

Određivanje kliničke važnosti strukturnih varijacija evolucijski novih gena kod ispitanika s neurorazvojnim poremećajima

Klein, Valentina

Master's thesis / Diplomski rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:922631>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-03-29**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)



Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Valentina Klein

**Određivanje kliničke važnosti strukturnih
varijacija evolucijski novih gena kod
ispitanika s neurorazvojnim poremećajima**

Diplomski rad

Zagreb, 2024.

University of Zagreb
Faculty of Science
Department of Biology

Valentina Klein

**Determination of clinical significance of
structural variations of evolutionary novel
genes in patients with neurodevelopmental
disorders**

Master thesis

Zagreb, 2024.

Ovaj rad je izrađen u Laboratoriju za medicinsku genetiku na Hrvatskom institutu za istraživanje mozga Medicinskog fakulteta, Sveučilišta u Zagrebu, u sklopu projekta „Koliko smo različiti kao ljudi? Raznolikost evolucijski novih gena u neurorazvojnim poremećajima“ pod mentorstvom dr. sc. Martine Rinčić, više znanstvene suradnice, te komentorstvom prof. dr. sc. Petre Korać. Rad je predan na ocjenu Biološkom odsjeku Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu radi stjecanja zvanja magistra Eksperimentalne biologije, modul: Fiziologija i imunobiologija.

ZAHVALE

Ovaj diplomski rad predstavlja završetak jednog putovanja, ali ne bih stigla ovdje bez pomoći i podrške onih najvažnijih.

Prije svega ogromno hvala mojim roditeljima koji su mi omogućili sve ove godine učenja i istraživanja. Hvala na ljubavi, podršci i riječima utjehe u teškim trenucima. Hvala vam što ste uvijek bili tu, s toplim obrokom i mudrim savjetom, čak i kada je sve izgledalo kao nemoguća misija. I za kraj, hvala što ste se strpljivo igrali tišine u vrijeme mojih ispita.

Dragi Dominik, ti si moja stijena i moja inspiracija. Hvala ti što si me trpio, podržavao i uvijek bio tu s riječima ohrabrenja. Tvoja ljubav mi je dala snagu kada sam je najviše trebala. Moje rame za plakanje i moj izvor smijeha, veliko hvala!

Svim prijateljima hvala što ste svaki teški trenutak zalili smijehom. Hvala za sva druženja, kave i zajednička ponavljanja za ispite. Bez vas bi ovaj put bio puno teži.

Mentorici dr.sc. Martini Rinčić veliko hvala na svim savjetima, strpljenju i razumijevanju. Vaše vodstvo i stručnost bili su od neprocjenjive vrijednosti. Hvala Vam što ste me usmjeravali i poticali da dam svoj maksimum.

Komentorici prof.dr.sc. Petri Korać hvala na dodatnim perspektivama i usmjeravanju. Vaše povratne informacije pomogle su mi i ohrabrine me da usavršim svoj rad.

I tako, nakon svih uzbuđenja, stresa, smijeha i suza, ovo poglavlje se zatvara. Hvala svima što ste bili dio mog putovanja!

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Diplomski rad

Određivanje kliničke važnosti strukturnih varijacija evolucijski novih gena kod ispitanika s neurorazvojnim poremećajima

Valentina Klein

Rooseveltov trg 6, 10000 Zagreb, Hrvatska

Neurorazvojni poremećaji predstavljaju heterogenu skupinu stanja koja utječu na normalan razvoj mozga, uključujući oštećenja kognitivnih i/ili motoričkih funkcija. U posljednjim desetljećima, strukturne varijacije genoma postale su dobro utvrđena etiologija neurorazvojnih poremećaja. Međutim zbog svoje kompleksnosti sama sekvenca ljudskog genoma nije bila u potpunosti poznata do unatrag nekoliko godina. Kompleksnost ljudskog genoma moguće je dokučiti upotrebom tehnologije sekvenciranja 3. generacije. Ovim istraživanjem analizirani su podaci dobiveni sekvenciranjem nanoporama dugih odsječaka DNA. U radu su vizualizirane strukturne varijacije u evolucijski novim genima 7 pacijenata s neutvrđenim uzrokom neurorazvojnih poremećaja. Dodatno su pretražene javno dostupne baze podataka koje sadrže informacije o odabranim genima i njihovim strukturnim varijantama s ciljem dobivanja cjelovite slike o pojedinom genu u ulozi razvoja i održavanja funkcija mozga. Detaljna analiza takvih sekvenci pomaže u evaluaciji rezultata i određivanju kliničkog značaja strukturnih varijacija evolucijski novih gena što je ključno za dijagnostiku, terapiju i prevenciju neurorazvojnih poremećaja. Rezultati ovog rada dali su nove kandidat-gene potencijalno uključene u patogenezu neurorazvojnih poremećaja.

Ključne riječi: ljudski genom, NRP, baze podataka, Nanopore sekvenciranje, SV, patogeneza (59 stranica, 18 slika, 4 tablice, 28 literaturnih navoda, jezik izvornika: hrvatski)

Rad je pohranjen u Središnjoj biološkoj knjižnici

Mentor: dr.sc. Martina Rinčić
Komentor: prof.dr.sc. Petra Korać

Ocjenitelji:
Prof.dr.sc. Petra Korać
Prof. dr.sc. Inga Urlić
Prof. dr.sc. Vesna Benković

Rad prihvaćen: 05.09.2024.

BASIC DOCUMENTATION CARD

University of Zagreb
Faculty of Science
Department of Biology

Master thesis

Determination of clinical significance of structural variations of evolutionary novel genes in patients with neurodevelopmental disorders

Valentina Klein

Rooseveltova trg 6, 10000 Zagreb, Croatia

Neurodevelopmental disorders represent a heterogeneous group of conditions that affect normal brain development, including cognitive and/or motor impairments. In recent decades, genomic structural variations have become a well-established etiology of neurodevelopmental disorders. However, due to its complexity, the sequence of the human genome itself was not fully known until a few years ago. The complexity of the human genome can be understood using 3rd generation sequencing technology. This research analyzed data obtained by long-read nanopore DNA sequencing. Structural variations in evolutionarily novel genes were visualized in seven patients with neurodevelopmental disorders of unknown cause. Additionally, publicly available databases containing information on selected genes and their structural variants were searched to provide a comprehensive understanding of each gene's role in brain development and function. Detailed analysis of such sequences helps in evaluating the results and determining the clinical significance of structural variations in evolutionarily novel genes, which is essential for the diagnosis, therapy and prevention of neurodevelopmental disorders. The results of this study identified new candidate-genes potentially involved in the pathogenesis of neurodevelopmental disorders.

Keywords: human genome, NRP, databases, Nanopore sequencing, SV, pathogenesis
(59 pages, 18 figures, 4 tables, 28 references, original in: Croatian)

Thesis is deposited in Central Biological Library.

Mentor: Martina Rinčić, PhD
Komentor: Prof. Petra Korać

Reviewers:

Assoc. Prof. Petra Korać
Assoc. Prof. Inga Urlić
Assoc. Prof. Vesna Benković

Thesis accepted: 05.09.2024.

SADRŽAJ

1.	UVOD.....	1
1.1.	Ljudski genom	2
1.2.	Strukturne varijacije	3
1.3.	Sekvenciranje.....	5
1.4.	Evolucijski novi geni kod čovjeka.....	9
1.5.	Neurorazvojni poremećaji	11
2.	CILJEVI ISTRAŽIVANJA	13
3.	MATERIJALI I METODE.....	14
3.1.	Materijali	14
3.2.	Obrada podataka dobivenih sekvenciranjem.....	14
3.3.	Pregled podataka o odabranim genima iz dostupnih baza podataka	15
3.3.1.	Baza Database of genomic variants	16
3.3.2.	Baza Decipher	17
3.3.3.	Baza Human brain transcriptome	19
3.3.4.	Baza Gene2Phenotype.....	19
3.3.5.	Baza ClinGen	20
3.3.6.	Baza Human protein atlas	20
3.3.7.	Baza SynGo.....	22
3.3.8.	Platforma SFARI Gene	23
3.3.9.	Baza AutDB	24
3.4.	Pregled literature.....	25
4.	REZULTATI	26
4.1.	Karakterizacija podataka dobivenih sekvenciranjem	26
4.2.	Vizualizacija strukturnih varijacija.....	26

4.3.	Provjera delecija i insercija.....	45
4.4.	Analiza baza podataka	47
5.	RASPRAVA.....	51
6.	ZAKLJUČAK.....	55
7.	LITERATURA	56

POPIS KRATICA

ADHD – poremećaj pozornosti s hiperaktivnošću, od engl. *Attention deficit hyperactivity disorder*

ASD – poremećaji iz spektra autizma, od engl. *Autism spectrum disorders*

ClinGen – baza podataka, od engl. *Clinical Genome Resource*

CNV – promjena broja kopija DNA, od engl. *Copy number variants*

DECIPHER – baza podataka, od engl. *Database of genomic variation and Phenotype in Humans using Ensembl Resources*

DGV – baza genomskih varijantni, od engl. *Database of genomic variants*

FISH – metoda fluorescentne in situ hibridizacije, od engl. *Fluorescence in situ hybridization*

G2P – od engl. *Gene2Phenotype*

GTEX – projekt ekspresije genotipa u tkivu, od engl. *Genotype-Tissue Expression*

HBT – baza podataka Transkriptom ljudskog mozga, od engl. *Human Brain Transcriptome*

HC-NDD – geni koji su povezani s pojavom neurorazvojnih poremećaja s visokom pouzdanošću, od engl. *High confidence neurodevelopmental disorders genes*

HPA – baza podataka Atlas ljudskih proteina, od engl. *Human protein atlas*

HPM – mapa ljudskog proteoma, od engl. *Human Proteome Map*

IGV – program Integrativni genomski preglednik, od engl. *Integrative Genomics Viewer*

LOEUF – od engl. *Loss of function observed*

LRS – sekvenciranje dugih odsječaka, od engl. *Long-read sequencing*

NCBI – Nacionalni centar za biotehnoške informacije, od engl. *National Center for Biotechnology Information*

NDD – neurorazvojni poremećaji, od engl. *Neurodevelopmental disorders*

ONT – metoda sekvenciranja, od engl. *Oxford Nanopore Technologies*

PCR – lančana reakcija polimerazom, od engl. *Polymerase chain reaction*

pHaplo – predviđena vjerojatnost haploinsuficijencije, od engl. *Predicted Probability of Haploinsufficiency*

pLI – vjerojatnost netolerancije na gubitak funkcije gena, od engl. *Probability of Loss-of-function Intolerance*

pTriplo – predviđena vjerojatnost triplosenzitivnosti, od engl. *Predicted Probability of Triplosensitivity*

sHet – selekcijski koeficijent varijanti heterozigotnog gubitka funkcije gena, od engl. *Selection coefficient of heterozygous loss-of-function variants*

SNP – polimorfizam jednog nukleotida, od engl. *Single-nucleotide polymorphism*

SNV – varijanta jednog nukleotida, od engl. *Single nucleotide variant*

SOLiD – metoda sekvenciranja, od engl. *Sequencing by Oligonucleotide Ligation and Detection*

SV – strukturna varijacija

SynGO – od engl. *Synaptic Gene Ontologies*

UCSC – od engl. *The University of California Santa Cruz*

1. UVOD

Po završetku projekta sekvenciranja ljudskog genoma (*Human genome project*), 2003. godine, omogućena je identifikacija milijuna varijacija u ljudskom genomu, od polimorfizama jedne baze (SNP od engl. *single-nucleotide polymorphism*) do strukturnih varijacija veličine preko nekoliko tisuća baza, što je otvorilo put novim istraživanjima u području medicinske genetike. Projekt *Human genome project* rezultirao je stvaranjem referentne sekvence ljudskog genoma, koja je postala temeljni alat u otkrivanju genetske podloge mnogih bolesti. Unatoč očekivanjima i predviđanjima prije sekvenciranja ljudskog genoma, objavom rezultata 2003. godine otkriveno je kako ljudski genom ima ukupnu veličinu od 3 milijarde nukleotida, čime stvara digitalnu sliku od približno 750 megabajta, te sadrži znatno manje očekivanih gena koji kodiraju proteine, prema procjeni 20 000 - 25 000 (Moraes i Góes, 2016). Tek 1,5% ljudskog genoma direktno kodira proteine, dok se preostali dio sastoji od regulacijskih sekvenci, pokretnih genetičkih elemenata, introna, gena za RNA, pseudogena te nekodirajuće DNA. Povijesno su ove sekvence poznate i kao „junk DNA“. Analiza ovako velikog skupa podataka omogućena je napretkom tehnologija sekvenciranja. Nedavno su nove tehnologije sekvenciranja dugih odsječaka DNA, razvijene od strane tvrtki *Oxford Nanopore Technologies* i *Pacific BioSciences*, pružile novi način sekvenciranja ljudskog genoma. Konkretno, ove tehnologije omogućavaju otkrivanje složenih ponavljajućih struktura prisutnih na različitim lokacijama unutar ljudskog genoma, uključujući: centromere, nizove ribosomske RNA i pericentromerne ponavljajuće sekvence koje se nisu mogle detektirati prijašnjim tehnologijama (Nurk i sur., 2022). Nedavno je objavljena potpuna „od telomere do telomere“ (od engl. T2T – *telomere to telomere*) sekvenca za jedan ljudski haplotip. Sekvenciranje i sastavljanje kompletnog ljudskog genoma od telomere do telomere omogućilo je uvid u potpunu, neprekidnu sekvencu DNA.

S dostupnošću genomske analize cijelog genoma, postalo je očito da složena arhitektura ljudskog genoma obuhvaća širi spektar varijacija poznatih kao strukturne varijacije, koje uključuju insercije, delecije, inverzije, translokacije i promjena broja kopija (CNV od engl. *copy number variants*). Upravo ove promjene u genomu mogu uzrokovati razne bolesti i patološka stanja, poput neurorazvojnih poremećaja. Tijekom desetljeća nakon projekta sekvenciranja ljudskog genoma tehnološki razvoj omogućio je sekvenciranje pojedinačne molekule DNA u stvarnom vremenu (sekvenciranje 3. generacije) čime je započela nova era za ljudsku genomiku. Napredne tehnologije

sekvenciranja omogućavaju detaljnu analizu strukturnih varijacija, a posebice su zanimljive evolucijski nove regije, koje se pojavljuju u genomu roda *Homo*, karakteristične za čovjeka.

Danas su raznovrsni podaci o genima i njihovim proteinskim produktima dostupni pretraživanjem mrežnih baza podataka, što doprinosi razumijevanju kliničke slike pacijenta i kreiranju najpovoljnije terapije za pojedinca. Precizna genomska analiza pospješuje pravovremeno postavljanje dijagnoze uz ranu mogućnost intervencije, a napretkom alata i tehnika sekvenciranja te informacije postaju brzo i lako dostupne. Tehnike sekvenciranja treće generacije zaobilaze dosadašnje prepreke i ograničenja analize podataka te otvaraju vrata otkrivanju kliničke važnosti strukturnih varijacija evolucijski novih gena kod ispitanika s neurorazvojnim poremećajima.

1.1. Ljudski genom

Genom je kompletna sekvenca DNA koja se nalazi u svakoj stanici organizma i sadrži sve genetske informacije potrebne za rast, razvoj, funkcioniranje i reprodukciju organizma. Ljudski genom, koji sadrži 3 200 000 000 nukleotida u haploidnom setu kromosoma, nosi nasljednu informaciju koja određuje nasljedne osobine i podložnost bolestima. Ljudski genom je dinamičan i promjenjiv, a na njega djeluju vanjski faktori i životni stil pojedinca. Upravo genomska varijabilnost, definirana kao razlika u slijedu DNA u genomima pojedinaca neke vrste, omogućava jedinstvenost svakoj osobi, dok istovremeno otežava identifikaciju patoloških promjena. Promjene u genomu mogu biti raznovrsne, od numeričkih do strukturnih, uključujući polimorfizme jednog nukleotida, tandemska ponavljanja, kratke insercije i/ili delecije, promjene broja kopija (CNV), uz kromosomske reanžmane uključujući: inverzije, translokacije i amplifikacije do kojih najčešće dolazi zbog pogreški tijekom mejoze. Promjene u genomu mogu biti korisne, štetne ili neutralne. Aberacija može zahvatiti jedan ili više kromosoma, a razina posljedica ovisi o zahvaćenih regijama. Promjene u genomu čovjeka nisu rezultat samo mejotičke nepravilnosti i pogreški tijekom replikacije DNA, već na njega utječu i okolišni faktori poput infekcija, zračenja, kemijskih mutagena, hormona te bolesti majke, manjak nutrijenata ili hipoksija u ranim stadijima razvoja (Talseth-Palmer i Scott, 2011). Većina bolesti rezultat je promjena i u genima i u okolišu, a njihov udio varira ovisno o dijagnozi pacijenta. Ljudska populacija obiluje polimorfizmima u obliku promjena slijeda DNA, međutim takve promjene imaju slab ili nikakav učinak na protein kojeg kodira određeni gen. S druge strane mutacije, definirane kao varijacije u slijedu DNA prisutne u manje od 1% populacije često uzrokuju genetske poremećaje jer promjene u broju kopija gena ili regulacijskih sekvenci

često značajno utječu na biološku funkciju pojedinih vrsta stanica, što rezultira raznim fenotipskim karakteristikama. Otkriće promjena u broju kopija (CNV) u ljudskom genomu značajno je proširilo naše razumijevanje uloge strukturnih varijacija molekule DNA u podlozi bolesti.

1.2. Strukturne varijacije

Strukturne varijacije (SV) u ljudskom genomu odnose se na promjene u organizaciji DNA koja uključuje gubitak, dodatak ili zamjenu fragmenata DNA većih od 50 parova baza (Mastrososa i sur., 2023). Strukturne varijacije karakterizira promjena u broju kopija DNA ili promjena genomskog položaja, a uključuju: varijante broja kopija, insercije, inverzije, delecije, translokacije, pokretne genetičke elemente, ekspanzije repetitivnih sekvenci i složene rekombinacije.

Ove varijacije mogu nastati kao rezultat različitih mehanizama, uključujući događaje tijekom replikacije DNA (primjerice smanjenjem ili povećanjem broja kratkih tandemskih ponavljanja) nehomologno rekombinacijsko spajanje, ili mogu biti posljedica utjecaja pokretnih genetičkih elemenata, a nastaju lomom fosfodieterske veze u molekuli DNA. Osim toga strukturne varijacije mogu nastati i *de novo*, što znači da ih pojedinac ne nasljeđuje od roditelja već nastaju spontano tijekom razvoja zametnih stanica ili kasnije tijekom života. Strukturne varijacije koje dovode do promjene u broju kopija DNA nazivaju se promjene broja kopija (CNV). Definirane su kao posebna podvrsta SV, najčešće nastale duplikacijama i/ili delecijama. Amplifikacije imaju za posljedicu povećanje broja kopija DNA segmenta koji je zahvaćen, dok delecije imaju za posljedicu gubitak broja kopija DNA u zahvaćenom segmentu. S obzirom na referentni broj kopija amplifikacije mogu biti pojedinačne pa ih u tom slučaju nazivamo duplikacije, dok u slučaju povećanja broja kopija nastaju multiplikacije. Kada govorimo o delecijama one mogu biti homozigotne ili heterozigotne ovisno o gubitku broja kopija. Strukturne varijacije obično se opisuju kao pojedinačne promjene, ali mogu uključivati i kombinacije različitih tipova varijacija (Mahmoud, 2019).

Istraživanje genomske varijabilnosti intenzivno je započelo objavljivanjem rezultata Projekta sekvenciranja ljudskog genoma, nakon čega je uslijedio projekt HapMap na koji se nastavlja projekt *The 1000 Genomes Project*, s ciljem stvaranja najvećeg javno dostupnog kataloga ljudske genomske varijabilnosti. Glavni cilj projekta uključivao je identifikaciju većine genomskih varijacija s frekvencijom od najmanje 1% u proučavanim populacijama, čime se pruža uvid u genomsku raznolikost i, potencijalno, razumijevanje biologije bolesti (Auton i sur., 2015). U

sklopu projekta sekvencirani su genomi 2504 pojedinaca iz 26 populacija pri čemu su sve dobivene informacije postale dostupne javnosti putem besplatnih baza podataka. Osim toga, ciljevi su uključivali pronalaženje kratkih insercija i/ili delecija i strukturnih varijanti, određivanje genotipa i haplotipa svih uzoraka te razvijanje metoda analize DNA-sekvenci, alata i resursa koji će biti korišteni u budućim istraživanjima. Rezultati projekta *The 1000 Genomes project* pokazali su da prosječan ljudski genom obuhvaća od 2100 do 2500 strukturnih varijacija koje uključuju oko 20 milijuna parova baza (Auton i sur., 2015). Strukturne varijacije mogu imati različite posljedice na funkciju gena i fenotip organizma, a u posljednje vrijeme otkrivena je njihova povezanost s etiologijom neurorazvojnih poremećaja. Strukturne varijacije djeluju na funkciju gena kroz nekoliko mehanizama. Najjednostavniji mehanizam uključuje dodatak sekvence dupliciranjem ili gubitak sekvence brisanjem cijelih gena, što rezultira efektom doziranja gena, povećavajući ili smanjujući razinu ekspresije gena i količinu kodiranog proteina. Ekstreman primjer ovog efekta predstavlja homozigotna delecija gena, koja uzrokuje potpuni gubitak transkripta i kodiranog proteina. Alternativno, strukturna varijacija može indirektno utjecati na razine ekspresije gena mijenjanjem prostornog odnosa između regulacijskih elemenata i gena. Strukturna varijacija može uzrokovati i nastanak fuzijskog proteina, ili može djelovati na promjenu okvira čitanja uz mogućnost dodavanja ili gubitka dijelova proteina (Hollox i sur., 2022). Iako strukturne varijacije i genomska raznolikost predstavljaju podlogu evolucije, poremećaj funkcije gena može uzrokovati bolest. Napretkom tehnologija detekcije sekvence DNA, uočeno je kako su strukturne varijacije vrlo zastupljene u neurodegenerativnim bolestima mozga, poput Alzheimerove bolesti i frontotemporalne demencije (De Coster i Van Broeckhoven, 2019). Uz to, približno 15-20% osoba s intelektualnim teškoćama i poremećajima iz autističnog spektra pokazuje klinički značajne strukturne varijacije (Weckselblatt i Rudd, 2015), stoga je važna njihova pravovremena detekcija.

Standardne metode detekcije strukturnih varijacija danas uključuju pruganje kromosoma, fluorescentnu hibridizaciju *in situ* (FISH, od engl. *Fluorescence in situ hybridization*) i kromosomske mikročipove. Dok posljednja tehnika otkriva promjene broja kopija (CNV) u obliku duplikacija i delecija, pruganje kromosoma i FISH također mogu detektirati strukturne varijacije koje nisu samo promjene u broju kopija, poput uravnoteženih translokacija. Najnovije tehnike sekvenciranja 3. generacije omogućavaju dublju analizu strukturnih varijacija uz mogućnost istraživanja molekularnih mehanizama njihova nastanka.

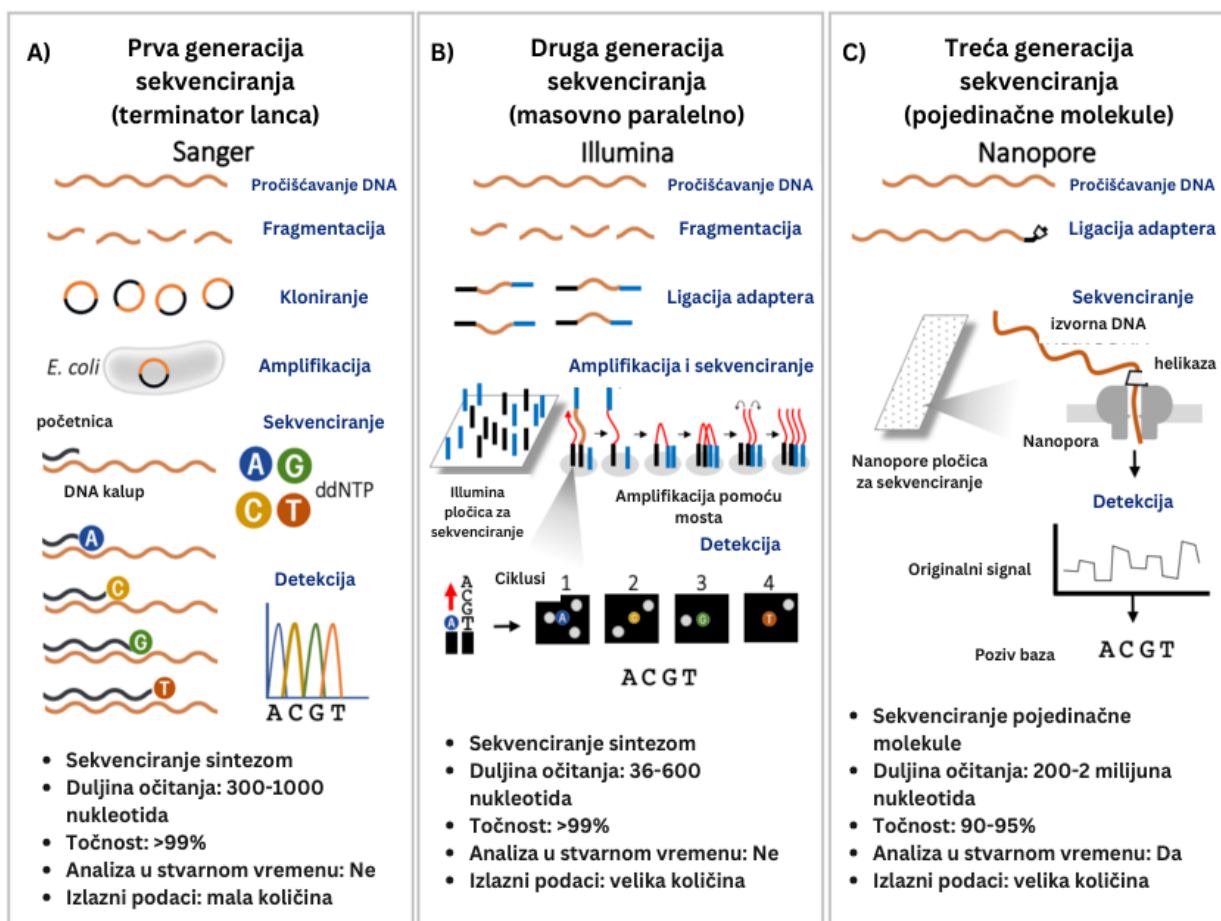
1.3. Sekvenciranje

Sekvenciranje je proces određivanja slijeda nukleotida (A, T/U, C i G) u molekuli DNA ili RNA. Kroz povijest, razvoj tehnologija sekvenciranja imao je ključnu ulogu u napretku razumijevanja ljudskog genoma. Od prvih metoda poput Sangerovog sekvenciranja, preko masovnog paralelnog sekvenciranja, do modernih tehnologija treće generacije, svaki je korak omogućio sve dublje uvide u genetske uzroke bolesti. Automatizacija Sangerovog sekvenciranja, razvijena u 1980-ima, predstavlja prvu generaciju tehnologija sekvenciranja DNA. Sangerova tehnika temeljila se na upotrebi lančane reakcije polimeraze (PCR od engl. *polymerase chain reaction*) za sintezu komplementarnih lanaca DNA u prisutnosti dideoksi-nukleotida, koji svojim vezanjem sprečavaju daljnju elongaciju. U samom procesu provodile su se 4 odvojene reakcije elongacije, do trenutka vezanja određenog dideoksi-nukleotida koji zaustavlja sintezu lanaca, nakon čega je potrebno provesti gel elektroforezu i fragmente poredati po veličini. Glavna ograničenja Sangerovog sekvenciranja uključivala su visoke troškove, nisku brzinu sekvenciranja i ograničenja u duljini očitavanja. Pojavom fluorescentno obilježenih signala i automatiziranih uređaja za kapilarnu elektroforezu reakcija se mogla provoditi u jednoj smjesi, ali duljina očitanih sljedova (engl. *reads*) nije bila veća od 1 kb (Heather i Chain, 2016).

Tehnološkim napretkom kompanija poput Pyrosequencing AB, Illumina i SOLiD, pojavljuje se druga generacija tehnologija sekvenciranja, nazvana još generacija masovnog paralelnog sekvenciranja. Fragmenti DNA prvo se ligiraju s adapterima, zatim se amplificiraju i sekvenciraju u masivno paralelnim reakcijama. Ove tehnike omogućile su istovremeno sekvenciranje milijuna kratkih fragmenata DNA, koje se naknadno sastavljaju u cjelovit slijed, povećavajući time brzinu i smanjujući troškove sekvenciranja. Međutim, ove tehnologije imaju ograničenja u duljini očitavanja i nisu bile idealne za detekciju strukturnih varijacija zbog veće mogućnosti pogreški. Sekvenciranje kratkih odsječaka DNA neće pouzdano mapirati sekvence u repetitivnim regijama genoma, kao što su regije nedavne segmentne duplikacije, tandemska ponavljanja ili regije obogaćene baznim parovima CG ili AT (Mastrososa i sur., 2023). Detekcija strukturnih varijacija metodom sekvenciranja kratkih odsječaka DNA pokazuje nisku osjetljivost (30-70%) te visoku stopu lažno pozitivnih rezultata (De Coster i Van Broeckhoven, 2019).

Najnoviji tehnološki napredak u ovom području dolazi s razvojem treće generacije sekvenciranja, poznate i kao sekvenciranje dugih odsječaka DNA (LRS od engl. *long-read sequencing*), koje su

u mogućnosti proizvesti kontinuirane sekvence u rasponu duljine od 10 kilobaza do nekoliko megabaza. Danas su na tržištu dostupne 2 varijante: tehnologija sekvenciranja koju je razvila tvrtka Pacific Bio-sciences (PacBio) i sekvencer tvrtke Oxford Nanopore Technologies (ONT). U radu Mastrososa i sur., 2023 istaknuto je kako LRS tehnologija može identificirati prosječno 25 000 strukturnih varijacija u ljudskom genomu za razliku od tehnologije sekvenciranja kratkih odsječaka istih uzoraka, koji, ovisno o alatima za otkrivanje SV-a, generiraju samo 3000 do 10 000 SV-a. Dodatno, otkrivanje strukturnih varijacija pomoću druge generacije sekvenciranja ograničeno je osjetljivošću i specifičnošću metode, te ga čini nepouzdanim kliničkim testom, stoga primjena modernih tehnologija sekvenciranja i analize podataka pruža novi uvid u prethodno propuštene varijante uzročnika bolesti. Glavne osobine pojedine generacije sekvenciranja uz pojedinačne korake prikazane su na Slici 1.



Slika 1. Shematski prikaz usporedbe glavnih karakteristika tri generacije sekvenciranja: a) Sangerovo sekvenciranje b) druga generacija sekvenciranja c) treća generacija sekvenciranja s glavnim predstavnicima i koracima sekvenciranja. Preuzeto i prilagođeno prema (Patterson i sur., 2020).

Nanoporsko sekvenciranje dugih odsječaka DNA razvijeno je s ciljem očitavanja signala u stvarnom vremenu bez prekida sinteze (poznato kao sekvenciranje u stvarnom vremenu) uz mogućnost sekvenciranja bez prethodnog umnožavanja uzorka. Ova tehnologija omogućuje neprekidno očitavanje pojedinačnih molekula DNA, čime se eliminira potreba za fragmentacijom DNA i omogućuje detaljnija analiza strukturnih varijacija. Tehnologija sekvenciranja ONT, za razliku od većine drugih metoda, ne ovisi o djelovanju DNA-polimeraze, već koristi enzim za odmatanje i proteinske pore kroz koje se provlače jednolančane molekule DNA ili RNA kroz sintetičke membrane. Kada lanac DNA prolazi kroz nanoporu, sekvencer prepoznaje bazu prema promjenama u električnom potencijalu. Promjene u ionskoj struji, koje su jedinstvene za svaku bazu, prevode se u nukleotide putem algoritama strojnog učenja što omogućava očitavanje sekvence u stvarnom vremenu te automatsku usporedbu sa referentnim genomom.

Prije sekvenciranja potrebno je pripremiti biblioteku, što se radi jednim od dva glavna protokola. Prvi protokol uključuje ligaciju adaptera te omogućava završetak pripreme unutar 10 minuta, ali ima nedostatak da se adapteri mogu nasumično ugraditi što skraćuje fragmente DNA prije samog sekvenciranja. Drugi protokol, temeljen na ligaciji, očuva duljinu fragmenata DNA i može se dovršiti za 1-2 sata. Oba protokola uključuju učitavanje pripremljenih knjižnica u pločicu za sekvenciranje, gdje mogu ostati funkcionalne do 72 sata. Razlog tome je što proteinske pore unutar pločice za sekvenciranje s vremenom denaturiraju zbog zagrijavanja, što smanjuje kvalitetu izlaznih podataka. Stoga se preporuča pranje pločice za sekvenciranje i ponovno učitavanje novih biblioteka tijekom trajanja eksperimenta kako bi se optimizirali rezultati sekvenciranja.

Određivanje baza je prvi korak potreban za pretvaranje sirovih podataka sekvenciranja (od engl. *raw data*) u sekvencu nukleotida procesom poziva baze (od engl. *base calling*). Sirovi podaci u obliku impulsa pretvaraju se u baze, stvarajući dugi niz. Trenutačno korišteni softver za proces poziva baze zove se Guppy, koji koristi neuronske mreže za određivanje sekvence iz sirovog signala (Mastrososa i sur., 2023). Na taj način omogućeno je cijelogenomsko sekvenciranje unutar 8 sati, za razliku od Sangerovog sekvenciranja ljudskog genoma koje bi trajalo godinama.

Glavne prednosti nanoporskog sekvenciranja uključuju brzinu, neprekidno očitavanje dugih odsječaka DNA, mogućnost sekvenciranja bez prethodnog umnažanja DNA te prenosivost uređaja poput MinION-a. Međutim, postoje i određeni nedostaci uključujući: 1) točnost – iako se kontinuirano poboljšava, točnost nanoporskog sekvenciranja još uvijek zaostaje u usporedbi s

drugim tehnologijama. Sekvenciranje ONT metodom pokazuje pogreške pozivanja baza u telomernim područjima i teško razdvaja homopolimerere dulje od 5-7 nukleotida (Mahmoud, 2019 i Mastroso i sur., 2023), 2) troškovi – početni troškovi opreme za Nanopore sekvenciranje predstavljaju značajnu investiciju, iako zanemarivu u odnosu na prethodnu generaciju Illumina tehnologije i 3) analiza podataka dobivena ovom tehnologijom sekvenciranja može biti složena zbog velike količine podataka i potrebe za sofisticiranim algoritmima. Bez obzira na sve navedeno, uvođenjem nanoporskog sekvenciranja, omogućava se brže i učinkovitije sekvenciranje DNA s mogućnošću očitavanja dugih odsječaka i detekcije strukturnih varijacija koje nije bilo moguće detektirati sekvenciranjem kratkih odsječaka.

Razvojem tehnologije 3. generacije sekvenciranja prethodno nedostupna područja genoma sada se mogu istražiti. Konkretno, sekvenciranje upotrebom tehnologije LRS-a omogućava detekciju 3 do 4 puta više strukturnih varijacija, osobito u rasponu od 50 do 1000 bp, u usporedbi sa tehnologijom sekvenciranja kratkih odsječaka (De Coster i Van Broeckhoven, 2019). Uz to, sekvenciranje dugih odsječaka olakšava sklapanje genoma jer se dulja očitavanja međusobno preklapaju, čime se eliminiraju praznine i pokriva se cijeli genom. LRS tehnologija omogućava precizno sekvenciranje ponavljajućih regija te otkrivanje epigenetskih modifikacija poput metilacije. Dodatno, nove tehnologije sekvenciranja pomažu u haplotipizaciji jer se umjesto kombiniranja majčinske i očeve kopije kromosoma, alati za LRS mogu koristiti za njihovo odvojeno sastavljanje; na taj način moguće je odrediti je li alel naslijeđen s majčine/očeve strane i nalazi li se varijanta u *cis* ili *trans* obliku. Slijedom toga, sekvenciranje tehnologijom treće generacije omogućava sveobuhvatniju molekularnu dijagnozu rijetkih genetskih bolesti i predstavlja iskorak u medicini.

Godine 2022. časopis Nature proglasio je sekvenciranje dugih odsječaka metodom godine, upravo zbog brojnih mogućnosti koje upotreba ove tehnologije pruža znanstvenoj zajednici. Nanopore tehnologija nudi najduža očitavanja, dok se PacBio sustav ističe visokom točnošću. LRS tehnologija posebno se ističe u identificiranju novih izoformi i varijanti transkripata velikih gena, te kao takva predstavlja moćan istraživački alat. Konačno, LRS tehnologija predstavlja obećavajući učinak u otkrivanju strukturnih promjena u mitohondrijskom genomu te omogućava direktno sekvenciranje nativnih mRNA molekula (Mastroso i sur., 2023).

1.4. Evolucijski novi geni kod čovjeka

Evolucijski novi geni, definirani kao geni koji su specifični za čovjeka ili imaju značajne razlike u ekspresiji i funkciji u usporedbi s drugim vrstama, postali su predmet intenzivnog istraživanja u kontekstu neurorazvojnih poremećaja. Navedeni geni imaju ključnu ulogu u regulaciji procesa razvoja centralnog živčanog sustava i predstavljaju poveznicu s pojavnim oblicima neurorazvojnih poremećaja. Ovi geni nastaju kao rezultat evolucijskih procesa poput dupliciranja gena, genomske reorganizacije i divergentne evolucije te postaju ključan element u razumijevanju evolucije i biologije čovjeka. Evolucijski novi geni dijelovi su regija nedavne segmentne duplikacije, što znači da su kopije određenih dijelova gena ili čitavih gena umetnute u genom. Ove segmentne duplikacije mogu obuhvatiti više tisuća parova baza. Takve duplikacije mogu biti posljedica različitih molekularnih mehanizama, poput nehomologne rekombinacije ili retrotranspozicije, a uzrokuju promjene u funkciji gena. Ljudski genom sadrži velik broj porodica gena koje brzo evoluiraju, a danas je poznato više od 400 segmenata ljudskog genoma koji su višestruko duplicirani tijekom evolucije hominoida. Ukupno, segmentne duplikacije čine približno 5% ljudskog genoma (Bekpen i Tautz, 2019).

Evolucijski novi geni često pokazuju specifične uzorke ekspresije u različitim tkivima i stadijima razvoja, sugerirajući njihovu važnu ulogu u adaptaciji na okolišne uvjete i intelektualnim naprecima čovjeka. Mnogi od ovih gena sudjeluju u razvoju i regulaciji funkcija povezanih s mozgom. Primjerice porodica gena *NBPF*, distribuirana je duž kromosoma 1 u 16 kopija od kojih su 6 u tandemu, a 10 ih je raspršeno, a uključeni su u moždane i neuronske funkcije. U prilog tome, otkriveno je kako porodica *NBPF* ima važnu ulogu u razvoju mozga te je povezana s neurološkim poremećajima poput mikro- i makrocefalije, autizma i shizofrenije (Bekpen i Tautz, 2019). Snažni dokaz evolucijskog razvoja ove skupine gena karakteristične za hominine predstavlja činjenica da genom makakija sadrži 3 kopije navedenog gena, dok genom miša i drugih sisavaca nema jasne ortologe (Bekpen i Tautz, 2019). Također, postoji snažna povezanost broja kopija gena *NBPF* s povećanjem veličina mozga kod primata i ljudi. Genska porodica *NBPF* nalazi se unutar domene *Olduvai* ljudskog genoma, važne za stabilnost genoma. Promjene u broju kopija domene *Olduvai* mogu imati posljedice na uravnoteženost genoma i funkcije gena, potencijalno korelirajući s različitim genetskim poremećajima ili promijenjenim karakteristikama razvoja i funkcije mozga, uključujući nastanak neuroloških poremećaja, razvojnih anomalija te varijacija u kognitivnim

sposobnostima. Stoga, istraživanja domene *Olduvai* i promjene broja kopija (CNV) pruža uvide u genetsku raznolikost i razumijevanje genetskih bolesti.

Još jedan primjer evolucijski novih gena sklonih duplikaciji uključuje gensku porodicu *GOLGA*, čije promjene su povezane s intelektualnim poteškoćama i razvojem shizofrenije, autizma i epilepsije (Bekpen i Tautz, 2019). S obzirom na specifičnost i kompleksnost takvih sekvenci, klasično eksomsko/genomsko sekvenciranje nije primjenjivo kod njihove detekcije, posebice u istraživanjima i otkrivanjima strukturnih varijanti povezanih s neurorazvojnim poremećajima. Jedan potencijalni pristup prevladavanju tih ograničenja je korištenje tehnologije LRS.

Dokazano je kako su regije segmentne duplikacije podložnije nastanku varijacija u broju kopija, u odnosu na ostatak genoma. Za dublje razumijevanje uloge promjena broja kopija DNA u podlozi ljudskih bolesti, neophodno je analizirati fenotipske varijacije koje proizlaze iz tih promjena. Duplikacije i delecije, putem različitih molekularnih mehanizama, mogu rezultirati promjenom fenotipa. Primarni mehanizam putem kojeg varijacije u broju kopija utječu na fenotip uključuju promjenu ekspresije okolnih gena unutar iste genomske regije, što može rezultirati različitim fenotipovima. Dodatno, može se dogoditi promjena u broju kopija regulacijskih regija gena koje utječu na ekspresiju. Osim toga, CNV-ovi mogu imati pozicijski učinak, formirajući se u različitim udaljenostima od heterokromatskih regija (de Smith i sur., 2010). Nedavna istraživanja identificirala su vezu između CNV-a i mnogih kompleksnih bolesti i poremećaja. Na primjer, pacijenti oboljeli od autizma, Parkinsonove i Alzheimerove bolesti, shizofrenije, epilepsije i autoimunih bolesti imaju duplikacije i/ili delecije brojnih regija genoma.

Neovisno o tome radi li se o varijacijama u broju kopija, tandemskim ponavljanjima, malim insercijama i/ili delcijama ili točkastim mutacijama, promjene u evolucijski novim genima pružaju uvid u fenotipske jedinstvenosti čovjeka te omogućavaju istraživanje patoloških stanja. Analizirajući navedene promjene, moguće je odrediti molekularne karakteristike staničnih, fizioloških, morfoloških, bihevioralnih i reproduktivnih svojstava čovjeka. Osim toga evolucijski novi geni doprinose razvoju strukturnih i funkcionalnih karakteristika mozga te utječu na kognitivne sposobnosti, a njihove promjene često su povezane s predispozicijom za neurorazvojne poremećaje.

1.5. Neurorazvojni poremećaji

Neurorazvojni poremećaji (NDD, od engl. *neurodevelopmental disorders*) predstavljaju složenu grupu stanja koja značajno utječe na normalan razvoj mozga i povezani su s brojnim kognitivnim, senzornim i motoričkim ograničenjima. Promjene u strukturi i funkciji započinju u ranim stadijima razvoja, a njihov stupanj ovisi o individualnim karakteristikama pacijenta i pravovremenom početku terapije. Prosječno, neurorazvojni poremećaji pogađaju 1-3% djece (Hiatt i sur., 2021). Međutim postoje i razlike među spolovima, pri čemu su muškarci podložnija skupina razvoju NDD-a (May i sur., 2019). Neurorazvojni poremećaji manifestiraju se različitim simptomima, uključujući probleme s komunikacijom, socijalnom interakcijom, pažnjom, motoričkim vještinama, intelektualnim ograničenjima i dr. Uz navedeno, oboljeli od neurorazvojnih poremećaja pokazuju smanjenu plodnost (Morris-Rosendahl i Crocq, 2020).

Etiologija neurorazvojnih poremećaja je složena i multifaktorska, uključujući genetske, epigenetske, neurološke, okolišne i prenatalne faktore. Genetski faktori imaju ključnu ulogu u podlozi nastanka neurorazvojnih poremećaja, s kombinacijom nasljednih i sporadičnih genetskih varijacija. Brojni neurorazvojni poremećaji rezultat su velikih strukturalnih varijacija *de novo*, a uključuju promjene u više stotina gena (Hiatt i sur., 2021). Osim toga, promjene epigenetskih modifikacija, poput metilacije DNA i modifikacija histona, mogu utjecati na ekspresiju gena tijekom razvoja mozga. Okolišni faktori, poput izloženosti toksinima, infekcijama, traumama i stresu, također mogu doprinijeti riziku od razvoja neurorazvojnih poremećaja. Sama dijagnoza NDD-a temelji se na multidisciplinarnom pristupu koji uključuje kliničku procjenu, neuropsihološko testiranje, procjenu razvoja i ponašanja te procjenu funkcionalnih sposobnosti. Za mnoge neurorazvojne poremećaje ne postoji jednostavan dijagnostički test, pa se dijagnoza postavlja na temelju promatranja simptoma i kliničkih kriterija definiranih u medicinsko-dijagnostičkim priručnicima. Primjer takvog priručnika je „Dijagnostički i statistički priručnik za mentalne poremećaje“ koji objedinjuje 6 kategorija koje pobliže definiraju specifične poremećaje, a to su: poremećaji iz spektra autizma, komunikacijski poremećaji, poremećaj pozornosti s hiperaktivnošću, poremećaji intelektualnog razvoja, poremećaj motorike uz specifične poremećaje učenja. Terapija oboljelih od neurorazvojnih poremećaja ovisi o vrsti i težini poremećaja, a uključuje terapiju ponašanja, logopedsku terapiju, fizioterapiju, te terapiju lijekovima.

Različite vrste genetičkih promjena povezane su s pojavom neurorazvojnih poremećaja, uključujući rearanžmane dijelova kromosoma, promjene u broju kopija DNA, male insercije i/ili delecije te točkaste mutacije. Međutim, intenzitet fenotipa bolesti ovisi i o osjetljivosti uključenih gena. Osjetljivost gena označava koliko su geni tolerantni na mutacije; niža tolerancija označava višu osjetljivost. Geni povezani s neurorazvojnim poremećajima često su visoko osjetljivi i njihova mutacija može značajno povećati rizik od bolesti. Takvi geni, poznati kao haploinsuficijentni, obično su rijetki i podložni jakom negativnom selektivnom pritisku. Nasuprot tome, manje osjetljivi geni često prenose varijante kroz generacije bez izazivanja bolesti, iako novija istraživanja pokazuju da poligenski neurorazvojni poremećaji mogu biti povezani s uobičajenim varijantama, gdje je razvoj bolesti rezultat kombiniranih učinaka više mutacija i njihovih interakcija.

Pretraživanjem baza podataka nedavno su utvrđeni jaki kandidatni geni povezani s neurorazvojnim poremećajima. Prema radu Leblond i sur., 2021 ukupan broj visokopouzdanih NDD gena (HC-NDD, od engl. *high confidence NDD genes*) iznosi 1586. Kod ljudi, 68% HC-NDD gena s dominantnim modelom nasljeđivanja snažno je netolerantno na mutacije koje uzrokuju gubitak funkcije, uzimajući u obzir pLI ili LOEUF parametre. Ovi geni eksprimirani su u ranim fazama razvoja fetusnog mozga, a imaju ulogu u staničnom ciklusu, metabolizmu i sinaptičkim funkcijama (Leblond i sur., 2021). Stoga, napretkom u tehnologijama sekvenciranja genoma i metodama analize podataka danas je moguće odrediti precizniju genetičku dijagnozu podloge neurorazvojnih poremećaja te otkriti dodatne gene uključene u razvoj ove kronične bolesti.

2. CILJEVI ISTRAŽIVANJA

Glavni cilj istraživanja je utvrditi kliničku značajnost strukturnih varijacija u evolucijski novim genima kod ispitanika s neurorazvojnim poremećajima upotrebom sekvenciranja dugih odsječaka DNA.

Specifični ciljevi su:

- Vizualizacija podataka dobivenih sekvenciranjem
- Pretraživanje javno dostupnih baza podataka kako bi se dobile dodatne informacije o evolucijski novim genima, što pomaže stvaranju cjelovite slike o ulozi pojedinog gena (i njegovog proteinskog produkta) u razvoju i održanju funkcija mozga te uočavanju promjena koje se dešavaju u podlozi neurorazvojjih poremećaja.
- Evaluacija strukturnih varijacija te određivanje kliničkog značaja uočenih promjena

3. MATERIJALI I METODE

3.1. Materijali

Diplomski rad izrađen je u sklopu projekta „Koliko smo različiti kao ljudi? Raznolikost evolucijski novih gena u neurorazvojnim poremećajima“ financiran od Zaklade ADRIS, pri čemu su prikupljeni uzorci 7 pacijenata. Istraživanje je odobreno od Etičkog povjerenstva Medicinskog fakulteta Sveučilišta u Zagrebu, Etičkog povjerenstva Centra za rehabilitaciju Stančić i Etičkog povjerenstva Centra za rehabilitaciju Zagreb uz informirani pristanak roditelja/staratelja za sudjelovanje u istraživanju.

U istraživanje su uključeni pacijenti s prethodnom genetičkom obradom koja nije utvrdila uzrok neurorazvojnih poremećaja. Obrada je uključivala analizu kariotipa metodom G-pruganja, analizu MLPA, te analizu DNA na kromosomskim mikročipovima. Nalazi svih navedenih pretraga pokazali su normalne rezultate. U svrhu anonimnosti podataka pacijenti su označeni šiframa CR2-1, CR2-2, CR2-4, ST-1, ST-2, ST-3 i ST-4. Svakom od 7 pacijenata uzeti su uzorci krvi i provedeno je sekvenciranje nanoporama dugih odsječaka DNA. Sekvenciranje je napravljeno u Institutu za humanu genetiku i genomsku medicinu, Sveučilišta RWTH Aachen, Njemačka. Učinjeno je sekvenciranje cijelog genoma, pri čemu su za potrebe izrade diplomskog rada izdvojene evolucijski nove regije, prisutne isključivo u genomu čovjeka. Genomske regije od interesa nađene su pretraživanjem baze *GenTree* (<http://gentree.ioz.ac.cn/>). Integrirana mrežna baza podataka *GenTree* pruža informacije o evoluciji i funkciji kodirajućih gena specifičnih za primata. Objedinjuje tri tipa podataka: (1) podatke o anotaciji gena iz baze *Ensembl* (2) zaključak o starosti gena i mehanizmu nastanka i (3) podatke o ulozi gena temeljene na funkcionalnim genomskim podacima iz baza *Human protein atlas* (HPA), *Genotype-Tissue Expression* (GTEx), *Human Proteome Map* (HPM) i *BrainSpan* uz kvantitativne genetičke podatke (*genome-wide association data from GWASdb2*) i selektivne podatke (*positive selection tracks from PopHuman*). Popis gena izdvojenih iz baze *GenTree* nalazi se u Prilogu 1.

3.2. Obrada podataka dobivenih sekvenciranjem

Nakon sekvenciranja dobiven je velik skup podataka, od čega su za potrebe ovog istraživanja korišteni sljedeći podaci: strukturne varijacije (SV) pojedinog pacijenta, kromosom na kojem se

SV nalazi, početnu i završnu koordinatu segmenta DNA, duljinu segmenta DNA izraženu u parovima baza, tip promjene (delecija, insercija ili inverzija), točan slijed baza, broj i oznaku gena koji SV uključuje, postojanje pomaka okvira čitanja, položaj unutar eksona ili introna uz druge dodatne informacije. U prvom koraku izdvojila sam podatke koji se nalaze unutar eksona s ciljem pronalaska kodirajućih regija. Potom sam provela izdvajanje velikih strukturnih varijacija, veličine više od 800 parova baza.

Podatke o strukturnim varijacijama koji se nalaze unutar eksona pregledala sam u računalnom programu *Integrative Genomics Viewer* (IGV) što je omogućilo dobivanje vizualne informacije. IGV, razvijen od strane instituta Broad, je jednostavan i multifunkcionalan softverski alat prilagođen za vizualizaciju i analizu genomskih podataka. *Integrative Genomics Viewer* dopušta unos velikog seta podataka, uključujući podatke sekvenciranja treće generacije, a moguće je analizirati gensku ekspresiju, mutacije, genomske anotacije, metilaciju, poravnate sekvence očitavanja, broj kopija te RNA-interferenciju. Unosom točnih koordinata strukturne varijacije pojedinog pacijenta analizirala sam delecije i insercije kako bih dobila podatke o pokrivenosti i kvaliteti sekvenciranja. Analizu sam provela unosom kromosoma na kojem se nalazi strukturna varijacija uz start- i stop-poziciju (primjer: chr4:190175199-190176043). Prije analize sekvence sam poredala po kvaliteti (opcija „sort alignment by mapping quality“). Na taj način olakšana je interpretacija strukturnih varijanti i funkcionalna karakterizacija. Kao referentni genom koristila sam Human GRCh38/hg38. Pregled podataka unutar IGV-a omogućava uočavanje grešaka tijekom sekvenciranja te izdvajanje stvarnih promjena. Iz podataka od interesa izdvojila sam aberacije koje se preklapaju kod više pacijenata kako bih odredila ima li učestalih promjena.

3.3. Pregled podataka o odabranim genima iz dostupnih baza podataka

U cilju pronalaska gena povezanih s neurorazvojnim poremećajima, unutar analiziranih podataka izdvojila sam gene zahvaćene strukturnom varijacijom. Geni na kojima sam provodila daljnju analizu su: *DUX4*, *NBPF26*, *GOLGA6L1*, *LOC102723623*, *POTEH*, *FAM86B1*, *FAM86B2*, *GOLGA6L22*, *RGPD6*, *NPIP9*, *NBPF14*, *ANKRD20A1*, *NBPF10*, *CLEC18A*, *NBPF19*, *FAM90A7P*, *ARHGAP11B* i *GOLGA6L6*.

3.3.1. Baza Database of genomic variants

Pregled baze *Database of genomic variants* (DGV) preko UCSC alata *Data integrator* (<https://genome.ucsc.edu/cgi-bin/hgIntegrator>) omogućava uvid u katalog strukturnih varijacija unutar ljudskog genoma. Unosom definiranih regija od interesa (primjer: chr1 120824880 120829778) uz odabir izlaznih podataka analizirala sam strukturne varijacije prisutne u općoj populaciji, koje je moguće usporediti s pacijentima čije uzorke sam analizirala. Samu analizu provela sam uz upotrebu standarda *DGV Gold*, kako bi se postavili strogi kriteriji i izdvojili podaci iz odabranih studija. Varijante *DGV Gold* služe kao referenca visokokvalitetnih CNV-ova, prikupljenih iz visokorezolucijskih studija s niskim stopama lažno pozitivnih rezultata. Koristila sam *DGV Gold Standard* za usporedbu i verifikaciju podataka, fokusirajući se na varijante s niskim rangom kako bih identificirala pouzdane strukturne varijacije. U bazi sam pregledala podatke o samoj varijanti (dodatak ili gubitak baza), broj obuhvaćenih istraživanja, broj ispitanika u pojedinom istraživanju te frekvenciju unutar određene populacije (afrička, europska, azijska, mješovita, meksička, bliskoistočna, nepoznata, indijanska, sjevernoamerička, oceanijska, južnoamerička te turska populacija). Naknadno sam izdvojila samo istraživanja koja uključuju gubitak baza. Kao referentni genom koristila sam Human GRCh38/hg38.

Sustav *The University of California Santa Cruz (UCSC) Genome Browser* (<https://genome.ucsc.edu/>) predstavlja interaktivnu mrežnu platformu koja pruža pristup genomskim informacijama raznih vrsta kralježnjaka i beskrležnjaka te glavnih modelnih organizama. Osnovni cilj je omogućiti znanstvenoj zajednici pristup detaljnim genomskim podacima, uključujući: podatke o genskoj ekspresiji, anotaciji gena, fenotipu, SNP-ove i druge varijacije, epigenetske oznake, regulacijske elemente, transkriptome te olakšati njihovu analizu i interpretaciju.

U svrhu ovog istraživanja pretražila sam *Human BLAT Search*. Polje *Human BLAT Search* u bazi *UCSC Genome Browser* omogućava pretraživanje specifičnih sekvenci u ljudskom genomu (*Assembly: Dec. 2013 GRCh38/hg38*) korištenjem algoritma BLAT (*BLAST-like Alignment Tool*). Ovim alatom ostvaruje se brzo i učinkovito pronalaženje koordinata sekvenci koje se podudaraju s ulaznim podacima. Koristeći ovaj alat, uvrstila sam točan slijed baza insercija pacijenata, s ciljem identifikacije područja u genomu koja se podudaraju s unesenom sekvencom. Analizu sam provela kako bih istražila podrijetlo sekvenci u genomu naših pacijenata.

3.3.2. Baza Decipher

Za utvrđivanje postojećih strukturnih varijanti u ciljanim genima pretražila sam mrežnu bazu *Decipher – DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources* (<https://www.deciphergenomics.org/>) unosom točnog imena gena. Baza *Decipher* uključuje podatke o 45837 pacijenata s različitim genetskim poremećajima i rijetkim bolestima, uključujući razvojne poremećaje, intelektualne teškoće, kronične i maligne bolesti, kongenitalne anomalije i druge rijetke sindrome. Baza *Decipher* opisuje 184814 fenotipskih karakteristika tih pacijenata uz 46800 zapisa o promjeni broja kopija DNA i 12244 varijanti sekvenci. Analiza pojedinih genskih varijanti i njihova posljedica (gubitak ili promjena funkcije) omogućava uvid u mehanizam bolesti i predstavlja prvi korak u određivanju kliničkog značaja pojedine promjene. Baza *Decipher* omogućava uvid u kromosomske aberacije, uključujući mikrolelecije i duplikacije te varijante patogenih sekvenci poput varijanti jednog nukleotida (SNV od engl. *single nucleotide variant*), insercije, delecije i malih insercija i/ili delecija te ih mapira u ljudski genom pomoću resursa *Ensembl* ili pretraživača *UCSC Genome*. *Ensembl* koristi najnoviju verziju ljudskog genoma, koju je napravio NCBI (trenutno NCBI-36). Baza *Decipher* objedinjuje niz alata dizajniranih za podršku u interpretaciji genskih varijanti, a informacije su prikupljene iz raznih bionformatičkih izvora. Za potrebe izrade diplomskog rada iz *Deciphera* sam preuzela slijedeće podatke: pLI, LOEUF, sHEt, pHaplo, pTriplo.

Vjerojatnost netolerancije na gubitak funkcije gena (pLI, od engl. *Probability of Loss-of-function Intolerance*) prikazuje koliko je pojedini gen osjetljiv na mutacije koje dovode do gubitka funkcije. Unutar baze *Decipher* vrijednosti se kreću u rangu od 0 do 1. Geni povišene vrijednosti pLI (≥ 0.9) pokazuju jaku netoleranciju na gubitak funkcije gena, dok su geni niske vrijednosti pLI (≤ 0.1) tolerantni na gubitak funkcije gena. Podaci pLI vrijednosti izvedeni su iz 125 748 eksoma jedinstvenih nesrodnih pojedinaca (*gnomAD v2.1.1*).

LOEUF (od engl. *Loss of function Observed*) predstavlja kvantitativnu mjeru koliko su česte ili rijetke mutacije koje dovode do gubitka funkcije gena u *gnomAD-u* u usporedbi s modelom bez mutacija. Budući da *gnomAD* nije savršena preslika populacije, te ne predstavlja pravo smanjenje gubitka funkcije gena, postavljena je maksimalna granica procjene kako bi se učinila konzervativnijom. Vrijednost može biti minimalno 0, ali nema gornje granice. Geni s manjim vrijednostima (bliže nuli) su osjetljiviji na mutacije.

Selekcijski koeficijent varijanti heterozigotnog gubitka funkcije gena (sHet, od engl. *Selection coefficient of heterozygous loss-of-function variants*) predstavlja procjenu utjecaja na fitnes (npr. preživljavanje i reprodukciju) za osobe koje imaju heterozigotne varijante gubitka funkcije u određenom genu. Unutar baze *Decipher* vrijednosti se kreću od 0 do 1. Geni s većim vrijednostima manje su tolerantni na mutacije. Ova mjera temelji se na 33 370 pojedinačnih eksoma iz podskupa nefinskih Europljana iz skupa podataka *ExAC*.

Vrijednost pHaplo (od engl. *Predicted Probability of Haploinsufficiency*) pokazuje vjerojatnost haploinsuficijencije predviđene analizom velikih rijetkih varijanti broja kopija utvrđenih mikročipovima u 950 278 osoba. Haploinsuficijencija je stanje u kojem je jedna funkcionalna kopija gena nedovoljna za održavanje normalne funkcije, obično zbog inaktivacije ili brisanja druge kopije. Vrijednosti pHaplo kvantificiraju vjerojatnost da određeni gen pokazuje ovu karakteristiku, a generirane su modelom strojnog učenja obučenog na 145 značajki na razini gena kako bi se predvidjela vjerojatnost da će potpuno brisanje svakog gena biti češće u skupini osoba pogođenih teškim, progresivnim bolestima, u usporedbi s općom populacijom. Vrijednost pHaplo ≥ 0.86 ukazuje na snažnu vjerojatnost da je gen osjetljiv na dozu, što znači da je učinak delecije jednako jak kao gubitak funkcije gena te može dovesti do patoloških stanja. Za one pacijente koji pokazuju delecije unutar ciljanih gena izdvojila sam pHaplo vrijednosti.

Vrijednost pTriplo je metrički pokazatelj koji odražava vjerojatnost triplosenzitivnosti (pTriplo; tj. netolerancije na duplikaciju) predviđenu analizom velikih rijetkih varijanti broja kopija utvrđenih mikročipovima u 950 278 pojedinaca. Vrijednosti pTriplo generirane su modelom strujnog učenja uvježbanog na 145 značajki na razini gena kako bi se predvidjela vjerojatnost da će potpuni dobitak kopije (tj. potpuna duplikacija) svakog gena biti obogaćena u skupini pojedinaca pogođenih teškim, rano počinjućim bolestima u usporedbi s općom populacijom. Vrijednost pTriplo ≥ 0.94 ukazuje na visoku vjerojatnost da je gen osjetljiv na povećanu dozu, što znači da duplikacije mogu značajno utjecati na funkciju gena i potencijalno dovesti do bolesti. Za pacijente koji pokazuju duplikacije unutar ciljanih gena izdvojila sam pTriplo vrijednosti.

Ovakva analiza pruža dublji uvid u pitanje koliko su pojedini geni osjetljivi na mutacije i je li promjena broja kopija gena posljedica normalne polimorfne varijante ili potencijalni uzrok kliničkih simptoma pacijenta.

3.3.3. Baza Human brain transcriptome

Slijedeći korak uključivao je pretraživanje baze *Human brain transcriptome* (HBT) (<https://hbatlas.org/>). Projekt HBT na Odjelu za Neurobiologiju Medicinskog Fakulteta Sveučilišta Yale predstavlja javno dostupnu bazu podataka koja sadrži podatke o transkriptomu i povezane metapodatke ljudskog mozga u razvoju i odraslog mozga. Baza obuhvaća informacije o transkriptomima na razini eksona generirane korištenjem metode *Affymetrix GeneChip Human Exon 1.0 ST Arrays* iz više od 1340 tkivnih uzoraka uzetih s obje hemisfere ljudskih mozgova *post mortem*. Uzorci obuhvaćaju razdoblje od embrijskog razvoja do odrasle dobi i predstavljaju muškarce i žene različitih etničkih skupina. Ukupno je uzorkovano 16 područja mozga: moždana kora, mediodorzalna jezgra talamusa, prugasto moždano tijelo (strijatum), amigdala, hipokampus i 11 područja neokorteksa. Podaci o genotipizaciji na razini genoma za 2,5 milijuna oznaka (*Illumina Human Omni 2.5-Quad Bead Chips*) dostupni su za sve uzorke na zahtjev. Osnovna pretraga unutar baze omogućava pretraživanje po imenu gena i području mozga, dok napredne funkcije nude popise 10-100 gena obogaćenih u jednoj od regija mozga. Gene od interesa upisala sam u bazu *Human brain transcriptome* kako bih dobila dodatne informacije o genskoj ekspresiji u pojedinim regijama ljudskog mozga te u određenim stadijima života.

3.3.4. Baza Gene2Phenotype

Baza *Gene2Phenotype* (G2P) (<https://www.ebi.ac.uk/gene2phenotype>) je javno dostupan online sustav dizajniran za razvoj, provjeru, uređivanje i distribuciju velikih skupova podataka temeljenih na dokazima, korisnih za filtriranje dijagnostičkih varijanti. Svaki unos u G2P povezuje specifični gen s određenom bolešću te ima propisani stupanj pouzdanosti i poveznicu na dokaze.

Baza omogućava pretraživanje unutar određenog panela: kardiološki panel (od engl. *Cardiac Panel*), panel rak (od engl. *Cancer panel*), panel razvojnih poremećaja (od engl. *Developmental Disorders-DD Panel*), panel oko (od engl. *Eye Panel*), panel koža (od engl. *Skin Panel*), panel koštanog sustava (od engl. *Skeletal Panel*) a moguće je pretraživanje po genima, fenotipovima, bolestima ili specifičnim varijantama. Unosom gena od interesa (primjerice *ARHGAP11B*) ostvarivo je povezati određenu mutaciju s njegovom fenotipskom posljedicom. Bazu sam pretražila za svih 18 gena od interesa.

3.3.5. Baza ClinGen

Za dobivanje dodatnih informacija o genima i njihovom kliničkom značaju pretražila sam bazu *ClinGen – Clinical Genome Resource* (<https://clinicalgenome.org/>). Baza *ClinGen* predstavlja središnji izvor u području medicinske genetike zadužen za informiranje o klinički relevantnim varijantama gena. Njena uloga je pružiti strukturiranu i ažuriranu bazu podataka o genetičkim varijantama i njihovim kliničkim značajkama, kako bi se olakšalo tumačenje genetičkih testova i dijagnostičkih rezultata. Baza *ClinGen* omogućuje istraživačima, kliničarima i drugim korisnicima pristup informacijama o varijantama gena, uključujući njihovu patogenost, fenotipske efekte i kliničke preporuke za upravljanje pacijentima. Unosom simbola pojedinog gena alatna traka nudi: općenite informacije o genu, povezanost gena i bolesti, osjetljivost na dozu, kliničku primjenjivost, patogenost varijante, status, budući rad i poveznicu na *ClinVar*. Javna baza podataka *ClinVar* prikuplja i pohranjuje informacije o genskim varijantama, uključujući njihovu povezanost s bolestima i odgovorima na lijekove. Omogućuje sistematsko katalogiziranje varijanti u ljudskom genomu i integrira informacije o njihovom kliničkom značaju.

3.3.6. Baza Human protein atlas

Unos odabranih gena u mrežnu bazu podataka *Human Protein Atlas* (HPA) (<https://www.proteinatlas.org/>) omogućava mapiranje gena te njihovih proteinskih produkata, u raznim stanicama, tkivima i organima. Mapiranje se temelji na integriranoj tehnologiji *-omics*, uključujući slikovne tehnike temeljene na antitijelima, proteomiku baziranu na masenoj spektrometriji, transkriptomiku te sistemsku biologiju. Pregledom baze stječu se informacije o lokaciji, razini ekspresije i funkciji proteina u različitim tkivima i stanicama ljudskog organizma. Na temelju tih podataka moguće je odrediti gdje su i koliko eksprimirani traženi genski produkti. Osim toga baza pruža informacije o interakcijama s drugim proteinima, broju transkripata, izgledu samog proteina i brojne druge podatke. Takvo profiliranje proteina omogućava razumijevanje njihove uloge u normalnim fiziološkim procesima, što olakšava daljnju analizu i usporedbu s ekspresijom proteina u patološkim stanjima.

Baza *Human Protein Atlas* sastoji se od dvanaest zasebnih dijelova, svaki usmjeren na određeni aspekt analize ljudskih proteina. U alatnoj traci moguće je pregledati: odjeljak Tkivo (od engl. *Tissue section*), odjeljak Mozak (od engl. *Brain section*), odjeljak Pojedinačni stanični tip (od engl.

Single Cell Type section), odjeljak Tkivni stanični tip (od engl. *Tissue Cell Type section*), odjeljak Patologija (od engl. *Pathology section*), odjeljak Atlas krvnih bolesti (od engl. *Disease Blood Atlas section*), odjeljak Imunološke stanice (od engl. *Immune Cell section*), odjeljak Krvni proteini (od engl. *Blood Protein section*), odjeljak Substanično (od engl. *Subcellular section*), Odjeljak Stanična linija (od engl. *Cell Line section*), odjeljak Struktura (od engl. *Structure section*), odjeljak Interakcija (od engl. *Interaction section*). U ovom istraživanju pretražila sam odjeljke:

1) Tkivo – prikazuje ekspresijski profil gena u svim glavnim tkivima i organima ljudskog tijela na razini mRNA i na razini proteina. Podaci o ekspresiji proteina dobiveni su iz 44 različita nepatološka tkiva čovjeka upotrebom profiliranja proteina temeljenog na antitijelima, koristeći metode konvencionalnog i višestrukog imunohistokemijskog bojenja. Temeljne slike nepatoloških tkiva obojenih imunohistokemijskom metodom dostupne su zajedno s oznakama razine ekspresije proteina. Podaci o proteinima obuhvaćaju 15 303 gena (76%) za koje su dostupna antitijela. Podaci o ekspresiji mRNA dobiveni su iz dubinskog sekvenciranja RNA (RNA-seq) iz 40 različitih normaliziranih tkiva. Normalizirana tkiva odnose se na biološke uzorke tkiva koji su obrađeni na način da se smanjuju ili otklanjaju varijacije unutar i između uzoraka koje nisu povezane s osnovnim biološkim procesima ili stanjima, omogućujući tako dosljedne i pouzdane rezultate u istraživačkim studijama. Normalizacija tkiva uključuje standardizaciju metode prikupljanja, obrade i analize tkiva kako bi se smanjila tehnička varijabilnost i osigurala komparabilnost podataka. Uz to, u bazi su grafičkim prikazom izdvojeni podaci o ekspresiji RNA, uz podatke o proteinskoj ekspresiji. Kategorija specifičnosti RNA (od engl. *Tissue specificity RNA*) temelji se na razinama ekspresije mRNA u konsenzusnom skupu podataka (od engl. *Consensus dataset*) koji se izračunava kombiniranjem razina ekspresije RNA u uzorcima iz transkriptomskih skupova podataka HPA i GTEx korištenjem internog kanala normalizacije. Kategorije specifičnosti RNA uključuju: tkivno obogaćeno (od engl. *tissue enriched*), grupno obogaćeno (od engl. *group enriched*), tkivno pojačan (od engl. *tissue enhanced*), niska tkivna specifičnost (od engl. *low tissue specificity*) uz kategoriju nije otkriveno (od engl. *not detected*). Tkivno obogaćeno kategorija opisuje ekspresiju u pojedinom tkivu/regiji ili stanici 4 puta većom u odnosu na druga tkiva/regije ili stanice. Grupno obogaćeno kategorija opisuje ekspresiju unutar grupe tkiva, primjerice 2-5 srodnih tkiva, regija mozga ili staničnih linija tako da je ekspresija u grupi četiri puta veća u odnosu na druga tkiva/regije/stanične linije. Kategoriji tkivno pojačan pridružuju se ona tkiva, moždane regije, stanične linije ili pojedinačni stanični tipovi u kojima je razina ekspresije 4 puta veća u

odnosu na prosjek. Kategorija niska tkivna specifičnost dodijeljena je ako je razina ekspresije ≥ 1 u najmanje jednom tipu tkiva, regije ili stanice, ali nije povišena ni u jednom tipu tkiva, regije ili stanica. Nije otkriveno kategorija dodjeljuje se ako je razina ekspresije manja od 1 u svim vrstama tkiva/regija/stanica.

2) Mozak – istražuje distribuciju i ekspresiju proteina u različitim regijama mozga sisavaca. Podaci o potpunom profiliranju mozga dobiveni su integriranjem informacija iz ljudskog, svinjskog i mišjeg mozga. Prikazani podaci odnose se na ljudske gene i njihove istovjetne ortologe u svinje i miša. Uzorci mozga grupirani su u 14 anatomskih regija, pružajući regionalnu klasifikaciju svih gena na temelju ekspresije RNA, pokazujući koji su proteini više eksprimirani u jednoj regiji u usporedbi s drugom. Unosom željenog gena moguće je provjeriti njegovu ekspresiju kod glavnih anatomskih regija, uključujući: moždanu koru, olfaktorni bulbus, hipokampusnu formaciju, amigdalnu, bazalne ganglije, talamus, hipotalamus, srednji mozak, mali mozak, pons, produženu moždinu, leđnu moždinu, bijelu tvar i koroidni pleksus, sve do pojedinih jezgara i potpodručja.

3) Pojedinačni stanični tip – odjeljak sadrži informacije o ekspresiji gena koji kodiraju proteine u pojedinačnim ljudskim stanicama temeljene na podacima o sekvenciranju RNA pojedinačnih stanica (scRNAseq, od engl. *Single-cell RNA sequencing*) dobivenom iz 31 ljudskog tkiva, uključujući mononuklearne stanice periferne krvi. Analiza scRNAseq-a temelji se na javno dostupnim podacima o ekspresiji na razini genoma i obuhvaća sve gene koji kodiraju proteine u 557 pojedinačnih staničnih klastera, što odgovara 15 različitih grupa stanica. Geni izraženi u svakom tipu stanica mogu se istražiti putem interaktivnih grafikona i stupčastih dijagrama, s poveznicama na odgovarajuća imunohistokemijska bojanja u ljudskim tkivima.

Nakon pregleda svih podataka, izdvojila sam podatke od interesa važne za razumijevanje kliničkog značaja pojedinog gena i njegove poveznice s nastankom neurorazvojnih poremećaja.

3.3.7. Baza SynGo

Unosom gena u mrežnu bazu *SynGO* – *Synaptic Gene Ontologies* (<https://www.syngoportal.org/>) moguće je procijeniti jesu li ciljani geni uključeni u funkcioniranje sinapsi. Baza pruža informacije o lokaciji i funkciji pojedinih sinaptičkih gena/proteina te opisuje njihove međusobne odnose. Mrežni alati baze pomažu analizirati, vizualizirati i strukturirati unesene podatke. U okviru baze *SynGO* dosad je anotirano 1112 gena koji su povezani s lokalizacijom i/ili funkcijom u sinapsama.

Geni unutar baze su izuzetno veliki, dobro očuvani, netolerantni na mutacije i povezani s poremećajima centralnog živčanog sustava. Bazu sam pregledala za svih 18 gena od interesa.

3.3.8. Platforma SFARI Gene

Platforma *SFARI Gene* (<https://gene.sfari.org/>) koristi se za istraživanje gena povezanih s predispozicijom za poremećaje iz spektra autizma (ASD, od engl. *autism spectrum disorders*). Primjenjujući pristup systemske biologije, platforma *SFARI Gene* povezuje informacije o kandidat-genima za autizam unutar svog originalnog modula *Human Gene* s podacima iz niza dodatnih modula. Nakon unosa imena gena moguće je pretraživati podatke u 4 modula. Modul Ljudski gen (od engl. *Human Gene*) sadrži temeljito anotiran popis gena koji su povezani s autizmom, dok modul Promjena broja kopija (od engl. *Copy Number Variant*) bilježi delecije i duplikacije jednog ili više gena u genomu i opisuje njihovu potencijalnu povezanost s autizmom. Uz njih, modul Životinjski modeli (od engl. *Animal Models*) sadrži informacije o linijama genetski modificiranih miševa koji predstavljaju potencijalne modele autizma.

Geni povezani s ASD-om ocjenjuju se primjenom skupa pravila za anotaciju i klasificiraju se u specifične kategorije na temelju dokaza koji podupiru njihovu povezanost s autizmom. Na taj način genu može biti dodijeljena kategorija:

- Kategorija Sindromski (od engl. *Syndromic*): Obuhvaća gene koji su povezani s rijetkim genetskim sindromima, a ti sindromi uključuju autistične osobine kao jedan od simptoma. Autizam u ovom slučaju nije izoliran poremećaj, već je dio kompleksnijeg sindroma s više kliničkih značajki. Mutacije samo u ovom genu nisu dovoljne za dijagnozu ASD-a
- Kategorija 1 – Visoko pouzdani (od engl. *High Confidence*): Ova kategorija označava gene koji su jasno povezani s autizmom, obično zbog prisutnosti barem tri mutacije *de novo* koje narušavaju funkciju gena, a opisane su u znanstvenoj literaturi.
- Kategorija 2 – Jaki kandidati (od engl. *Strong Candidate*): Kategorija jakih kandidata uključuje gene sa dvije mutacije *de novo* koje vjerojatno narušavaju funkciju gena, a rizična varijanta ima funkcionalni učinak.
- Kategorija 3 – Indikativni dokazi (od engl. *Suggestive Evidence*): Uključuje gene s jednom prijavljenom mutacijom *de novo* koja možda narušava funkciju gena. Informacije o ovim

genima dolaze iz istraživanja rijetkih nasljednih bolesti, a same nisu detaljno uspoređene s kontrolnom skupinom.

Sadržaj baze podataka *SFARI Gene* potječe iz recentnih i provjerenih znanstvenih istraživanja, a podaci se redovno ažuriraju kako bi im uvijek bila dodijeljena prava kategorija. Također, najnovija verzija platforme *SFARI Gene* uključuje niz posebnih alata za vizualizaciju podataka, što omogućuje bolji prikaz genetičkih podataka iz baze. U bazi je moguće naći i vanjske poveznice, koje pružaju dodatne informacije o traženom genu. U bazi sam istražila sve gene od interesa.

3.3.9. Baza AutDB

Unosom podataka u bazu *AutDB* (<http://autism.mindspec.org/autdb/Welcome.do>) istražuje se povezanost gena s poremećajima iz spektra autizma. Baza predstavlja integrirani model za prikupljanje, anotaciju, pohranu i vizualizaciju kandidatnih gena za ASD. Prvi put objavljena 2007. baza *AutDB* se razvila u složeni izvor različitih vrsta genetičkih i funkcionalnih dokaza povezanih s poremećajima iz spektra autizma. Sadašnji moduli uključuju: 1) Ljudski gen (od engl. *Human gene*) koji obuhvaća sve gene povezane s ASD-om i njihove varijante, 2) Životinjski modeli (od engl. *Animal models*) koji dokumentira bihevioralne, anatomske i fiziološke podatke o linijama genetski modificiranih miševa koji predstavljaju modele autizma, 3) Proteinska interakcija (od engl. *Protein interaction – PIN*) koja rekonstruira mreže interakcija između proteina povezanih s ASD-om i 4) Model promjene broja kopija DNA (od engl. *Copy Number Variant – CNV module*) koji prikazuje duplikacije i delecije kromosomskih regija identificiranih u ASD-u. Od ožujka 2024. godine *AutDB* sadrži anotaciju za 1440 gena, 2274 lokusa CNV-a, 3573 modela glodavaca i 38 296 PIN-ova. Podaci o kandidatnim genima prikupljeni su pretraživanjem znanstvene literature, posebice baze *PubMed*, nakon čega su klasificirani u 6 kategorija: funkcionalni (od eng. *functional*), rijetka varijanta pojedinačnog gena (od engl. *rare single gene variant*), povezanost (od engl. *association*), bez povezanosti (od engl. *no association*), multigenički CNV (od engl. *multigenic CNV*) i sindromski (od engl. *syndromic*). Također, genima je unutar baze dodijeljena ocjena (od engl. *gene ranking*) ovisno o snazi dokaza povezanosti određene genetske varijante s razvojem poremećaja iz spektra autizma. Ocjene su kreirane na temelju skupa dokaza od strane znanstvenika iz tog područja. Nakon unosa gena, rezultate pretraživanja pratila sam na dvije razine. Na prvoj razini prikazani su sažeti podaci o genu uključujući: jedinstveni identifikacijski broj,

simbol gena, naziv gena, kromosomsku lokaciju, genetičku kategoriju, ocjenu gena, broj varijanti, popis povezanih poremećaja, broj znanstvenih studija koje potkrepljuju navedene podatke uz primarnu referencu *PubMed*-a te mogućnost uređivanja. Druga razina dodatno prikazuje značajnost za autizam s prikazom sažetka izvješća o kandidatnom genu, molekularnu funkciju gena, poveznice prema vanjskim bazama podataka poput *NCBI*, *UniProt* i *GeneCards*, te ostale reference. Na ovaj način baza *AutDB* pomaže u razumijevanju genetičkih osnova autizma te identificiranju potencijalnih terapijskih ciljeva i dijagnostičkih markera.

3.4. Pregled literature

Za kraj sam detaljno pregledala bazu *PubMed* s ciljem pribavljanja dodatnih informacija o ciljanim genima te o njihovim ulogama u neurorazvojnim poremećajima kao i prikupljanja dodatnog znanja o njihovom kliničkom značaju. Baza *PubMed* predstavlja javno dostupnu online knjižnicu koja sadrži veliku količinu literature iz područja biomedicine, zdravstva i srodnih znanosti, uključujući znanstvene članke, kliničke studije, meta analize, pregledne radove, knjige i druge dokumente.

PubMed nudi napredne mogućnosti pretraživanja koje omogućuju korisnicima da pretražuju literaturu po ključnim riječima, autorima, temama, godini izdanja i drugim kriterijima. Bazu sam pretražila prema ključnim riječima poput imena gena, „neurodevelopmental“, „brain“, „disorders“, „neocortex“, „intelligence“, „human“, „CNV“, „autism“, „clinical“ i dr. Uz navedeno, unutar baze *PubMed* prikupila sam informacije o molekularnim karakteristikama gena i njihovih proteinskih produkata, biološki značaj te promjene povezane s patološkim stanjima. Filtrirala sam literaturu unutar posljednjih 5 godina, s ciljem pribavljanja aktualnih informacija. Za one gene o kojima nije bilo radova iz tog razdoblja, proširila smo pretragu na radove objavljene u posljednjih 10 godina.

4. REZULTATI

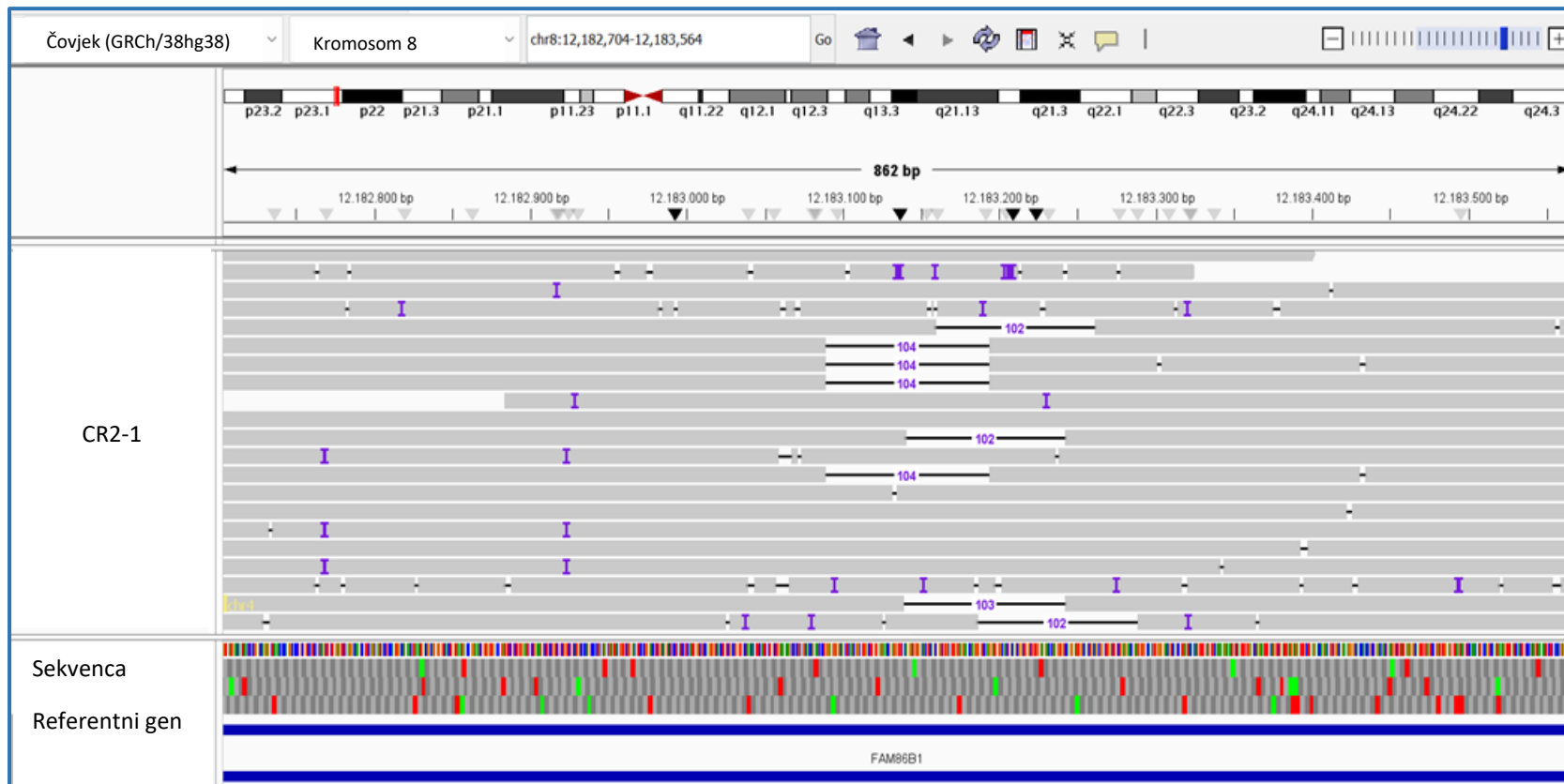
4.1. Karakterizacija podataka dobivenih sekvenciranjem

Iz podataka dobivenih sekvenciranjem u ovom istraživanju izdvojila sam detaljne informacije o strukturnim varijacijama unutar eksona u odabranim genima svih 7 pacijenata, kako je prikazano u Prilogu 2. Izdvojene informacije sadrže broj i točnu poziciju strukturnih varijacija u evolucijski novim regijama genoma pojedinog pacijenta uz uključene gene. Kod pacijenta CR2-1 pronašla sam 8 strukturnih varijacija, od kojih se 5 kategorizira kao delecija, dok su 3 insercije, a zahvaćaju između 54 – 4898 parova baza. U slučaju pacijenta CR2-2 pronašla sam 5 strukturnih varijacija u eksonima definiranih regijama, od kojih se 4 kategoriziraju kao delecija, a jedna je insercija. Strukturne varijacije zahvaćaju između 54 i 3185 parova baza. Kod pacijenta CR2-4 pronašla sam 8 strukturnih varijacija unutar definiranih regija, od kojih je 7 delecija, a jedna insercija, pri čemu je zahvaćeno između 53 - 6868 parova baza. U slučaju pacijenta ST-1 pronašla sam 10 strukturnih varijacija, pri čemu je 9 kategorizirano kao delecija, jedna insercija, a promjene zahvaćaju između 53 i 3196 parova baza. Kod pacijenata ST-2 i ST-4 pronašla sam po 4 strukturne varijacije koje zahvaćaju između 52 i 4898 parova baza, a sve su delecije. U slučaju pacijenta ST-3 pronašla sam 4 strukturne varijacije, od kojih se jedna kategorizira kao insercija, a ostale su delecije te zahvaćaju između 69 i 3197 parova baza. Ukupno sam izdvojila 43 strukturnih varijacija unutar 18 gena kod 7 pacijenata. U prilogu 3. prikazane su strukturne varijacije veće od 800 parova baza, uz broj gena zahvaćenih promjenom. Velike strukturne varijacije uključuju od 1 do 1217 gena, a ukupno sam pronašla 84 strukturne varijacije koje zahvaćaju tjelesne i spolne kromosome.

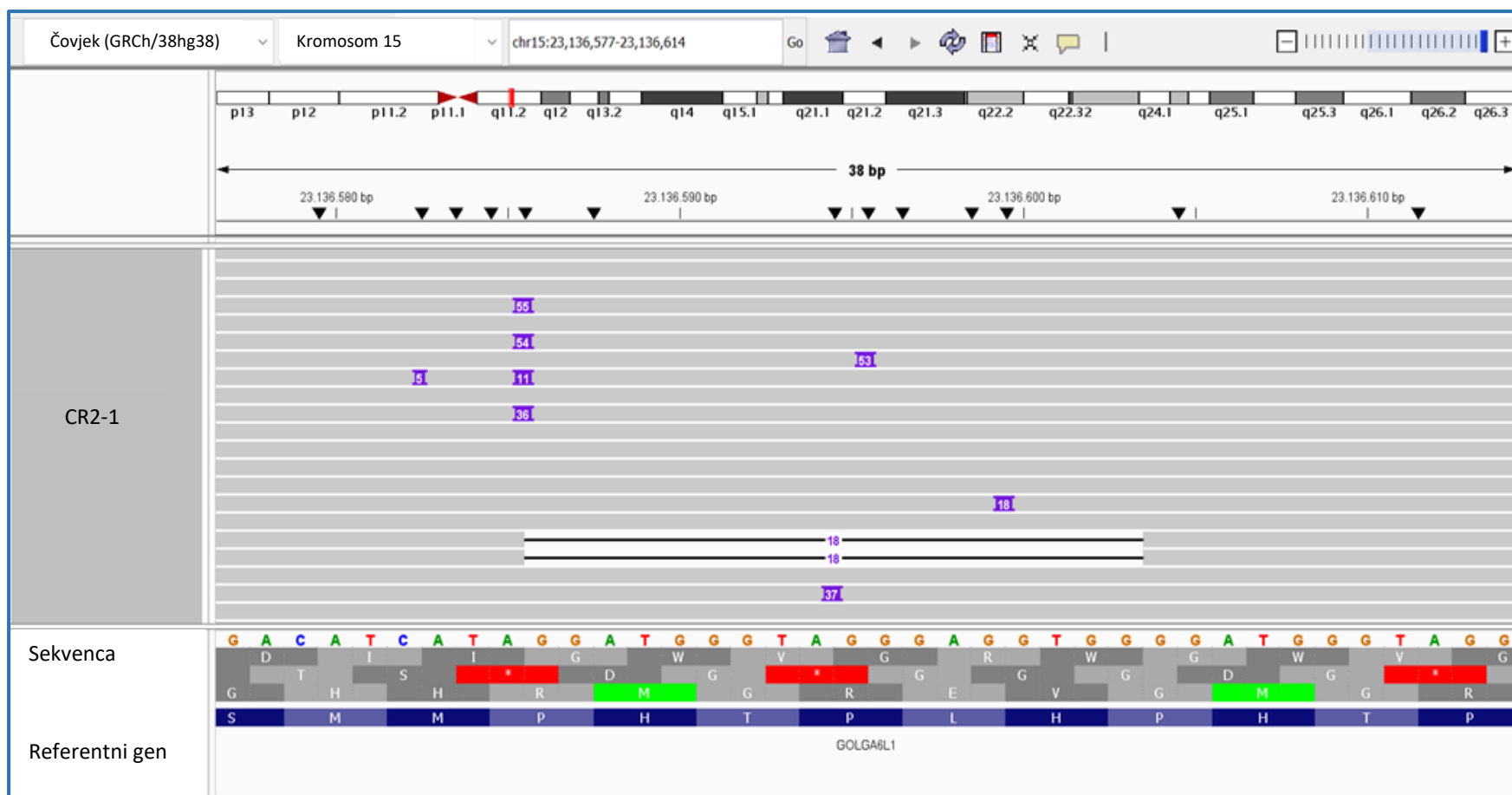
4.2. Vizualizacija strukturnih varijacija

Unosom koordinata strukturnih varijacija u program *Integrative Genomics Viewer* utvrdila sam stvarne strukturne aberacije, što je ostvarivo jedino dobivanjem vizualne informacije o pojedinoj SV. Slike 2-18 predstavljaju vizualizaciju sekvenci izabranih genomskih regija pojedinog pacijenta upotrebom programa IGV. Crne linije prikazuju delecije unutar ciljanih gena s brojem zahvaćenih baza, dok plavi pravokutnici prikazuju insercije s brojem umetnutih baza. Gornji dio slika predstavlja kromosom sa svijetlim i tamnim prugama uz crveno označen položaj strukturne varijacije. Sivo označeni stupci predstavljaju pokrivenost (engl. *coverage*) dok svijetlo sivi redci prikazuju poravnanja sekvenci, a crni trokutići razdvojena poravnanja (engl. *split alignments*).

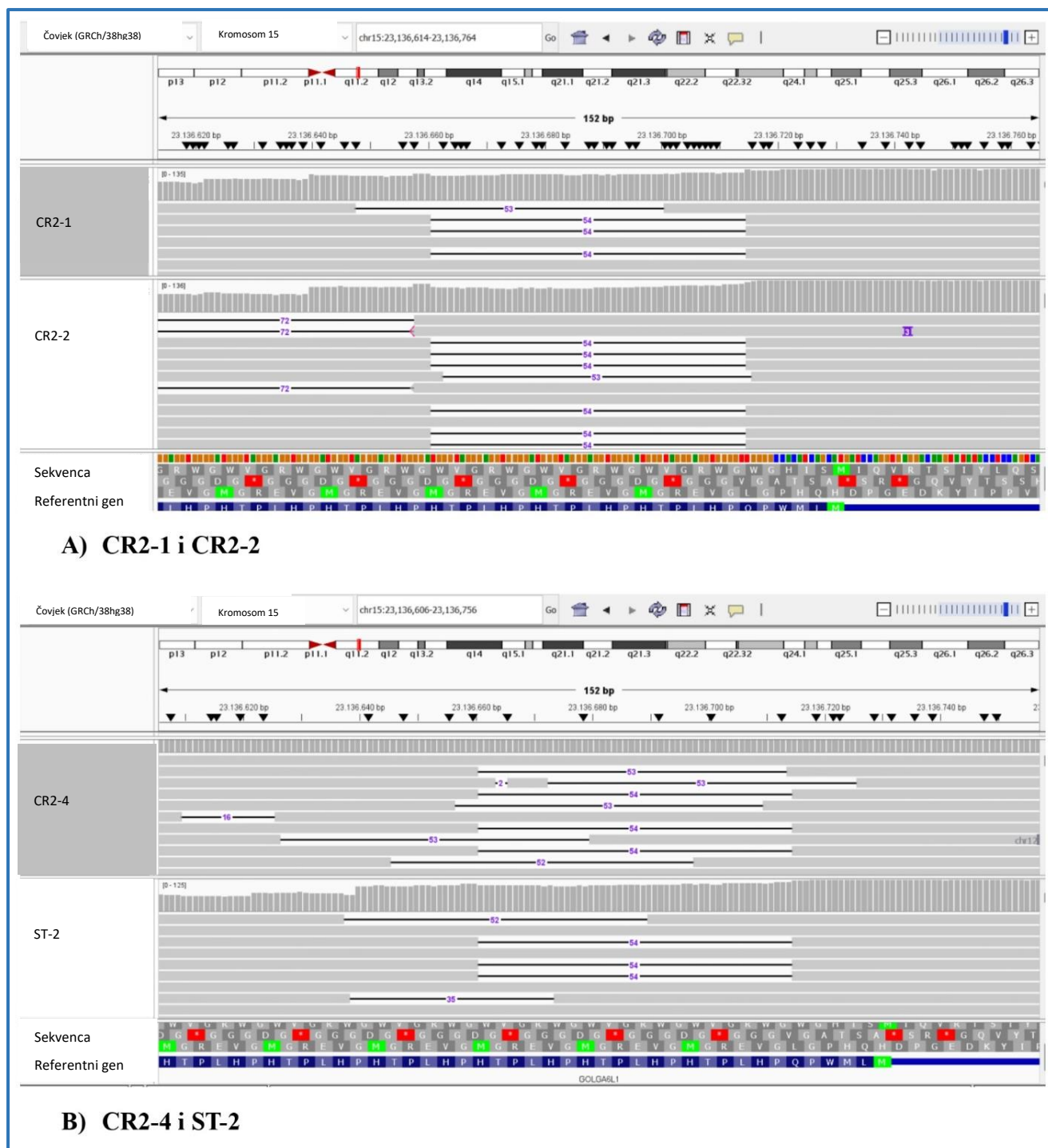
U donjem dijelu slika 2 do 18 prikazan je referentni gen sa pripadajućim eksonima. U gornjem lijevom kutu prikazana je inačica genoma: čovjek (GRCh/38hg38). Kod pacijenta CR2-1 pronašla sam ukupno 5 značajnih strukturnih varijacija, od kojih su one jedinstvene pacijentu CR2-1 prikazane na Slici 2, 3 i 4. Varijacije koje se potpuno ili djelomično preklapaju kod više pacijenata označene su posebno. Značajne SV izdvojene su temeljem visoke pokrivenosti i kontinuiranih poravnanja učitanih ulaznih podataka za svakog pacijenta.



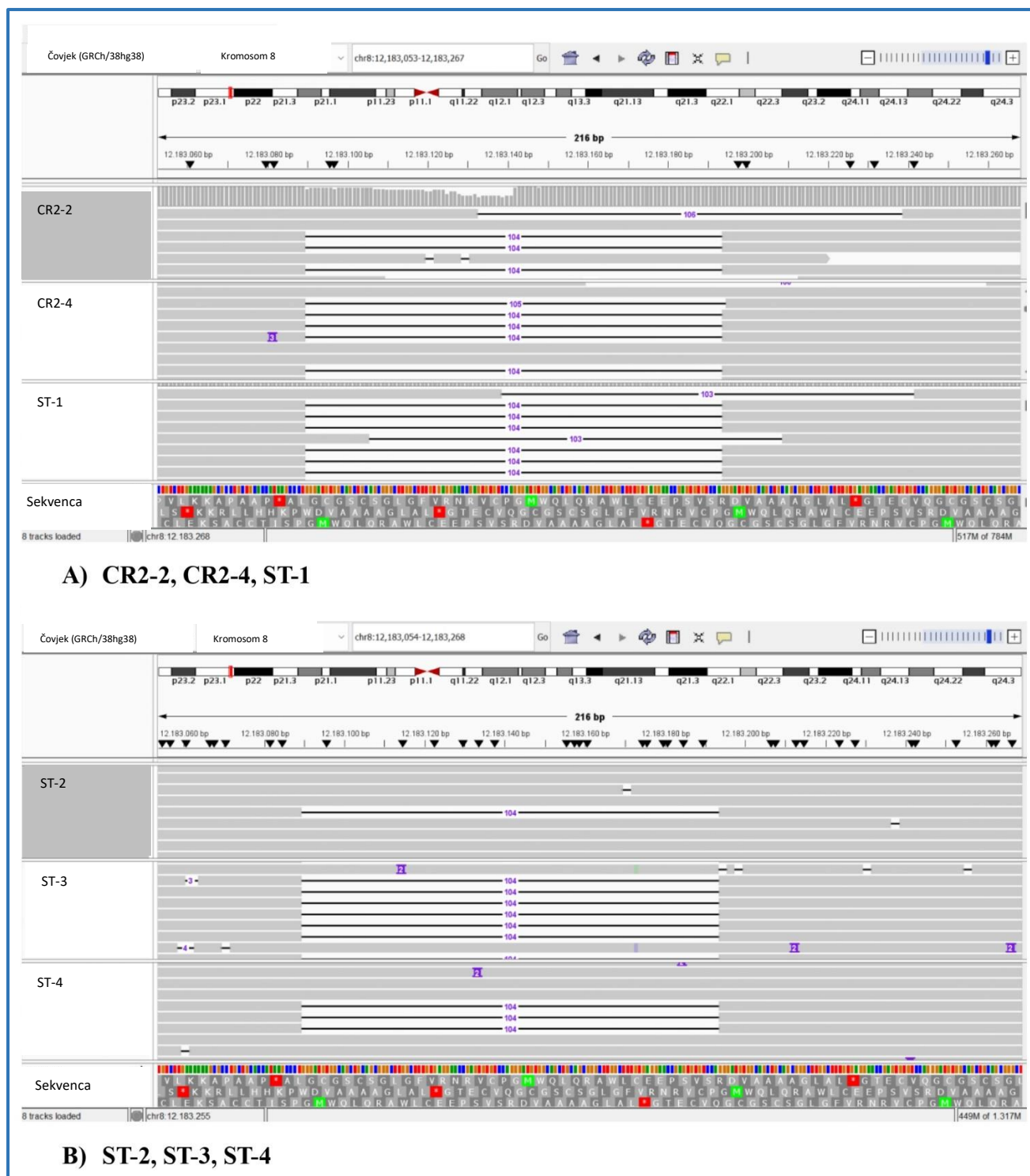
Slika 2. Vizualni prikaz strukturne varijacije kod pacijenta CR2-1 upotrebom programa *Integrative Genomics Viewer*. Slika prikazuje deleciju (crne linije) na kromosomu 8 veličine 104 parova baza u genu *FAM86B1*. Sivi redci predstavljaju poravnanja sekvenci, a crni trokutići razdvojena poravnanja. U gornjem dijelu slike prikazan je kromosom sa svijetlim i tamnim prugama uz crveno označen položaj strukturne varijacije.



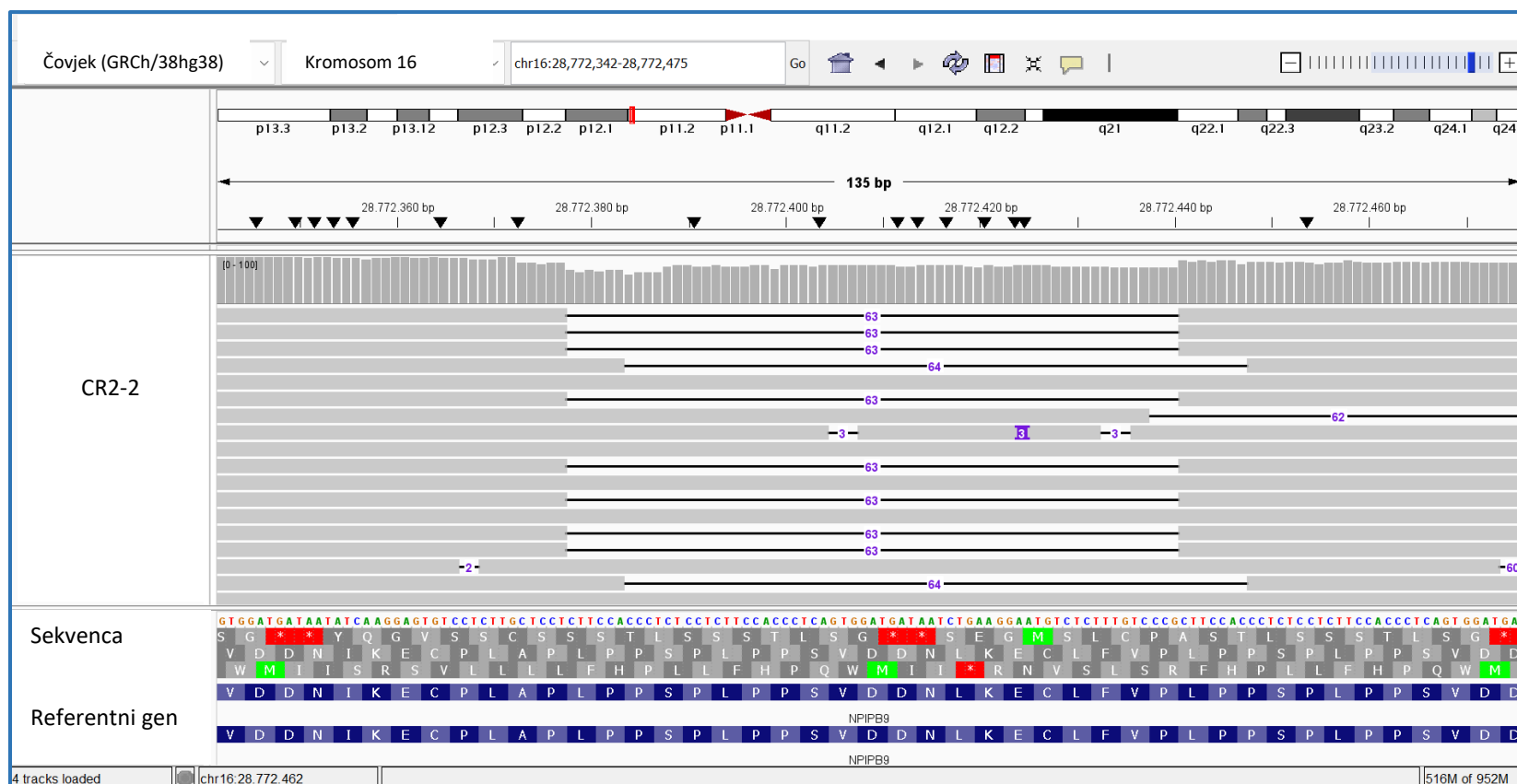
Slika 3. Vizualni prikaz strukturne varijacije kod pacijenta CR2-1 upotrebom programa *Integrative Genomics Viewer*. Slika prikazuje inserciju (plavi pravokutnici) na kromosomu 15 veličine 54 parova baza u genu *GOLGA6L*. Sivi redci predstavljaju poravnanja sekvenci, a crni trokutići razdvojena poravnanja. U gornjem dijelu slike prikazan je kromosom sa svijetlim i tamnim prugama uz crveno označen položaj strukturne varijacije.



Slika 5. Vizualni prikaz strukturne varijacije na kromosomu 15 zajedničke pacijentima CR2-1, CR2-2 (A), CR2-4 i ST-2 (B). Slika prikazuje deleciju (crne linije) veličine 54 parova baza koja zahvaća gen *GOLGA6L1* unutar eksona 1 (NM_001001413.3). Sivi redci predstavljaju poravnanja sekvenci, crni trokutići razdvojena poravnanja, a svijetlo sivi stupci dubinu pokrivenosti. U gornjem dijelu slike a i b prikazan je kromosom sa svijetlim i tamnim prugama uz crveno označen položaj strukturne varijacije.



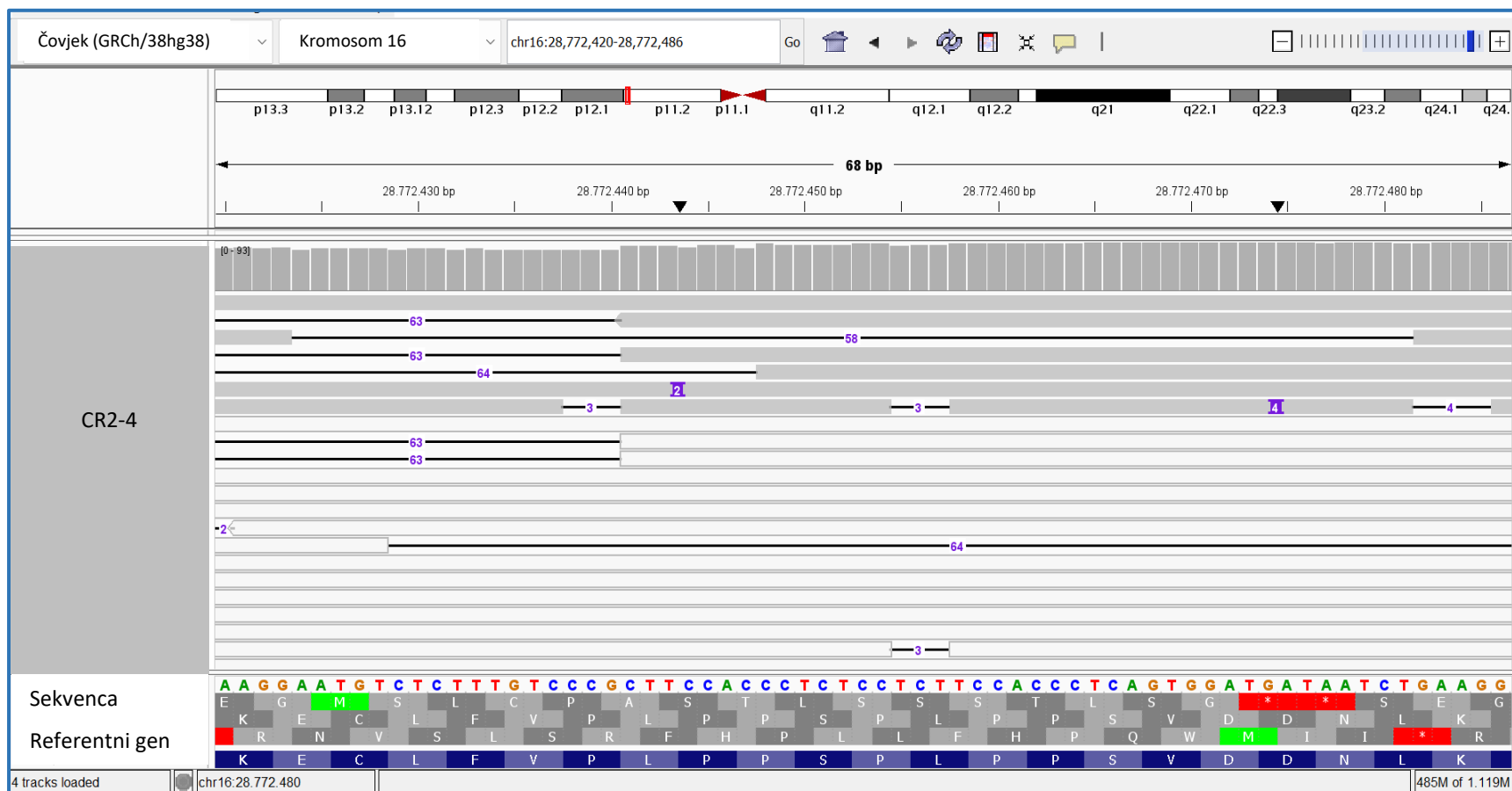
Slika 6. Vizualni prikaz strukturne varijacije upotrebom programa IGV na kromosomu 8 zajedničke pacijentima CR2-2, CR2-4, ST-1 (A), ST-2, ST-3 i ST-4 (B). Slika prikazuje deleciju (crne linije) veličine 104 parova baza koja zahvaća gen *FAM86B1* unutar eksona 6 (NR_003494.3). Sivi redci predstavljaju poravnanja sekvenci, crni trokutići razdvojena poravnanja, a svijetlo sivi stupci dubinu pokrivenosti.



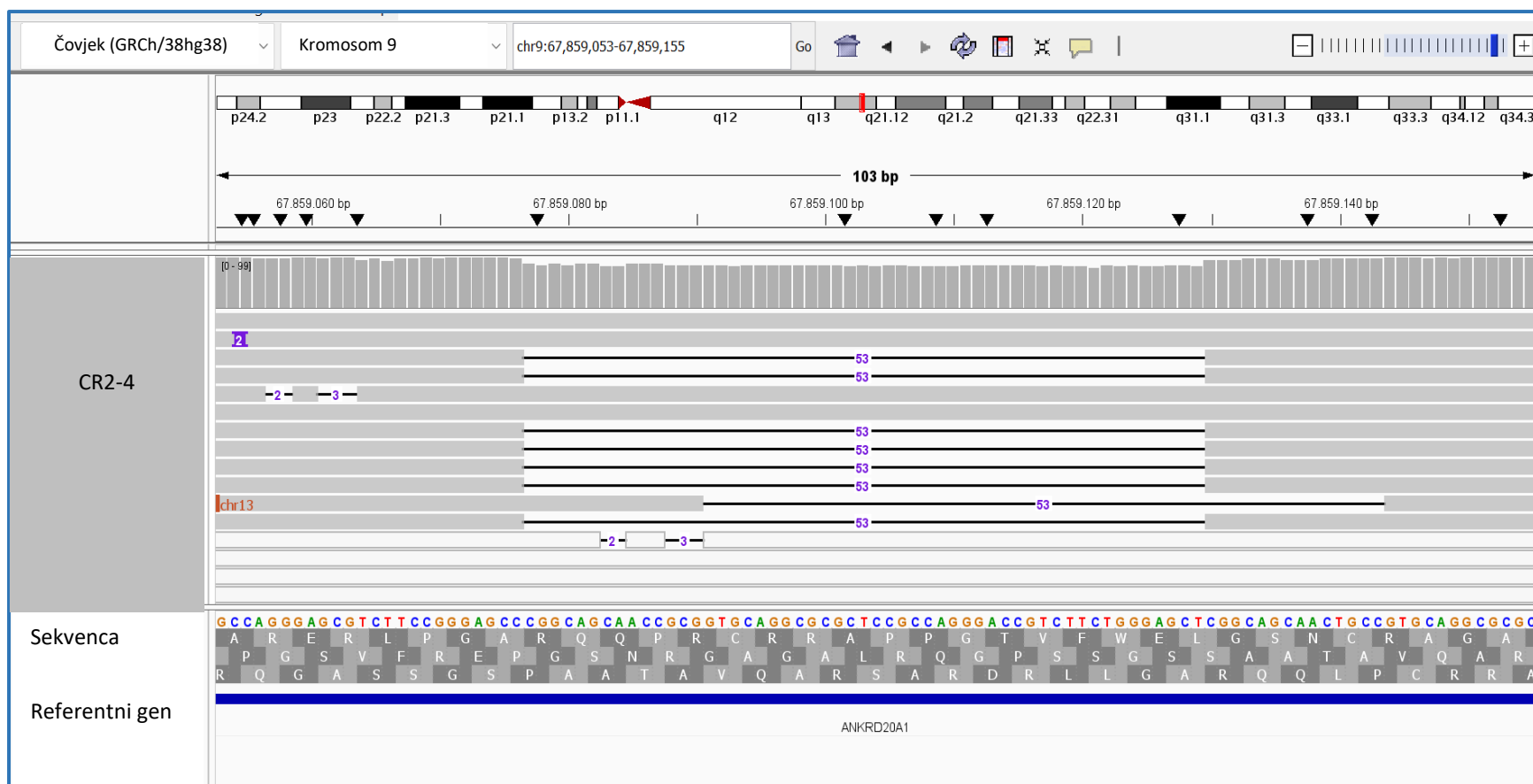
Slika 7. Vizualni prikaz strukturne varijacije kod pacijenta CR2-2 upotrebom programa *Integrative Genomics Viewer*. Slika prikazuje deleciju (crne linije) na kromosomu 16 veličine 63 parova baza u genu *NPIP9*. Sivi redci predstavljaju poravnanja sekvenci, crni trokutići razdvojena poravnanja, a svijetlo sivi stupci dubinu pokrivenosti. U gornjem dijelu slike prikazan je kromosom sa svijetlim i tamnim prugama uz crveno označen položaj strukturne varijacije.

Kod pacijenta CR2-4 izdvojila sam 6 značajnih strukturnih varijacija, od kojih je delecija na kromosomu 15 prikazana na Slici 5, a delecija na kromosomu 8 na Slici 6. Strukturne aberacije svojstvene samo pacijentu CR2-4 prikazane su na Slici 8 i 9. Na kromosomu 1 pacijenta CR2-4, ST-1, ST-2 i ST-3 izdvojila sam zajedničku aberaciju, prikazanu na Slici 10.

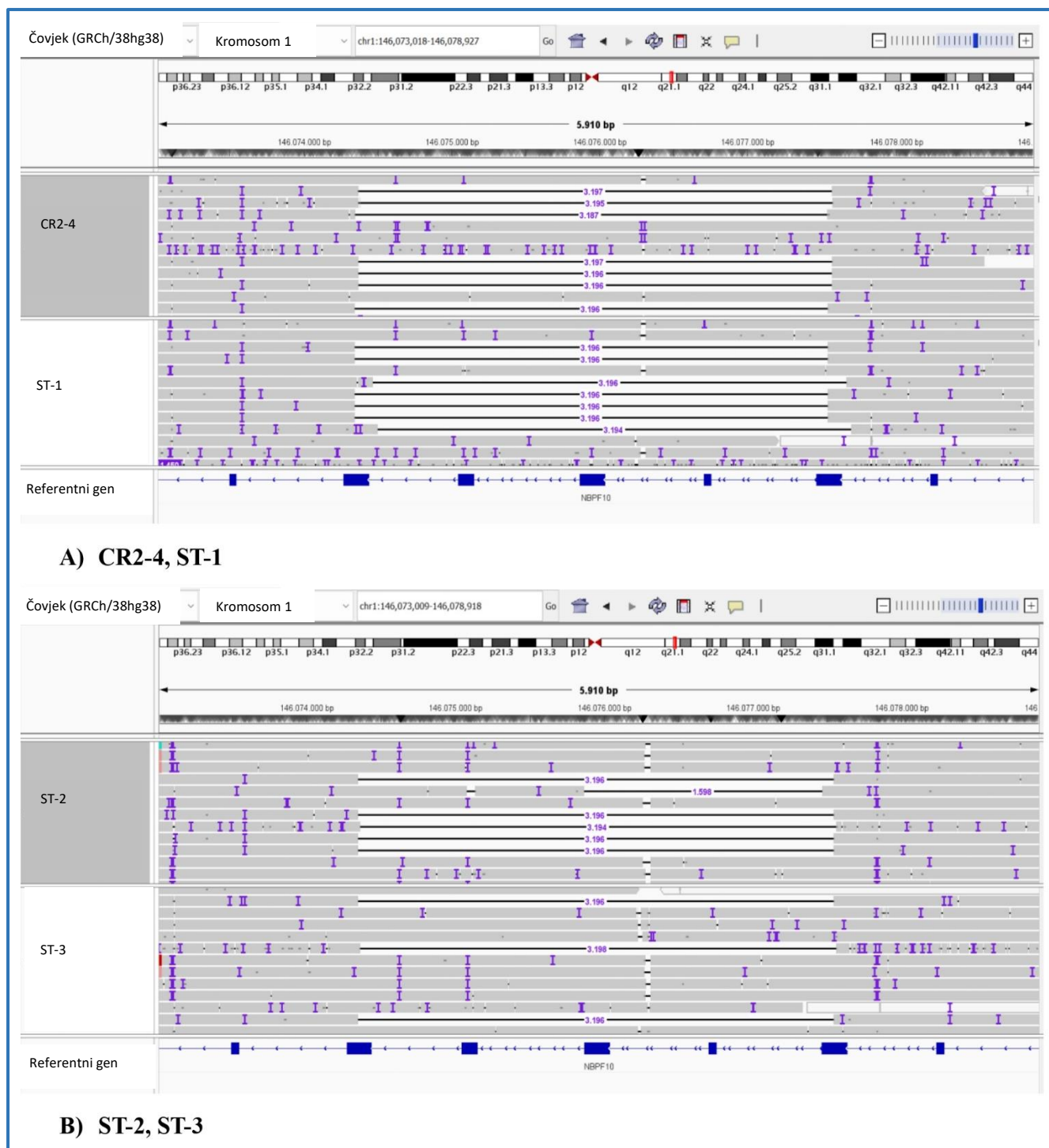
Uz to, pacijenti CR2-4 i ST-1 imaju istu aberaciju u genu NBPF10 na kromosomu 1, prikazanu na Slici 11. Ukupno, pacijent CR2-4 ima dvije velike strukturne varijacije, koje zahvaćaju više od 3000 parova baza.



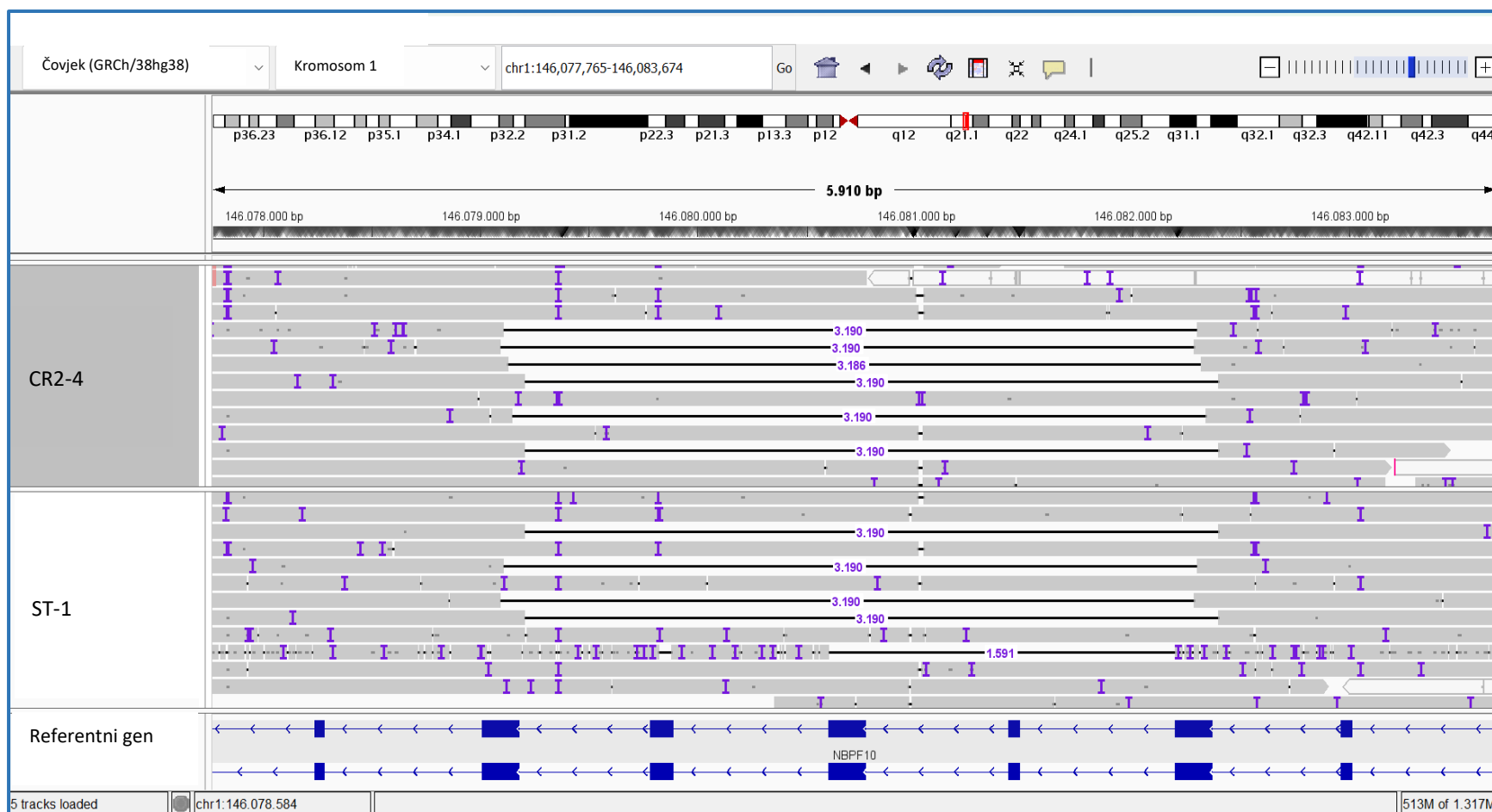
Slika 8. Vizualni prikaz strukturne aberacije kod pacijenta CR2-4 upotrebom programa *Integrative Genomics Viewer*. Slika prikazuje deleciju (crne linije) na kromosomu 16 veličine 58 parova baza u genu *NIPB9*. Sivi redci predstavljaju poravnanja sekvenci, crni trokutići razdvojena poravnanja, a svijetlo sivi stupci dubinu pokrivenosti. U gornjem dijelu slike prikazan je kromosom sa svijetlim i tamnim prugama uz crveno označen položaj strukturne varijacije.



Slika 9. Vizualni prikaz strukturne aberacije kod pacijenta CR2-4 upotrebom programa *Integrative Genomics Viewer*. Slika prikazuje deleciju (crne linije) na kromosomu 9 veličine 53 parova baza u genu *ANKRD20A1*. Sivi redci predstavljaju poravnanja sekvenci, crni trokutići razdvojena poravnanja, a svijetlo sivi stupci dubinu pokrivenosti. U gornjem dijelu slike prikazan je kromosom sa svijetlim i tamnim prugama uz crveno označen položaj strukturne varijacije.



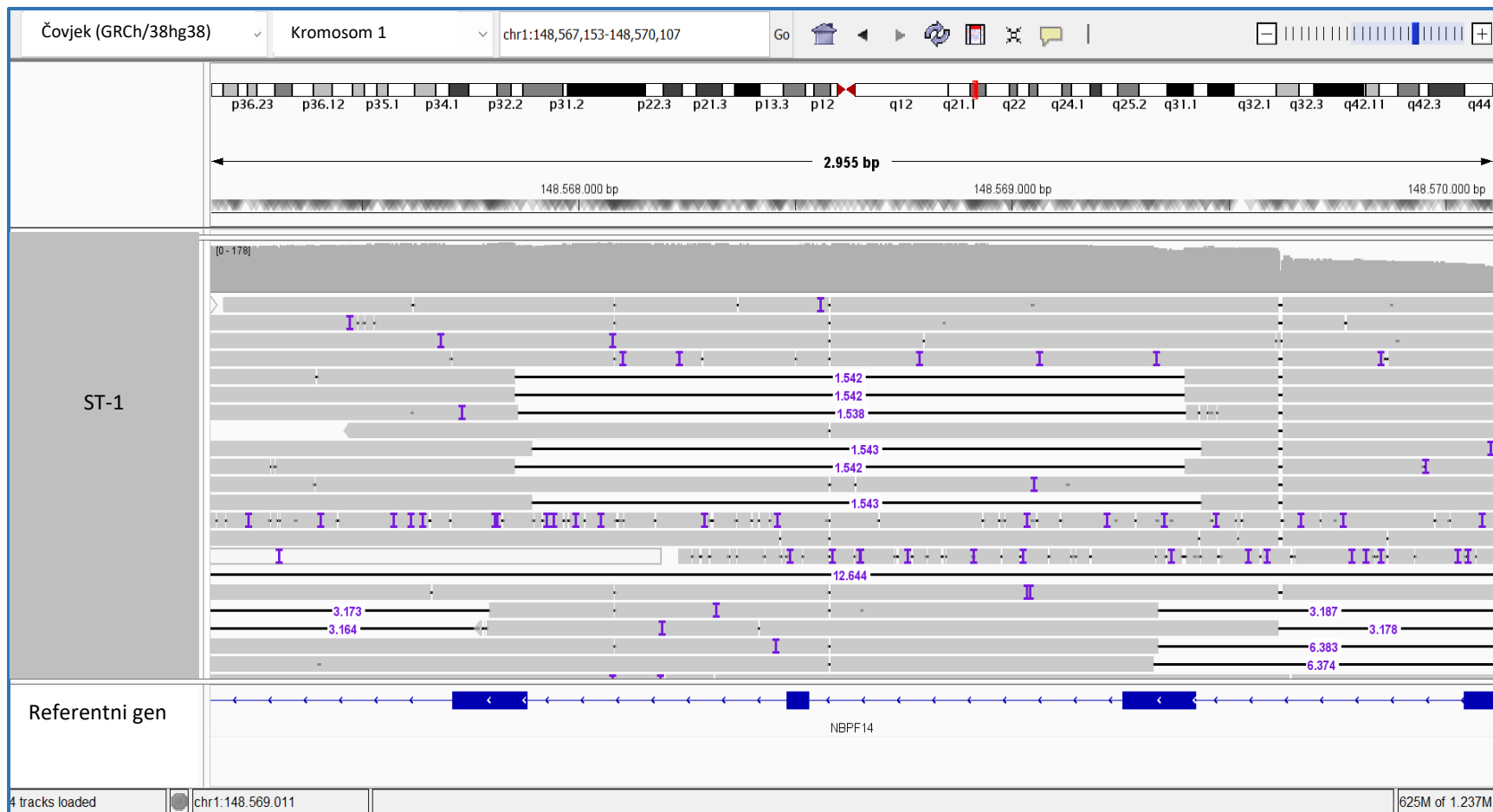
Slika 10. Vizualni prikaz velike strukturne aberacije upotrebom programa IGV na kromosomu 1 zajedničke pacijentima CR2-4, ST-1 (A), ST-2 i ST-3 (B). Slika prikazuje deleciju (crne linije) veličine ~ 3195 parova baza koja zahvaća gen *NBP10* uključujući eksone 76-80. Sivi redci predstavljaju poravnanja sekvenci. U donjem dijelu slike prikazan je referentni gen s uključenim eksonima.



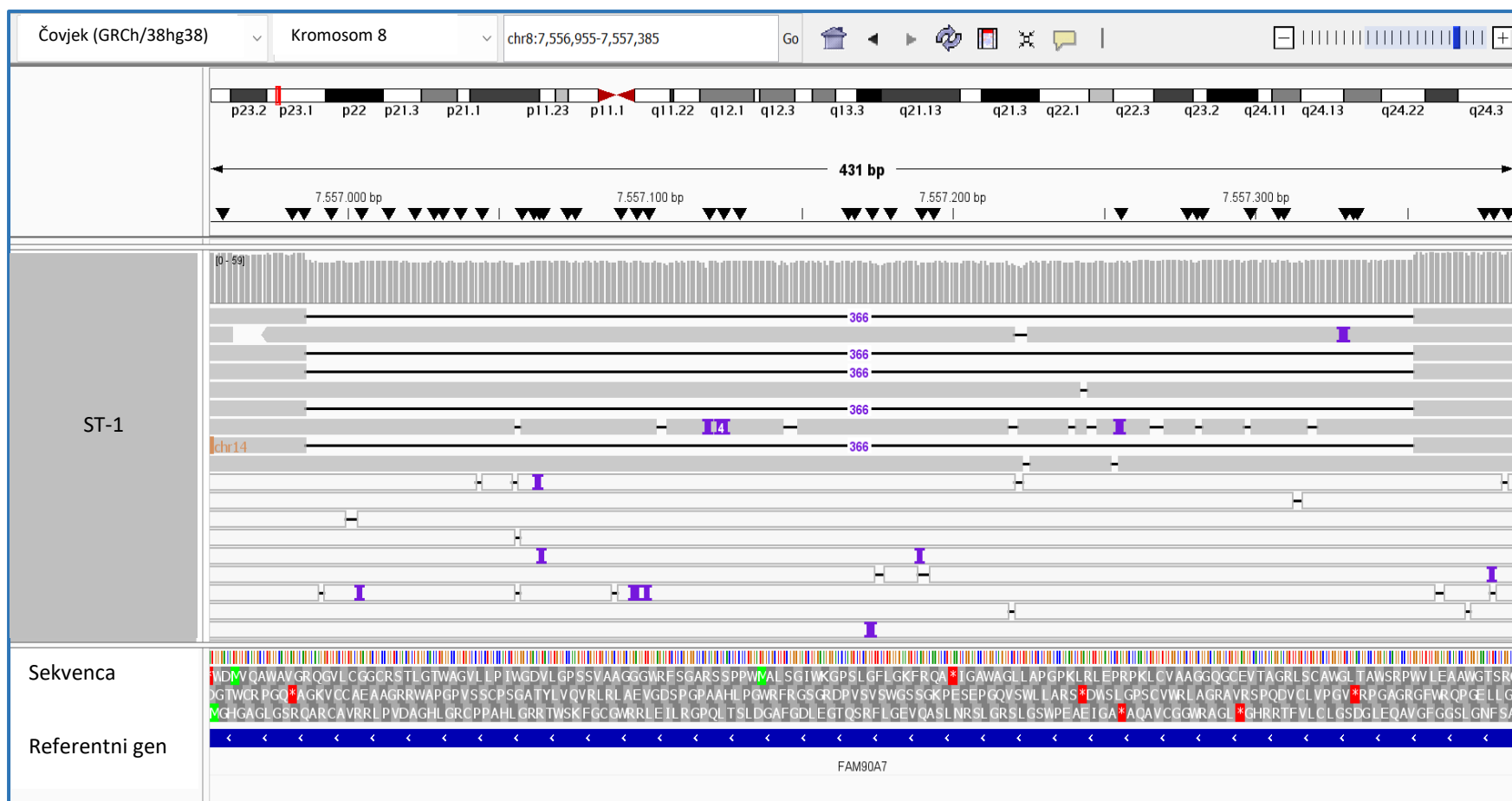
Slika 11. Vizualni prikaz strukturne aberacije upotrebom programa IGV na kromosomu 1 zajedničke pacijentima CR2-4 i ST-1. Slika prikazuje deleciju (crne linije) veličine ~ 3190 parova baza koja zahvaća eksone 70-74 gena *NBPF10*. Sivi redci predstavljaju poravnanja sekvenci.

Rezultati istraživanja sekvenci pacijenta ST-1 pokazuju prisutnost 9 značajnih strukturnih varijacija, od kojih 5 dijeli s drugim pacijentima (Slika 10, 11, 15, 16, 17). Pacijent ST-1 pokazuje značajno narušenu strukturu kromosoma 1, gdje se nalaze čak 4 od 9 pronađenih aberacija. Uz to, na kromosomima 8 i 15 nalaze se po dvije strukturne varijacije. One aberacije jedinstvene pacijentu ST-1

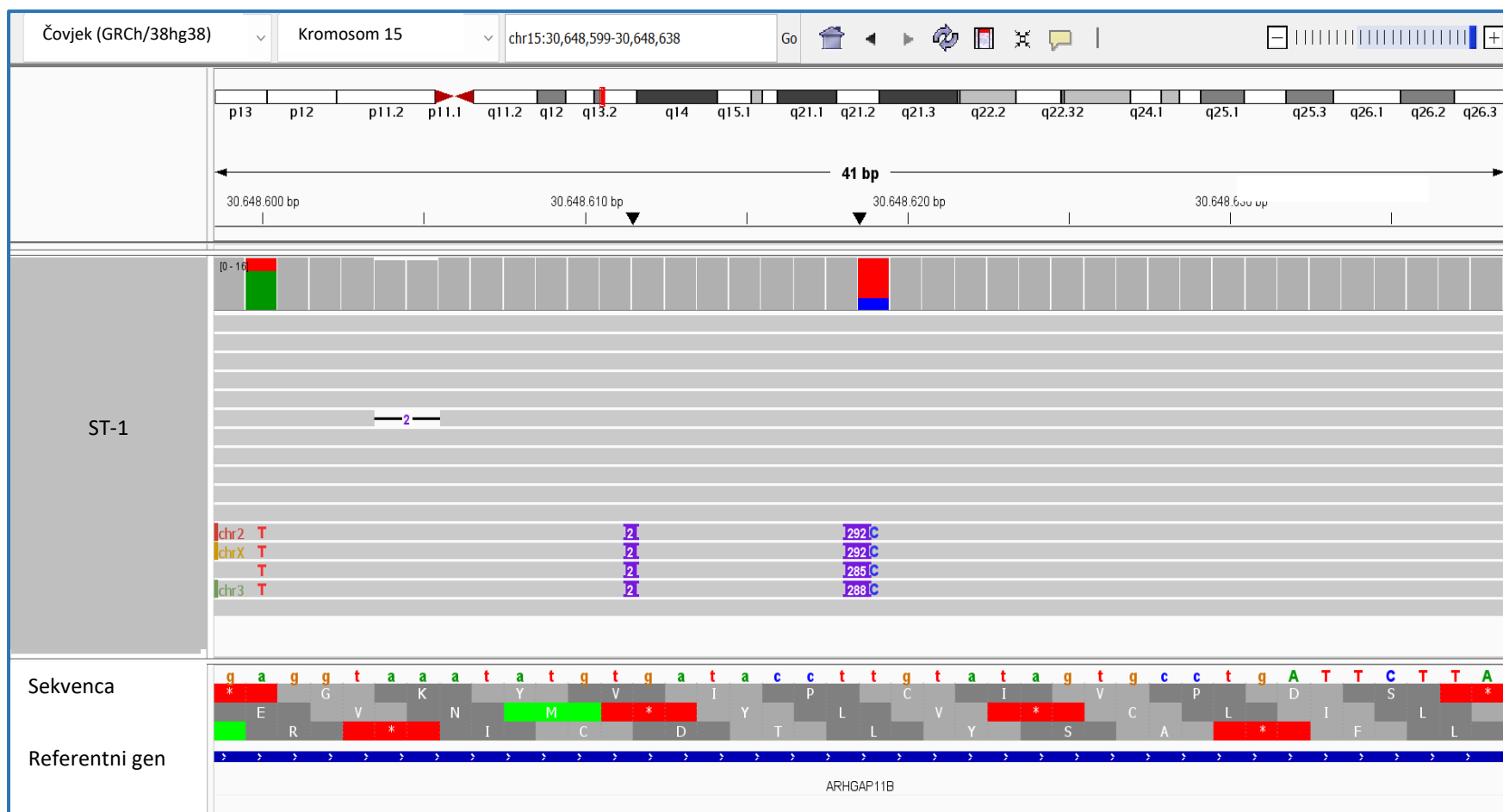
prikazane su na Slikama 12, 13 i 14. Na kromosomu 15 našla sam deleciju veličine 69 parova baza zajedničku pacijentima ST-1 i ST-3 (Slika 15), dok se na kromosomu 22 nalazi delecija koju imaju pacijenti ST-1, ST-4 i CR2-1 (Slika 16). Uz navedeno, velika aberaciju na kromosomu 1 zajednička je pacijentima ST-1 i ST-2, kako je prikazano na Slici 17.



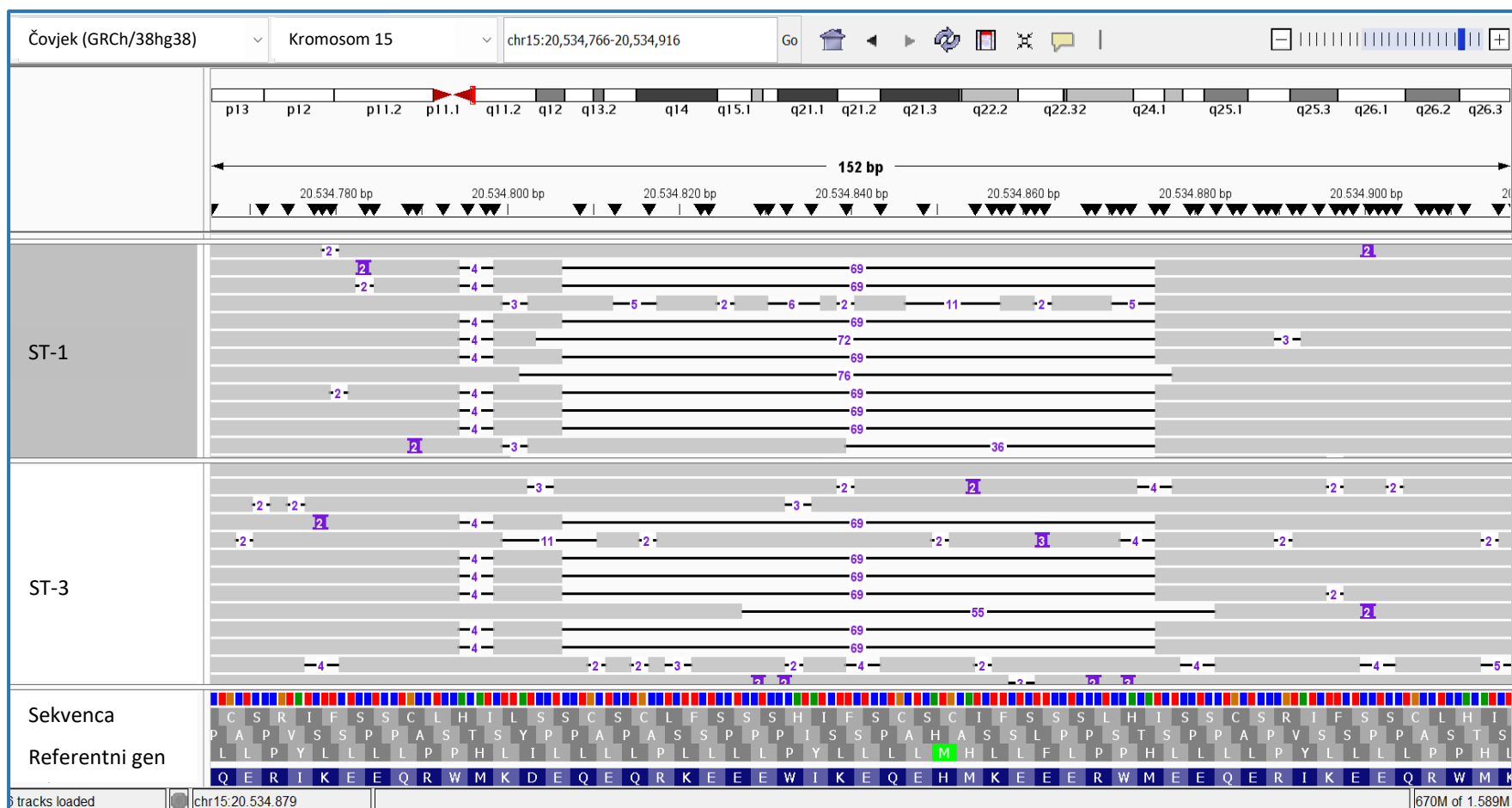
Slika 12. Vizualni prikaz velike strukturne aberacije na kromosomu 1 pacijenta ST-1 upotrebom programa *Integrative Genomics Viewer*. Slika prikazuje deleciju (crne linije) veličine 1542 parova baza koja zahvaća eksone 25-27 gena *NBPF14*. Sivi redci predstavljaju poravnanja sekvenci.



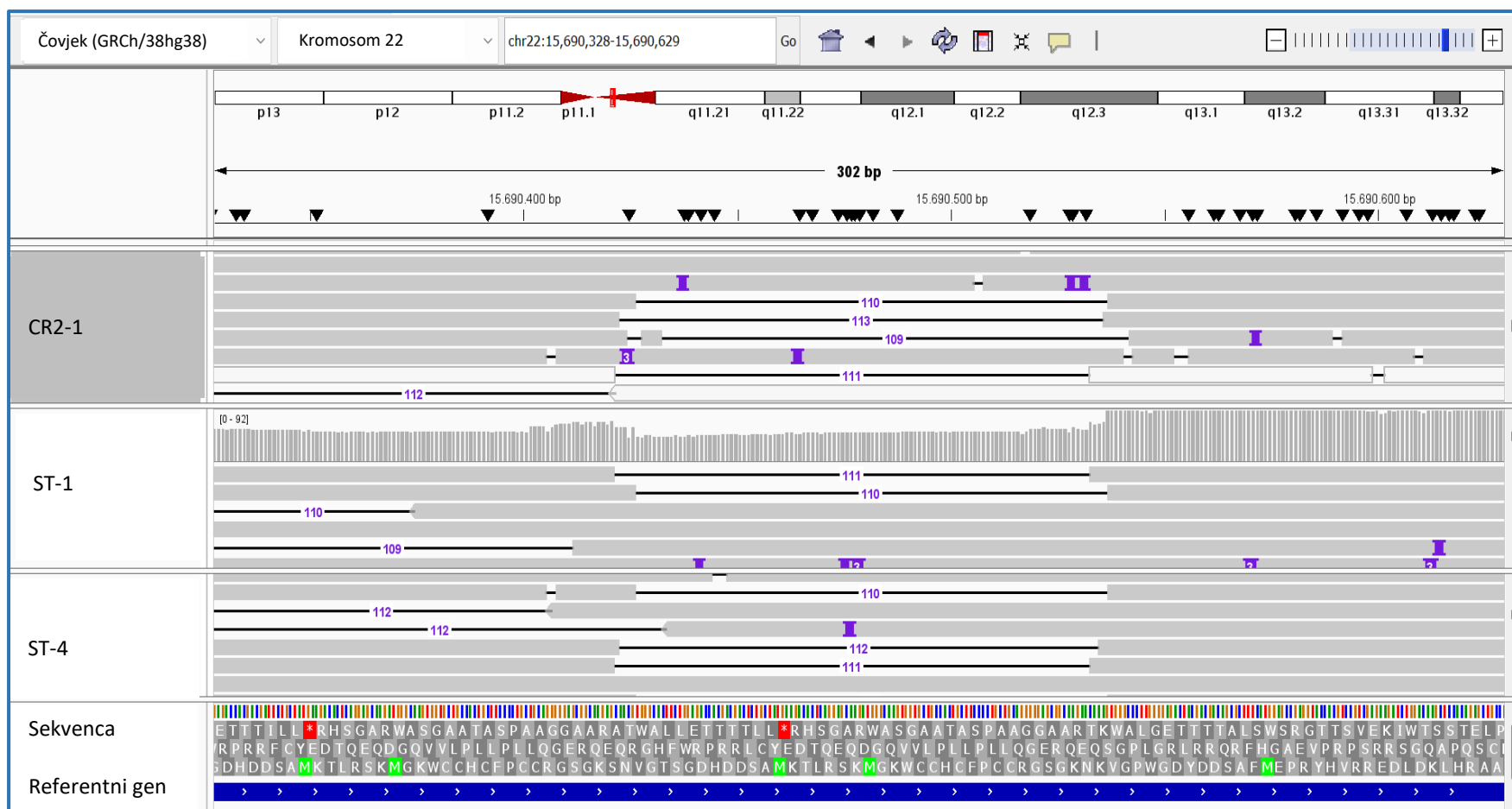
Slika 13. Vizualni prikaz strukturne aberacije na kromosomu 8 pacijenta ST-1 upotrebom programa *Integrative Genomics Viewer*. Slika prikazuje deleciju (crne linije) veličine 366 parova baza unutar gena *FAM90A7*. Sivi redci predstavljaju poravnanja sekvenci, crni trokutići razdvojena poravnanja, a svijetlo sivi stupci dubinu pokrivenosti. U gornjem dijelu slike prikazan je kromosom sa svijetlim i tamnim prugama uz crveno označen položaj strukturne varijacije.



Slika 14. Vizualni prikaz strukturne varijacije na kromosomu 15 pacijenta ST-1 upotrebom programa *Integrative Genomics Viewer*. Slika prikazuje inserciju (plavi pravokutnik) veličine 292 parova baza koja zahvaća gen *ARHGAP11B*. Sivi redci predstavljaju poravnanja sekvenci, crni trokutići razdvojena poravnanja, a svijetlo sivi stupci dubinu pokrivenosti. U gornjem dijelu slike prikazan je kromosom sa svijetlim i tamnim prugama uz crveno označen položaj strukturne varijacije.

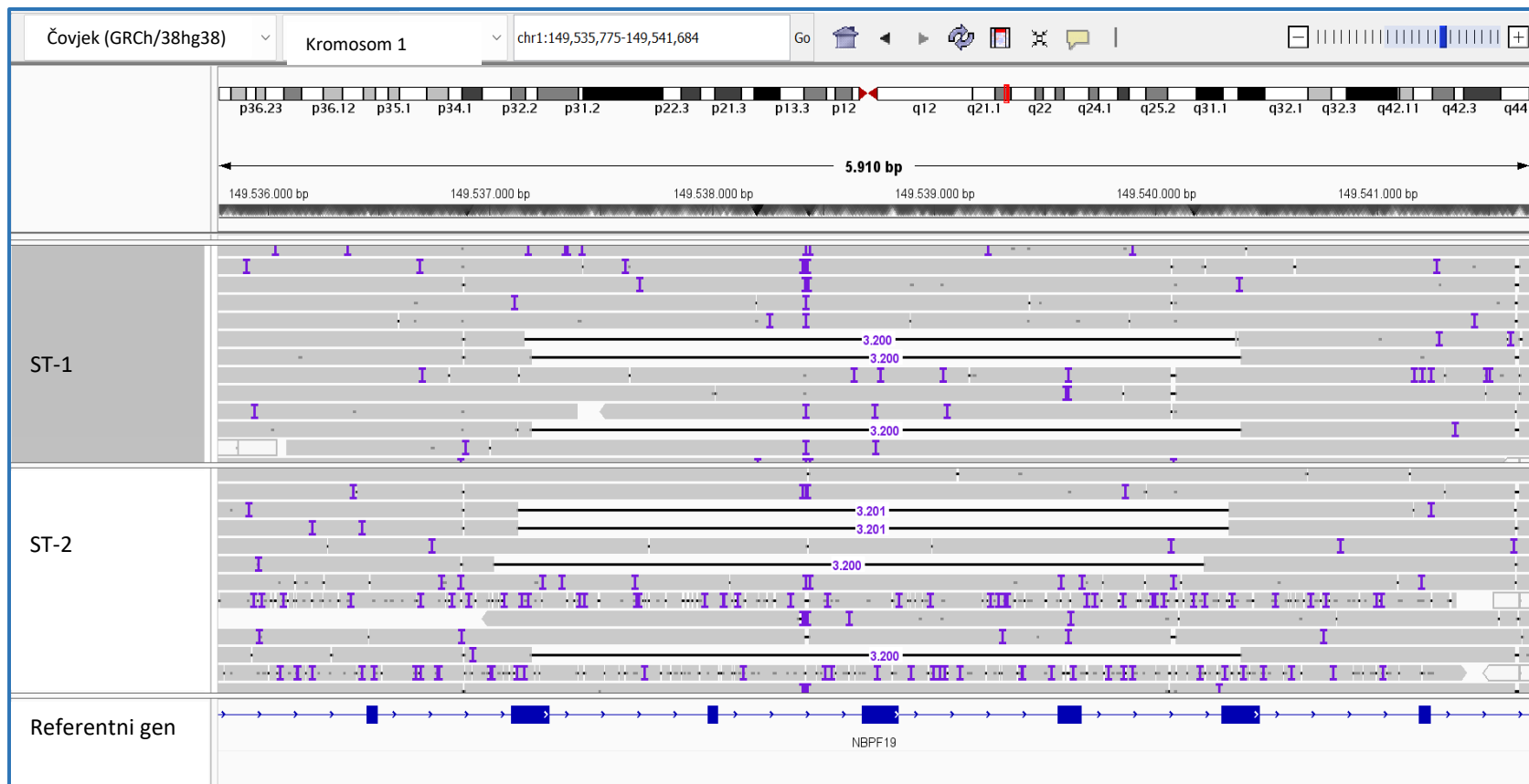


Slika 15. Vizualni prikaz strukturne varijacije na kromosomu 15 zajedničke pacijentima ST-1 i ST-3 upotrebom programa *Integrative Genomics Viewer*. Slika prikazuje deleciju (crne linije) veličine 69 parova baza koja zahvaća ekson 8 (NM_001145004.2) gena *GOLGA6L6*. Sivi redci predstavljaju poravnanja sekvenci, a crni trokutići razdvojena poravnanja. U gornjem dijelu slike prikazan je kromosom sa svijetlim i tamnim prugama uz crveno označen položaj strukturne varijacije.



Slika 16. Vizualni prikaz strukturne varijacije na kromosomu 22 zajedničke pacijentima CR2-1, ST-1 i ST-4 dobiven upotrebom programa *Integrative Genomics Viewer*. Slika prikazuje deleciju (crne linije) veličine ~ 110 parova baza koja zahvaća ekson 1 (NM_001136213.1) gena *POTEH*. Sivi redci predstavljaju poravnanja sekvenci, crni trokutići razdvojena poravnanja, a svijetlo sivi stupci dubinu pokrivenosti. U gornjem dijelu slike prikazan je kromosom sa svijetlim i tamnim prugama uz crveno označen položaj strukturne varijacije.

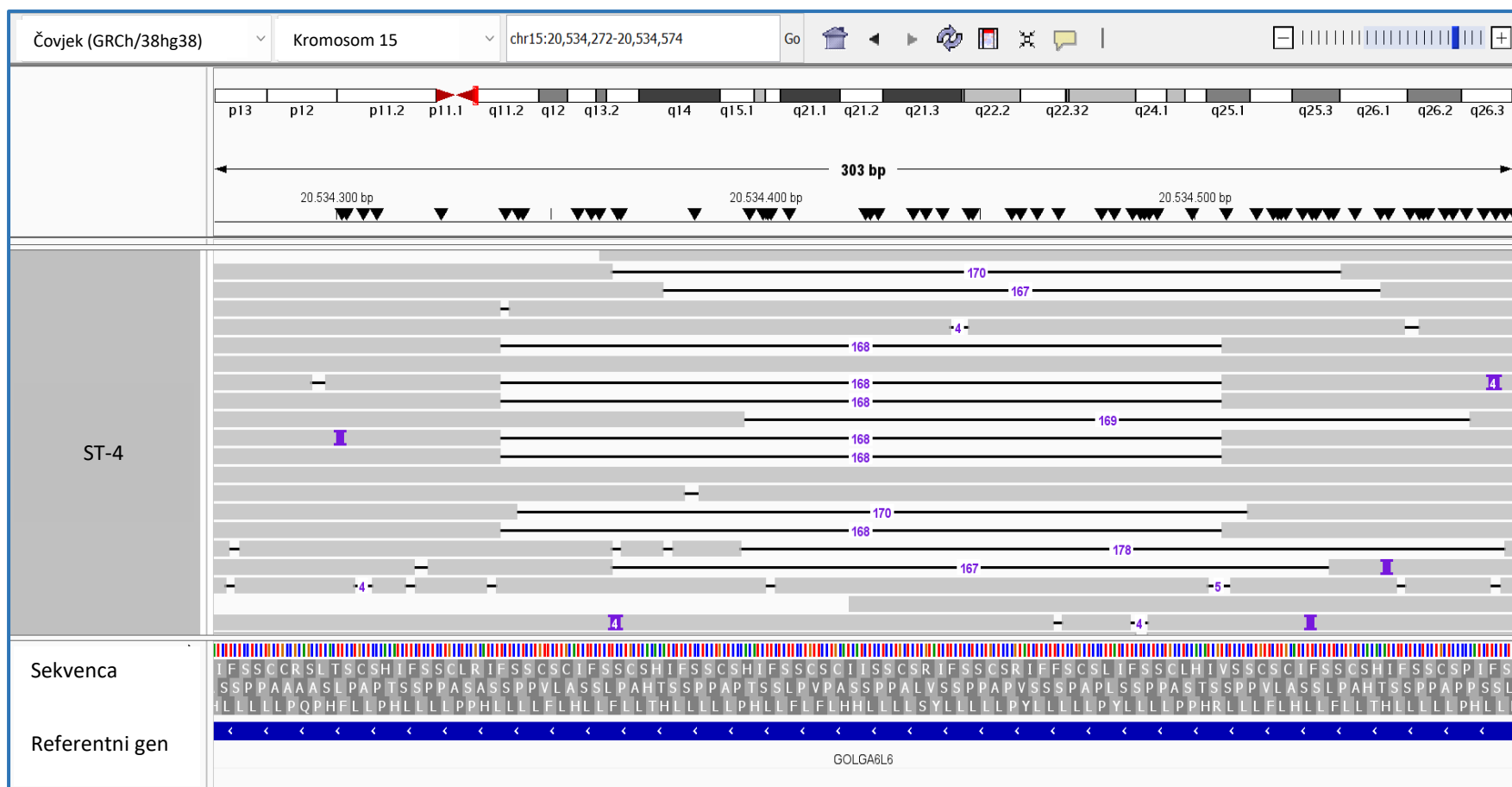
Pacijent ST-2 ima 4 značajne strukturne varijacije, od kojih je delecija na kromosomu 1 veličine 3196 parova baza prikazana na Slici 10, delecija na kromosomu 15 veličine 52 pb na Slici 5, a delecija na kromosomu 8, veličine 104 parova baza, prikazana je na Slici 6. Velika aberacija zajednička pacijentima ST-1 i ST-2 prikazana je na Slici 17.



Slika 17. Vizualni prikaz strukturne varijacije na kromosomu 1 zajedničke pacijentima ST-1 i ST-2 dobiven upotrebom programa *Integrative Genomics Viewer*. Slika prikazuje veliku deleciju (crne linije) veličine ~ 3200 parova baza koja zahvaća eksone 72-76 gena *NBPF19*. Sivi redci predstavljaju poravnanja sekvenci.

Vizualizacijom sekvenci pacijenta ST-3 pronašla sam 3 značajne strukturne varijacije, od kojih je delecija na kromosomu 1 unutar gena *NBPF10* zajednička pacijentima CR2-4, ST-1, ST-2 i ST-3 (Slika 10). Deleciju od 69 parova baza na kromosomu 15 imaju pacijenti ST-1 i ST-3 (Slika 15). Najučestalija strukturna varijacija, prisutna kod 6 od 7 ispitanika, uključujući CR2-2, CR2-4, ST-1, ST-2, ST-3 i ST-4, jest delecija na kromosomu 8 koja zahvaća 104 bazna para u genu *FAM86B1* (Slika 6).

Analizom podataka u programu *Integrative Genomics Viewer* u slučaju pacijenta ST-4 izdvojila sam 3 značajne strukturne varijacije, među kojima se nalazi delecija na kromosomu 22 veličine 111 parova baza unutar gena *POTEH*, zajednička pacijentima CR2-1, ST-1 i ST-4, prikazana na Slici 16. Uz to, pacijent ST-4 ima najprisutniju deleciju unutar ispitivane skupine, smještene na kromosomu 8 veličine 104 parova baza unutar gena *FAM86B1*, kako je prikazano na Slici 6. Strukturna varijacija u obliku delecije na kromosomu 15 jedinstvena pacijentu ST-4 prikazana je na Slici 18.



Slika 18. Vizualni prikaz strukturne varijacije na kromosomu 15 pacijenta ST-4 upotrebom programa *Integrative Genomics Viewer*. Slika prikazuje deleciju (crne linije) veličine 168 parova baza koja zahvaća gen *GOLGA6L6*. Sivi redci predstavljaju poravnanja sekvenci, a crni trokutići razdvojena poravnanja. U gornjem dijelu slike prikazan je kromosom sa svijetlim i tamnim prugama uz crveno označen položaj strukturne varijacije.

Strukturne aberacije 7 pacijenata zahvaćaju ukupno 18 gena. Među učestalim aberacijama ističu se delecija na kromosomu 8 veličine 104 parova baza unutar gena *FAM86B1* prisutna kod 6 od 7 ispitanika (Slika 6), delecija na kromosomu 15 veličine 54 pb unutar gena *GOLGA6L1* prisutna u 4 od 7 pacijenata (CR2-1, CR2-2, CR2-4 i ST-1) prikazana na Slici 5, delecija od ~ 3195 parova baza unutar gena *NBPF10* na kromosomu 1 (Slika 10) prisutna kod ispitanika CR2-4, ST-1, ST-2 i ST-3 te delecija na kromosomu 22 koja zahvaća 111 parova baza u genu *POTEH* pacijenata CR2-1, ST-1 i ST-4 (Slika 16). Ukupno, detektirala sam 15 delecija i 3 insercije.

4.3. Provjera delecija i insercija

Delecije nađene kod pacijenata uključenih u ovo istraživanje, uz njihovu distribuciju unutar opće populacije, izdvojene iz baze *Database of Genomic Variants* navedene su u Tablici 1. U DGV-u je moguće provesti populacijsku statistiku i za dodatnu kategoriju Mješovito koja uključuje: bliskoistočnu populaciju, indijansku, sjevernoameričku, oceanijsku te južnoameričku populaciju. Međutim, ove kategorije su isključene iz tablice jer nema pripadnika tih populacija u istraživanjima koja su uključena u *DGV gold standard* podatke. Za chr1:149537190-149540389 i chr9:67859078-67859130 regije u DGV-u nisu zabilježene delecije.

Tablica 1. Prikaz koordinata pojedine strukturne varijacije (delecije) pronađene kod pacijenata uključenih u istraživanje te njihova distribucija u općoj populaciji.

Početak	Kraj	DGV identifikacijski broj	Rang	Broj studija	Broj uzoraka	Frekvencija	Broj uzoraka iz specifične populacije					
							Afrička	Europska	Azijska	Meksička	Nepoznata	Turska
chr1:146074351-146077546												
145430994	149056513	gssvL5032	3	3	8	0,06%	1	0	0	1	5	0
chr1:146079112-146082316												
145430994	149056513	gssvL5032	3	3	8	0,06%	1	0	0	1	5	0
chr1:148567861-148569402												
145430994	149056513	gssvL5032	3	3	8	0,06%	1	0	0	1	5	0
chr8:7556988-7557353												
7087251	8197726	gssvL121757	12	3	289	2,33%	15	102	2	0	170	0
7137582	7680412	gssvL121761	8	3	13	0,09%	1	0	1	0	11	0
7172455	8078321	gssvL121758	14	2	561	4,40%	0	170	0	0	391	0
7329097	7598872	gssvL121767	7	6	76	0,59%	3	27	3	0	43	0
7399278	7788793	gssvL121762	8	2	19	0,17%	0	4	0	0	14	1
7409830	7598490	gssvL121774	14	3	88	0,69%	2	19	0	0	67	0
7477474	7598872	gssvL121779	7	3	9	0,07%	0	7	0	0	2	0
7545480	7587110	gssvL121814	7	4	10	0,64%	0	8	1	0	1	0
chr8:12183110-12183243												
11913670	12713055	gssvL122091	8	4	28	0,21%	2	8	4	1	13	0
12019685	12195891	gssvL122097	8	2	5	0,45%	4	0	0	0	0	1
12019685	12532643	gssvL122092	16	3	40	0,32%	1	15	1	0	23	0
12130216	12204348	gssvL122107	7	4	51	1,91%	21	28	2	0	0	0
chr15:20534808-20534876												
19985575	22472558	gssvL37539	12	5	1092	7,84%	57	280	2	0	753	0
19998695	20753972	gssvL37544	12	3	31	0,25%	3	7	0	0	21	0
20011689	21498353	gssvL37541	14	3	248	1,79%	8	69	0	0	171	0
20210877	20688382	gssvL37550	16	2	31	0,25%	0	9	0	0	22	0
20327351	20538669	gssvL37559	7	3	79	0,64%	1	38	0	0	40	0
20351176	21104690	gssvL37546	7	3	4	5,88%	3	1	0	0	0	0
20372246	20688370	gssvL37552	14	3	66	0,48%	0	24	0	0	42	0
20466234	21526215	gssvL37543	12	3	14	0,11%	0	4	1	0	9	0
chr15:23136662-23136715												
22358241	23226874	gssvL37542	8	3	52	0,43%	1	17	0	0	34	0
22738857	23226874	gssvL37548	8	4	13	0,32%	2	10	1	0	0	0
23130344	23226024	gssvL37582	12	2	2	0,13%	0	2	0	0	0	0
chr16:28772379-28772482												
28620263	28857362	gssvL43238	13	2	3	0,02%	0	2	0	0	1	0
chr22:15690292-15690402												
15280112	15927980	gssvL76181	7	6	125	1,00%	15	48	4	0	58	0
15518443	15741775	gssvL76184	7	2	4	0,35%	1	2	0	0	1	0
15521146	15922837	gssvL76182	12	3	34	0,27%	6	6	2	0	20	0
chr22:15690428-15690537												
15280112	15927980	gssvL76181	7	6	125	1,00%	15	48	4	0	58	0
15518443	15741775	gssvL76184	7	2	4	0,35%	1	2	0	0	1	0
15521146	15922837	gssvL76182	12	3	34	0,27%	6	6	2	0	20	0
							171	963	30	4	2001	2

U tablici 2. prikazani su rezultati dobiveni pomoću alata UCSC BLAT. Sekvence insercija koristila sam u UCSC BLAT alatu kako bih odredila koordinate u genomu na koje se mapira insercija.

Tablica 2. Prikaz rezultata pretraživanja BLAT za insercije. Na lijevoj strani su naznačeni rezultati sekvenciranja koji su korišteni za poravnavanje u referentni genom, uključujući oznaku ispitanika, koordinate insercije i duljina insercije. Na desnoj strani su prikazani rezultati pretraživanja BLAT. Prikazane su koordinate na koje je mapirana insercija, duljina te početak i kraj insercije. Prikazani su rezultati isključivo prva dva poravnavanja.

Rezultati nanoporskog sekvenciranja				UCSC BLAT rezultati pretraživanja						
Kromosom	Početak	Kraj	Duljina	Kromosom	Početak	Kraj	Duljina sekvence	Početak	Kraj	
CR2-1	15	23130374	23130374	82	15	28559815	28559896	82	1	82
					15	22465240	22465321	82	1	82
	15	23136596	23136596	54	15	28553652	28553704	53	1	53
					15	22459023	22459075	53	1	53
2	110576892	110576892	440	2	86914099	86914935	837	8	440	
				2	86913986	86914518	533	12	440	
CR2-2	2	110576897	110576897	438	2	86913984	86914835	852	3	430
					2	86913985	86914567	583	4	363
CR2-4	16	69963926	69963926	6868	16	74408693	74415224	6532	687	6868
					16	69957756	69963924	6169	699	6868
ST-1	15	30648619	30648619	288	15	30586833	30587121	289	1	288
					15	30379155	30379444	290	1	288
ST-3	15	23129676	23129676	126	15	20534465	20534590	126	1	126
					15	28561048	28561173	126	1	126

4.4. Analiza baza podataka

Rezultati pretraživanja mrežne baze *Decipher* sažeto su prikazani u Tablici 3. Za istraživanje kliničkog značaja strukturnih aberacija izdvojene su vrijednosti pLI, LOEUF, sHEt, pHaplo, pTriplo izvedene iz baze *Decipher* za svaki od ukupno 18 izabranih gena. Klinički značajni podaci uključuju vrijednost $pLI \geq 0.9$ zbog jake netolerancije na gubitak funkcije gena, vrijednost LOEUF što bliže 0 zbog povećane osjetljivosti na mutacije uz vrijednost sHEt što bliže 1 s obzirom da su geni s većim vrijednostima manje tolerantni na mutacije. Za gene kod kojih je nađena delecija kod svih 7 pacijenata analiziran je parametar pHaplo, a njegova vrijednost ≥ 0.86 ukazuje na klinički

značajne podatke, dok je za gene s nađenom amplifikacijom analiziran parametar pTriplo, s klinički značajnim podacima ≥ 0.94 .

Tablica 3. Popis gena s osnovnim podacima dobivenim iz mrežne baze *Decipher*. Kosa crta (/) označava nedostatak podataka.

Naziv gena	pLI	LOEUF	sHEt	pHaplo	pTriplo
<i>DUX4</i>	/	/	/	/	/
<i>NBPF26</i>	/	/	/	/	/
<i>GOLGA6L1</i>	/	/	/	0,03	0,16
<i>LOC102723623</i>	/	/	/	/	/
<i>POTEH</i>	0,00	1,67	/	0,06	0,22
<i>FAM86B1</i>	0,00	1,84	/	0,13	0,11
<i>FAM86B2</i>	0,00	1,74	/	0,08	0,13
<i>GOLGA6L22</i>	/	/	/	/	/
<i>RGPD6</i>	0,39	1,72	/	0,11	0,37
<i>NPIP9</i>	0,11	1,86	/	0,07	0,20
<i>NBPF14</i>	/	/	/	0,10	0,07
<i>ANKRD20A1</i>	/	/	/	0,12	0,24
<i>NBPF10</i>	/	/	/	0,07	0,08
<i>CLEC18A</i>	0,01	1,34	/	0,20	0,51
<i>NBPF19</i>	/	/	/	/	/
<i>FAM90A7P</i>	/	/	/	/	/
<i>ARHGAP11B</i>	0,00	1,56	0,136	/	/
<i>GOLGA6L6</i>	/	/	/	0,08	0,12

U bazama *Human brain transcriptome*, *Gene2Phenotype*, *ClinGen* i *SynGo* nisu pronađeni nikakvi podaci povezani s istraživanim genima.

Preuzimanjem podataka iz baza *SFARI Gene* i *AutDB* prikupljeni su podaci o genu *ARHGAP11B*. Prema rezultatima pretraživanja baze *SFARI Gene* genu *ARHGAP11B* dodijeljena je kategorija Jaki kandidat (od engl. *Strong Candidate*) s rezultatom 2.1. što ga svrstava u jake kandidat-gene potencijalno povezane s intelektualnim poteškoćama. Pretraživanjem i analizom baze *AutDB*

izdvojeni su podaci o povezanosti gena *ARHGAP11B* s poremećajima iz spektra autizma, intelektualnim poteškoćama i shizofrenijom uz genetsku kategoriju *Rare single gene variant*. Kategorija *Rare single gene variant* u bazi podataka AutDB označava rijetke varijante jednog gena koje se rijetko pojavljuju u populaciji, ali imaju značajan utjecaj na rizik za razvoj autizma. Za ostale gene od interesa unutar navedenih baza nije bilo podataka.

Iz baze podataka *Human protein atlas* izdvojila sam podatke o genima i njihovoj ekspresiji kako je prikazano u Tablici 4. Podaci povezani s mozgom i moždanim funkcijama dodatno su istaknuti (plava obojenja). Za gen *DUX4* zabilježena je ekspresija RNA u mozgu, gen *FAM86BI* pokazuje tkivno obogaćenje na razini ekspresije RNA u području mozga i hipofize, s naglaskom na cerebralni korteks. Genu *RGPDG* dodijeljena je kategorija ekspresije "grupno obogaćen" u ekscitacijskim i inhibitornim neuronima, dok je ekspresija gena *NBPF14* pojačana u astrocitima, a detektirana je u svih regijama mozga. Ekspresija gena *ANKRD20A1* utvrđena je unutar cerebralnog korteksa, a pojedinačne stanice s povišenom ekspresijom su ekscitacijski neuroni, inhibitorni neuroni te oligodendrocitske prekursorske stanice. Gen *CLEC18A* pokazuje pojačanu regionalnu ekspresiju na području koroidnog pleksusa, dok je za gen *NBPF19* prema podacima ekspresije iz baze HPA navedena mješovita funkcija unutar koroidnog pleksusa, a pojedinačne stanice s povišenom ekspresijom uključuju astrocite, oligodendrocitske prekursorske stanice, mikroglia-stanice i ekscitacijske neurone.

Tablica 4. Popis gena s podacima dobivenim iz baze *Human protein atlas*. Plava obojenja predstavljaju nalaze povezane s mozgom i moždanim funkcijama za pojedini gen.

Naziv gena	Tkivna ekspresija RNA	Proteinska ekspresija	Ekspresija RNA u mozgu	Ekspresija RNA u pojedinačnim stanicama
<i>DUX4</i>	Testisi	Nema informacija	Neuroni (mješovita funkcija), niska regionalna specifičnost	Grupno obogaćeni: kasne spermatide, rane spermatide
<i>NBPF26</i>	Citoplazma većine tkiva, najviše koža	U većini tkiva	Nespecifično	Pojačani stanični tipovi: skvamozne epitelne stanice, bazalne skvamozne epitelne stanice, eritroidne stanice, makrofagi, Panethove stanice, B-stanice
<i>GOLGA6L1</i>	Testisi	Nema informacija	Nije detektirano	Pojačani stanični tipovi: rane spermatide, kasne spermatide, spermatociti
<i>LOC102723623</i>	Nema informacija	Nema informacija	Nema informacija	Nema informacija
<i>POTEH</i>	Prostata	Nema informacija	Nije detektirano	Obogaćene stanice: prostatičke glandularne stanice
<i>FAM86B1</i>	Tkivno obogaćeni (mozak, hipofiza)	Testisi	Cerebralni korteks	Pojačani stanični tipovi: proksimalne tubularne stanice, skvamozne epitelne stanice, sekretorne stanice, trepetljikave stanice
<i>FAM86B2</i>	Nije detektirano	Testisi	Nije detektirano	Jajne stanice
<i>GOLGA6L22</i>	Tkivno obogaćeni (testisi)	Nema informacija	Nije detektirano	Grupno obogaćeni: rane spermatide, kasne spermatide
<i>RGPD6</i>	Testisi - niska tkiva specifičnost	Srednja u većini tkiva	Nespecifično	Grupno obogaćeni: ekscitacijski neuroni, inhibitorski neuroni
<i>NPIP9</i>	Tkivno obogaćeni (testisi)	Visoka u većini tkiva	Nespecifično	Plazma stanice
<i>NBPF14</i>	Koža	Citoplazmatska ekspresija u većini tkiva	Sve regije	Pojačani stanični tipovi: Panethove stanice, astrociti
<i>ANKRD20A1</i>	Tkivno pojačani (testisi)	Nema informacija	Cerebralni korteks, mješovita funkcija	Pojačani stanični tipovi: ekscitacijski neuroni, oligodendrocitske prekursorske stanice, inhibitorski neuroni, spermatociti, spermatogonije
<i>NBPF10</i>	Niska tkivna specifičnost, otkrivena u većini, visoka u koštanoj srži i koži	Citoplazmatska ekspresija u većini tkiva	Nespecifično	Pojačani stanični tipovi: Panethove stanice, skvamozne epitelne stanice, rane spermatide
<i>CLEC18A</i>	Grupno obogaćeni (koroidni pleksus, bubreg)	Nema informacija	Regionalno pojačan (koroidni pleksus)	Obogaćene stanice: proksimalne tubularne stanice
<i>NBPF19</i>	Koža, mrežnica, koštana srž	Nema informacija	Koroidni pleksus, mješovita funkcija	Pojačani stanični tipovi: astrociti, oligodendrocitske prekursorske stanice, mikroglia stanice, ekscitacijski neuroni
<i>FAM90A7P</i>	Nema informacija	Nema informacija	Nema informacija	Nema informacija
<i>ARHGAP11B</i>	Tkivno obogaćeni (koštana srž, limfoidna tkiva)	Nema informacija	Makrofagi i mikroglia stanice (imunološki odgovor)	Pojačani stanični tipovi: trofoblasti ekstraviloze, eritroidne stanice, rane spermatide, stanice koje izlučuju želučanu sluz, plazma stanice
<i>GOLGA6L6</i>	Tkivno obogaćeni (testisi)	Dokazi na razini transkripta	Nije detektirano	Obogaćene stanice: rane spermatide

5. RASPRAVA

U ovom radu istražene su različite strukturne varijacije koje se nalaze u evolucijski novim regijama genoma kod 7 ispitanika s teškom kliničkom slikom neurorazvojnih poremećaja. Svaki od 7 pacijenata imao je višestruke kromosomske aberacije, sugerirajući da velike i ponavljajuće SV znatno narušavaju osnovne životne funkcije, što je vidljivo iz kliničke slike pacijenata uključenih u studiju. Najčešće poremećene neurorazvojne karakteristike pacijenata uključivale su epilepsiju, nerazvijen govor, poremećaj ponašanja i intelektualnu zaostalost, uz promijenjene morfološke karakteristike poput nepokretnosti, niskog rasta i mikrocefalije. Prema bazi podataka *Human protein atlas* tkivna ekspresija većine istraživanih gena pojačana je u muškim spolnim organima (testisi i prostata) što bi moglo biti u korelaciji s većom pojavnosću neurorazvojnih poremećaja kod muškaraca, u odnosu na žene. Međutim potrebna su dodatna istraživanja koja bi testirala ovu pretpostavku.

Rezultati ovog istraživanja pokazali su da su u svim slučajevima insercije porijeklom s različitih područja u genomu, što sam potvrdila preciznim mapiranjem genomskih regija, upotrebom alata UCSC BLAT. Posebno je zanimljiva insercija od 6868 parova baza kod ispitanika CR2-4 koja je detektirana u koordinatama 16:69963926-69963926, što odgovara lokaciji na 16q22.1, a porijeklom može biti s dvije kromosomske regije: 16:69957756-69963924 što odgovara 16q22.1 i 16:74408693-74415224 što odgovara 16q23.1 regiji. Istraživanje podrijetla sekvenci insercija u genomima pacijenata otvorilo je vrata nastavku istraživanja molekularnih karakteristika ispitivanih gena.

Analizom podataka izvedenih iz baze DGV (Tablica 1) uočeno je da stupac označen kao „Rang“ pokazuje visoke brojeve, što ukazuje na manju pouzdanost identificiranih aberacija. Delecije koje imaju niske frekvencije također su nepouzdanе, jer su njima pridružene varijante s visokim rangom. Također, podaci u programu su većinom prikupljeni od nepoznatih populacija, odnosno porijeklo uzoraka nije strogo definirano, što dodatno smanjuje pouzdanost interpretacije. Zaključno, možemo reći da se delecije navedenih regija ne nalaze generalno prisutne u populaciji.

Pretragom literature unutar baze *Pubmed* povezala sam istraživane gene s poremećajima iz autističnog spektra, intelektualnim poteškoćama i shizofrenijom što sugerira da su ti geni uključeni u funkcije živčanog sustava, a njihove promjene rezultiraju neurološkim poremećajima. Kao geni od posebnog interesa istaknuli su se oni porodice *NBPF*, geni *ARHGAP11B* i *RGPD6* te porodica

gena *GOLGA* zbog širokog spektra nalaza koji ih povezuju s mozgom i bolestima živčanog sustava. Mnogi od njih podložni su duplikacijama i nalaze se u evolucijski novim regijama karakterističnim za vrstu *Homo sapiens* (Bekpen i Tautz, 2019). Chen i sur. su 2022. godine otkrili segmentnu duplikaciju 2q12.2/q13 *de novo* koja uključuje gen *RGPD6*, a koja zbog promjene funkcije dovodi do terminacije trudnoće, što potvrđuje značaj ovog gena tijekom fetusnog razdoblja. Također, ista je grupa nekoliko godina ranije detektirala mikroduplikaciju 2q13 koja uključuje gen *RGPD6*, povezanu s poremećajem autističnog spektra, intelektualnim nedostatkom i poremećajem jetre (Chen i sur., 2017). U sklopu ovog istraživanja detektirala sam insercija u genu *RGPD6* veličine 440 parova baza. Uz to, delecija razvojnih gena poput *RGPD6* dovodi do težih fenotipskih posljedica pacijenata sa trisomijom 18p, budući da su oni uključeni u nuklearno-citoplazmatski promet, staničnu signalizaciju i diferencijaciju tkiva (Jaiswal i sur., 2015). Dodatno, u ovom radu su otkrivene strukturne aberacije u centromernim područjima, unutar gena *GOLGA6L6* i *POTEH*, što prijašnjim metodama sekvenciranja 2. generacije nije bilo moguće.

Pretraživanjem literature otkrila sam kako je najveći broj gena istraživanih u ovom radu uključen je poremećaj autističnog spektra. Riječ je o čestom nasljednom poremećaju koji zahvaća otprilike jednog od 42 muškarca i jednu od 189 žena (Davis i sur., 2019). Upravo zbog učestalosti, pretpostavlja se da su strukturne varijacije gena povezanih s autizmom važne za bolje razumijevanje molekularne podloge NDD-a. Jedan takav kandidat s visokim i varijabilnim brojem kopija, koji je izostao u tradicionalnim genetskim studijama autizma, uključuje porodicu proteina *Olduvai*, primarno kodiranu genima *NBPF*. Domene *Olduvai*, koje se sastoje od približno 65 aminokiselina, pokazuju najveći porast broja kopija specifičan za ljude, u odnosu na bilo koju drugu regiju genoma koja kodira proteine. Broj kopija *Olduvai* također pokazuje poveznicu s veličinom mozga i brojem neurona među antropoidnim primatima, kao i s veličinom mozga i kognitivnim sposobnostima unutar ljudske populacije. Zanimljivo je da se većina ljudskih *Olduvai* kopija mapira na genomsku regiju 1q21, gdje su duplikacije povezane s autizmom i makrocefalijom dok su recipročne delecije povezane sa shizofrenijom i mikrocefalijom (Davis i sur., 2019). U ovom istraživanju, aberacije gena *NBPF* pronađene su kod pacijenata CR2-4, ST-1, ST-2 i ST-3. Prema dostupnim podacima iz medicinske dokumentacije vidljivo je da su svima zajedničke neurorazvojne karakteristike intelektualne teškoće i nerazvijen govor, što bi moglo biti u vezi s aberacijama gena *NBPF*. Kod tri od četiri pacijenta zabilježena je epilepsija, a kod pacijentice ST-3 utvrđena je mikrocefalija.

Značaj istraživanja evolucijski novih gena te njihovih strukturnih promjena ukazuje i rad Fischer i sur., 2022 gdje je pokazano kako gen *ARHGAP11B*, eksprimiran specifično u ljudi, sudjeluje u ekspanziji neokorteksa, utjecajem na bazalne radijalne glija-stanice stimulacijom glutaminolize u mitohondrijima. Gen *ARHGAP11B* nastao je djelomičnom duplikacijom gena *ARHGAP11A*, koji kodira protein za aktivaciju GTP-aze Rho, isključivo u ljudskoj lozi, nakon njenog razdvajanja od loze čimpanzi. Ekspresija gena *ARHGAP11B* u embrijskom neokorteksu miševa potiče proliferaciju bazalnih progenitora i njihovu samoobnovu, što može rezultirati povećanjem površine kortikalne ploče i iniciranjem girifikacije (Florio i sur., 2015). Ove funkcije upućuju na to da je gen *ARHGAP11B* imao značajnu ulogu u evolucijskom širenju ljudskog neokorteksa te zato predstavlja značajan izvor informacija o ljudskom zdravlju i bolesti. Promjene u ekspresiji i funkciji ovog gena potencijalno utječu na normalan razvoj mozga, što može imati za posljedicu pojavu neurorazvojnih poremećaja. U ovom istraživanju strukturna varijacija gena *ARHGAP11B* u obliku insercije pronađena je kod pacijenta ST-1. Dodatno, pregledni rad Huttner i sur., 2024 ističe mogućnost doprinosa gena *ARHGAP11B* kortikalnim malformacijama poput mikrocefalije. Na temelju toga, gen *ARHGAP11B* bi mogao biti značajan faktor u slučajevima sindroma mikrodelecije 15q13.3, koji se povezuje s mikrocefalijom. Osim navedenog, gen *ARHGAP11B*, induciran u odraslih miševa, uspješno povećava fleksibilnost pamćenja i povezan je s promjenama neuro-bihevioralnih značajki, što je ključno za poboljšanje kognitivnih sposobnosti ljudske vrste (Xing i sur., 2021). Dalje, indukcijom gena *ARHGAP11B* kod tvora Kalebic i sur., 2018 pokazali su postnatalno povećanje gustoće neurona u gornjim kortikalnim slojevima. S obzirom na ta istraživanja, kao i rezultate ovog rada, moguće je da promjene ovog gena utječu na rani razvoj mozga i predstavljaju potencijalni doprinos patogenezi neurorazvojnih poremećaja kod pacijenata uključenih u ovu studiju.

U ovom istraživanju otkriven je veliki broj CNV-ova, što je u skladu s pretpostavkom kako su varijacije u broju kopija ključan genetski faktor u etiologiji psihijatrijskih poremećaja. Važnost istraživanja CNV-ova u podlozi neurorazvojnih poremećaja pokazuje i rad Zarrei i sur., 2019 koji opisuje veliku pojavnost varijacija broja kopija DNA kod pacijenata oboljelih od poremećaja pozornosti s hiperaktivnošću (ADHD) te poremećaja autističnog spektra. Kako je istaknuto u radu 4,4% od 10,5% osoba s neurorazvojnim poremećajima bili su nositelji novih CNV-ova, dok su ostale osobe bile nositelji aneuploidija ili već poznatih varijanti genetičkih poremećaja. Općenito, geni povezani s neurorazvojnim poremećajima podvrgnuti su jakom selektivnom pritisku tijekom

evolucije, te su netolerantni na mutacije koje dovode do gubitka funkcije u ljudskoj populaciji (Leblond i sur., 2021). S obzirom na tkivnu distribuciju i vremenski uvjetovanu ekspresiju, većina gena HC-NDD eksprimirana je u vrlo ranom fetusnom razdoblju unutar amigdale, korteksa, hipokampusa i strijatuma. Uz to, geni sa sinaptičkim funkcijama, opisani u bazi *SynGO*, više su eksprimirani nakon rođenja u malom mozgu i korteksu (Leblond i sur., 2021). Međutim za odabrane gene u ovom istraživanju nije bilo podataka u bazi *SynGO*, što sugerira da su evolucijski novi geni još nedovoljno istraženi.

Pretraživanjem mrežne baze *Decipher* i praćenjem parametara pLI, LOEUF, sHet, pHaplo i pTriplo nisu zabilježeni jasni dokazi kliničkog značaja ispitivanih gena u podlozi neurorazvojnih poremećaja. Međutim rezultati analize mrežne baze *Human protein atlas* pokazali su prisutnost genske ekspresije na razini mRNA gena *DUX4*, *FAM86B1*, *RGPD6*, *NBPF14*, *ANKRD20A1*, *CLEC18A* i *NBPF19* u različitim regijama i stanicama mozga. Takvi rezultati upućuju na to da su evolucijski novi geni doprinijeli kognitivnim, fiziološkim i intelektualnim naprecima čovjeka i da su važan dio evolucije, a njihove promjene utječu na normalan razvoj i funkcije čovjeka.

Iako je ovo istraživanje doprinijelo povećanju općeg znanja o genima potencijalno povezanim sa neurorazvojnim poremećajima, u dijagnostičke svrhe mogu se koristiti samo geni s jasnim dokazima povezanosti, stoga su potrebna buduća istraživanja u tom smjeru kako bi se hipoteza u potpunosti testirala. Osim toga, istraživanje je provedeno uz određena ograničenja koje je potrebno uzeti u obzir pri interpretaciji rezultata. Prva značajna limitacija jest nedostatak uzoraka DNA roditelja ispitanika, što onemogućava precizno određivanje genetičke podloge i procjenu nasljednih svojstava ispitanih varijanti. Drugi izazov odnosi se na tehničke aspekte sekvenciranja, koje je provedeno uz nisku pokrivenost zbog financijskih ograničenja, što može utjecati na točnost i sveobuhvatnost dobivenih podataka. Treći nedostatak proizlazi iz metodološkog fokusa istraživanja, koji je bio usmjeren na specifične regije genoma, zbog čega nije moguće isključiti potencijalni utjecaj drugih aberacija u genomu koje nisu bile obuhvaćene analizom. Uz to, neurorazvojni poremećaji predstavljaju heterogenu, složenu grupu stanja karakteriziranu promjenama u brojnim oblicima funkcija i ponašanja, stoga je u pojedinim slučajevima teško odrediti točne uzroke nastanka ove multifaktorske bolesti. Međutim, danas medicina i zdravstvena skrb oboljelih značajno napreduju, pa rano otkrivanje genetskih znakova može osigurati bolji život oboljelima, a obiteljima pružiti informacije o uzroku neurorazvojnih poremećaja.

6. ZAKLJUČAK

Sekvenciranje tehnologijom treće generacije uz daljnju analizu i vizualizaciju podataka omogućava pronalazak novih strukturnih varijacija te njihovu detaljniju karakterizaciju koju nije bilo moguće utvrditi prethodnim metodama obrade uzoraka, međutim javno dostupne baze podataka još ne sadrže dovoljan broj informacija o evolucijski novim genima u podlozi neurorazvojnih poremećaja. Unatoč napretku tehnologija sekvenciranja i metoda detekcije brojnih strukturnih varijacija, potrebna su dodatna istraživanja kako bi se precizno utvrdila klinička važnost navedenih promjena. Rezultati ovog istraživanja pokazali su kako geni porodice *NBPF*, *GOLGA*, gen *ARHGAP11B* i genska porodica *RGPD6* predstavljaju potencijalne kandidate uključene u patogenezu neurorazvojnih poremećaja. Zaključak o novim kandidat-genima dobiven je temeljem već poznatih literaturnih navoda koji ih povezuju sa poremećajima moždanih funkcija i intelektualnim nedostatkom, temeljem njihove uključenosti u neurorazvoj te pronalaskom značajnih strukturnih varijacija u njima u sklopu ovog istraživanja. Konačno, ovo istraživanje doprinijelo je povećanju općeg znanja o podlozi neurorazvojnih poremećaja i omogućilo je daljnje istraživanje molekularnih karakteristika novih kandidat-gena.

7. LITERATURA

Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., Flicek, P., Gabriel, S.B., Gibbs, R.A., Green, E.D., Hurles, M.E., Knoppers, B.M., Korbel, J.O., Lander, E.S., Lee, C., Lehrach, H., Mardis, E.R., Marth, G.T., McVean, G.A., Nickerson, D.A., Schmidt, J.P., Sherry, S.T., Wang, J., Wilson, R.K., Gibbs, R.A., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., Kovar, C., Lee, S., Muzny, D., Reid, J.G., Zhu, Y., Wang, J., Chang, Y., Feng, Q., Fang, X., Guo, X., Jian, M., Jiang, H., Jin, X., Lan, T., Li, G., Li, J., Li, Y., Liu, S., Liu, X., Lu, Y., Ma, X., Tang, M., Wang, B., Wang, G., Wu, H., Wu, R., Xu, X., Yin, Y., Zhang, D., Zhang, W., Zhao, J., Zhao, M., Zheng, X., Lander, E.S., Altshuler, D.M., Gabriel, S.B., Gupta, N., Gharani, N., Toji, L.H., Gerry, N.P., Resch, A.M., Flicek, P., Barker, J., Clarke, L., Gil, L., Hunt, S.E., Kelman, G., Kulesha, E., Leinonen, R., McLaren, W.M., Radhakrishnan, R., Roa, A., Smirnov, D., Smith, R.E., Streeter, I., Thormann, A., Toneva, I., Vaughan, B., Zheng-Bradley, X., Bentley, D.R., Grocock, R., Humphray, S., James, T., Kingsbury, Z., Lehrach, H., Sudbrak, R., Albrecht, M.W., Amstislavskiy, V.S., Borodina, T.A., Lienhard, M., Mertes, F., Sultan, M., Timmermann, B., Yaspo, M.-L., Mardis, E.R., Wilson, R.K., Fulton, L., Fulton, R., Sherry, S.T., Ananiev, V., Belaia, Z., Beloslyudtsev, D., Bouk, N., Chen, C., Church, D., Cohen, R., Cook, C., Garner, J., Hefferon, T., Kimelman, M., Liu, C., Lopez, J., Meric, P., O'Sullivan, C., Ostapchuk, Y., Phan, L., Ponomarov, S., Schneider, V., Shekhtman, E., Sirotkin, K., Slotta, D., Zhang, H., McVean, G.A., Durbin, R.M., Balasubramaniam, S., Burton, J., Danecek, P., Keane, T.M., Kolb-Kokocinski, A., McCarthy, S., Stalker, J., Quail, M., Schmidt, J.P., Davies, C.J., Gollub, J., Webster, T., Wong, B., Zhan, Y., Auton, A., Campbell, C.L., Kong, Y., Marcketta, A., Gibbs, R.A., Yu, F., Antunes, L., Bainbridge, M., Muzny, D., Sabo, A., Huang, Z., Wang, J., Coin, L.J.M., Fang, L., Guo, X., Jin, X., Li, G., Li, Q., Li, Y., Li, Z., Lin, H., Liu, B., Luo, R., Shao, H., Xie, Y., Ye, C., Yu, C., Zhang, F., Zheng, H., Zhu, H., Alkan, C., Dal, E., Kahveci, F., Marth, G.T., Garrison, E.P., Kural, D., Lee, W.-P., Fung Leong, W., Stromberg, M., Ward, A.N., Wu, J., Zhang, M., Daly, M.J., DePristo, M.A., Handsaker, R.E., Altshuler, D.M., Banks, E., Bhatia, G., del Angel, G., Gabriel, S.B., Genovese, G., Gupta, N., Li, H., Kashin, S., Lander, E.S., McCarroll, S.A., Nemes, J.C., Poplin, R.E., Yoon, S.C., Lihm, J., Makarov, V., Clark, A.G., Gottipati, S., Keinan, A., Rodriguez-Flores, J.L., Korbel, J.O., Rausch, T., Fritz, M.H., Stütz, A.M., Flicek, P., Beal, K., Clarke, L., Datta, A., Herrero, J., McLaren, W.M., Ritchie, G.R.S., Smith, R.E., Zerbino, D., Zheng-Bradley, X., Sabeti, P.C., Shlyakhter, I., Schaffner, S.F., Vitti, J., Cooper, D.N., Ball, E.V., Stenson, P.D., Bentley, D.R., Barnes, B., Bauer, M., Keira Cheetham, R., Cox, A., Eberle, M., Humphray, S., Kahn, S., Murray, L., Peden, J., Shaw, R., Kenny, E.E., Batzer, M.A., Konkel, M.K., Walker, J.A., MacArthur, D.G., Lek, M., Sudbrak, R., Amstislavskiy, V.S., Herwig, R., Mardis, E.R., Ding, L., Koboldt, D.C., Larson, D., Ye, K., Gravel, S., The 1000 Genomes Project Consortium, Corresponding authors, Steering committee, Production group, Baylor College of Medicine, BGI-Shenzhen, Broad Institute of MIT and Harvard, Coriell Institute for Medical Research, European Molecular Biology Laboratory, E.B.I., Illumina, Max Planck Institute for Molecular Genetics, McDonnell Genome Institute at Washington University, US National Institutes of Health, University of Oxford,

Wellcome Trust Sanger Institute, Analysis group, Affymetrix, Albert Einstein College of Medicine, Bilkent University, Boston College, Cold Spring Harbor Laboratory, Cornell University, European Molecular Biology Laboratory, Harvard University, Human Gene Mutation Database, Icahn School of Medicine at Mount Sinai, Louisiana State University, Massachusetts General Hospital, McGill University, National Eye Institute, N., 2015. A global reference for human genetic variation. *Nature* 526, 68–74. <https://doi.org/10.1038/nature15393>

Bekpen, C., Tautz, D., 2019. Human core duplicon gene families: game changers or game players? *Briefings in Functional Genomics* 18, 402–411. <https://doi.org/10.1093/bfgp/elz016>

Chen, C.-P., Chen, S.-W., Chern, S.-R., Wu, P.-S., Wu, F.-T., Pan, Y.-T., Pan, C.-W., Wang, W., 2022. Prenatal diagnosis and molecular cytogenetic characterization of a *de novo* duplication of 2q12.2→q13 encompassing *MALL*, *NPHP1*, *RGPD6* and *BUB1*. *Taiwanese Journal of Obstetrics and Gynecology* 61, 1044–1047. <https://doi.org/10.1016/j.tjog.2022.08.007>

Chen, C.-P., Lin, S.-P., Lee, C.-L., Chern, S.-R., Wu, P.-S., Chen, Y.-N., Chen, S.-W., Wang, W., 2017. Recurrent 2q13 microduplication encompassing *MALL*, *NPHP1*, *RGPD6*, and *BUB1* associated with autism spectrum disorder, intellectual disability, and liver disorder. *Taiwan J Obstet Gynecol* 56, 98–101. <https://doi.org/10.1016/j.tjog.2016.12.003>

Davis, J.M., Heft, I., Scherer, S.W., Sikela, J.M., 2019. A Third Linear Association Between Olduvai (DUF1220) Copy Number and Severity of the Classic Symptoms of Inherited Autism. *Am J Psychiatry* 176, 643–650. <https://doi.org/10.1176/appi.ajp.2018.18080993>

De Coster, W., Van Broeckhoven, C., 2019. Newest Methods for Detecting Structural Variations. *Trends in Biotechnology* 37, 973–982. <https://doi.org/10.1016/j.tibtech.2019.02.003>

de Smith, A.J., Trewick, A.L., Blakemore, A.I.F., 2010. Implications of copy number variation in people with chromosomal abnormalities: potential for greater variation in copy number state may contribute to variability of phenotype. *Hugo J* 4, 1–9. <https://doi.org/10.1007/s11568-010-9144-z>

Fischer, J., Fernández Ortuño, E., Marsoner, F., Artioli, A., Peters, J., Namba, T., Eugster Oegema, C., Huttner, W.B., Ladewig, J., Heide, M., 2022. Human-specific *ARHGAP11B* ensures human-like basal progenitor levels in hominid cerebral organoids. *EMBO Rep* 23, e54728. <https://doi.org/10.15252/embr.202254728>

Florio, M., Albert, M., Taverna, E., Namba, T., Brandl, H., Lewitus, E., Haffner, C., Sykes, A., Wong, F.K., Peters, J., Guhr, E., Klemroth, S., Prüfer, K., Kelso, J., Naumann, R., Nüsslein, I., Dahl, A., Lachmann, R., Pääbo, S., Huttner, W.B., 2015. Human-specific gene *ARHGAP11B* promotes basal progenitor amplification and neocortex expansion. *Science* 347, 1465–1470. <https://doi.org/10.1126/science.aaa1975>

Heather, J.M., Chain, B., 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics* 107, 1–8. <https://doi.org/10.1016/j.ygeno.2015.11.003>

- Hiatt, S.M., Lawlor, J.M.J., Handley, L.H., Ramaker, R.C., Rogers, B.B., Partridge, E.C., Boston, L.B., Williams, M., Plott, C.B., Jenkins, J., Gray, D.E., Holt, J.M., Bowling, K.M., Bebin, E.M., Grimwood, J., Schmutz, J., Cooper, G.M., 2021. Long-read genome sequencing for the molecular diagnosis of neurodevelopmental disorders. *Human Genetics and Genomics Advances* 2, 100023. <https://doi.org/10.1016/j.xhgg.2021.100023>
- Hollox, E.J., Zuccherato, L.W., Tucci, S., 2022. Genome structural variation in human evolution. *Trends in Genetics* 38, 45–58. <https://doi.org/10.1016/j.tig.2021.06.015>
- Huttner, W.B., Heide, M., Mora-Bermúdez, F., Namba, T., 2024. Neocortical neurogenesis in development and evolution-Human-specific features. *J Comp Neurol* 532, e25576. <https://doi.org/10.1002/cne.25576>
- Jaiswal, S.K., Kumar, A., Ali, A., Rai, A.K., 2015. Co-occurrence of mosaic supernumerary isochromosome 18p and intermittent 2q13 deletions in a child with multiple congenital anomalies. *Gene* 559, 94–98. <https://doi.org/10.1016/j.gene.2015.01.042>
- Kalebic, N., Gilardi, C., Albert, M., Namba, T., Long, K.R., Kostic, M., Langen, B., Huttner, W.B., 2018. Human-specific ARHGAP11B induces hallmarks of neocortical expansion in developing ferret neocortex. *Elife* 7, e41241. <https://doi.org/10.7554/eLife.41241>
- Leblond, C.S., Le, T.-L., Malesys, S., Cliquet, F., Tabet, A.-C., Delorme, R., Rolland, T., Bourgeron, T., 2021. Operative list of genes associated with autism and neurodevelopmental disorders based on database review. *Mol Cell Neurosci* 113, 103623. <https://doi.org/10.1016/j.mcn.2021.103623>
- Mahmoud, M., 2019. Structural variant calling: the long and the short of it.
- Mastrososa, F.K., Miller, D.E., Eichler, E.E., 2023. Applications of long-read sequencing to Mendelian genetics. *Genome Med* 15, 42. <https://doi.org/10.1186/s13073-023-01194-3>
- May, T., Adesina, I., McGillivray, J., Rinehart, N.J., 2019. Sex differences in neurodevelopmental disorders. *Curr Opin Neurol* 32, 622–626. <https://doi.org/10.1097/WCO.0000000000000714>
- Moraes, F., Góes, A., 2016. A decade of human genome project conclusion: Scientific diffusion about our genome knowledge. *Biochem Mol Biol Educ* 44, 215–223. <https://doi.org/10.1002/bmb.20952>
- Morris-Rosendahl, D.J., Crocq, M.-A., 2020. Neurodevelopmental disorders—the history and future of a diagnostic concept. *Dialogues Clin Neurosci* 22, 65–72. <https://doi.org/10.31887/DCNS.2020.22.1/macrocq>
- Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bizkadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S.J., Diekhans, M., Logsdon, G.A., Alonge, M., Antonarakis, S.E., Borchers, M., Bouffard, G.G., Brooks, S.Y., Caldas, G.V., Chen, N.-C., Cheng, H., Chin, C.-S., Chow, W., de Lima, L.G., Dishuck, P.C., Durbin, R., Dvorkina, T.,

Fiddes, I.T., Formenti, G., Fulton, R.S., Fungtammasan, A., Garrison, E., Grady, P.G.S., Graves-Lindsay, T.A., Hall, I.M., Hansen, N.F., Hartley, G.A., Haukness, M., Howe, K., Hunkapiller, M.W., Jain, C., Jain, M., Jarvis, E.D., Kerpedjiev, P., Kirsche, M., Kolmogorov, M., Korlach, J., Kremitzki, M., Li, H., Maduro, V.V., Marschall, T., McCartney, A.M., McDaniel, J., Miller, D.E., Mullikin, J.C., Myers, E.W., Olson, N.D., Paten, B., Peluso, P., Pevzner, P.A., Porubsky, D., Potapova, T., Rogae, E.I., Rosenfeld, J.A., Salzberg, S.L., Schneider, V.A., Sedlazeck, F.J., Shafin, K., Shew, C.J., Shumate, A., Sims, Y., Smit, A.F.A., Soto, D.C., Sović, I., Storer, J.M., Streets, A., Sullivan, B.A., Thibaud-Nissen, F., Torrance, J., Wagner, J., Walenz, B.P., Wenger, A., Wood, J.M.D., Xiao, C., Yan, S.M., Young, A.C., Zarate, S., Surti, U., McCoy, R.C., Dennis, M.Y., Alexandrov, I.A., Gerton, J.L., O'Neill, R.J., Timp, W., Zook, J.M., Schatz, M.C., Eichler, E.E., Miga, K.H., Phillippy, A.M., 2022. The complete sequence of a human genome. *Science* 376, 44–53. <https://doi.org/10.1126/science.abj6987>

Patterson, J., Wongsurawat, T., Rodriguez, A., 2020. A Glioblastoma Genomics Primer for Clinicians. *MRAJ* 8. <https://doi.org/10.18103/mra.v8i2.2034>

Talseth-Palmer, B.A., Scott, R.J., 2011. Genetic Variation and its Role in Malignancy. *Int J Biomed Sci* 7, 158–171.

Weckselblatt, B., Rudd, M.K., 2015. Human Structural Variation: Mechanisms of Chromosome Rearrangements. *Trends Genet* 31, 587–599. <https://doi.org/10.1016/j.tig.2015.05.010>

Xing, L., Kubik-Zahorodna, A., Namba, T., Pinson, A., Florio, M., Prochazka, J., Sarov, M., Sedlacek, R., Huttner, W.B., 2021. Expression of human-specific ARHGAP11B in mice leads to neocortex expansion and increased memory flexibility. *EMBO J* 40, e107093. <https://doi.org/10.15252/embj.2020107093>

Zarrei, M., Burton, C.L., Engchuan, W., Young, E.J., Higginbotham, E.J., MacDonald, J.R., Trost, B., Chan, A.J.S., Walker, S., Lamoureux, S., Heung, T., Mojarad, B.A., Kellam, B., Paton, T., Faheem, M., Miron, K., Lu, C., Wang, T., Samler, K., Wang, X., Costain, G., Hoang, N., Pellicchia, G., Wei, J., Patel, R.V., Thiruvahindrapuram, B., Roifman, M., Merico, D., Goodale, T., Drmic, I., Speevak, M., Howe, J.L., Yuen, R.K.C., Buchanan, J.A., Vorstman, J.A.S., Marshall, C.R., Wintle, R.F., Rosenberg, D.R., Hanna, G.L., Woodbury-Smith, M., Cytrynbaum, C., Zwaigenbaum, L., Elsabbagh, M., Flanagan, J., Fernandez, B.A., Carter, M.T., Szatmari, P., Roberts, W., Lerch, J., Liu, X., Nicolson, R., Georgiades, S., Weksberg, R., Arnold, P.D., Bassett, A.S., Crosbie, J., Schachar, R., Stavropoulos, D.J., Anagnostou, E., Scherer, S.W., 2019. A large data resource of genomic copy number variation across neurodevelopmental disorders. *npj Genom. Med.* 4, 26. <https://doi.org/10.1038/s41525-019-0098>

ŽIVOTOPIS

Valentina Klein, rođena 21.5.1999. godine u Zagrebu, gdje sam provela veći dio svog života. Osnovnu i srednju školu pohađala sam u rodnom gradu, a zbog interesa za prirodne znanosti upisala sam Prirodoslovnu školu Vladimira Preloga, smjer: kemijski tehničar. Nakon završetka srednje škole započela sam studij na Odjelu za biologiju Sveučilišta Josipa Jurja Strossmayera u Osijeku radi stjecanja novih iskustava. Tijekom studija aktivno sam sudjelovala u raznim studentskim projektima, stručnim skupovima i na smotrama fakulteta. Uz to, bila sam članica studentske udruge ZOA, te volonter na 21. Tjednu mozga. Kao dodatak akademskom životu, uključila sam se u rad Studentskog zbora, s titulom Studentskog pravobranitelja. Moje zanimanje za znanost pokušala sam prenijeti i ostalim kolegama, odradivanjem demonstrature na kolegiju Mikrobiologija.

U rujnu 2022. godine završila sam preddiplomski studij s odličnim uspjehom i nagradama za najboljeg studenta 1., 2. i 3. godine studija. Završni rad pod temom "Uloga leptinskog receptora u maternici" izrađen je u suradnji s Medicinskim fakultetom u Osijeku.

Zbog osobnog razvoja nastavila sam svoje školovanje na diplomskom studiju Prirodoslovno-matematičkog fakulteta u Zagrebu, upisavši smjer Eksperimentalna biologija, modul: Fiziologija i imunobiologija. Uz redovite studentske obaveze volontirala sam na projektu Sjeti me se, u udruzi Bioteka te na Danu karijera PMF-a. Ljubav prema znanosti otvorila mi je mogućnost rada u STEM području kao voditelj dječjih znanstvenih radionica za tvrtku Blizu znanosti. Motivirana stjecanju novih znanja i vještina, tijekom petogodišnjeg studija odradila sam 3 stručne prakse koje su doprinijele rastu u mom akademskom, ali i privatnom životu. Posebno bih istaknula praksu u Zavodu za transfuzijsku i regenerativnu medicinu KBC Sestre milosrdnice koja me obogatila za iskustva prikupljanje, obrada i pohrana stanica/tkiva. Osim toga, aktivno sam sudjelovala na projektu "Evolution in the dark" te svojim radom doprinijela rješavanju znanstvenih problema i prikupljanju rezultata koji imaju potencijal objave u znanstvenom radu. Moju kompetentnost i marljivost dodatno ističe kategorija 10% najuspješnijih studenata za 4. i 5. studijsku godinu.

Kroz studij sam unaprijedila samostalno snalaženje u laboratoriju, rad u aseptičkim uvjetima, razvoj kritičkog razmišljanja te kontinuiranu želju za napretkom i razvojem. Dobro govorim engleski jezik uz samostalno korištenje u jeziku i pismu, uz osnovno poznavanje njemačkog jezika. Izvrsno se snalazim u radu s računalima i raznim softverskim alatima.

PRILOZI

Prilog 1. Popis gena jedinstvenih za čovjeka izdvojenih iz baze *GenTree*.

Gen	Identifikacijski broj gena	Naziv gena	Transkript
ENSG00000099290	ENSG00000099290	WASHC2A	ENST00000611324
ENSG00000103512	ENSG00000103512	NOMO1	ENST00000620755
ENSG00000108448	ENSG00000108448	TRIM16L	ENST00000395671
ENSG00000125498	ENSG00000125498	KIR2DL1	ENST00000336077
ENSG00000125551	ENSG00000125551	PLGLB2	ENST00000359481
ENSG00000130538	ENSG00000130538	OR11H1	ENST00000252835
ENSG00000135226	ENSG00000135226	UGT2B28	ENST00000511240
ENSG00000136682	ENSG00000136682	CBWD2	ENST00000433343
ENSG00000145736	ENSG00000145736	GTF2H2	ENST00000521602
ENSG00000147996	ENSG00000147996	CBWD5	ENST00000489273
ENSG00000153684	ENSG00000153684	GOLGA8F	ENST00000526619
ENSG00000157322	ENSG00000157322	CLEC18A	ENST00000449317
ENSG00000163046	ENSG00000163046	ANKRD30BL	ENST00000295181
ENSG00000166664	ENSG00000166664	CHRFAM7A	ENST00000397827
ENSG00000170160	ENSG00000170160	CCDC144A	ENST00000456009
ENSG00000171478	ENSG00000171478	SPACA5B	ENST00000304270
ENSG00000171928	ENSG00000171928	TVP23B	ENST00000581733
ENSG00000171931	ENSG00000171931	FBXW10	ENST00000301938
ENSG00000171943	ENSG00000171943	SRGAP2C	ENST00000304465
ENSG00000172014	ENSG00000172014	ANKRD20A4P	ENST00000615302
ENSG00000174876	ENSG00000174876	AMY1B	ENST00000330330
ENSG00000176269	ENSG00000176269	OR4F21	ENST00000320901
ENSG00000177144	ENSG00000177144	NUDT4B	ENST00000322209
ENSG00000179172	ENSG00000179172	HNRNPCL1	ENST00000317869
ENSG00000180658	ENSG00000180658	OR2A4	ENST00000315453
ENSG00000181625	ENSG00000181625	SLX1B	ENST00000567248
ENSG00000182974	ENSG00000182974	OR4M2B	ENST00000332663
ENSG00000183054	ENSG00000183054	RGPD6	ENST00000455695
ENSG00000183281	ENSG00000183281	PLGLB1	ENST00000355705
ENSG00000183336	ENSG00000183336	BOLA2	ENST00000330978
ENSG00000183632	ENSG00000183632	TP53TG3	ENST00000569420
ENSG00000183889	ENSG00000183889		ENST00000524823
ENSG00000184110	ENSG00000184110	EIF3C	ENST00000395587
ENSG00000184492	ENSG00000184492	FOXD4L1	ENST00000306507
ENSG00000184659	ENSG00000184659	FOXD4L4	ENST00000377413
ENSG00000185040	ENSG00000185040	SPDYE16	ENST00000632547
ENSG00000185164	ENSG00000185164	NOMO2	ENST00000562312
ENSG00000185304	ENSG00000185304	RGPD2	ENST00000398146
ENSG00000186092	ENSG00000186092	OR4F5	ENST00000641515
ENSG00000186146	ENSG00000186146	DEFB131A	ENST00000334879
ENSG00000186523	ENSG00000186523	FAM86B1	ENST00000534520

ENSG00000187627	ENSG00000187627	RGPD1	ENST00000409776
ENSG00000188092	ENSG00000188092	GPR89B	ENST00000461786
ENSG00000188219	ENSG00000188219	POTEE	ENST00000626191
ENSG00000188234	ENSG00000188234	AGAP4	ENST00000311652
ENSG00000188610	ENSG00000188610	FAM72B	ENST00000452190
ENSG00000188626	ENSG00000188626	GOLGA8M	ENST00000563027
ENSG00000189375	ENSG00000189375	TBC1D28	ENST00000405044
ENSG00000196369	ENSG00000196369	SRGAP2B	ENST00000619678
ENSG00000196862	ENSG00000196862	RGPD4	ENST00000408999
ENSG00000196993	ENSG00000196993	NPIP8	ENST00000357796
ENSG00000198062	ENSG00000198062	POTEH	ENST00000343518
ENSG00000198573	ENSG00000198573	SPANXC	ENST00000358993
ENSG00000204379	ENSG00000204379	XAGE1A	ENST00000375602
ENSG00000204382	ENSG00000204382	XAGE1B	ENST00000375613
ENSG00000204505	ENSG00000204505	PRAMEF9	ENST00000415919
ENSG00000204779	ENSG00000204779	FOX4L5	ENST00000377420
ENSG00000204918	ENSG00000204918	PRR20B	ENST00000377930
ENSG00000204919	ENSG00000204919	PRR20A	ENST00000377931
ENSG00000205456	ENSG00000205456	TP53TG3D	ENST00000568044
ENSG00000205571	ENSG00000205571	SMN2	ENST00000507458
ENSG00000205572	ENSG00000205572	SERF1B	ENST00000515588
ENSG00000212807	ENSG00000212807	OR2A2	ENST00000641810
ENSG00000213240			ENST00000612520
ENSG00000213648	ENSG00000213648	SULT1A4	ENST00000360423
ENSG00000214534	ENSG00000214534	ZNF705EP	ENST00000525199
ENSG00000214940	ENSG00000214940	NPIPA8	ENST00000541810
ENSG00000215126	ENSG00000215126	CBWD6	ENST00000467791
ENSG00000215269	ENSG00000215269	GAGE12G	ENST00000445148
ENSG00000215372	ENSG00000215372	ZNF705G	ENST00000400156
ENSG00000215784	ENSG00000215784	FAM72D	ENST00000400889
ENSG00000216649	ENSG00000216649	GAGE12E	ENST00000381698
ENSG00000223443	ENSG00000223443	USP17L2	ENST00000333796
ENSG00000223569	ENSG00000223569	USP17L15	ENST00000456464
ENSG00000224089	ENSG00000224089	CT47A10	ENST00000430448
ENSG00000224659	ENSG00000224659	GAGE12J	ENST00000442437
ENSG00000224902	ENSG00000224902	GAGE12H	ENST00000381722
ENSG00000225327	ENSG00000225327	USP17L3	ENST00000530484
ENSG00000225899	ENSG00000225899	FRG2B	ENST00000443774
ENSG00000226023	ENSG00000226023	CT47A6	ENST00000419194
ENSG00000226685	ENSG00000226685	CT47A12	ENST00000419982
ENSG00000226929	ENSG00000226929	CT47A11	ENST00000457020
ENSG00000227140	ENSG00000227140	USP17L5	ENST00000507227
ENSG00000227151	ENSG00000227151	PRR20D	ENST00000452123
ENSG00000227488	ENSG00000227488	GAGE12D	ENST00000405679
ENSG00000227551	ENSG00000227551	USP17L12	ENST00000415041

ENSG00000228856	ENSG00000228856	USP17L30	ENST00000504104
ENSG00000229579	ENSG00000229579	USP17L26	ENST00000610514
ENSG00000229665	ENSG00000229665	PRR20C	ENST00000544357
ENSG00000229676	ENSG00000229676	ZNF492	ENST00000456783
ENSG00000230178	ENSG00000230178	OR4F3	ENST00000456475
ENSG00000230430	ENSG00000230430	USP17L25	ENST00000509271
ENSG00000230549	ENSG00000230549	USP17L1	ENST00000529559
ENSG00000230594	ENSG00000230594	CT47A4	ENST00000415858
ENSG00000231051	ENSG00000231051	USP17L28	ENST00000504543
ENSG00000231396	ENSG00000231396	USP17L10	ENST00000417945
ENSG00000231637	ENSG00000231637	USP17L29	ENST00000511681
ENSG00000232264	ENSG00000232264	USP17L24	ENST00000504481
ENSG00000232399	ENSG00000232399	USP17L13	ENST00000616303
ENSG00000232948	ENSG00000232948	DEFB130A	ENST00000400079
ENSG00000233024	ENSG00000233024	NPIPA9	ENST00000531453
ENSG00000233136	ENSG00000233136	USP17L11	ENST00000457736
ENSG00000233232	ENSG00000233232	NPIP B7	ENST00000435324
ENSG00000233917	ENSG00000233917	POTEB	ENST00000435397
ENSG00000234289	ENSG00000234289	H2BC12L	ENST00000599962
ENSG00000235780	ENSG00000235780	USP17L27	ENST00000515574
ENSG00000236125	ENSG00000236125	USP17L4	ENST00000526929
ENSG00000236362	ENSG00000236362	GAGE12F	ENST00000440137
ENSG00000236396	ENSG00000236396	SLC35G4	ENST00000588001
ENSG00000236737	ENSG00000236737	GAGE12B	ENST00000361446
ENSG00000236761	ENSG00000236761	CTAGE9	ENST00000314099
ENSG00000237038	ENSG00000237038	USP17L8	ENST00000527080
ENSG00000237671	ENSG00000237671	GAGE12C	ENST00000420398
ENSG00000237957	ENSG00000237957	CT47A5	ENST00000439466
ENSG00000244693			ENST00000487179
ENSG00000248920	ENSG00000248920	USP17L19	ENST00000515566
ENSG00000248933	ENSG00000248933	USP17L22	ENST00000511280
ENSG00000249104	ENSG00000249104	USP17L17	ENST00000511568
ENSG00000249811	ENSG00000249811	USP17L21	ENST00000506414
ENSG00000250745	ENSG00000250745	USP17L20	ENST00000506151
ENSG00000250844	ENSG00000250844	USP17L18	ENST00000504209
ENSG00000250913	ENSG00000250913	USP17L23	ENST00000506619
ENSG00000254598	ENSG00000254598	CSNK2A3	ENST00000528848
ENSG00000255009	ENSG00000255009	UBTFL1	ENST00000530464
ENSG00000255192	ENSG00000255192	NANO GP8	ENST00000528386
ENSG00000255251	ENSG00000255251	PRR23D1	ENST00000533250
ENSG00000255524	ENSG00000255524	NPIP B8	ENST00000529716
ENSG00000256671	ENSG00000256671	LIMS4	ENST00000632758
ENSG00000260596	ENSG00000260596	DUX4	ENST00000616166
ENSG00000260691	ENSG00000260691	ANKRD20A1	ENST00000620792

ENSG00000261739	ENSG00000261739	GOLGA8S	ENST00000562295
ENSG00000261740	ENSG00000261740	BOLA2-SMG1P6	ENST00000569622
ENSG00000261794	ENSG00000261794	GOLGA8H	ENST00000566740
ENSG00000263353	ENSG00000263353	PPIAL4A	ENST00000577856
ENSG00000263464			ENST00000585245
ENSG00000263513	ENSG00000263513	FAM72C	ENST00000584486
ENSG00000266338	ENSG00000266338	NBPF15	ENST00000577412
ENSG00000268221	ENSG00000268221	OPN1MW	ENST00000596998
ENSG00000268320	ENSG00000268320	SCGB1C2	ENST00000595228
ENSG00000268738	ENSG00000268738	HSFX2	ENST00000598963
ENSG00000269028		HSFX3	ENST00000598964
ENSG00000269433	ENSG00000269433	OPN1MW3	ENST00000599405
ENSG00000270339	ENSG00000270339	PPIAL4H	ENST00000584068
ENSG00000270629	ENSG00000270629	NBPF14	ENST00000593495
ENSG00000271321	ENSG00000271321	CTAGE6	ENST00000470691
ENSG00000271383	ENSG00000271383	NBPF19	ENST00000426482
ENSG00000271425	ENSG00000271425	NBPF10	ENST00000583866
ENSG00000271449	ENSG00000271449	CT45A2	ENST00000612907
ENSG00000271567	ENSG00000271567	PPIAL4E	ENST00000581164
ENSG00000273136	ENSG00000273136	NBPF26	ENST00000619748
ENSG00000273514	ENSG00000273514	FOXD4L6	ENST00000622588
ENSG00000273590	ENSG00000273590		ENST00000623661
ENSG00000273696	ENSG00000273696	CT45A7	ENST00000620885
ENSG00000273976	ENSG00000273976	GOLGA6L1	ENST00000614055
ENSG00000274252	ENSG00000274252	GGTLC3	ENST00000619998
ENSG00000274274	ENSG00000274274	GAGE13	ENST00000612958
ENSG00000274276	ENSG00000274276		ENST00000617706
ENSG00000274391	ENSG00000274391	TPTE	ENST00000612746
ENSG00000274419	ENSG00000274419	TBC1D3D	ENST00000616101
ENSG00000274512	ENSG00000274512	TBC1D3L	ENST00000617678
ENSG00000274559	ENSG00000274559		ENST00000464664
ENSG00000274611	ENSG00000274611	TBC1D3	ENST00000620215
ENSG00000274808	ENSG00000274808	TBC1D3B	ENST00000611257
ENSG00000274933	ENSG00000274933	TBC1D3I	ENST00000616671
ENSG00000275113	ENSG00000275113	GAGE2E	ENST00000621907
ENSG00000275464	ENSG00000275464		ENST00000632537
ENSG00000275553	ENSG00000275553	ELOA3CP	ENST00000620688
ENSG00000275793	ENSG00000275793	RIMBP3	ENST00000619918
ENSG00000275895	ENSG00000275895		ENST00000636315
ENSG00000275954	ENSG00000275954	TBC1D3F	ENST00000620210
ENSG00000275969	ENSG00000275969	SPATA31A3	ENST00000428649
ENSG00000275993	ENSG00000275993		ENST00000613488
ENSG00000276040	ENSG00000276040	SPATA31A7	ENST00000619167

ENSG00000276076	ENSG00000276076		ENST00000624932
ENSG00000276203	ENSG00000276203	ANKRD20A3P	ENST00000614733
ENSG00000276289	ENSG00000276289		ENST00000618699
ENSG00000276581	ENSG00000276581	SPATA31A5	ENST00000437823
ENSG00000276612	ENSG00000276612		ENST00000623476
ENSG00000277058	ENSG00000277058	HNRNPCL3	ENST00000617807
ENSG00000277117	ENSG00000277117		ENST00000620481
ENSG00000277277	ENSG00000277277		ENST00000613894
ENSG00000277322	ENSG00000277322	GOLGA6L6	ENST00000619213
ENSG00000277865	ENSG00000277865	GOLGA6L22	ENST00000620691
ENSG00000278085	ENSG00000278085	CT45A8	ENST00000611660
ENSG00000278289	ENSG00000278289	CT45A6	ENST00000612878
ENSG00000278299	ENSG00000278299	TBC1D3C	ENST00000622206
ENSG00000278558	ENSG00000278558	TMEM191B	ENST00000613597
ENSG00000278599	ENSG00000278599	TBC1D3E	ENST00000621587
ENSG00000278646	ENSG00000278646		ENST00000613352
ENSG00000278961	ENSG00000278961		ENST00000646133
ENSG00000279408	ENSG00000279408	OR4N4C	ENST00000623441
ENSG00000279493	ENSG00000279493		ENST00000624081
ENSG00000280071	ENSG00000280071		ENST00000623810
ENSG00000280433	ENSG00000280433		ENST00000623998
ENSG00000283378	ENSG00000283378	CNTNAP3C	ENST00000636699
ENSG00000283967	ENSG00000283967	TAF11L8	ENST00000640803
ENSG00000283988	ENSG00000283988	TAF11L9	ENST00000639969
ENSG00000284042	ENSG00000284042	TAF11L6	ENST00000640558
ENSG00000284234	ENSG00000284234	TAF11L5	ENST00000638483
ENSG00000284283	ENSG00000284283	TAF11L4	ENST00000638408
ENSG00000284356	ENSG00000284356	TAF11L10	ENST00000638842
ENSG00000284439	ENSG00000284439	TAF11L3	ENST00000638668
ENSG00000284465	ENSG00000284465	TAF11L7	ENST00000640846
ENSG00000284662	ENSG00000284662	OR4F16	ENST00000332831
ENSG00000284733	ENSG00000284733	OR4F29	ENST00000426406
ENSG00000284906	ENSG00000284906		ENST00000563110
ENSG00000285077	ENSG00000285077	ARHGAP11B	ENST00000428041
ENSG00000285607	ENSG00000285607	FAM90A9	ENST00000648344
ENSG00000285620	ENSG00000285620	FAM90A16	ENST00000648174
ENSG00000285657	ENSG00000285657	FAM90A19	ENST00000647903
ENSG00000285687	ENSG00000285687	FAM90A22	ENST00000648590
ENSG00000285720	ENSG00000285720	FAM90A17	ENST00000650426
ENSG00000285765	ENSG00000285765	FAM90A23	ENST00000648435
ENSG00000285814	ENSG00000285814	FAM90A14	ENST00000648315
ENSG00000285913	ENSG00000285913	FAM90A18	ENST00000647769
ENSG00000285937	ENSG00000285937	FAM90A8	ENST00000647775
ENSG00000285950	ENSG00000285950	FAM90A10	ENST00000533716
ENSG00000285975	ENSG00000285975	FAM90A7	ENST00000650288

Prilog 2. Popis strukturnih varijacija unutar eksona s detaljnim informacijama kod pacijenata CR2-1, CR2-2, CR2-4, ST-1, ST-2, ST-3 i ST-4.

Šifra pacijenta	Kromosom SV	Početak SV	Kraj SV	Duljina SV	Tip SV	Kvaliteta	Filter	Kromosomska pruga	Naziv gena	Transkript	Pomak okvira čitanja	Broj eksona	Lokacija
CR2-1	4	190175199	190176043	-844	DEL	50	GT	q35.2	DUX4	NR_137167	no	6	exon2-intron2
	1	120824880	120829778	-4898	DEL	58	PASS	p11.2	NBPF26	NM_001351372	yes	30	exon15-intron19
	15	23136661	23136715	-54	DEL	45	PASS	q11.2	GOLGA6L1	NM_001001413	yes	9	exon1-exon1
	22	15690291	15690402	-111	DEL	54	GT	q11.1	POTEH	NM_001136213	yes	11	exon1-exon1
	8	12183141	12183243	-102	DEL	31	PASS	p23.1	FAM86B1	NM_001083537	no	7	exon7-exon7
	15	23130374	23130374	82	INS	46	PASS	q11.2	GOLGA6L1	NM_001001413	yes	9	exon8-exon8
	15	23136596	23136596	54	INS	41	GT	q11.2	GOLGA6L1	NM_001001413	yes	9	exon1-exon1
2	110576892	110576892	440	INS	28	PASS	q13	RGPD6	NM_001384364	no	23	exon1-exon1	
CR2-2	Kromosom SV	Početak SV	Kraj SV	Duljina SV	Tip SV	Kvaliteta	Filter	Kromosomska pruga	Naziv gena	Transkript	Pomak okvira čitanja	Broj eksona	Lokacija
	16	28772378	28772441	-63	DEL	53	PASS	p11.2	NPIP9	NM_001287250	yes	7	exon7-exon7
	1	148570081	148573266	-3185	DEL	60	GT	q21.2	NBPF14	NM_015383	yes	66	exon19-intron19
	15	23136661	23136715	-54	DEL	52	PASS	q11.2	GOLGA6L1	NM_001001413	yes	9	exon1-exon1
	8	12183109	12183213	-104	DEL	44	PASS	p23.1	FAM86B1	NM_001083537	no	7	exon7-exon7
	2	110576897	110576897	438	INS	23	PASS	q13	RGPD6	NM_001384364	no	23	exon1-exon1
CR2-4	Kromosom SV	Početak SV	Kraj SV	Duljina SV	Tip SV	Kvaliteta	Filter	Kromosomska pruga	Naziv gena	Transkript	Pomak okvira čitanja	Broj eksona	Lokacija
	16	28772424	28772482	-58	DEL	37	PASS	p11.2	NPIP9	NM_001287250	yes	7	exon7-exon7
	9	67859077	67859130	-53	DEL	31	PASS	q21.11	ANKRD20A1	NM_032250	no	15	exon1-exon1
	1	120824913	120829811	-4898	DEL	54	PASS	p11.2	NBPF26	NM_001351372	yes	30	exon15-intron19
	1	146074376	146077571	-3195	DEL	56	PASS	q21.1	NBPF10	NM_001302371	no	90	exon76-exon80
	1	146079126	146082316	-3190	DEL	56	PASS	q21.1	NBPF10	NM_001039703	yes	86	exon70-exon74
	15	23136661	23136714	-53	DEL	42	PASS	q11.2	GOLGA6L1	NM_001001413	no	9	exon1-exon1
	8	12183090	12183194	-104	DEL	35	PASS	p23.1	FAM86B1	NM_001083537	no	7	exon7-exon7
	16	69963926	69963926	6868	INS	22	PASS	q22.1	CLEC18A	NM_001136214	no	13	exon13-exon13
ST-1	Kromosom SV	Početak SV	Kraj SV	Duljina SV	Tip SV	Kvaliteta	Filter	Kromosomska pruga	Naziv gena	Transkript	Pomak okvira čitanja	Broj eksona	Lokacija
	1	149537189	149540389	-3200	DEL	55	GT	q21.2	NBPF19	NM_001351365	no	93	exon71-exon75
	1	146074350	146077546	-3196	DEL	55	PASS	q21.1	NBPF10	NM_001302371	yes	90	exon76-exon80
	1	146079111	146082301	-3190	DEL	60	GT	q21.1	NBPF10	NM_001039703	yes	86	exon70-exon74
	1	148567860	148569402	-1542	DEL	60	PASS	q21.2	NBPF14	NM_015383	yes	66	exon20-exon22
	8	7556987	7557353	-366	DEL	54	PASS	p23.1	FAM90A7P	NR_046343	no	3	exon3-exon3
	8	12183091	12183194	-103	DEL	28	PASS	p23.1	FAM86B1	NM_001083537	no	7	exon7-exon7
	15	30648619	30648619	288	INS	60	PASS	q13.2	ARHGAP11B	NR_148423	no	11	exon11-exon11
	15	20534807	20534876	-69	DEL	27	PASS	q11.2	GOLGA6L6	NM_001145004	yes	9	exon8-exon8
	15	23136639	23136692	-53	DEL	60	GT	q11.2	GOLGA6L1	NM_001001413	no	9	exon1-exon1
	22	15690427	15690537	-110	DEL	44	PASS	q11.1	POTEH	NM_001136213	no	11	exon1-exon1
ST-2	Kromosom SV	Početak SV	Kraj SV	Duljina SV	Tip SV	Kvaliteta	Filter	Kromosomska pruga	Naziv gena	Transkript	Pomak okvira čitanja	Broj eksona	Lokacija
	1	149537131	149540331	-3200	DEL	37	GT	q21.2	NBPF19	NM_001351365	no	93	exon71-exon75
	1	146074350	146077546	-3196	DEL	55	PASS	q21.1	NBPF10	NM_001302371	yes	90	exon76-exon80
	15	23136663	23136715	-52	DEL	44	PASS	q11.2	GOLGA6L1	NM_001001413	yes	9	exon1-exon1
	8	12183090	12183194	-104	DEL	40	PASS	p23.1	FAM86B1	NM_001083537	no	7	exon7-exon7
ST-3	Kromosom SV	Početak SV	Kraj SV	Duljina SV	Tip SV	Kvaliteta	Filter	Kromosomska pruga	Naziv gena	Transkript	Pomak okvira čitanja	Broj eksona	Lokacija
	1	146074356	146077553	-3197	DEL	58	PASS	q21.1	NBPF10	NM_001302371	yes	90	exon76-exon80
	15	20534807	20534876	-69	DEL	26	PASS	q11.2	GOLGA6L6	NM_001145004	yes	9	exon8-exon8
	8	12183090	12183194	-104	DEL	35	PASS	p23.1	FAM86B1	NM_001083537	no	7	exon7-exon7
	15	23129676	23129676	126	INS	60	GT	q11.2	GOLGA6L1	NM_001001413	yes	9	exon8-exon8
ST-4	Kromosom SV	Početak SV	Kraj SV	Duljina SV	Tip SV	Kvaliteta	Filter	Kromosomska pruga	Naziv gena	Transkript	Pomak okvira čitanja	Broj eksona	Lokacija
	1	120824898	120829796	-4898	DEL	60	PASS	p11.2	NBPF26	NM_001351372	yes	30	exon15-intron19
	15	20534339	20534507	-168	DEL	46	PASS	q11.2	GOLGA6L6	NM_001145004	yes	9	exon8-exon8
	22	15690422	15690533	-111	DEL	44	PASS	q11.1	POTEH	NM_001136213	yes	11	exon1-exon1
	8	12183090	12183194	-104	DEL	42	PASS	p23.1	FAM86B1	NM_001083537	no	7	exon7-exon7

Prilog 3. Popis strukturnih varijacija većih od 800 parova baza kod pacijenata CR2-1, CR2-2, CR2-4, ST-1, ST-2, ST-3 i ST-4.

Šifra pacijenta	Kromosom SV	Početak SV	Kraj SV	Duljina SV	Tip SV	Broj gena
CR2-1	7	68170787	98453528	30282741	INV	222
	4	190175199	190176043	-844	DEL	1
	1	120835013	120836007	-994	DEL	1
	1	149551318	149552878	-1560	DEL	1
	1	146066843	146069986	-3143	DEL	1
	1	146088693	146091871	-3178	DEL	1
	1	146083942	146087128	-3186	DEL	1
	8	7582317	7585973	-3656	DEL	0
	1	120824880	120829778	-4898	DEL	1
	1	146136990	146143286	-6296	DEL	1
	5	70072862	70079172	-6310	DEL	2
CR2-2	7	68170787	98453528	30282741	INV	222
	17	36109418	36187342	77924	DUP	6
	16	14894753	14937510	42757	INV	23
	2	86936276	86936276	3576	INS	2
	1	120835013	120836007	-994	DEL	1
	1	120827523	120829341	-1818	DEL	1
	1	146066760	146069910	-3150	DEL	1
	1	148570081	148573266	-3185	DEL	1
	1	120824840	120829738	-4898	DEL	1
	16	28698464	28709344	-10880	DEL	1
	9	66129156	66144584	-15428	DEL	0
CR2-4	X	52433310	55652522	3219212	INV	52
	X	141012592	141244560	231968	INV	7
	16	14894753	14937510	42757	INV	23
	2	86936276	86936276	7094	INS	2
	16	69963926	69963926	6868	INS	1
	16	69962420	69962420	6846	INS	1
	1	120812372	120812372	6307	INS	1
	2	86936276	86936276	3581	INS	2
	1	120835026	120836019	-993	DEL	1
	1	120827570	120829388	-1818	DEL	1
	1	120833985	120836007	-2022	DEL	1
	1	146066773	146069916	-3143	DEL	1
	1	146079126	146082316	-3190	DEL	1
	1	146074376	146077571	-3195	DEL	1
	8	12135286	12140022	-4736	DEL	2
	1	120824913	120829811	-4898	DEL	1
	1	146136990	146143286	-6296	DEL	1
15	21642492	21653114	-10622	DEL	1	

Šifra pacijenta	Kromosom SV	Početak SV	Kraj SV	Duljina SV	Tip SV	Broj gena
ST-1	7	68170787	98453528	30282741	INV	222
	16	14894753	14937510	42757	INV	23
	X	52510824	52523688	12864	DUP	2
	2	86936276	86936276	3580	INS	
	2	86936276	86936276	3580	INS	2
	1	120835013	120836007	-994	DEL	
	1	120835013	120836007	-994	DEL	1
	1	148567860	148569402	-1542	DEL	
	1	148567860	148569402	-1542	DEL	1
	1	120827523	120829341	-1818	DEL	
	1	120827523	120829341	-1818	DEL	1
	1	146067005	146070148	-3143	DEL	
	1	146067005	146070148	-3143	DEL	1
	1	146088711	146091885	-3174	DEL	
	1	146088711	146091885	-3174	DEL	1
	1	146083915	146087102	-3187	DEL	
	1	146083915	146087102	-3187	DEL	1
	1	146079111	146082301	-3190	DEL	
	1	146079111	146082301	-3190	DEL	1
	1	146074350	146077546	-3196	DEL	
	1	146074350	146077546	-3196	DEL	1
	1	149537189	149540389	-3200	DEL	1
	1	149537189	149540389	-3200	DEL	
	1	120824840	120829738	-4898	DEL	
	1	120824840	120829738	-4898	DEL	1
	1	146136990	146143286	-6296	DEL	
	1	146136990	146143286	-6296	DEL	1
	15	21642491	21653114	-10623	DEL	1
	15	21642491	21653114	-10623	DEL	
	ST-2	7	68170787	98453528	30282741	INV
16		32285314	33340036	1054722	DUP	9
16		14894753	14937510	42757	INV	23
X		52502066	52523688	21622	INV	2
1		120835035	120836027	-992	DEL	1
1		146066770	146069910	-3140	DEL	1
1		149546552	149549737	-3185	DEL	1
1		146074350	146077546	-3196	DEL	1
1		149537131	149540331	-3200	DEL	1
1		146136990	146143286	-6296	DEL	1
5		70072862	70079172	-6310	DEL	2

Šifra pacijenta	Kromosom SV	Početak SV	Kraj SV	Duljina SV	Tip SV	Broj gena
ST-3	7	68170787	98453528	30282741	INV	222
	8	7607927	12611663	5003736	INV	85
	7	143742323	143874697	132374	DUP	5
	X	52489208	52502073	12865	DUP	2
	1	146066769	146069916	-3147	DEL	1
	1	146074356	146077553	-3197	DEL	1
	1	146136990	146143286	-6296	DEL	1
ST-4	1	16680842	146052340	129371498	INV	1217
	7	68170787	98453528	30282741	INV	222
	16	69964152	69964152	6867	INS	0
	2	86936276	86936276	3589	INS	2
	1	120827622	120829439	-1817	DEL	1
	1	146067068	146070210	-3142	DEL	1
	1	120824898	120829796	-4898	DEL	1
	15	21642491	21653114	-10623	DEL	1
	9	66129156	66144584	-15428	DEL	0
	7	65509484	126874698	-61365214	DEL	523