirmation of copy number variable regions among Mexican Americans. *Eur J Hum Genet*. 2013;21(4):404-9.
26. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*. 2007;131(7):1235-47.
27. Hastings PJ, Ira G, Lupski JR. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet*. 2009;5(1):e1000327.
28. Arlt MF, Wilson TE, Glover TW. Replication stress and mechanisms of CNV formation. *Curr Opin Genet Dev*. 2012;22(3):204-10.
29. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. *Science*. 1991;251(4995):767-73.
30. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*. 1992;258(5083):818-21.
31. Baldwin EL, Lee JY, Blake DM, Bunke BP, Alexander CR, Kogan AL, et al. Enhanced detection of clinically relevant genomic imbalances using a targeted plus whole genome oligonucleotide microarray. *Genet Med*. 2008;10(6):415-29.

32. Xu L, Hou Y, Bickhart DM, Song J, Liu GE. Comparative Analysis of CNV Calling Algorithms: Literature Survey and a Case Study Using Bovine High-Density SNP Data. *Microarrays (Basel)*. 2013;2(3):171-85.
33. Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, et al. Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am J Hum Genet*. 2006;79(2):275-90.
34. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 2007;17(11):1665-74.
35. Winchester L, Yau C, Ragoussis J. Comparing CNV detection methods for SNP arrays. *Brief Funct Genomic Proteomic*. 2009;8(5):353-66.
36. Li TY, Zhang F. Screening of lung cancer related SNPs and CNVs with SNP microarrays. *Eur Rev Med Pharmacol Sci*. 2015;19(2):225-34.
37. Zhang X, Du R, Li S, Zhang F, Jin L, Wang H. Evaluation of copy number variation detection for a SNP array platform. *BMC Bioinformatics*. 2014;15:50.
38. Seiser EL, Innocenti F. Hidden Markov Model-Based CNV Detection Algorithms for Illumina Genotyping Microarrays. *Cancer Inform*. 2014;13(Suppl 7):77-83.
39. Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, et al. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res*. 2007;35(6):2013-25.
40. Pang AW, MacDonald JR, Pinto D, Wei J, Rafiq MA, Conrad DF, et al. Towards a comprehensive structural variation map of an individual human genome. *Genome Biol*. 2010;11(5):R52.
41. Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, et al. The generic genome browser: a building block for a model organism system database. *Genome Res*. 2002;12(10):1599-610.
42. Erikson GA, Deshpande N, Kesavan BG, Torkamani A. SG-ADVISED CNV: copy-number variant annotation and interpretation. *Genet Med*. 2015;17(9):714-8.
43. Teo SM, Ku CS, Naidoo N, Hall P, Chia KS, Salim A, et al. A population-based study of copy number variants and regions of homozygosity in healthy Swedish individuals. *J Hum Genet*. 2011;56(7):524-33.
44. Chen W, Hayward C, Wright AF, Hicks AA, Vitart V, Knott S, et al. Copy number variation across European populations. *PLoS One*. 2011;6(8):e23087.
45. Nguyen DQ, Webber C, Ponting CP. Bias of selection on human copy-number variants. *PLoS Genet*. 2006;2(2):e20.

46. de Smith AJ, Walters RG, Froguel P, Blakemore AI. Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease. *Cytogenet Genome Res.* 2008;123(1-4):17-26.

PRIEDAI

1 lentelė. Kopijų skaičiaus pokyčių regionai ir su jais susieti genai.

Chromosoma	Pradžia	Pabaiga	Ilgis	Varianto tipas	Geno pavadinimas
chr1	12866590	12880356	13766	gain	PRAMEF11
chr1	53803639	53832824	29185	loss	N
chr1	88657007	88909218	252211	loss	N
chr1	149039031	149201987	162956	loss	N
chr1	239093239	239097408	4169	loss	N
chr1	2784092	2850181	66089	loss	ACTRT2
chr1	7048779	7114641	65862	loss	CAMTA1
chr1	12866590	12880356	13766	loss	PRAMEF11
chr1	45610193	45716285	106092	loss	ZSWIM5
chr1	62666518	62680981	14463	loss	L1TD1
chr1	78585086	78611730	26644	loss	GIPC2
chr1	178984086	179022236	38150	loss	FAM20B
chr1	188735336	188826462	91126	loss	C1orf99
chr1	220849101	220879617	30516	loss	C1orf115
chr1	228699477	228805855	106378	loss	DUSP5P
chr1	34210532	34217930	7398	loss	CSMD2
chr1	46182670	46221440	38770	loss	IPP
chr1	244490369	244497339	6970	loss	C1orf100
chr1	7748525	7784902	36377	gain	CAMTA1
chr1	196823299	196855303	32004	gain	CFHR4
chr1	33394530	33450631	56101	loss	RNF19B
chr1	196823299	196855303	32004	loss	CFHR4
chr1	224300549	224346959	46410	loss	FBXO28
chr1	249170711	249199102	28391	loss	PGBD2
chr1	197482265	197644773	162508	loss	DENND1B
chr1	18543783	18897505	353722	loss	IGSF21
chr1	1026707	1066403	39696	loss	UNQ2998
chr1	169761107	169773317	12210	loss	C1orf112
chr1	147763232	147911246	148014	loss	DKFZp451B1418
chr1	882033	904165	22132	loss	NOC2L
chr1	154225344	154270877	45533	loss	UBAP2L
chr1	1646574	1691050	44476	gain	CDC2L1
chr1	1646574	1691050	44476	loss	CDC2L1
chr1	1257593	1288823	31230	gain	CPSF3L
chr1	1257593	1288823	31230	loss	CPSF3L
chr1	2473258	2532416	59158	loss	N
chr1	150664077	150679358	15281	loss	GOLPH3L
chr1	246968455	247008439	39984	loss	elys

chr1	247305873	247340784	34911	loss	AK131017
chr1	225944165	226015358	71193	loss	DKFZp564M2223
chr1	48588004	48721158	133154	loss	SKINTL
chr1	151999346	152152329	152983	loss	AK125884
chr1	189100707	189189282	88575	loss	DBCCR1L
chr1	37957375	38000184	42809	loss	C1orf149
chr1	63283005	63336271	53266	loss	ATG4C
chr1	168118292	168167139	48847	loss	TIPRL
chr1	3645051	3654746	9695	loss	TP73
chr1	46923102	46967890	44788	loss	DMBX1
chr1	157145225	157161346	16121	loss	ETV3
chr1	231727447	231813133	85686	gain	DISC1
chr1	1981117	2027900	46783	loss	PRKCZ
chr1	41866315	41962681	96366	loss	EDN2
chr1	108316189	108756690	440501	gain	VAV3
chr1	22203788	22233669	29881	loss	HSPG2
chr1	150799754	150864164	64410	loss	ARNT
chr1	240127293	240147503	20210	gain	CHRM3
chr1	100327026	100388748	61722	loss	AGL
chr1	9321241	9400868	79627	gain	H6PD
chr1	100458417	100609639	151222	loss	SLC35A3
chr1	97743804	97761972	18168	loss	DPYD
chr1	97862237	97873007	10770	loss	DPYD
chr1	98273735	98316285	42550	loss	DPYD
chr1	194288364	194351596	63232	loss	CDC73
chr1	25239288	25297216	57928	loss	RUNX3
chr1	247149996	247167077	17081	loss	ZNF695
chr1	28636692	28780454	143762	gain	MED18
chr1	104113158	104129880	16722	gain	RNPC3
chr1	150270209	150448688	178479	loss	MRPS21
chr1	156273061	156333831	60770	loss	CCT3
chr1	17597720	17610899	13179	gain	PADI3
chr1	94352380	94375735	23355	loss	GCLM
chr1	6440287	6520311	80024	loss	ACOT7
chr1	8989235	9061382	72147	gain	CA6
chr1	16371931	16495998	124067	loss	CLCNKB
chr1	25889421	25900240	10819	loss	LDLRAP1
chr1	1210471	1233941	23470	loss	SCNN1D
chr1	198389958	198433606	43648	loss	ATP6V1G3
chr1	219865236	219914351	49115	loss	N
chr1	36581791	36639778	57987	loss	TRAPPC3
chr1	3362256	3464451	102195	loss	ARHGEF16

chr1	247559156	247568051	8895	loss	NLRP3
chr1	245636915	245647726	10811	loss	KIF26B
chr1	225090870	225211934	121064	loss	DNAH14
chr1	163348444	163359017	10573	loss	NUF2
chr1	212943581	212978197	34616	loss	NSL1
chr1	151428605	151464911	36306	loss	POGZ
chr1	115537933	115591058	53125	loss	SYCP1
chr1	72818226	72838529	20303	loss	NEGR1
chr1	79501129	79519555	18426	loss	ELTD1
chr1	223003498	223068480	64982	loss	DISP1
chr1	51423567	51441961	18394	loss	FAF1
chr1	35341981	35428427	86446	loss	DLGAP3
chr1	33850889	33877807	26918	loss	PHC2
chr1	247828797	247845256	16459	loss	OR13G1
chr1	247929510	247977420	47910	loss	OR14A16
chr1	248426484	248515194	88710	loss	OR2T33
chr1	29268715	29363984	95269	loss	EPB41
chr1	3479742	3510633	30891	loss	MEGF6
chr1	2366316	2404255	37939	loss	PLCH2
chr1	120113492	120145883	32391	loss	ZNF697
chr1	120162787	120188180	25393	loss	ZNF697
chr1	191833819	191887735	53916	loss	RGS18
chr1	153041903	153069584	27681	loss	N
chr1	8278899	8390888	111989	loss	SLC45A1
chr1	113455099	113494171	39072	loss	SLC16A1
chr1	174468208	174507529	39321	loss	RABGAP1L
chr1	33719641	33790495	70854	loss	ZNF362
chr1	10762432	10796865	34433	loss	CASZ1
chr1	22760325	22807171	46846	loss	ZBTB40
chr1	41346428	41378076	31648	gain	CITED4
chr1	47496737	47532614	35877	loss	CYP4X1
chr10	39088829	42509794	3420965	gain	N
chr10	42725663	42827951	102288	gain	N
chr10	47555294	47619334	64040	gain	N
chr10	47630260	47662831	32571	gain	N
chr10	135298293	135336514	38221	gain	N
chr10	2067049	2093110	26061	loss	N
chr10	26931729	26950217	18488	loss	N
chr10	39088829	42509794	3420965	loss	N
chr10	42725663	42827951	102288	loss	N
chr10	47630260	47662831	32571	loss	N
chr10	65550686	65587937	37251	loss	N

chr10	66665315	66729281	63966	loss	N
chr10	67067094	67105065	37971	loss	N
chr10	107618048	107682005	63957	loss	N
chr10	107984319	108015495	31176	loss	N
chr10	130128297	130152724	24427	loss	N
chr10	130197453	130214365	16912	loss	N
chr10	135298293	135336514	38221	loss	N
chr10	800213	813426	13213	loss	LARP4B
chr10	19796290	19830436	34146	loss	AK297683
chr10	132626327	132653724	27397	loss	TCERG1L
chr10	133932946	133953468	20522	loss	JAKMIP3
chr10	134244680	134267284	22604	loss	C10orf91
chr10	45666965	45709496	42531	loss	N
chr10	65274927	65359701	84774	loss	REEP3
chr10	20449266	20827105	377839	loss	PLXDC2
chr10	124660458	124704148	43690	loss	FAM24A
chr10	5642758	5652277	9519	gain	ASB13
chr10	13056587	13060410	3823	loss	N
chr10	588126	754033	165907	loss	DIP2C
chr10	82875729	82890180	14451	loss	SH2D4B
chr10	102668856	102707526	38670	loss	FAM178A
chr10	95431535	95459817	28282	loss	C10orf4
chr10	134724151	134784486	60335	loss	DKFZp434A1721
chr10	1171353	1250184	78831	loss	WDR37
chr10	8086169	8108591	22422	loss	N
chr10	1084359	1132851	48492	loss	C10orf110
chr10	22607432	22639707	32275	loss	COMMD3
chr10	21796081	21838380	42299	loss	C10orf140
chr10	133997031	134069930	72899	loss	JAKMIP3
chr10	991737	1034412	42675	gain	CRFG
chr10	5019319	5036440	17121	loss	AKR1CL2
chr10	57736009	58063346	327337	loss	ZWINT
chr10	58904859	58936939	32080	loss	ZWINT
chr10	91468924	91498254	29330	loss	KIF20B
chr10	31662485	31747464	84979	loss	ZEB1
chr10	135656	230625	94969	loss	ZMYND11
chr10	49723037	49735563	12526	loss	ARHGAP22
chr10	83153009	83178907	25898	loss	NRG3
chr10	80964520	81021841	57321	loss	ZMIZ1
chr10	5348337	5389209	40872	gain	UCN3
chr10	88496391	88555186	58795	loss	BMPR1A
chr10	53557566	53593978	36412	loss	PRKG1

chr10	112929753	112979082	49329	gain	ADRA2A
chr10	70431420	70577222	145802	loss	TET1
chr10	120927214	120952674	25460	loss	PRDX3
chr10	3114645	3219445	104800	loss	PFKP
chr10	75937553	76373903	436350	loss	ADK
chr10	70174706	70271353	96647	loss	DNA2
chr10	135053604	135089034	35430	gain	VENTX
chr10	135053604	135089034	35430	loss	VENTX
chr10	1292991	1310657	17666	loss	ADARB2
chr10	1653960	1664242	10282	loss	ADARB2
chr10	1521858	1569586	47728	loss	ADARB2
chr10	129692293	129753711	61418	loss	PTPRE
chr10	12501397	12512388	10991	loss	CAMK1D
chr10	112231445	112285812	54367	loss	DUSP5
chr10	17623775	17636308	12533	loss	PTPLA
chr10	72431601	72445630	14029	loss	ADAMTS14
chr10	120453168	120494372	41204	loss	C10orf46
chr10	71959947	71992518	32571	loss	PPA1
chr10	27524480	27577773	53293	loss	ACBD5
chr10	74141852	74175914	34062	loss	CBARA1
chr10	18250004	18825693	575689	gain	SLC39A12
chr10	94660133	94765333	105200	loss	EXOC6
chr10	27155368	27201386	46018	loss	ABII
chr10	134557068	134609699	52631	loss	INPP5A
chr10	56989733	57008142	18409	loss	PCDH15
chr10	55667100	55742717	75617	loss	PCDH15
chr10	134935916	134968464	32548	gain	GPR123
chr10	134884595	134902795	18200	loss	GPR123
chr10	134935916	134968464	32548	loss	GPR123
chr10	5517065	5552475	35410	loss	CALML5
chr10	32070330	32085248	14918	loss	N
chr10	117121966	117339523	217557	loss	ATRNL1
chr10	129330043	129350641	20598	loss	NPS
chr10	105331761	105467895	136134	loss	NEURL
chr10	126020788	126049426	28638	loss	OAT
chr10	135002837	135053239	50402	loss	KNDC1
chr10	96499710	96557335	57625	loss	CYP2C19
chr10	104843148	104939232	96084	loss	NT5C2
chr10	68078481	68107483	29002	loss	CTNNA3
chr10	68350382	68374084	23702	loss	CTNNA3
chr10	68382135	68412958	30823	loss	CTNNA3
chr10	68827014	68922242	95228	gain	CTNNA3

chr10	68827014	68922242	95228	loss	CTNNA3
chr10	126629204	126676686	47482	loss	ZRANB1
chr10	70941113	70976702	35589	gain	SUPV3L1
chr10	43414098	43444534	30436	loss	BMS1
chr10	21907016	21974979	67963	loss	MLLT10
chr10	64560470	64572169	11699	loss	ADO
chr11	37746335	37834436	88101	gain	N
chr11	2034098	2057646	23548	gain	H19
chr11	2066031	2076042	10011	gain	H19
chr11	23386182	23416800	30618	loss	N
chr11	37501843	37532517	30674	loss	N
chr11	80916305	80921481	5176	loss	N
chr11	81503992	81517261	13269	loss	N
chr11	134916745	134934063	17318	loss	N
chr11	1051714	1101077	49363	loss	MUC2
chr11	2066031	2076042	10011	loss	H19
chr11	72265848	72279549	13701	loss	AL832797
chr11	85220773	85292310	71537	loss	DLG2
chr11	133809032	133821131	12099	loss	IGSF9B
chr11	50487842	50697498	209656	gain	N
chr11	50487842	50697498	209656	loss	N
chr11	25453442	25507340	53898	loss	LUZP2
chr11	25610986	25642425	31439	loss	LUZP2
chr11	86825869	86884115	58246	loss	TMEM135
chr11	128868090	129090852	222762	loss	RICS
chr11	17306222	17339127	32905	loss	NUCB2
chr11	723428	816016	92588	loss	EPS8L2
chr11	1515270	1607345	92075	loss	HCCA2
chr11	63859120	63871549	12429	loss	MACROD1
chr11	249097	274780	25683	gain	S11
chr11	64538346	64584986	46640	loss	SF1
chr11	76805777	76836438	30661	loss	CAPN5
chr11	75884832	75915138	30306	loss	WNT11
chr11	108158133	108219338	61205	loss	ATM
chr11	57377991	57434122	56131	loss	SERPING1
chr11	67171736	67280620	108884	loss	PPP1CA
chr11	134157072	134208908	51836	gain	GLB1L3
chr11	118550601	118579747	29146	gain	TREH
chr11	67793014	67821035	28021	gain	ALDH3B1
chr11	65431360	65481165	49805	loss	KAT5
chr11	236091	248016	11925	loss	SIRT3
chr11	2544870	2612560	67690	gain	KCNQ1

chr11	2617781	2631302	13521	loss	KCNQ1
chr11	8959020	8967510	8490	loss	ASCL3
chr11	127857444	127895471	38027	gain	ETS1
chr11	502721	567390	64669	loss	RNH1
chr11	1401889	1462216	60327	loss	BRSK2
chr11	126242697	126263885	21188	loss	ST3GAL4
chr11	130352545	130375349	22804	loss	ADAMTS15
chr11	104041862	105273706	1231844	gain	CASP12
chr11	9405882	9444199	38317	loss	IPO7
chr11	120557343	120630452	73109	loss	GRIK4
chr11	107659024	107670228	11204	gain	SLC35F2
chr11	69992264	70009720	17456	loss	ANO1
chr11	62901531	62916395	14864	loss	SLC22A24
chr11	63483576	63516825	33249	loss	RTN3
chr11	46777958	46872873	94915	loss	CKAP5
chr11	65338944	65382564	43620	loss	SSSCA1
chr11	1126978	1162401	35423	loss	MUC5AC
chr11	99525126	99566782	41656	loss	CNTN5
chr11	131324796	131341372	16576	gain	NTM
chr11	131651470	131658127	6657	loss	NTM
chr11	833667	840033	6366	loss	CD151
chr11	392953	408352	15399	gain	PKP3
chr11	392953	408352	15399	loss	PKP3
chr11	92648639	92868270	219631	loss	MTNR1B
chr11	87946467	87966174	19707	loss	RAB38
chr11	198510	219089	20579	gain	ODF3
chr11	8324546	8330352	5806	loss	LMO1
chr11	96212445	96267523	55078	loss	JRKL
chr11	96804466	96823102	18636	loss	JRKL
chr11	97399271	97443131	43860	loss	JRKL
chr11	48747611	48878863	131252	gain	OR4A47
chr11	48747611	48878863	131252	loss	OR4A47
chr11	55370325	55427700	57375	loss	OR4C11
chr11	47659134	48133526	474392	gain	MTCH2
chr12	32024256	32056577	32321	gain	C12orf35
chr12	74307539	74417792	110253	loss	N
chr12	94881995	94893462	11467	loss	N
chr12	131730317	131825125	94808	loss	N
chr12	13505640	13517574	11934	loss	C12orf36
chr12	31053979	31074862	20883	loss	TSPAN11
chr12	133140840	133164690	23850	loss	FBRSL1
chr12	63979105	64086156	107051	gain	DPY19L2

chr12	74681897	74714268	32371	loss	N
chr12	128398477	128414474	15997	loss	N
chr12	10094643	10123677	29034	loss	CR625615
chr12	40025034	40118248	93214	loss	C12orf40
chr12	63979105	64086156	107051	loss	DPY19L2
chr12	80746885	80775055	28170	loss	C12orf64
chr12	31259371	31385093	125722	gain	N
chr12	59312113	59398014	85901	loss	LRIG3
chr12	97013604	97047347	33743	loss	AX747187
chr12	99742239	99813939	71700	loss	ANKS1B
chr12	10680216	10747697	67481	loss	N
chr12	83179677	83206342	26665	loss	TMTC2
chr12	50206723	50247467	40744	loss	NCKAP5L
chr12	25648676	25654445	5769	loss	IFLTD1
chr12	31408696	31413325	4629	gain	FAM60A
chr12	31408696	31413325	4629	loss	FAM60A
chr12	89885923	89914260	28337	loss	TUWD12
chr12	99007355	99077742	70387	loss	IKIP
chr12	9799126	9834754	35628	loss	N
chr12	72088303	72209525	121222	loss	TMEM19
chr12	124773340	124872587	99247	loss	FAM101A
chr12	1272363	1510640	238277	gain	ELKS
chr12	1272363	1510640	238277	loss	ELKS
chr12	133481917	133501212	19295	loss	AK055957
chr12	64786176	64875548	89372	loss	BC042855
chr12	52348258	52423179	74921	loss	ACVR1B
chr12	118622836	118749798	126962	loss	TAOK3
chr12	26928425	26951251	22826	loss	ITPR2
chr12	91726350	91804509	78159	gain	DCN
chr12	6238592	6263819	25227	loss	VWF
chr12	123413311	123464279	50968	loss	ABCB9
chr12	63586581	63633448	46867	loss	AVPR1A
chr12	21686526	21761315	74789	loss	GYS2
chr12	12179741	12209844	30103	loss	ETV6
chr12	133729999	133777645	47646	loss	ZNF10
chr12	51511264	51561056	49792	loss	TFCP2
chr12	70672316	70713522	41206	loss	CNOT2
chr12	56497625	56607927	110302	loss	PA2G4
chr12	132619099	132654262	35163	loss	DDX51
chr12	113495646	113541940	46294	loss	DTX1
chr12	7756706	7782625	25919	loss	APOBEC1
chr12	32468481	32545352	76871	loss	BICD1

chr12	69645384	69663047	17663	loss	CPSF6
chr12	34403574	34565016	161442	gain	ALG10
chr12	34642579	34853011	210432	gain	ALG10
chr12	38072773	38155198	82425	gain	ALG10B
chr12	34403574	34565016	161442	loss	ALG10
chr12	38072773	38155198	82425	loss	ALG10B
chr12	56876372	56935652	59280	loss	GLS2
chr12	20986061	21032242	46181	loss	SLCO1B3
chr12	20899716	20928803	29087	loss	SLCO1C1
chr12	129226773	129266496	39723	gain	SLC15A4
chr12	32758226	32879346	121120	loss	FGD4
chr12	7617083	7645776	28693	loss	CD163
chr12	52845952	52869980	24028	loss	KRT6C
chr12	49569311	49586489	17178	loss	TUBA1B
chr12	102477708	102527494	49786	loss	NUP37
chr12	27050208	27199180	148972	loss	C12orf11
chr12	18443809	18463608	19799	loss	PIK3C2G
chr12	65504969	65579872	74903	loss	WIF1
chr12	8000911	8044997	44086	gain	SLC2A14
chr12	8000911	8044997	44086	loss	SLC2A14
chr12	55843589	55864695	21106	loss	OR6C2
chr12	129802877	129866293	63416	gain	TMEM132D
chr12	129935733	130480186	544453	gain	TMEM132D
chr12	112425746	112471090	45344	loss	TMEM116
chr12	58137563	58178162	40599	loss	TSPAN31
chr12	118984590	119039724	55134	loss	SUDS3
chr12	191619	218719	27100	loss	IQSEC3
chr13	19259939	19377679	117740	gain	N
chr13	23399100	23437981	38881	gain	N
chr13	104772962	104810437	37475	gain	N
chr13	19683883	19719615	35732	loss	N
chr13	23399100	23437981	38881	loss	N
chr13	55145502	55468512	323010	loss	N
chr13	56279111	56363447	84336	loss	N
chr13	69247022	69267981	20959	loss	N
chr13	70744882	70773398	28516	loss	N
chr13	76726422	76739423	13001	loss	N
chr13	57630924	57704823	73899	loss	PRR20
chr13	73005122	73053409	48287	loss	C13orf37
chr13	112181401	112225200	43799	loss	C13orf16
chr13	112307701	112325591	17890	loss	C13orf16
chr13	63250900	63328523	77623	loss	N

chr13	63922734	63972811	50077	loss	N
chr13	30800292	30874175	73883	loss	KATNAL1
chr13	115075317	115103529	28212	loss	ZNF828
chr13	51119737	51481769	362032	gain	BCMS
chr13	22007076	22159277	152201	loss	ZDHHC20
chr13	50806640	50933929	127289	gain	N
chr13	30997830	31041594	43764	loss	HMGIL10
chr13	114506082	114527839	21757	loss	FAM70B
chr13	59364350	59424460	60110	loss	N
chr13	41215695	41330170	114475	loss	FOXO1
chr13	100615239	100638392	23153	loss	ZIC5
chr13	50025331	50086266	60935	loss	SETDB2
chr13	110290826	110325749	34923	gain	IRS2
chr13	73340141	73596658	256517	loss	DIS3
chr13	103194007	103200341	6334	loss	TPP2
chr13	24387412	24436475	49063	loss	MIPEP
chr13	20949049	20958746	9697	loss	CRYL1
chr13	25201908	25250059	48151	gain	ATP12A
chr13	103692842	103699213	6371	gain	SLC10A2
chr13	104062178	104137962	75784	loss	SLC10A2
chr13	99355106	99358478	3372	gain	SLC15A1
chr13	95937148	95983387	46239	gain	ABCC4
chr13	96004493	96030662	26169	gain	ABCC4
chr13	95748220	95752224	4004	loss	ABCC4
chr13	95867802	95870021	2219	loss	ABCC4
chr13	20847138	20875548	28410	gain	GJB6
chr13	113683045	113774467	91422	loss	MCF2L
chr13	33174224	33321044	146820	loss	PDS5B
chr13	50134099	50187949	53850	loss	RCBTB1
chr13	62935904	62955214	19310	loss	PCDH20
chr13	65909171	65939433	30262	loss	N
chr13	80909943	80924165	14222	loss	SPRY2
chr13	114900150	114941946	41796	loss	RASA3
chr13	114817068	114847914	30846	loss	RASA3
chr13	42767185	42812355	45170	loss	DGKH
chr13	83786517	83799793	13276	loss	SLITRK1
chr13	84103707	84157926	54219	loss	SLITRK1
chr13	84434888	84459077	24189	loss	SLITRK1
chr13	41663555	41778889	115334	loss	KBTBD6
chr13	113994246	114023401	29155	loss	GRTP1
chr13	24881184	24980949	99765	loss	SPATA13
chr14	27462794	27517393	54599	loss	N

chr14	40929306	41298408	369102	loss	N
chr14	101094176	101112530	18354	loss	N
chr14	101130576	101142889	12313	loss	N
chr14	104872237	104919105	46868	loss	N
chr14	51866997	51903109	36112	loss	FRMD6
chr14	77560478	77583216	22738	loss	KIAA1737
chr14	107212287	107287663	75376	loss	N
chr14	39702329	39802186	99857	loss	MIA2
chr14	106004323	106199579	195256	loss	N
chr14	105871769	105906073	34304	loss	BC127913
chr14	63775684	63857808	82124	gain	GPHB5
chr14	34921900	35110512	188612	loss	C14orf147
chr14	64923976	64988924	64948	loss	MTHFD1
chr14	38058065	38088472	30407	loss	FOXA1
chr14	31380191	31392763	12572	loss	STRN3
chr14	105623498	105646179	22681	loss	JAG2
chr14	104624689	104630711	6022	loss	KIF26A
chr14	103232425	103306215	73790	loss	TRAF3
chr14	20780366	20802314	21948	loss	CCNB1IP1
chr14	37000781	37044193	43412	loss	NKX2-8
chr14	35221860	35347512	125652	loss	BAZ1A
chr14	105673692	105685865	12173	loss	BRF1
chr14	21358155	21405885	47730	gain	RNASE3
chr14	23148504	23181524	33020	loss	OXA1L
chr14	103862321	103918028	55707	loss	MARK3
chr14	35477499	35523266	45767	loss	SRP54
chr14	54996630	55094787	98157	loss	CGRRF1
chr14	22237187	22262039	24852	loss	TCRA
chr14	22599159	22611787	12628	loss	TCRA
chr14	22749742	22764078	14336	loss	TCRA
chr14	51074460	51101641	27181	loss	ATL1
chr14	86287347	86310164	22817	loss	FLRT2
chr14	53265770	53324891	59121	loss	FERMT2
chr14	52604957	52673160	68203	loss	PTGDR
chr14	22091105	22189305	98200	loss	OR10G2
chr14	20213937	20420338	206401	gain	OR4Q3
chr14	20213937	20420338	206401	loss	OR4Q3
chr14	35703521	35717282	13761	loss	KIAA0391
chr14	105243930	105314093	70163	loss	INF2
chr14	105100021	105148654	48633	loss	INF2
chr14	25757475	25795870	38395	loss	STXBP6
chr14	43939771	44000927	61156	gain	FSCB

chr14	44150766	44292051	141285	gain	FSCB
chr14	43939771	44000927	61156	loss	FSCB
chr14	24453366	24484060	30694	loss	DHRS4
chr15	24521050	24668007	146957	gain	N
chr15	32509892	32514341	4449	gain	N
chr15	24521050	24668007	146957	loss	N
chr15	32509892	32514341	4449	loss	N
chr15	34718594	34800595	82001	loss	N
chr15	36018978	36036487	17509	loss	N
chr15	93262252	93310575	48323	loss	N
chr15	93986669	93991307	4638	loss	N
chr15	37922267	37956447	34180	loss	TMCO5A
chr15	45202919	45268182	65263	loss	C15orf43
chr15	59736724	59779534	42810	loss	FAM81A
chr15	81960409	81989193	28784	gain	TMC3
chr15	57644456	57775826	131370	gain	CGNL1
chr15	20997359	21143735	146376	loss	POTEB
chr15	81960409	81989193	28784	loss	TMC3
chr15	25414436	25465998	51562	gain	N
chr15	58034646	58183524	148878	gain	GCOM1
chr15	50847401	50952909	105508	loss	CHAK1
chr15	79167117	79190131	23014	loss	MORF4L1
chr15	34929955	34979705	49750	loss	GJD2
chr15	58949991	59003311	53320	loss	ADAM10
chr15	29261716	29309931	48215	loss	APBA2
chr15	24056437	24063516	7079	loss	NDN
chr15	41682094	41739356	57262	loss	NDUFAF1
chr15	41948046	42065422	117376	loss	MGA
chr15	57225766	57279185	53419	loss	TCF12
chr15	50564884	50704367	139483	loss	GABPB1
chr15	71135323	71309592	174269	loss	LARP6
chr15	82009329	82088114	78785	gain	MEX3B
chr15	40555065	40673124	118059	loss	PAK6
chr15	55645661	55752669	107008	loss	PIGB
chr15	87831146	87848502	17356	loss	AGBL1
chr15	43286746	43485930	199184	loss	UBR1
chr15	27857708	27880721	23013	loss	GABRG3
chr15	59971635	60030307	58672	gain	BNIP2
chr15	59189972	59205253	15281	loss	SLTM
chr15	99159955	99214985	55030	loss	IGF1R
chr15	98615560	98676558	60998	loss	ARRDC4
chr15	30936285	30968006	31721	loss	N

chr15	41159606	41194151	34545	loss	RHOV
chr15	52077041	52112359	35318	loss	TMOD2
chr15	22924097	22949828	25731	gain	CYFIP1
chr15	22924097	22949828	25731	loss	CYFIP1
chr15	47406224	47451005	44781	loss	SEMA6D
chr15	47640978	47661431	20453	loss	SEMA6D
chr15	43893817	43939642	45825	loss	STRC
chr15	102338427	102366849	28422	gain	OR4F6
chr15	102338427	102366849	28422	loss	OR4F6
chr15	62243196	62340126	96930	loss	VPS13C
chr15	64347946	64382738	34792	loss	FAM96A
chr15	101874762	101890811	16049	loss	PCSK6
chr15	52257874	52319695	61821	gain	LEO1
chr15	94790863	94818923	28060	loss	MCTP2
chr15	46661811	46682794	20983	loss	SQRDL
chr16	11547488	11573613	26125	loss	N
chr16	1134818	1170512	35694	loss	C1QTNF8
chr16	9118815	9201969	83154	loss	C16orf72
chr16	34516074	34687052	170978	gain	N
chr16	18970621	19022447	51826	gain	TMC7
chr16	5885940	6028451	142511	loss	A2BP1
chr16	29262404	29412503	150099	gain	NPIPL2
chr16	6736475	6812292	75817	loss	A2BP1
chr16	7115289	7188406	73117	loss	A2BP1
chr16	767519	787827	20308	gain	AL360260
chr16	767519	787827	20308	loss	AL360260
chr16	2945090	3036551	91461	loss	FLYWCH2
chr16	163666	212648	48982	loss	C16orf35
chr16	4751045	4812209	61164	loss	ANKS3
chr16	1468750	1496005	27255	loss	C16orf91
chr16	88916851	88987861	71010	loss	GALNS
chr16	21578388	21740231	161843	gain	LOC23117
chr16	21578388	21740231	161843	loss	LOC23117
chr16	90074001	90097748	23747	loss	DBNDD1
chr16	69666939	69734562	67623	loss	KIAA0827
chr16	4296785	4408267	111482	loss	AK094828
chr16	575312	618966	43654	loss	C16orf10
chr16	30147264	30215620	68356	loss	AK303166
chr16	29927702	29968015	40313	loss	AK303166
chr16	28930759	28952063	21304	loss	NPIPL2
chr16	87301475	87342510	41035	loss	AK125749
chr16	55843251	55864452	21201	loss	CES hBr1

chr16	88868252	88880480	12228	loss	CDT1
chr16	8853787	9035124	181337	gain	ABAT
chr16	348222	433623	85401	loss	AXIN1
chr16	2125787	2144176	18389	loss	TSC2
chr16	75548080	75580985	32905	gain	CHST5
chr16	17077040	17135449	58409	loss	XYLT1
chr16	80822749	80861676	38927	loss	CDYL2
chr16	90121131	90163275	42144	loss	PRDM7
chr16	2022074	2085996	63922	loss	TBL3
chr16	88164	131108	42944	loss	POLR3K
chr16	1374817	1448439	73622	gain	UBE2I
chr16	46760715	46889354	128639	loss	MYLK3
chr16	5123474	5176946	53472	loss	ALG1
chr16	84131705	84146860	15155	loss	MBTPS1
chr16	2296850	2319716	22866	loss	DCI
chr16	74829373	74873927	44554	loss	FA2H
chr16	1184967	1201854	16887	loss	CACNA1H
chr16	1268828	1311333	42505	loss	CACNA1H
chr16	2387259	2471582	84323	loss	ABCA3
chr16	3115627	3140616	24989	loss	IL32
chr16	1893882	1916103	22221	loss	C16orf73
chr16	824097	845819	21722	loss	MSLNL
chr16	678877	749057	70180	loss	RAB40C
chr16	83990089	84031722	41633	loss	OSGIN1
chr16	88595391	88616697	21306	gain	ZFPM1
chr16	88595391	88616697	21306	loss	ZFPM1
chr16	1025982	1047729	21747	loss	LMF1
chr16	1802231	1824075	21844	gain	MAPK8IP3
chr16	1802231	1824075	21844	loss	MAPK8IP3
chr16	25954603	26023509	68906	loss	HS3ST4
chr16	87727737	87769951	42214	loss	JPH3
chr16	13616086	13643974	27888	loss	SHISA9
chr17	20699545	20766280	66735	loss	N
chr17	67873145	67972274	99129	loss	N
chr17	2624906	2670267	45361	loss	KIAA0664
chr17	77367697	77394037	26340	gain	HRNBP3
chr17	21363577	21528062	164485	loss	C17orf51
chr17	77141711	77192452	50741	loss	HRNBP3
chr17	80980140	81006628	26488	loss	B3GNTL1
chr17	65065890	65072560	6670	loss	HELZ
chr17	45482857	45504869	22012	loss	C17orf57
chr17	3715038	3730978	15940	loss	C17orf85

chr17	44165803	44189066	23263	gain	KIAA1267
chr17	28457048	28497259	40211	loss	CCDC55
chr17	58955084	59047387	92303	loss	BCAS3
chr17	42712760	42775217	62457	loss	C17orf104
chr17	33684035	33760166	76131	loss	SLFN11
chr17	3343406	3378876	35470	loss	SPATA22
chr17	42219773	42305706	85933	loss	C17orf53
chr17	79899899	79953622	53723	loss	MYADML2
chr17	7766228	7801509	35281	loss	AK128353
chr17	4519139	4564690	45551	loss	BC043554
chr17	77654710	77708263	53553	loss	LOC651900
chr17	58015258	58048915	33657	loss	RPS6KB1
chr17	76208850	76228629	19779	loss	BIRC5
chr17	77756363	77798847	42484	loss	CBX2
chr17	64236477	64288637	52160	loss	PRKCA
chr17	64410915	64482150	71235	loss	PRKCA
chr17	56056706	56080449	23743	loss	VEZF1
chr17	46657927	46688371	30444	loss	HOXB3
chr17	76578308	76595130	16822	gain	DNAH17
chr17	39743081	39874075	130994	loss	JUP
chr17	1267857	1447756	179899	loss	YWHAE
chr17	9001372	9029118	27746	loss	NTN1
chr17	6295959	6307248	11289	gain	AIPL1
chr17	80330675	80377689	47014	loss	UTS2R
chr17	7190435	7233011	42576	loss	SLC2A4
chr17	19510081	19528641	18560	loss	ALDH3A2
chr17	74313085	74358542	45457	loss	PRPSAP1
chr17	80453932	80538945	85013	loss	FOXK2
chr17	71836273	72569644	733371	gain	RPL38
chr17	33244038	33297064	53026	loss	CCT6B
chr17	27869968	27935688	65720	loss	TAOK1
chr17	66270082	66289041	18959	loss	ARSG
chr17	6850368	6867601	17233	loss	ALOX12
chr17	21267589	21303007	35418	gain	KCNJ12
chr17	21267589	21303007	35418	loss	KCNJ12
chr17	73033628	73052648	19020	loss	KCTD2
chr17	4357773	4375506	17733	loss	SPNS3
chr17	173906	212731	38825	loss	RPH3AL
chr17	12344	19389	7045	loss	DOC2B
chr17	66460315	66483524	23209	loss	WIPI1
chr17	50949615	51314354	364739	loss	KIF2B
chr17	75330920	75352172	21252	loss	9-Sep

chr17	79079292	79127077	47785	loss	BAIAP2
chr17	37767453	37784164	16711	loss	PPP1R1B
chr17	37790370	37843858	53488	loss	PPP1R1B
chr17	72630430	72667519	37089	gain	RAB37
chr17	873571	936300	62729	loss	NXN
chr17	3111045	3119767	8722	gain	OR1A1
chr17	45919465	45942176	22711	loss	SP6
chr17	1130470	1168676	38206	loss	TUSC5
chr17	40868087	40904514	36427	loss	EZH1
chr17	34461869	34476396	14527	gain	TBC1D3B
chr17	15644506	15695001	50495	loss	TBC1D26
chr18	12040933	12096249	55316	loss	BX648696
chr18	67209141	67217271	8130	loss	DOK6
chr18	66747568	66755508	7940	loss	CCDC102B
chr18	46667299	46786058	118759	loss	DYM
chr18	31386615	31413448	26833	loss	NOL4
chr18	44540969	44578794	37825	loss	DKFZp667C165
chr18	77900829	77931598	30769	loss	LOC100130522
chr18	24735468	24746598	11130	loss	C18orf16
chr18	51725404	51765409	40005	loss	MBD2
chr18	46210577	46221152	10575	loss	KIAA0427
chr18	7070642	7083007	12365	loss	LAMA1
chr18	29203385	29226444	23059	loss	B4GALT6
chr18	19177285	19235006	57721	loss	ESCO1
chr18	77241665	77279749	38084	loss	NFATC1
chr18	23831649	23927627	95978	loss	TAF4B
chr18	6424209	6440712	16503	loss	L3MBTL4
chr18	716739	724611	7872	loss	YES1
chr18	9625750	9642637	16887	loss	PPP4R1
chr18	159382	220658	61276	loss	USP14
chr18	77545200	77661170	115970	loss	KCNG2
chr18	60084646	60162791	78145	gain	ZCCHC2
chr18	8664894	8679822	14928	loss	RAB12
chr18	4325322	4330817	5495	loss	DLGAP1
chr18	1590043	1625222	35179	loss	N
chr18	1725758	1839387	113629	loss	N
chr18	1931566	1967534	35968	loss	N
chr18	13510974	13564877	53903	loss	C18orf1
chr18	1265932	1361783	95851	loss	N
chr18	76503998	76513505	9507	loss	SALL3
chr18	76521952	76570960	49008	loss	SALL3
chr18	76667529	76751058	83529	loss	SALL3

chr18	74110563	74201771	91208	loss	ZNF516
chr18	31146910	31226573	79663	loss	ASXL3
chr18	65746807	65783645	36838	loss	TMX3
chr18	65848177	65891421	43244	loss	TMX3
chr18	66352301	66373235	20934	loss	TMX3
chr18	2728328	2792659	64331	loss	SMCHD1
chr19	24403047	24562565	159518	gain	N
chr19	27763440	27987224	223784	gain	N
chr19	53529789	53552296	22507	gain	env
chr19	24403047	24562565	159518	loss	N
chr19	27763440	27987224	223784	loss	N
chr19	43802290	43835770	33480	loss	N
chr19	3804701	3842327	37626	gain	ZFR2
chr19	23087679	23153750	66071	loss	N
chr19	650931	665186	14255	loss	RNF126
chr19	13942221	13975377	33156	loss	ZSWIM4
chr19	7543113	7575408	32295	loss	PEX11G
chr19	4493707	4530261	36554	loss	HDGF2
chr19	1562643	1593540	30897	loss	MEX3D
chr19	1512093	1526613	14520	loss	ADAMTSL5
chr19	59079933	59097160	17227	loss	LOC100131691
chr19	49111237	49196721	85484	loss	FAM83E
chr19	1396462	1436873	40411	loss	AK126693
chr19	16017006	16027067	10061	loss	DKFZp761K058
chr19	16040292	16053249	12957	loss	DKFZp761K058
chr19	814023	830854	16831	loss	LPPR3
chr19	55245738	55332600	86862	gain	KIR3DL3
chr19	3591589	3599983	8394	loss	GIPC3
chr19	1266334	1290235	23901	loss	C19orf23
chr19	44043447	44076153	32706	loss	XRCC1
chr19	3440080	3586470	146390	loss	NFIC
chr19	17354586	17428045	73459	loss	NR2F6
chr19	2248683	2275079	26396	loss	AMH
chr19	2660939	2748127	87188	loss	GNG7
chr19	1200712	1235071	34359	gain	STK11
chr19	1200712	1235071	34359	loss	STK11
chr19	5864624	6193599	328975	loss	FUT5
chr19	36643694	36675065	31371	loss	COX7A1
chr19	10520513	10546770	26257	loss	PDE4A
chr19	1313763	1357082	43319	loss	MUM1
chr19	13061983	13122612	60629	gain	RAD23A
chr19	20621828	20707568	85740	loss	ZNF737

chr19	21455154	21570590	115436	loss	ZNF708
chr19	48329545	48341815	12270	loss	CRX
chr19	53581151	53612989	31838	loss	ZNF160
chr19	58338213	58453342	115129	loss	ZNF587
chr19	55983854	56009881	26027	loss	ZNF628
chr19	21206463	21249429	42966	loss	ZNF430
chr19	20835290	20925934	90644	gain	ZNF626
chr19	19776814	19795559	18745	loss	ZNF101
chr19	44352639	44454909	102270	loss	ZNF283
chr19	20116641	20136704	20063	loss	ZNF682
chr19	53960154	53992143	31989	loss	ZNF761
chr19	53300984	53344701	43717	loss	ZNF28
chr19	58968875	59000273	31398	loss	ZNF324
chr19	11979164	12130477	151313	loss	ZNF439
chr19	36939572	37010552	70980	loss	ZNF566
chr19	52528144	52580626	52482	gain	ZNF614
chr19	53085659	53117809	32150	loss	ZNF701
chr19	887127	907673	20546	loss	MED16
chr19	1856186	1876566	20380	loss	KLF16
chr19	3615164	3647901	32737	loss	C19orf29
chr19	2028984	2050823	21839	loss	MKNK2
chr19	55912332	55947296	34964	loss	UBE2S
chr19	1936848	1989867	53019	loss	CSNK1G2
chr19	15772414	15773914	1500	loss	CYP4F3
chr19	1781879	1784944	3065	loss	ATP8B3
chr19	1007353	1031212	23859	loss	GRIN3B
chr19	41082868	41114065	31197	loss	SHKBP1
chr19	52283382	52365654	82272	gain	FPR3
chr19	699109	734851	35742	gain	PALM
chr19	699109	734851	35742	loss	PALM
chr19	43328006	43458563	130557	gain	PSG3
chr19	43328006	43458563	130557	loss	PSG3
chr19	55084512	55128339	43827	gain	LILRA2
chr19	54729050	54740720	11670	gain	LILRB3
chr19	54729050	54740720	11670	loss	LILRB3
chr19	51186081	51217178	31097	loss	SHANK1
chr19	577665	585805	8140	loss	BSG
chr19	8564523	8579011	14488	loss	PRAM1
chr19	55874641	55887665	13024	loss	IL11
chr19	41141434	41192715	51281	loss	NUMBL
chr19	16967738	17009371	41633	loss	SIN3B
chr19	41355998	41365094	9096	gain	CYP2A6

chr19	41355998	41365094	9096	loss	CYP2A6
chr19	1457874	1481480	23606	loss	APC2
chr19	7921623	7988117	66494	loss	EVI5L
chr19	301639	364694	63055	loss	MIER2
chr19	56105161	56130705	25544	loss	FIZ1
chr2	228243310	228258288	14978	gain	TM4SF20
chr2	13716541	13745887	29346	loss	N
chr2	35675611	35713309	37698	loss	N
chr2	111970330	112010486	40156	loss	N
chr2	47011	105493	58482	loss	FAM110C
chr2	32509330	32550485	41155	loss	YIPF4
chr2	34042593	34048462	5869	loss	tmp_locus_39
chr2	37247717	37277811	30094	loss	HEATR5B
chr2	154582564	154661248	78684	loss	GALNT13
chr2	185017926	185056639	38713	loss	ZNF804A
chr2	207145650	207178872	33222	loss	ZDBF2
chr2	240625255	240640112	14857	loss	BC132948
chr2	242929164	242955426	26262	loss	BC101234
chr2	132137239	132161303	24064	loss	N
chr2	44652343	44735438	83095	loss	C2orf34
chr2	76941049	76949101	8052	loss	LRRTM4
chr2	191018607	191029604	10997	loss	C2orf88
chr2	239020497	239035852	15355	loss	ESPNL
chr2	132410985	132617789	206804	loss	BX648270
chr2	126806674	127048167	241493	loss	GYPC
chr2	159723506	159778783	55277	loss	TANC1
chr2	229362382	229457449	95067	loss	SPHKAP
chr2	87768679	87965359	196680	gain	RMND5A
chr2	242996474	243048760	52286	loss	N
chr2	23947589	24071427	123838	loss	ATAD2B
chr2	39785624	39911538	125914	loss	N
chr2	47337000	47398394	61394	loss	CALM2
chr2	53872721	53968774	96053	loss	ASB3
chr2	20109061	20135283	26222	loss	KIAA1336
chr2	131055763	131082639	26876	loss	CCDC115
chr2	55680496	55798108	117612	gain	CCDC104
chr2	159259303	159300522	41219	loss	CCDC148
chr2	131833758	131898811	65053	loss	FAM168B
chr2	71852034	71865287	13253	loss	DYSF
chr2	85077774	85140165	62391	loss	C2orf89
chr2	11317226	11359119	41893	loss	PQLC3
chr2	55514668	55591166	76498	loss	KIAA1212

chr2	198477620	198614612	136992	loss	RFTN2
chr2	242750743	242824974	74231	loss	PABL
chr2	105455303	105481442	26139	loss	AK096498
chr2	114447184	114493315	46131	loss	NS5ATP3
chr2	74531420	74583478	52058	loss	NBC4
chr2	241913553	241989722	76169	loss	AK055890
chr2	3501050	3519739	18689	loss	HMFT1638
chr2	477364	535043	57679	loss	N
chr2	225346646	225534192	187546	loss	CUL3
chr2	239935542	239957392	21850	loss	HDAC4
chr2	129086164	129106110	19946	loss	HS6ST1
chr2	41240080	41267941	27861	loss	SLC8A1
chr2	203246410	203315611	69201	loss	BMPR2
chr2	173703518	173727773	24255	loss	RAPGEF4
chr2	170344060	170397635	53575	loss	BBS5
chr2	38956947	38964658	7711	gain	GALM
chr2	28144067	28205581	61514	loss	BRE
chr2	208465777	208480504	14727	loss	CREB1
chr2	27568565	27586155	17590	loss	GTF3C2
chr2	176927667	177081309	153642	loss	EVX2
chr2	175259064	175304048	44984	loss	CIR1
chr2	102527975	102569723	41748	loss	MAP4K4
chr2	220337196	220372633	35437	loss	SPEG
chr2	57416841	57445471	28630	gain	VRK2
chr2	134845085	134895113	50028	loss	MGAT5
chr2	207386623	207470884	84261	loss	ADAM23
chr2	241508389	241545400	37011	loss	RNPEPL1
chr2	61540642	61629635	88993	loss	USP34
chr2	162164385	162211730	47345	loss	PSMD14
chr2	61665336	61825142	159806	loss	USP34
chr2	177137489	177170168	32679	loss	MTX2
chr2	54366449	54427209	60760	loss	ACYP2
chr2	18327703	18340920	13217	gain	KCNS3
chr2	97441266	97527545	86279	loss	CNNM4
chr2	162580384	162640725	60341	loss	SLC4A10
chr2	32437338	32478629	41291	gain	SLC30A6
chr2	234266842	234299919	33077	loss	DGKD
chr2	140713355	140774793	61438	loss	LRP1B
chr2	72505475	73131167	625692	loss	EXOC6B
chr2	20694771	20708719	13948	loss	RHOB
chr2	39261251	39305488	44237	loss	SOS1
chr2	42419829	42642531	222702	loss	EML4

chr2	241740907	241826517	85610	loss	KIF1A
chr2	201403980	201445151	41171	loss	SGOL2
chr2	154411685	154423181	11496	loss	RPRM
chr2	50882657	50947044	64387	loss	NRXN1
chr2	51926599	51931744	5145	loss	NRXN1
chr2	51975932	52012903	36971	loss	NRXN1
chr2	52297892	52384668	86776	loss	NRXN1
chr2	213187034	213191389	4355	loss	ERBB4
chr2	213275304	213336689	61385	loss	ERBB4
chr2	110885620	110980346	94726	loss	NPHP1
chr2	144112684	144135930	23246	loss	ARHGAP15
chr2	99651963	99735052	83089	loss	TSGA10
chr2	949196	974720	25524	loss	SNTG2
chr2	242710662	242744413	33751	gain	GAL3ST2
chr2	73339430	73427222	87792	loss	RAB11FIP5
chr2	113042446	113094793	52347	loss	ZC3H6
chr2	32861950	33163650	301700	gain	TTC27
chr2	71615534	71658541	43007	loss	ZNF638
chr2	158839672	158876018	36346	loss	UPP2
chr2	231720362	231752542	32180	loss	ITM2C
chr2	233320884	233360507	39623	loss	ALPI
chr2	167575861	167620935	45074	loss	XIRP2
chr2	231940959	231990370	49411	loss	PSMD1
chr2	9543221	9758621	215400	loss	ASAP2
chr20	46592955	46626758	33803	loss	N
chr20	36574948	36596817	21869	loss	VSTM2L
chr20	58098827	58118612	19785	loss	PHACTR3
chr20	61665921	61688888	22967	loss	N
chr20	13977991	14036568	58577	loss	MACROD2
chr20	14794361	14823472	29111	loss	MACROD2
chr20	61906188	61983755	77567	gain	ARF1GAP
chr20	61906188	61983755	77567	loss	ARF1GAP
chr20	62468726	62506947	38221	loss	C20orf135
chr20	62654992	62743322	88330	loss	PRPF6
chr20	44607601	44668008	60407	loss	MMP9
chr20	60890375	60992402	102027	loss	LAMA5
chr20	62860980	62886216	25236	loss	MYT1
chr20	61024062	61088166	64104	loss	GATA5
chr20	55297487	55351869	54382	loss	TFAP2C
chr20	60681037	60765422	84385	loss	LSM14B
chr20	351944	391025	39081	loss	TRIB3
chr20	17949452	17993460	44008	loss	SNX5

chr20	59569516	59586563	17047	gain	CDH4
chr20	59629381	59868406	239025	gain	CDH4
chr20	62341930	62391013	49083	loss	ZGPAT
chr20	61331445	61352671	21226	loss	NTSR1
chr20	62141509	62164304	22795	gain	PPDPF
chr20	62141509	62164304	22795	loss	PPDPF
chr20	35274672	35379741	105069	gain	NDRG3
chr20	17552349	17587933	35584	loss	DSTN
chr20	12624679	12634690	10011	gain	N
chr20	12819817	12889314	69497	gain	N
chr20	32251396	32265838	14442	loss	NECAB3
chr20	222836	237497	14661	loss	DEFB132
chr20	50817761	51245688	427927	gain	ZFP64
chr20	62193445	62216808	23363	loss	PRIC285
chr21	24492713	24550449	57736	loss	N
chr21	24667856	24680450	12594	loss	N
chr21	26145577	26174872	29295	loss	N
chr21	47439473	47481083	41610	loss	N
chr21	14613203	15011015	397812	loss	POTED
chr21	18689984	18729311	39327	loss	N
chr21	15165769	15332033	166264	loss	N
chr21	33065597	33096866	31269	loss	SFRS15
chr21	35281113	35434373	153260	gain	C21orf60
chr21	44830955	44839153	8198	gain	SIK1
chr21	44830955	44839153	8198	loss	SIK1
chr21	47800592	47910523	109931	loss	PCNT
chr21	46817380	46821907	4527	loss	COL18A1
chr21	46888468	46933163	44695	loss	COL18A1
chr21	45681579	45755360	73781	loss	DNMT3L
chr21	34590089	34605530	15441	loss	IFNAR2
chr21	48042513	48055376	12863	loss	PRMT2
chr21	42912973	42926621	13648	loss	TMPRSS2
chr21	46221166	46255595	34429	loss	UBE2G2
chr21	47566828	47599176	32348	loss	FTCD
chr21	46283558	46298869	15311	loss	PTTG1IP
chr21	46945023	46953291	8268	gain	SLC19A1
chr21	46945023	46953291	8268	loss	SLC19A1
chr21	45764396	45878637	114241	loss	TRPM2
chr21	45180708	45206355	25647	gain	PDXK
chr21	45180708	45206355	25647	loss	PDXK
chr21	47494789	47545768	50979	loss	COL6A2
chr21	40520783	40568528	47745	loss	PSMG1

chr22	48719380	48730913	11533	loss	FAM19A5
chr22	48784158	48811946	27788	loss	FAM19A5
chr22	37719004	37749613	30609	loss	N
chr22	50077362	50100662	23300	loss	N
chr22	48956316	48988885	32569	loss	FAM19A5
chr22	51101899	51151350	49451	loss	SHANK3
chr22	25842086	25914592	72506	gain	N
chr22	25842086	25914592	72506	loss	N
chr22	46421842	46459132	37290	loss	LOC100271722
chr22	39428919	39445556	16637	loss	APOBEC3D
chr22	16855618	17259954	404336	gain	DQ578965
chr22	18844632	18981562	136930	gain	N
chr22	25686040	25809960	123920	gain	IGLL3
chr22	50435086	50472042	36956	loss	IL17REL
chr22	46311903	46363739	51836	loss	WNT7B
chr22	46875743	46919541	43798	loss	CELSR1
chr22	46942068	46980242	38174	loss	CELSR1
chr22	21980894	22055394	74500	loss	YDJC
chr22	22307381	22555078	247697	gain	TOP3B
chr22	39279441	39316732	37291	loss	CBX6
chr22	19721103	19743640	22537	loss	TBX1
chr22	38415720	38477930	62210	loss	PICK1
chr22	39941154	39966766	25612	loss	CACNAII
chr22	50609653	50633686	24033	loss	PANX2
chr22	17586470	17624072	37602	loss	IL17RA
chr22	37826951	37959214	132263	loss	MFNG
chr22	46106914	46205231	98317	loss	ATXN10
chr22	20220764	20249918	29154	loss	RTN4R
chr22	50947611	50980818	33207	loss	NCAPH2
chr22	20098359	20189202	90843	loss	DGCR8
chr22	47464861	47595959	131098	loss	TBC1D22A
chr22	40876234	40959920	83686	loss	MKL1
chr22	44107669	44188948	81279	loss	EFCAB6
chr3	165204343	165406029	201686	gain	N
chr3	21841043	21865234	24191	loss	N
chr3	22827129	22841826	14697	loss	N
chr3	104078453	104270022	191569	loss	N
chr3	106217161	106233426	16265	loss	N
chr3	120572921	120606438	33517	loss	N
chr3	147623914	147805662	181748	loss	N
chr3	162305516	162376870	71354	loss	N
chr3	165041116	165073006	31890	loss	N

chr3	154186021	154490612	304591	loss	GPR149
chr3	188682281	188699808	17527	loss	TPRG1
chr3	191286526	191315365	28839	loss	PYDC2
chr3	62966870	62987808	20938	loss	N
chr3	197780227	197817764	37537	loss	N
chr3	155433793	155520927	87134	loss	C3orf33
chr3	13582768	13630819	48051	loss	FBLN2
chr3	101445118	101489629	44511	loss	CEP97
chr3	113604898	113654705	49807	gain	DKFZp434C0328
chr3	75453100	75636449	183349	loss	N
chr3	51802912	51917015	114103	loss	IQCF6
chr3	69047141	69103000	55859	loss	AER61
chr3	195644331	195777210	132879	loss	SDHALP1
chr3	183380034	183491044	111010	loss	KLHL24
chr3	126105301	126156742	51441	loss	BC033989
chr3	178833722	178893028	59306	loss	BC032034
chr3	171779873	171821513	41640	loss	fad104
chr3	12322627	12386336	63709	loss	PPARG
chr3	90313394	90442925	129531	loss	EPHA3
chr3	89402447	89417171	14724	loss	EPHA3
chr3	144782354	144993976	211622	loss	PLOD2
chr3	8833877	8857962	24085	gain	OXTR
chr3	196767854	196827248	59394	loss	DLG1
chr3	173246813	173289281	42468	gain	NLGN1
chr3	81052986	81099000	46014	loss	GBE1
chr3	163885473	163930675	45202	loss	SI
chr3	50340996	50393640	52644	loss	HYAL1
chr3	60888974	61021526	132552	loss	FHIT
chr3	141743951	141845000	101049	loss	TFDP2
chr3	65191847	65212596	20749	loss	MAGI1
chr3	65358578	65624729	266151	gain	MAGI1
chr3	12649575	12709166	59591	loss	RAF1
chr3	184211984	184242853	30869	loss	EPHB3
chr3	62004450	62016422	11972	loss	PTPRG
chr3	175893082	175907342	14260	loss	NAALADL2
chr3	174567744	174582723	14979	loss	NAALADL2
chr3	174124164	174260303	136139	loss	N
chr3	175187978	175267563	79585	loss	NAALADL2
chr3	126227185	126264060	36875	loss	UROC1
chr3	193252723	193303534	50811	loss	ATP13A4
chr3	155775874	155849212	73338	loss	KCNAB1
chr3	39924900	39966748	41848	loss	MYRIP

chr3	120977383	121084467	107084	loss	STXBP5L
chr3	122894253	122995630	101377	loss	SEC22A
chr3	42594019	42832978	238959	loss	SEC22C
chr3	88442852	88536336	93484	loss	C3orf38
chr3	5410364	5438374	28010	loss	EDEM1
chr3	57431985	57510407	78422	loss	DNAH12
chr3	33204539	33281293	76754	loss	SUSD5
chr3	61495	89511	28016	gain	CHL1
chr3	61495	89511	28016	loss	CHL1
chr3	234725	243307	8582	loss	CHL1
chr3	104773065	104833842	60777	loss	ALCAM
chr3	118730933	118803278	72345	gain	IGSF11
chr3	152878719	153041225	162506	loss	RAP2B
chr3	100351696	100398405	46709	gain	GPR128
chr3	172740452	172786148	45696	loss	SPATA16
chr3	169887114	169958551	71437	loss	PHC3
chr3	145909152	145989009	79857	loss	PLSCR4
chr3	97983942	98027888	43946	loss	OR5H6
chr3	151325890	151371825	45935	loss	AADACL2
chr3	151514590	151546041	31451	loss	AADAC
chr3	66319062	66403355	84293	loss	SLC25A26
chr3	73287838	73298666	10828	loss	PDZRN3
chr3	17447410	17486643	39233	loss	TBC1D5
chr3	179698241	179778368	80127	loss	PEX5L
chr3	112610825	112661354	50529	loss	CD200R1
chr3	53060637	53098775	38138	loss	SFMBT1
chr3	107266759	107337979	71220	loss	BBX
chr3	121526107	121610395	84288	loss	IQCB1
chr4	34900365	34917769	17404	gain	N
chr4	58064464	58098554	34090	gain	N
chr4	63856049	63901897	45848	gain	N
chr4	161952228	162003624	51396	gain	N
chr4	179247331	179380467	133136	gain	N
chr4	190685303	190710253	24950	gain	N
chr4	92200866	92251839	50973	gain	FAM190A
chr4	59667025	59710106	43081	loss	N
chr4	68154530	68207449	52919	loss	N
chr4	105507674	105519735	12061	loss	N
chr4	132961872	133093090	131218	loss	N
chr4	142246430	142296923	50493	loss	N
chr4	150060844	150184393	123549	loss	N
chr4	158538338	158562472	24134	loss	N

chr4	161837715	161864128	26413	loss	N
chr4	178951714	178988148	36434	loss	N
chr4	181310122	181316450	6328	loss	N
chr4	182655999	182692368	36369	loss	N
chr4	189194661	189247773	53112	loss	N
chr4	36915898	36931883	15985	loss	KIAA1239
chr4	56986929	57009127	22198	loss	KIAA1211
chr4	172147796	173089205	941409	loss	GALNTL6
chr4	91941029	92009492	68463	loss	FAM190A
chr4	127745441	127788899	43458	loss	INTU
chr4	1619530	1684457	64927	loss	FAM53A
chr4	110486492	110555678	69186	loss	CCDC109B
chr4	144782841	144896250	113409	gain	GYPE
chr4	40042666	40121561	78895	loss	ACOT7L
chr4	113533796	113579162	45366	loss	C4orf21
chr4	1363886	1416863	52977	loss	KIAA1530
chr4	64138341	64184644	46303	gain	N
chr4	64138341	64184644	46303	loss	N
chr4	1786701	1815989	29288	loss	FGFR3
chr4	160979988	161041088	61100	loss	RAPGEF2
chr4	95713679	95750866	37187	gain	BMPR1B
chr4	95892127	95933336	41209	loss	BMPR1B
chr4	167222523	167329523	107000	loss	TLL1
chr4	25554332	25577184	22852	gain	SLC34A2
chr4	38807653	38859515	51862	loss	TLR6
chr4	992235	1009419	17184	loss	IDUA
chr4	37856777	37885195	28418	loss	PGM2
chr4	71566	150683	79117	loss	ZNF718
chr4	3525057	3666471	141414	loss	LRPAP1
chr4	56231118	56285304	54186	loss	SRD5A3
chr4	49053522	49109992	56470	loss	CWH43
chr4	70584038	70634942	50904	loss	SULT1B1
chr4	94227970	94319668	91698	gain	GRID2
chr4	93344512	93403678	59166	loss	GRID2
chr4	22044372	22069379	25007	loss	KCNIP4
chr4	43634967	43666139	31172	loss	KCTD8
chr4	809593	880094	70501	loss	CPLX1
chr4	110388034	110466758	78724	loss	SEC24B
chr4	190903688	190939665	35977	loss	TUBB4Q
chr4	41218908	41231007	12099	loss	APBB2
chr4	127037075	127171396	134321	gain	FAT4
chr4	36113030	36159118	46088	loss	ARAP2

chr4	60239646	60283466	43820	loss	LPHN3
chr4	61107712	61363382	255670	loss	LPHN3
chr4	70713065	70716626	3561	gain	SULT1E1
chr4	41956313	42099799	143486	loss	TMEM33
chr4	3419677	3496940	77263	loss	RGS12
chr4	73757679	74058406	300727	gain	COX18
chr4	100799961	100880420	80459	loss	MAPKSP1
chr4	71849697	71937444	87747	loss	MOBK1A
chr4	9371116	9456977	85861	gain	DEFB131
chr4	153669008	153696634	27626	loss	TIGD4
chr4	215351	283118	67767	loss	ZNF876P
chr4	759374	786401	27027	loss	PCGF3
chr5	116821562	116916674	95112	gain	N
chr5	7178644	7191074	12430	loss	N
chr5	8708019	8744557	36538	loss	N
chr5	104452655	104501134	48479	loss	N
chr5	62549338	62567045	17707	loss	ISCA1L
chr5	84679869	84822809	142940	loss	NBPF22P
chr5	113160490	113171928	11438	loss	YTHDC2
chr5	115494919	115571703	76784	loss	COMMD10
chr5	124989000	125172583	183583	loss	GRAMD3
chr5	177612034	177620189	8155	loss	FLJ13057
chr5	177727491	177744687	17196	loss	COL23A1
chr5	177772890	177780183	7293	loss	COL23A1
chr5	119410495	119518377	107882	loss	PRR16
chr5	10067090	10099927	32837	gain	FAM173B
chr5	120244182	120345122	100940	loss	PRR16
chr5	72334678	72387486	52808	loss	FCHO2
chr5	101701924	101762909	60985	loss	SLCO6A1
chr5	37138845	37173930	35085	loss	C5orf42
chr5	40726495	40774673	48178	loss	TTC33
chr5	79851241	79962438	111197	loss	ANKRD34B
chr5	5116544	5154458	37914	loss	AK094462
chr5	170814023	170843198	29175	loss	NPM1
chr5	176569187	176700884	131697	loss	NSD1
chr5	76310780	76343998	33218	loss	AGGF1
chr5	109363175	109375480	12305	loss	MAN2A1
chr5	78629346	78763345	133999	loss	HOMER1
chr5	1815949	1856595	40646	loss	NDUFS6
chr5	35578461	35621098	42637	loss	SPEF2
chr5	149548822	149582659	33837	loss	CDX1
chr5	2041665	2064968	23303	loss	IRX4

chr5	336951	412693	75742	loss	AHRR
chr5	438101	560922	122821	loss	AHRR
chr5	178286356	178341750	55394	loss	ZNF354B
chr5	132214330	132361894	147564	loss	AFF4
chr5	179687013	179704740	17727	loss	MAPK9
chr5	97049480	97081520	32040	loss	RIOK2
chr5	97108559	97165802	57243	loss	RIOK2
chr5	128876639	128978534	101895	loss	ADAMTS19
chr5	64625511	64659167	33656	loss	ADAMTS6
chr5	7985855	8015429	29574	loss	MTRR
chr5	77296281	77359088	62807	loss	AP3B1
chr5	54790559	54850717	60158	loss	PPAP2A
chr5	46127095	46399093	271998	gain	HCN1
chr5	46127095	46399093	271998	loss	HCN1
chr5	54954478	55016916	62438	loss	SLC38A9
chr5	101060923	101096015	35092	gain	SLCO4C1
chr5	170213756	170223209	9453	gain	GABRP
chr5	1211556	1317948	106392	loss	SLC6A19
chr5	75436389	75450373	13984	gain	SV2C
chr5	40929543	40943442	13899	loss	C7
chr5	133942478	134063041	120563	loss	SAR1B
chr5	77669593	77759362	89769	loss	SCAMP1
chr5	66489051	66502865	13814	gain	CD180
chr5	176788570	176832165	43595	gain	RGS14
chr5	140473137	140504432	31295	loss	PCDHB2
chr5	140137644	140192643	54999	loss	PCDHA1
chr5	140210616	140233387	22771	loss	PCDHA1
chr5	18852401	18895407	43006	gain	CDH18
chr5	140723174	140772427	49253	loss	PCDHGA1
chr5	84192194	84222269	30075	loss	EDIL3
chr5	49694961	49748280	53319	loss	EMB
chr5	24541668	24568262	26594	loss	CDH10
chr5	24779565	24794619	15054	loss	CDH10
chr5	25698030	25723615	25585	loss	CDH10
chr5	142174078	142217069	42991	loss	ARHGAP26
chr5	178627708	178637124	9416	loss	ADAMTS2
chr5	178741383	178827859	86476	loss	ADAMTS2
chr5	33954511	33983044	28533	gain	SLC45A2
chr5	1517198	1536257	19059	loss	LPCAT1
chr5	32151119	32193664	42545	gain	GOLPH3
chr5	32151119	32193664	42545	loss	GOLPH3
chr5	25328	156287	130959	loss	PLEKHG4B

chr5	33998768	34749416	750648	gain	AMACR
chr5	70762935	70862874	99939	loss	BDP1
chr5	179611713	179623637	11924	loss	RASGEF1C
chr6	795414	927700	132286	gain	N
chr6	168538682	168577460	38778	gain	FRMD1
chr6	1528282	1569619	41337	loss	N
chr6	67017494	67039739	22245	loss	N
chr6	75542900	75610722	67822	loss	N
chr6	140779173	140849701	70528	loss	N
chr6	170414285	170439873	25588	loss	N
chr6	2573436	2585649	12213	loss	C6orf195
chr6	30765578	30770732	5154	loss	C6orf214
chr6	32268079	32278791	10712	loss	C6orf10
chr6	89799737	89854436	54699	loss	SFRS13B
chr6	112961006	112965827	4821	loss	RFPL4B
chr6	127903054	127965850	62796	loss	C6orf58
chr6	168538682	168577460	38778	loss	FRMD1
chr6	168620659	168654978	34319	loss	DACT2
chr6	68663856	68699843	35987	loss	N
chr6	30942283	30996362	54079	loss	MUC21
chr6	32319363	32324944	5581	loss	C6orf10
chr6	74123363	74136427	13064	loss	DDX43
chr6	95421874	95504340	82466	loss	MANEA
chr6	971454	1038522	67068	gain	N
chr6	107174515	107329420	154905	gain	N
chr6	163171970	163203212	31242	loss	PACRG
chr6	163486937	163530506	43569	loss	PACRG
chr6	124434003	124468892	34889	gain	TCBA1
chr6	170099397	170122234	22837	loss	WDR27
chr6	31272773	31277987	5214	gain	N
chr6	105596610	105630330	33720	loss	RP11-99L11.2
chr6	17364146	17446523	82377	loss	DKFZp686D0714
chr6	69323661	69358140	34479	loss	CR595314
chr6	106815036	106841948	26912	gain	ATG5
chr6	108666	181348	72682	loss	DUSP22
chr6	162933252	163072782	139530	gain	PARK2
chr6	162789196	162847889	58693	gain	PARK2
chr6	162933252	163072782	139530	loss	PARK2
chr6	162789196	162847889	58693	loss	PARK2
chr6	29536684	29546911	10227	loss	GABBR1
chr6	103157617	103198856	41239	loss	GRIK2
chr6	24303195	24433483	130288	loss	DCDC2

chr6	31360254	31367864	7610	gain	MICA
chr6	31360254	31367864	7610	loss	MICA
chr6	31457633	31473286	15653	gain	MICB
chr6	31206205	31211049	4844	gain	HLA-C
chr6	31338528	31340806	2278	loss	HLA-B
chr6	29817244	29825846	8602	gain	HLA-G
chr6	29801110	29816420	15310	loss	HLA-G
chr6	29842600	29909558	66958	loss	HLA-G
chr6	32498834	32519004	20170	loss	HLA-DRB1
chr6	32670978	32682308	11330	loss	HLA-DQA2
chr6	33001734	33016886	15152	loss	HLA-DPA1
chr6	32620282	32640336	20054	gain	HLA-DQB1
chr6	32837695	32844032	6337	loss	N
chr6	26155684	26175852	20168	gain	HIST1H1E
chr6	110516180	110570460	54280	loss	CDC40
chr6	31872383	31894625	22242	loss	C2
chr6	63543538	63617477	73939	gain	LGSN
chr6	63543538	63617477	73939	loss	LGSN
chr6	111155400	111309591	154191	loss	AMD1
chr6	33324526	33340509	15983	loss	LYPLA2
chr6	46587520	46594121	6601	loss	CYP39A1
chr6	107671156	107721094	49938	loss	PDSS2
chr6	161244545	161267335	22790	loss	PLG
chr6	31424756	31432124	7368	gain	HCP5
chr6	56209919	56268471	58552	gain	COL21A1
chr6	169589925	169621135	31210	loss	THBS2
chr6	32056579	32072939	16360	loss	TNXB
chr6	87395567	87430465	34898	loss	HTR1E
chr6	110255274	110296455	41181	loss	GPR6
chr6	78972930	79029367	56437	loss	IRAK1BP1
chr6	42954428	43083774	129346	loss	MEA1
chr6	157059428	157108163	48735	loss	ARID1B
chr6	30144419	30147689	3270	loss	TRIM15
chr6	65828852	65910220	81368	loss	EYS
chr6	66277243	66366331	89088	loss	EYS
chr6	29094695	29161434	66739	loss	OR2J2
chr6	29344036	29346363	2327	loss	OR12D3
chr6	52623806	52657551	33745	loss	GSTA2
chr6	17661619	17711490	49871	loss	NUP153
chr6	142455730	142531036	75306	loss	VTA1
chr6	160570623	160574534	3911	loss	SLC22A1
chr6	44226194	44235689	9495	loss	NFKBIE

chr7	44935	62643	17708	gain	N
chr7	25653642	25775690	122048	gain	N
chr7	11740020	11753281	13261	gain	THSD7A
chr7	62113011	62241040	128029	gain	LOC643955
chr7	88848044	88924014	75970	gain	ZNF804B
chr7	108339862	108415940	76078	gain	C7orf66
chr7	44935	62643	17708	loss	N
chr7	83185	113319	30134	loss	N
chr7	7292458	7324800	32342	loss	N
chr7	67187678	67283756	96078	loss	N
chr7	89553014	89695380	142366	loss	N
chr7	136811372	136882925	71553	loss	N
chr7	155146549	155169830	23281	loss	N
chr7	6652740	6690240	37500	loss	ZNF853
chr7	12580007	12604371	24364	loss	BC075797
chr7	53465744	53590250	124506	loss	POM121L12
chr7	62113011	62241040	128029	loss	LOC643955
chr7	88848044	88924014	75970	loss	ZNF804B
chr7	155264377	155312325	47948	loss	CNPY1
chr7	32665229	32711465	46236	loss	AVL9
chr7	39849482	39894883	45401	loss	LOC401327
chr7	150455295	150486169	30874	loss	LOC100128542
chr7	8800804	8817762	16958	loss	NXPH1
chr7	8830494	8856074	25580	loss	NXPH1
chr7	155416363	155470821	54458	loss	AX747269
chr7	1151066	1181139	30073	loss	C7orf50
chr7	29690658	29778251	87593	gain	N
chr7	38319339	38341226	21887	loss	N
chr7	825494	872848	47354	loss	HEATR2
chr7	5534598	5608931	74333	loss	FBXL18
chr7	1866916	2094340	227424	loss	MAD1
chr7	39646505	39666267	19762	loss	LOC646999
chr7	73139762	73226807	87045	loss	WBSCR26
chr7	64679561	65088807	409246	loss	INTS4L1
chr7	128500139	128530275	30136	loss	KCP
chr7	1458931	1478972	20041	loss	FLJ00139
chr7	100256227	100406144	149917	loss	AK055267
chr7	99310180	99368633	58453	gain	FLJ00284
chr7	139935860	139984240	48380	gain	DKFZp761N0624
chr7	1507011	1542697	35686	loss	KIAA1440
chr7	44313930	44342969	29039	loss	CAMK2B
chr7	101864836	101921289	56453	loss	CUX1

chr7	104948076	105076977	128901	loss	SRPK2
chr7	100779243	100783735	4492	loss	SERPINE1
chr7	154782113	154823360	41247	loss	PAXIP1
chr7	27223563	27264337	40774	loss	HOXA11
chr7	1271819	1307284	35465	loss	UNCX
chr7	119383120	119469564	86444	loss	KCND2
chr7	111023322	111122474	99152	loss	IMMP2L
chr7	2528599	2613424	84825	loss	LFNG
chr7	150706382	150719024	12642	loss	NOS3
chr7	150699876	150703914	4038	loss	NOS3
chr7	46072124	46135621	63497	loss	IGFBP3
chr7	20468718	20576180	107462	gain	ITGB8
chr7	20371758	20389202	17444	loss	ITGB8
chr7	30529411	30554997	25586	loss	GGCT
chr7	561070	687469	126399	loss	PRKAR1B
chr7	90988012	90999109	11097	loss	FZD1
chr7	83794756	83841443	46687	gain	SEMA3A
chr7	83433435	83471330	37895	loss	SEMA3A
chr7	10794857	10821385	26528	loss	NDUFA4
chr7	76432653	76615349	182696	gain	N
chr7	76432653	76615349	182696	loss	N
chr7	57481104	57642390	161286	loss	ZNF716
chr7	17425895	17453479	27584	gain	AHR
chr7	17425895	17453479	27584	loss	AHR
chr7	149563550	149585639	22089	gain	ZNF862
chr7	55999702	56077235	77533	loss	ZNF713
chr7	157208617	157217055	8438	loss	DNAJB6
chr7	72722981	72783382	60401	loss	FKBP6
chr7	23767480	23811800	44320	loss	STK31
chr7	157754482	157791029	36547	gain	PTPRN2
chr7	157543338	157583166	39828	loss	PTPRN2
chr7	157754482	157791029	36547	loss	PTPRN2
chr7	158113390	158170316	56926	loss	PTPRN2
chr7	158276925	158350292	73367	loss	PTPRN2
chr7	153478281	153734322	256041	gain	DPP6
chr7	65449716	65600463	150747	loss	ASL
chr7	91588035	91674302	86267	loss	AKAP9
chr7	87259384	87271837	12453	loss	ABCB1
chr7	48587909	48652670	64761	loss	ABCA13
chr7	100795046	100818846	23800	loss	AP1S1
chr7	6410527	6443838	33311	loss	RAC1
chr7	142377014	142456928	79914	gain	N

chr7	17599438	17626941	27503	loss	SNX13
chr7	17667816	17681020	13204	loss	SNX13
chr7	149472312	149513846	41534	loss	SSPO
chr7	29322256	29407487	85231	gain	CHN2
chr7	29322256	29407487	85231	loss	CHN2
chr7	1009337	1053958	44621	loss	ADAP1
chr7	66637001	66689729	52728	loss	TYW1
chr7	156754981	156772542	17561	loss	NOM1
chr7	70325633	70364102	38469	gain	AUTS2
chr7	142835239	142881540	46301	loss	PIP
chr7	140067575	140213880	146305	loss	SLC37A3
chr7	150553475	150560322	6847	loss	ABP1
chr7	100968363	101063059	94696	gain	EMID2
chr7	20731918	20750849	18931	loss	ABCB5
chr7	99429842	99450699	20857	loss	CYP3A43
chr8	2346867	2423004	76137	gain	N
chr8	2517411	2582764	65353	gain	N
chr8	43447235	43657438	210203	gain	N
chr8	96466242	96727560	261318	gain	C8orf37
chr8	1027787	1074425	46638	loss	N
chr8	1165147	1218258	53111	loss	N
chr8	5601191	5605087	3896	loss	N
chr8	5749849	5785930	36081	loss	N
chr8	9321780	9343978	22198	loss	N
chr8	58024168	58173412	149244	loss	N
chr8	64878574	64960855	82281	loss	N
chr8	90066677	90087094	20417	loss	N
chr8	102387329	102402505	15176	loss	N
chr8	143240263	143259732	19469	loss	N
chr8	1937286	1978115	40829	loss	KBTBD11
chr8	3624896	3660565	35669	loss	CSMD1
chr8	3984758	3988435	3677	loss	CSMD1
chr8	4681869	4697447	15578	loss	CSMD1
chr8	73239680	73298506	58826	loss	AK309726
chr8	913042	959338	46296	loss	N
chr8	47022970	47075668	52698	loss	N
chr8	2135762	2181540	45778	loss	N
chr8	125117277	125180387	63110	loss	FER1L6
chr8	146238281	146293414	55133	loss	C8orf33
chr8	142476529	142509280	32751	loss	FLJ43860
chr8	633684	657650	23966	loss	ERICH1
chr8	92128840	92183658	54818	loss	LRRC69

chr8	113357431	113390874	33443	loss	CSMD3
chr8	115320834	115404627	83793	loss	CSMD3
chr8	144984956	145042001	57045	loss	PLEC1
chr8	42704924	43024625	319701	loss	RNF170
chr8	29910706	30128040	217334	loss	TMEM66
chr8	145574467	145621556	47089	loss	C8orfK29
chr8	10395495	10530750	135255	loss	AK307207
chr8	86027232	86102223	74991	loss	KIAA1764
chr8	124259431	124342221	82790	loss	C8orf76
chr8	98708032	98738361	30329	loss	MTDH
chr8	145495578	145519034	23456	loss	BOP1
chr8	97516983	97522542	5559	loss	SDC2
chr8	144799465	144913873	114408	loss	MAPK15
chr8	144508768	144548806	40038	loss	MAFA
chr8	21976710	22029128	52418	loss	HR
chr8	196274	209152	12878	gain	ZNF596
chr8	196274	209152	12878	loss	ZNF596
chr8	30462149	30536792	74643	loss	GTF2E2
chr8	137687538	137857327	169789	loss	KHDRBS3
chr8	66995147	67118307	123160	loss	DNAJC5B
chr8	10297819	10316236	18417	loss	MSRA
chr8	11391650	11430485	38835	gain	BLK
chr8	120909236	121009824	100588	loss	DEPDC6
chr8	25522134	25581875	59741	loss	PPP2R2A
chr8	9482398	9578982	96584	loss	TNKS
chr8	38966013	39021272	55259	loss	ADAM32
chr8	73602555	73635954	33399	gain	KCNB2
chr8	145639476	145666578	27102	loss	SLC39A4
chr8	104078514	104102910	24396	loss	ATP6V1C1
chr8	15948235	16021467	73232	gain	MSR1
chr8	15948235	16021467	73232	loss	MSR1
chr8	50462970	50528355	65385	loss	SNTG1
chr8	83271395	83293872	22477	loss	SNX16
chr8	143514434	143620399	105965	loss	BAI1
chr8	131761998	131879067	117069	gain	ADCY8
chr8	131921955	132256944	334989	gain	ADCY8
chr8	17739537	17744768	5231	loss	FGL1
chr8	144309595	144425247	115652	loss	ZFP41
chr8	38190769	38240007	49238	loss	WHSC1L1
chr8	15402936	15410250	7314	loss	TUSC3
chr8	140882516	140892871	10355	loss	TRAPPC9
chr8	18852474	18858239	5765	loss	PSD3

chr8	1865205	1892069	26864	loss	ARHGEF10
chr8	146200945	146225280	24335	loss	ZNF252
chr8	135661484	135678804	17320	loss	ZFAT
chr9	139511247	139522101	10854	gain	N
chr9	138164038	138298164	134126	gain	C9orf62
chr9	107704228	107722578	18350	loss	N
chr9	139511247	139522101	10854	loss	N
chr9	185632	213149	27517	loss	C9orf66
chr9	8009428	8014674	5246	loss	C9orf123
chr9	25314595	25336447	21852	loss	TUSC1
chr9	25502753	25540603	37850	loss	TUSC1
chr9	30253168	30338229	85061	loss	LINGO2
chr9	90759374	90792863	33489	loss	DKFZp434M131
chr9	132268125	132305053	36928	loss	N
chr9	138999710	139051477	51767	loss	C9orf69
chr9	28280653	28299437	18784	gain	LINGO2
chr9	124992671	125006622	13951	loss	RBM18
chr9	15585395	15634325	48930	loss	DKFZp686I0230
chr9	139607966	139643023	35057	loss	FAM69B
chr9	133995353	134107816	112463	gain	AIF1L
chr9	34197250	34334126	136876	loss	UBAP1
chr9	95580375	95652950	72575	loss	ANKRD19
chr9	18930880	19098712	167832	loss	FAM154A
chr9	95144650	95309243	164593	loss	CENPP
chr9	137290724	137302938	12214	gain	FLJ00318
chr9	137210862	137253446	42584	loss	FLJ00318
chr9	30493749	30641824	148075	loss	IRP1
chr9	30903823	30979640	75817	loss	IRP1
chr9	31394025	31573759	179734	loss	IRP1
chr9	136281965	136307407	25442	loss	XPMC2H
chr9	114393235	114470138	76903	loss	C9orf29
chr9	140786918	140972887	185969	gain	AK128414
chr9	21352861	21405362	52501	loss	IFNA13
chr9	125601801	125828592	226791	loss	AL833455
chr9	140075719	140088630	12911	loss	ANAPC2
chr9	36979395	36999023	19628	loss	PAX5
chr9	140636332	140734160	97828	loss	EHMT1
chr9	94588498	94651414	62916	loss	ROR2
chr9	127242082	127261893	19811	loss	NR5A1
chr9	75597830	75979673	381843	gain	ALDH1A1
chr9	21902354	22029546	127192	loss	MTAP
chr9	14297978	14334138	36160	loss	NFIB

chr9	1565742	1592554	26812	loss	SMARCA2
chr9	826701	874129	47428	loss	DMRT1
chr9	20518923	20648623	129700	loss	MLLT3
chr9	139077342	139107879	30537	loss	LHX3
chr9	97024027	97081010	56983	gain	ZNF169
chr9	139921813	139932210	10397	loss	ABCA2
chr9	99148152	99210583	62431	loss	ZNF367
chr9	4784932	4809157	24225	loss	RCL1
chr9	113552952	113573073	20121	loss	MUSK
chr9	95898531	96126074	227543	loss	WNK2
chr9	10661677	10686202	24525	loss	PTPRD
chr9	8110767	8161748	50981	gain	PTPRD
chr9	9057632	9061819	4187	loss	PTPRD
chr9	139316744	139404895	88151	loss	PMPCA
chr9	136550817	136608669	57852	loss	SARDH
chr9	135947165	135966362	19197	loss	CEL
chr9	140408189	140436548	28359	loss	PNPLA7
chr9	4438240	4443085	4845	loss	SLC1A1
chr9	138595441	138662309	66868	loss	KCNT1
chr9	140301447	140333858	32411	loss	GRIN1
chr9	17697735	17751559	53824	loss	SH3GL2
chr9	17788028	17923017	134989	loss	SH3GL2
chr9	95683366	95700701	17335	loss	FGD3
chr9	21157372	21255150	97778	loss	IFNA21
chr9	134436280	134449466	13186	loss	RAPGEF1
chr9	139234512	139279173	44661	loss	GPSM1
chr9	86368105	86448944	80839	loss	GKAP1
chr9	17261209	17280616	19407	loss	CNTLN
chr9	105763301	105822396	59095	loss	CYLC2
chr9	5299385	5335470	36085	loss	RLN2
chr9	107265452	107281379	15927	gain	OR13F1
chr9	5140115	5181610	41495	loss	INSL6
chr9	136938961	136965553	26592	loss	BRD3
chr9	127781259	127801065	19806	loss	SCAI
chr9	140164250	140266928	102678	loss	COBRA1
chr9	6695397	6910916	215519	gain	KDM4C
chr9	11995297	12124410	129113	gain	TYRP1
chr9	11995297	12124410	129113	loss	TYRP1
chr9	623848	725147	101299	gain	KANK1

„N“ - nėra