

# Reevaluacija genetičkih nalaza dobivenih kromosomskim čipovima kod bolesnika s neurorazvojnim poremećajima

---

**Bogner, Ana-Marija**

**Master's thesis / Diplomski rad**

**2023**

*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:217:218150>

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

*Download date / Datum preuzimanja:* **2024-12-01**



*Repository / Repozitorij:*

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



Sveučilište u Zagrebu  
Prirodoslovno-matematički fakultet  
Biološki odsjek

Ana-Marija Bogner

**Reevaluacija genetičkih nalaza dobivenih  
kromosomskim čipovima kod bolesnika s  
neurorazvojnim poremećajima**

Diplomski rad

Zagreb, 2023.

University of Zagreb  
Faculty of Science  
Department of Biology

Ana-Marija Bogner

**Reevaluation of genetic findings obtained by  
chromosomal chips for patients with  
neurodevelopmental disorders**

Master thesis

Zagreb, 2023.

Ovaj rad je izrađen u Laboratoriju za medicinsku genetiku na Hrvatskom institutu za istraživanje mozga Medicinskog fakulteta, Sveučilišta u Zagrebu pod mentorstvom dr. sc. Martine Rinčić, više znanstvene suradnice, te komentorstvom izv. prof. dr. sc. Petre Korać. Rad je predan na ocjenu Biološkom odsjeku Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu radi stjecanja zvanja magistra molekularne biologije.

## ZAHVALE

Prvo se iskreno želim zahvaliti svojoj mentorici dr.sc. Martini Rinčić na davanju šanse, susretljivosti, iznimnom strpljenju, mentorstvu i svim savjetima danim tijekom izrade i pisanja ovog diplomskog rada. Ogromno hvala i komentorici izv.prof.dr.sc. Petra Korać na spremnosti za preuzimanje uloge, pomoći i razumijevanju.

Posebno se zahvaljujem cijeloj svojoj obitelji na podršci od osnovnoškolskih dana. Hvala Vam što ste trpili svaki moj živčani slom, što ste me uveseljavali i bodrili.

Posebno mjesto i zahvalu zaslužuju i moji prijatelji, *Dušmani*, *Bake* i *HAKK*.

*Dušmani*, hvala Vam što ste obilježili ovaj period mog života neizmjernom količinom smijeha, pomoći, odgovorima na glupa pitanja, razgovorima, raspravama, druženjima, kavama i izletima. *Bake*, hvala na podršci od mladih ludih dana, za svaku avanturu i kafu. Poznajte me dublje nego što ja poznajem samu sebe!

I šta sad? Sad zovi *HAKK*!

*Za Miška.*

---

# TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu  
Prirodoslovno-matematički fakultet  
Biološki odsjek

Diplomski rad

## Reevaluacija genetičkih nalaza dobivenih kromosomskim čipovima kod bolesnika s neurorazvojnim poremećajima

Ana-Marija Bogner

Rooseveltov trg 6, 10000 Zagreb, Hrvatska

Genetički testovi u širokoj su upotrebi već dugi niz godina, ali uvođenjem genetičkih testova visoke rezolucije nastala je prekretnica u medicinskoj genetici. Za ispitanike s neurorazvojnim poremećajima najčešće korišteni genetički test u proteklom desetljeću je analiza kromosoma na mikročipovima. Iako nam danas genetički testovi visoke rezolucije omogućavaju otkrivanje vrlo malih promjena u ljudskom genomu, određivanje patogenosti detektiranih varijanti ostaje izazov u donošenju dijagnostičkih odluka. Novim spoznajama dolazi do promjena u razumijevanja ljudskog genoma zbog čega je interpretacija genetičkih nalaza podložna promjeni tijekom vremena. Ovim radom detektirane su promjene u interpretaciji genetičkih nalaza dobivenih kromosomskim mikročipovima kod ispitanika s neurorazvojnim poremećajima uzimajući u obzir nove znanstvene spoznaje. Reevaluirani su rezultati inicijalno klasificirani kao varijante nesigurne važnosti koristeći javno dostupne baze podataka. Optimiziran je hodogram reevaluacije podataka i dobivena lista 29 potencijalnih kandidata gena koji su od kliničke važnosti za razvoj neurorazvojnih poremećaja. Istraženi su biološki putevi u koje su geni-kandidati uključeni te je zabilježena jedna lncRNA povišene razine ekspresije u mozgu. Ponovnom analizom podataka uspostavljena je dosad nepoznata veza između detektiranih promjena u broju kopija DNA i fenotipa na osnovi novih znanstvenih saznanja o ljudskom genomu.

**Ključne riječi:** promjene broja kopija, CNV, CMA, NDD

(65 stranica, 12 slika, 5 tablica, 84 literaturnih navoda, jezik izvornika: hrvatski)

Rad je pohranjen u Središnjoj biološkoj knjižnici

Mentor: dr.sc. Martina Rinčić

Komentor: izv.prof.dr.sc. Petra Korać

Ocjenitelji:

izv.prof.dr.sc. Petra Korać

izv.prof.dr.sc. Inga Urlić

izv.prof.dr.sc. Maja Matulić

Rad prihvaćen: 07.09.2023.

# BASIC DOCUMENTATION CARD

---

University of Zagreb  
Faculty of Science  
Department of Biology

Master thesis

Reevaluation of genetic findings obtained by chromosomal chips for patients with  
neurodevelopmental disorders

Ana-Marija Bogner

Rooseveltova trg 6, 10000 Zagreb, Croatia

Genetic testing has been widely used for many years, but introducing high-resolution genetic methods made a turning point in medical genetics. For patients with neurodevelopmental disorders, the most commonly used genetic test in the past decade has been chromosomal microarray analysis. Although today's high-resolution genetic tests detect very small changes in the human genome, determining the pathogenicity of detected variants remains challenging in diagnostic decision-making. Accumulating information and a better understanding of the human genome resulted in the interpretation of genetic findings being subject to change over time. Considering new scientific knowledge, this work detected changes in the interpretation of genetic findings obtained by chromosomal microarray in subjects with neurodevelopmental disorders. The results initially classified as of uncertain significance have been re-evaluated using publicly available databases. A workflow was optimized, and a list of 29 potential candidate genes clinically relevant for neurodevelopmental disorders was obtained. The biological pathways in which the candidates are involved have been investigated, and one lncRNA has been recorded with elevated expression in the brain. Reanalysis of the data has established a previously unknown link between the detected changes in DNA copy number and the phenotype based on new scientific knowledge about the human genome.

Keywords: copy number variant, CNV, NDD, CMA  
(65 pages, 12 figures, 5 tables, 84 references, original in: Croatian)  
Thesis is deposited in Central Biological Library.

Mentor: Martina Rinčić, PhD  
Komentor: Assoc. Prof. Petra Korac

Reviewers:

Assoc. Prof. Petra Korac  
Assoc. Prof. Inga Urlic  
Assoc. Prof. Maja Matulic

Thesis accepted: 07.09.2023.

## SADRŽAJ

<b>1.</b>	<b>UVOD</b> .....	<b>1</b>
1.1	LJUDSKI GENOM.....	1
1.2	PROMJENA U BROJU KOPIJA DNA.....	3
1.3	GENETIČKO TESTIRANJE .....	6
1.4	ANALIZA KROMOSOMA NA MIKROČIPOVIMA .....	6
1.5	METODA ISTOVREMENOG UMNAŽANJA VEZANIH PROBA .....	8
1.6	NEURORAZVOJNI POREMEĆAJI .....	10
1.7	NEURORAZVOJNI POREMEĆAJI I PROMJENE BROJA KOPIJA DNA .....	11
1.8	REEVALUACIJA REZULTATA .....	14
<b>2.</b>	<b>CILJEVI ISTRAŽIVANJA</b> .....	<b>16</b>
<b>3.</b>	<b>MATERIJALI I METODE</b> .....	<b>17</b>
3.1	UZORCI .....	17
3.2	BAZE PODATAKA .....	18
3.3	ANOTACIJA PROMJENA BROJA KOPIJA NA NOVU INAČICU GENOMA .....	25
3.4	ISTRAŽIVANJE POTENCIJALNIH GENA KANDIDATA .....	26
3.5	PROVJERA EKSPRESIJE IZABRANIH GENA .....	27
3.6	ODGOVARAJUĆI MODELNI ORGANIZMI.....	28
3.7	ONTOLOGIJA GENA .....	28
3.8	OSNOVNE INFORMACIJE O GENIMA I DEFINIRANJE KLINIČKE VAŽNOSTI .....	28
3.9	NEKODIRAJUĆE SEKVENCE UNUTAR DETEKTIRANIH ABERACIJA.....	29
3.10	<i>DATABASE OF GENOMIC VARIANTS</i> .....	29
3.11	VIZUALIZACIJA SIGNALNIH I METABOLIČKIH PUTEVA .....	29
3.12	LITERATURNI PREGLED.....	30
<b>4.</b>	<b>REZULTATI</b> .....	<b>31</b>
4.1	USPJEŠNOST ANOTACIJE NA NOVU INAČICU GENOMA.....	31
4.2	KANDIDATSKI GENI .....	31
4.3	ONTOLOGIJA GENA .....	32
4.4	EKSPRESIJA KANDIDATSKIH GENA I ODGOVARAJUĆI MIŠJI MODELI .....	32
4.5	KLINIČKA VAŽNOST .....	32
4.6	NEKODIRAJUĆE SEKVENCE UNUTAR DETEKTIRANIH ABERACIJA.....	35
4.7	STRUKTURNE ABERACIJE PRISUTNE U OPĆOJ POPULACIJI .....	36
4.8	SIGNALNI I METABOLIČKI PUTOVI .....	36



4.9	LITERATURNI PREGLED I OSNOVNE INFORMACIJE O GENIMA .....	45
4.10	KANDIDATSKI GENI .....	45
4.11	HODOGRAM REEVALUACIJE PODATAKA .....	47
<b>5.</b>	<b>RASPRAVA .....</b>	<b>48</b>
5.1	ANOTACIJA ABERACIJA NA NOVU INAČICU GENOMA.....	48
5.2	OPTIMIZACIJA HODOGRAMA.....	49
5.3	PRENATALNA AKTIVNOST GENA .....	49
5.4	INTERAKCIJE PROTEINA .....	51
5.5	NEURORAZVOJNI POREMEĆAJI, IMUNOSNI SUSTAV I TUMORI.....	52
5.6	DUGE NEKODIRAJUĆE MOLEKULE RNA .....	53
<b>6.</b>	<b>ZAKLJUČAK.....</b>	<b>55</b>
<b>7.</b>	<b>LITERATURA.....</b>	<b>56</b>

## POPIS KRATICA

aCGH – komparativna genomska hibridizacija na mikročipu, od engl. *Array comparative genomic hybridization*

ADHD - poremećaj sa deficitom pažnje i hiperaktivnošću, od engl. *Attention deficit/hyperactivity disorder*

ASD – poremećaj iz spektra autizma, od engl. *Autism spectrum disorder*

ClinGen – Baza *Clinical Genome Resource*

CMA – analiza kromosoma na mikročipovima, od engl. *Chromosomal microchip analysis*

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

DD/ID – zastoj u razvoju/intelektualne poteškoće, od engl. *Developmental delay/intellectual disability*

DECIPHER – baza *DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources*

DGV – projekt *Database of Genomic Variants*

FISH – fluorescentna hibridizacija *in situ*, od engl. *Fluorescence in situ hybridization*

GTE<sub>x</sub> – projekt *The Genotype-Tissue Expression*

HBT – baza podataka *The Human Brain Transcriptome*

HPA – baza podataka *The Human Protein Atlas*

HPO – baza podataka *Human Phenotype Ontology*

LOEUF – opaženi gubitak funkcije gena, od engl. *Loss of function Observed*

MGI – baza podataka *Mouse Genome Informatics*

MLPA – metoda istovremenog umnažanja vezanih proba, od engl. *Multiplex Ligation-dependent Probe Amplification assay*

NCBI - *National Center for Biotechnology Information*

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

NDD-CNV – promjene broja kopija povezane s povećanim rizikom za nastanak neurorazvojnih poremećaja

OMIM – *Online Mendelian Inheritance in Man*

PANTHER – softverski sustav *Protein ANalysis THrough Evolutionary Relationships*

PCR – lančana reakcija polimerazom, od engl. *Polymerase Chain Reaction*

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

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

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

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

SIGNOR – baza podataka *The SIGnaling Network Open Resource*

UCSC – mrežni alat *The University of California Santa Cruz Genome Browser*

UniProt – baza podataka *The Universal Protein Resource*

VUS – varijante nepoznate/nesigurne važnosti, od engl. *variants of uncertain significance*

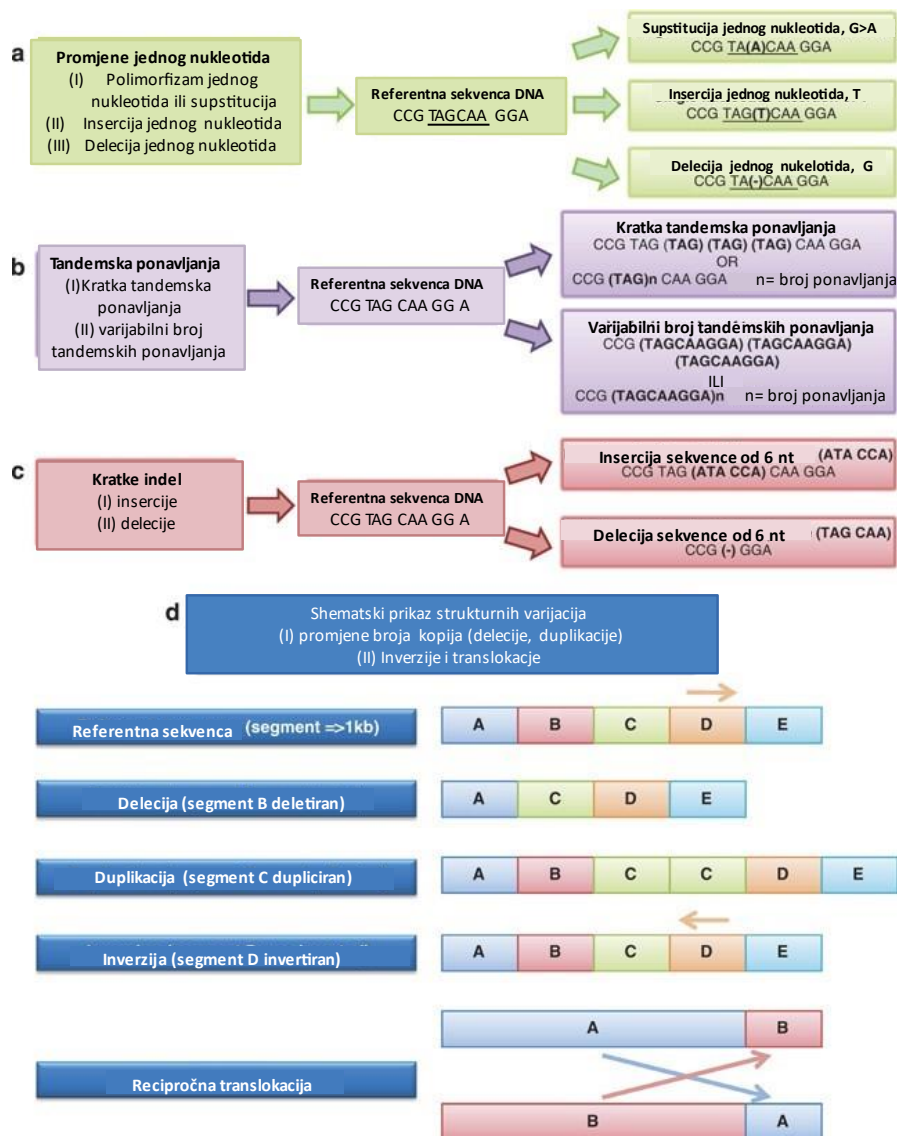
## 1. UVOD

Teorije i studije o ljudskoj genetici imaju dugu povijest koja seže i do starogrčke literature gdje se mogu naći zapisi o nasljeđivanju fizičkih osobina kod ljudi. Već su u osamnaestom i devetnaestom stoljeću objavljena zapažanja o nasljeđivanju brojnih bolesti. Međutim, povijest humane genetike temeljene na znanstvenim otkrićima započinje 1865. godine objavom Mendelovih zakona nasljeđivanja. Vrlo važan korak u razvoju humane genetike u medicini napravljen je početkom 20. stoljeća kada se dolazi do spoznaje da Mendelovi zakoni mogu objasniti nasljeđivanje alkaptonurije i drugih urođenih bolesti metabolizma kod čovjeka. Kroz zadnjih 100 godina, medicinska genetika je prerasla iz subspecijalnosti koju zanima nekoliko rijetkih nasljednih poremećaja u prepoznatu medicinsku specijalnost. Danas je medicinska genetika neophodna u dijagnostici i liječenju mnogih poremećaja, kako onih rijetkih tako i onih učestalih (Nussbaum i sur. 2007). Procjenjuje se da danas na području Europske unije 36 milijuna ljudi boluje od rijetkih bolesti. Među njima je do 50 posto djece za čiju su dijagnostiku i liječenje neophodna znanja medicinske genetike (health.ec.europa.eu).

### 1.1 LJUDSKI GENOM

Kompletni sadržaj ljudskog genoma definiran je kao potpuna suma genetičke informacije svakog pojedinca. Ljudski genom nije uniforman. Osim jednojajčanih (monozigotnih) blizanaca, ne postoje dva čovjeka na Zemlji koja dijele potpuno isti genom. Nadalje, ljudski genom nije statičan. Završetkom projekta *Human Genome Project*, koji je za cilj imao odrediti redosljed sekvence ljudskog genoma omogućeno je njegovo istraživanje kao cjelokupnog entiteta, a ne gena po gena. Znanje kompletne sekvence ljudske DNA omogućava identifikaciju svih ljudskih gena, određivanje mjere varijabilnosti gena u različitim populacijama i prepoznavanje kako promjene u genima doprinose zdravlju i bolesti. Jedna od glavnih značajki genetike je rasvjetljavanje uloge genetičke varijabilnosti i mutacija kao predispozicija za nastanak bolesti, modifikaciji tijeka bolesti i uzrokovanju same bolesti. Skoro svaka bolest rezultat je međudjelovanja gena i okoliša, a doprinos genetičke komponente može varirati (Nussbaum i sur. 2007). Genetičku varijabilnost definiramo kao promjene u sekvenci DNA unutar genoma pojedinaca u populaciji (Ku i sur. 2010). Zbog velike raznolikosti u ljudskom genomu, genetička varijabilnost smatra se parametrom koji kontrolira fenotip pojedinca. Promjene u genomu mogu zahvaćati dijelove DNA različitih veličina, od nekoliko nukleotida do više megabaza i pojavljivati se u različitim oblicima poput polimorfizama jednog nukleotida, uzastopnih ponavljanja, malih insercija i delecija (indels), amplifikacija ili delecija koje izazivaju promjenu broja kopija

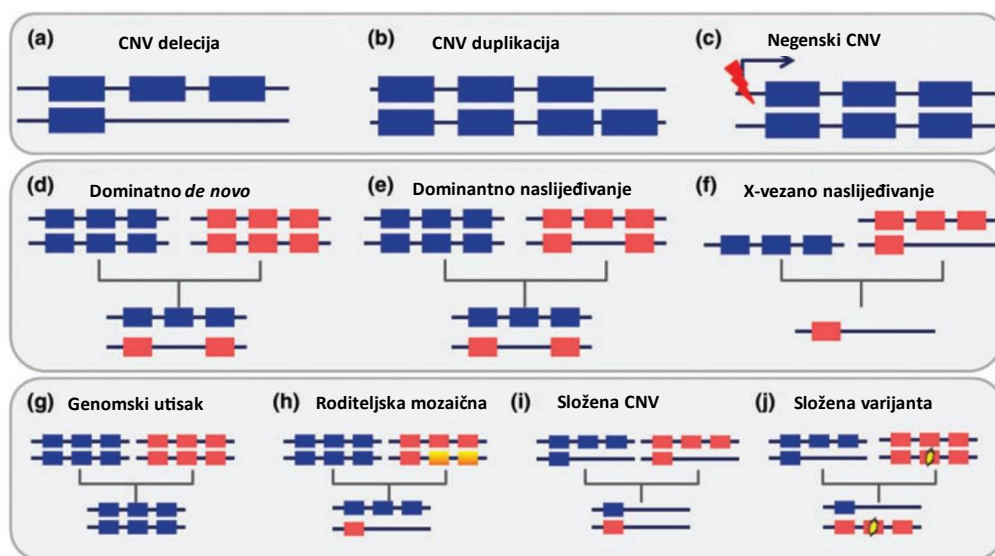
većeg segmenta DNA sekvence (CNV, od engl. *copy number variants*), kromosomskih rearanžmana poput inverzija i translokacija, *copy neutral* gubitka heterozigotnosti ili homozigotnosti, transpozona, fragilnih mjesta i *null* alela (Slika 1). Promjene u ljudskom genomu se prirodno pojavljuju i predstavljaju ostatke grešaka koje su se dogodile prilikom replikacije DNA ili su inducirane od strane vanjskih faktora poput virusa, slobodnih radikala, UV zračenja i kemijskih mutagena (Bente i Scott 2011, Ku i sur. 2010).



Slika 1. Shematski prikaz nekkih od oblika genetičke varijabilnosti: promjene jednog nukleotida (a), tandemskih ponavljanja (b), kratkih indels (c) i strukturnih promjena (d). Preuzeto i prilagođeno iz Ku i sur. 2010.

## 1.2 PROMJENA U BROJU KOPIJA DNA

Promjene u broja kopija DNA odnose se na vrstu genetske varijacije koja uključuje promjene u broju kopija određenog segmenta DNA u genomu pojedinca. Pojedinačne varijante mogu biti kratke ili uključivati tisuće baza. CNV-ovi mogu imati alele koji su tandem duplikacije 2, 3, 4 ili više identičnih kopija DNA segmenta uzastopno. Ako je učestalost CNV-a u populaciji  $>1\%$  onda ga smatramo polimorfizmom, bilo da je prisutan kao višak ili manjak određenog segmenta DNA (Nussbaum i sur. 2007). CNV-ovi su uobičajeno prisutni u genomu ljudi i modelnih organizama poput miša, štakora i vinske mušice. Kod ljudi je za sad pronađeno više od 2.5 milijuna različitih CNV-ova mapiranih na približno 200 000 genomskih regija što utvrđuje njihov značajni doprinos genetičkoj varijabilnosti (Heckel i sur. 2015, Macdonald i sur. 2013). Uloga CNV-ova u genetičkim sindromima odavno je poznata, s ponavljajućim mikrodelecijama/mikroduplicacijama koje su detektirane kod sindroma poput Prader-Willi, Smith-Magenis i Williams-Beuren (Cahan i sur. 2009, Henrichsen i sur. 2009). CNV-ovi mogu imati različit doprinos sindromima, a modeli kojima se objašnjava utjecaj CNV-ova na ispoljavanje fenotipa bolesti su; pojava *de novo*, X-vezano i recesivno nasljeđivanje, mozaicizam, genomski utisak i digensko nasljeđivanje (Slika 2) (Hehir-Kwa i sur. 2013).



Preuzeto i prilagođeno iz Hehir-Kwa i sur. 2013.

CNV-ove možemo klasificirati na temelju nekoliko karakteristika kao što su sadržaj gena, veličina ili pozicija CNV-a, broj kopija te klinička značajnost.

S obzirom na napredak tehnologije koja omogućava detekciju sve većeg broja sve manjih CNV-ova, u medicinskoj genetici od presudne je važnosti definirati njihovo kliničko značenje. S obzirom na posljedice koje možemo očekivati kod pacijenata koji imaju CNV-ove možemo ih kategorizirati kao benigne varijante na jednoj strani spektra te patogene varijante na drugoj strani spektra. Kao prethodno navedeno, CNV-ovi su važan izvor normalne i patogene varijabilnosti. Patogeni CNV su tipično veći i obuhvaćaju više gena, značajno su obogaćeni genima uključenima u razvoj i onima s većom evolucijskom konzerviranošću broja kopija kod sisavaca. Kod benignih CNV-ova pronađeni geni imaju varijabilnije brojeve kopija što upućuje na to da je dominantni uzročnik patogenosti CNV-a osjetljivost gena na dozu (Pös i sur. 2021, Riggs i sur. 2019).

Kako bi se pravilno mogle razvrstati i interpretirati CNV-ovi razvijen je uniformni sustav preporuka od strane *American College of Medical Genetics and Genomics (ACMG)* i *Clinical Genome Resources (ClinGen)* prema kojima se CNV-ove razvrstava u 5 kategorija: benigne, vjerojatno benigne, varijante nepoznate/nesigurne važnosti (VUS, od engl. *variant of uncertain significance*), vjerojatno patogene i patogene varijante. Sam napredak u detekciji CNV-ova doveo je do velikog broja novih CNV-ova za čije određivanje kliničkog učinka su potrebne daljnje analize.

1. **Patogene varijante** – smatra se da su povezane s bolestima, uključuju CNV-ove koje objašnjavaju pacijentov fenotip, povezane su sa statusom nositelja recesivnog stanja i one koje indiciraju rizik za nastanak bolesti nepovezanog fenotipa. Patogeni CNV-ovi koji objašnjavaju neurorazvojnu povijest pacijenta mogu uključivati opetovane delecije ili duplikacije poput 22q11.2 delecijskog sindroma ili 17p11.2 mikrodelecije koja uzrokuje Smith-Magenis sindrom te novonastale delecije ili amplifikacije koje utječu na gene osjetljive na dozu.
2. **Vjerojatno patogene varijante** – uključuju CNV-ovi za koje postoji dovoljno dokaza koji sugeriraju povezanost s bolesti, ali nema dodatnih dokaza koji bi mogli pojasniti patogenost.
3. **Varijante nepoznate/nesigurne važnosti** – uključuju CNV-ove za koje nema dovoljno dokaza da bi se utvrdilo jesu li patogene ili benigne.
4. **Vjerojatno benigne varijante** – uključuje CNV-ove za koje postoje dokazi koji sugeriraju da nisu povezane s bolestima, ali je potrebno još dokaza za razjašnjenje.

5. **Benigne varijante** – uključuju CNV-ove za koje se smatra da nisu povezane s bolestima, često su zastupljene u općoj populaciji ili detektirane kod pojedincima bez bolesnog fenotipa (Savatt i Myers, 2021).



### 1.3 GENETIČKO TESTIRANJE

Genetički testovi su u širokoj upotrebi već dugi niz godina, a od 1960-ih konvencionalna analiza kariotipa G-pruganjem postala je zlatni standard za detekciju genetičkih etiologija pacijenata s neobjašnjivim zastajanjem u razvoju/intelektualnim poteškoćama i višestrukim kongenitalnim anomalijama. Glavni nedostatak te metode je nemogućnost vizualizacije kromosomskih promjena manjih od 5 Mb, dok je glavna prednost mogućnost prikaza strukture genoma. Razvojem metode analize kromosoma na mikročipovima došlo je do revolucije u polju kliničke citogenetike zbog povećanja rezolucije koje je doprinijelo otkrivanju sve suptilnijih sub-kromosomskih mikrodelecija i mikroduplikacija koje mijenjaju više od 50 pb, a manje od 5 Mb. Ako su ponavljajuće mikrodelecije i mikroduplikacije povezane sa specifičnom kliničkom slikom, nazivaju se mikrodelecijski/mikroduplikacijski sindromi. Danas u literaturi postoji definirano više od 200 mikrodelecijskih, te ~ 80 mikroduplikacijskih sindroma (MMS) (Highsmith Jr. 2013, Wetzel i Darbro 2022.). Metode genetičkog testiranja koje omogućavaju otkrivanje MMS-ova su FISH (fluorescentna *in situ* hibridizacija), kromosomski mikročipovi, te sekvenciranje sljedeće generacije. Metoda FISH omogućava analizu specifičnih regija u genomu – ciljano testiranje, dok kromosomski mikročipovi i sekvenciranje sljedeće generacije omogućavaju detekciju promjena u broju kopija DNA cijelom dužinom genoma. U zadnjih desetak godina analiza kromosoma na mikročipovima postala je metoda izbora genetičkog testiranja velikog broja pacijenata jer pruža pretragu cjelokupnog genoma u visokoj rezoluciji uz značajno manji trošak od sekvenciranja sljedeće generacije.

### 1.4 ANALIZA KROMOSOMA NA MIKROČIPOVIMA

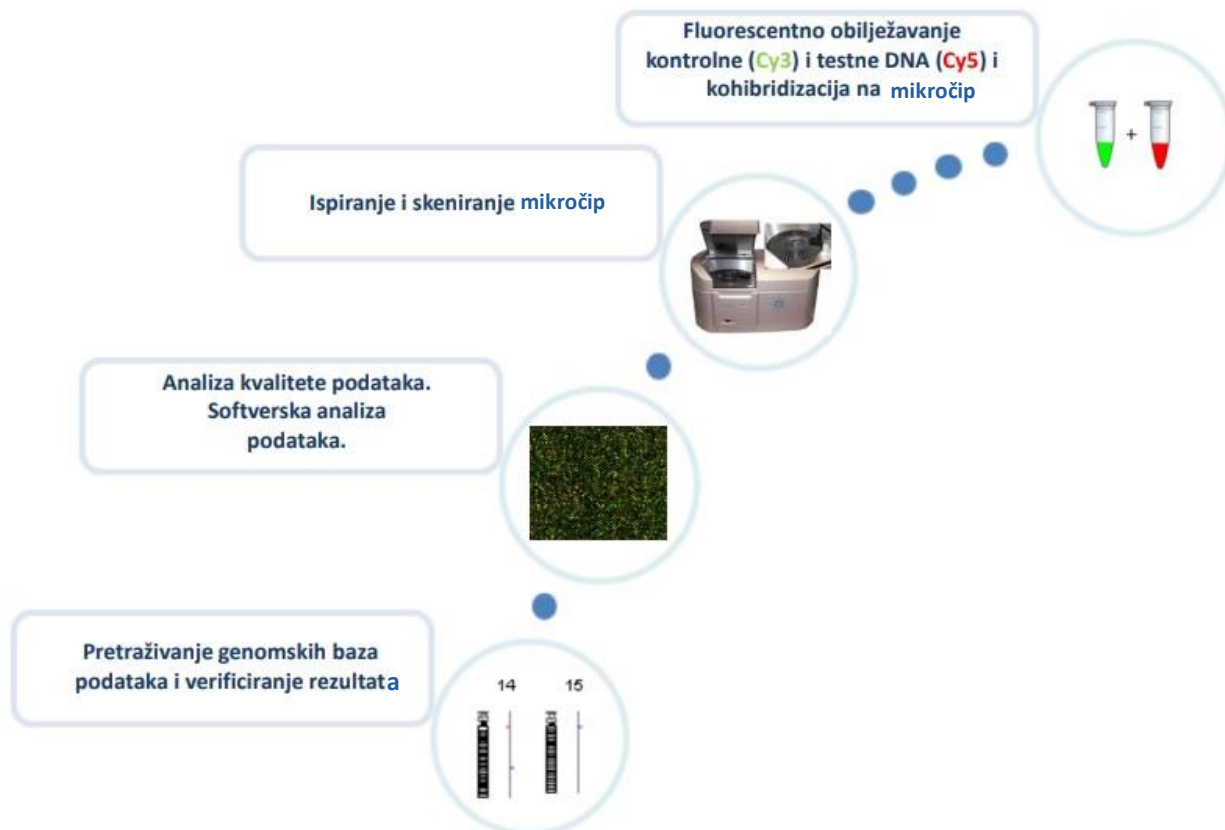
Analiza kromosoma na mikročipovima (molekularna kariotipizacija) je molekularna citogenetička metoda koja se temelji na metodi komparativne genomske hibridizacije (Highsmith Jr. 2013).

Metoda CMA (analiza kromosoma na mikročipovima, od engl. *Chromosomal microchip analysis*) trenutno predstavlja primarni dijagnostički test za višestruke kongenitalne anomalije (MCA, od engl. *Multiple congenital anomalies*), zastoj u razvoju/intelektualne poteškoće (DD/ID, od engl. *Developmental delay/intellectual disability*) i poremećaje iz spektra autizma (ASD, od engl. *Autism spectrum disorder*) (Henderson i sur. 2014).

Uvođenje tehnike mikročipova dovodi do napretka na području rezolucije detekcije genomskih aberacija jer omogućava pretraživanje čitavog genoma te detekciju suptilnih submikroskopskih aberacija značajno manjih od 5 Mb. Osnove komparativne genomske hibridizacije na mikročipu postavljene su 1997. objavom prvih radova u kojima se opisuje protokol tehnike *Matrix DNA*,

kohibridizacije cjelokupne genomske DNA pacijenta i kontrolne DNA obilježene različitim fluorescentno obilježenim nukleotidima na staklenu pločicu koja sadrži imobilizirane probe. Daljnjim napretkom tehnike, rezolucija aCGH (komparativna genomska hibridizacija na mikročipu, od engl. *Array comparative genomic hybridization*) za detekciju submikroskopskih promjena broja kopija DNA smanjila se na manje od 1 Mb (Pinkel i sur. 1998, Rinčić 2014, Solinas-Toldo i sur. 1997). Rezolucija ovisi o dizajnu pločice i broju probi koji može varirati od 40 tisuća do 1 milijuna te o ciljanom razmještanju proba, pokrivaju li cijeli genom (od engl. *whole genomic array*) ili obuhvaćaju li lokuse poznatih mikrodelecijskih/duplikacijskih sindroma (od engl. *targeted array*). Proizvodnja mikročipova temelji se na sintezi oligonukleotida *in situ* na temelju zadane sekvence proba direktno na površinu stakalca čime se dobiva velika gustoća proba na stakalcu, a takvi mikročipovi se nazivaju oligonukleotidni čipovi visoke rezolucije.

Osnova metode aCHG je hibridizacija cjelokupne genomske DNA na stakalce s imobiliziranim probama. Genomska DNA uzorka i kontrole se prije nanošenja na stakalce kemijski ili mehanički fragmentira i odvojeno obilježava fluorescentno modificiranim nukleotidima, Cy5-dUTP ili Cy3-dUTP. Nakon nanošenja na stakalce dolazi do hibridizacije nakon čega se pranjem uklanja višak nehibridizirane DNA. Skeniranjem hibridiziranih pločica bilježi se razina fluorescencije svake probe na čipu. Razina fluorescencije softverskim paketom za analizu prevodi se u logaritamsku vrijednost ( $\log_2$ ) što omogućava mjerenje broja kopija DNA u uzorku (Slika 3).



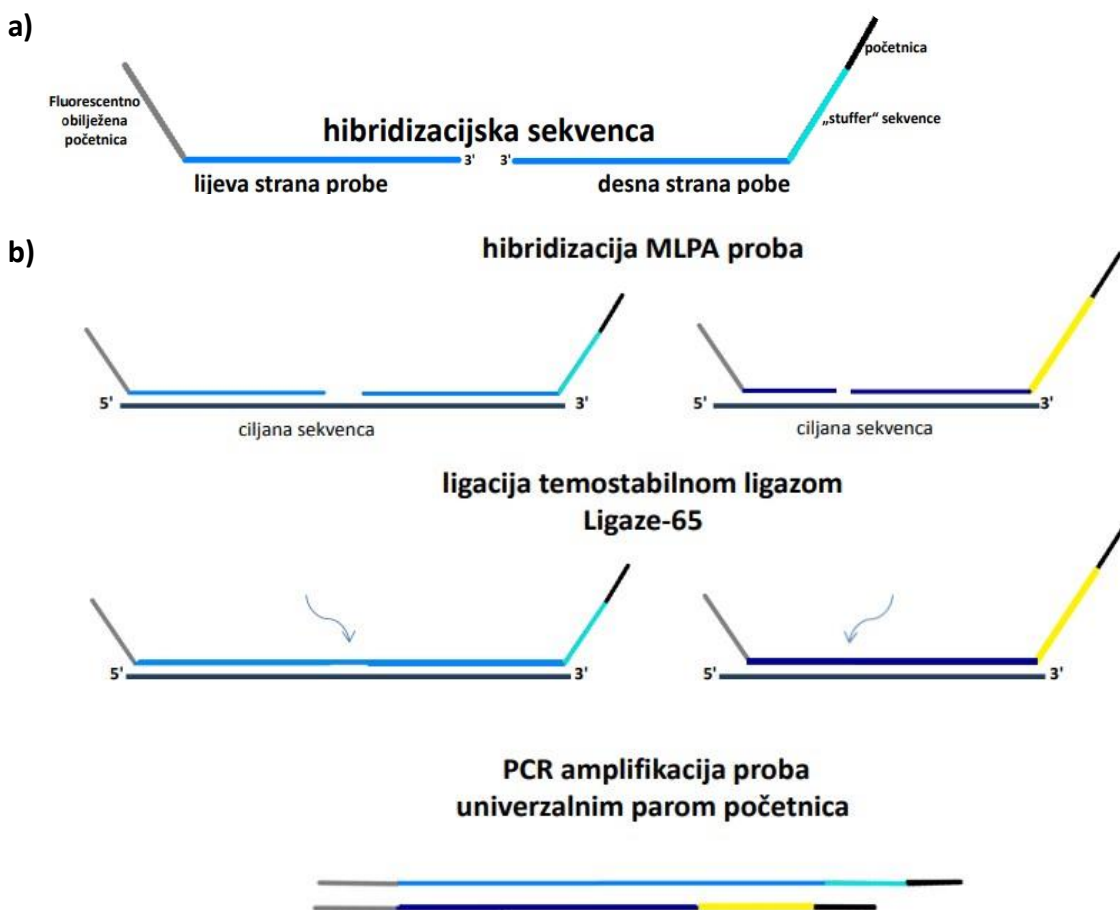
Slika 3. Shematski prikaz analize kromosoma na mikročipovima. Preuzeto i prilagođeno iz Rinčić, 2014.

### 1.5 METODA ISTOVREMENOG UMNAŽANJA VEZANIH PROBA

Novija metoda analize genoma koja je vrlo brzo postala široko rasprostranjena za genetičko testiranje bolesti je **metoda istovremenog umnažanja vezanih proba** (MLPA, od engl. *Multiplex Ligation-dependent Probe Amplification assay*) (Stuppia i sur. 2012).

Metoda MLPA temelji se višestrukom PCR-u (Lančana reakcija polimerazom, od engl. *Polymerase Chain Reaction*). Može se koristiti 50-60 proba, svaka specifična za određenu sekvencu DNA (većinom eksoni gena od interesa) kako bi se odredio relativni broj kopija DNA svake sekvence. Tijekom PCR-a dolazi do umnažanja probi dodanih uzorku, a ne DNA uzorka. Svaka proba MLPA sastoji se od dva oligonukleotida, 5' i 3' polu-probe koje se sastoje od specifične ciljne sekvence i univerzalne početnice koja omogućava umnažanje svih proba MLPA simultano. Kraća, 5' proba sastoji se od fluorescentno obilježene početnice duljine 19 nt na svom 5' kraju i specifične ciljne sekvence duljine 21-30 nt na svom 3' kraju. Dulja polu-proba na svom 5' fosforiliranom kraju sadrži specifičnu ciljnu sekvencu duljine 25-43 nt i neobilježenu sekvencu komplementarnu drugoj početnici (duljine 36 nt). Na jednoj ili obje polu-probe nalaze se i *stuffer* sekvence

različitih duljina (19-370 nt) koje omogućavaju diferencijaciju proba na temelju duljine tijekom elektroforeze te kasnije kvantifikaciju (Slika 4). Dulja polu-proba nastaje kloniranjem u specifični vektor (poput M13-derived SALSA vektor) koji sadrži restrikcijska mjesta i jedinstvene *stuffer* sekvence. Svaka dulja polu-proba produkt je različitog M13 vektora koji sadrži različitu *stuffer* sekvencu.



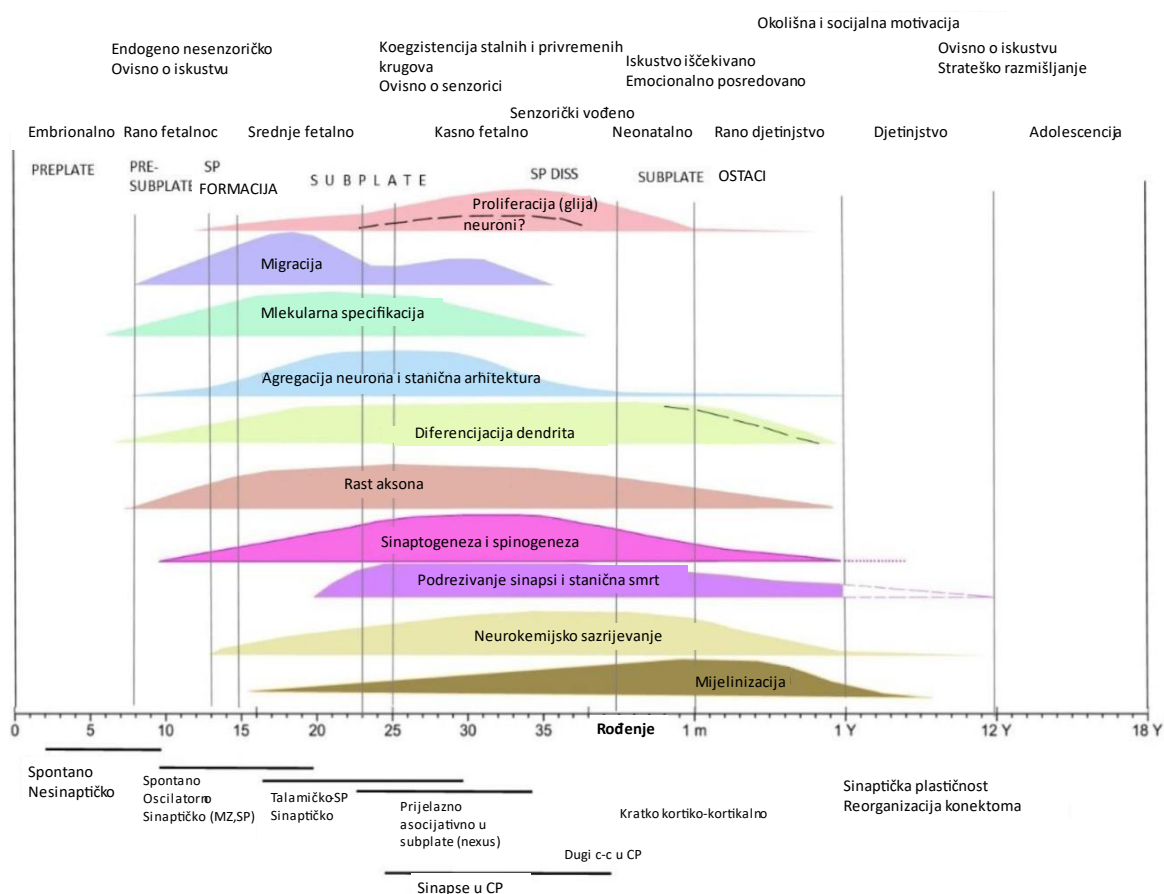
Slika 4. Shema osnovnog principa metode MLPA i dizajna probe za metodu MLPA: a) shematski prikaz dizajna probe MLPA, b) osnovni princip metode MLPA koji obuhvaća korake: hibridizacija MLPA probe, ligacije dva dijela probe ligazom Ligaze-65 i PCR amplifikacija univerzalnim parom početnica. Preuzeto i prilagođeno iz Rinčić, 2014.

Do ligacije i umnažanja dolazi tek nakon potpunog komplementarnog sparivanja obje polu-probe na ciljanoj sekvenci i to bez praznina. Umnažanje PCR-om se odvija pomoću samo jednog para fluorescentno obilježenih početnica. Kako će se umnožiti samo spojene probe tijekom PCR-a, broj produkata nakon ligacije proba je mjera ciljane sekvence u uzorku. Produkti PCR-a se razdvajaju po veličini kapilarnom elektroforezom u denaturirajućim uvjetima. Visina ili površina pikova fluorescencije se mjeri i na taj način se kvantificira

količina produkata PCR-a. Nakon normalizacije i usporedbe s kontrolnim uzorkom DNA dobiva se relativna količina ciljne sekvence DNA u početnom uzorku DNA (Rinčić 2014, Schouten, 2002 ).

## 1.6 NEURORAZVOJNI POREMEĆAJI

Neurorazvoj je kompleksni i dinamični proces koji uključuje precizno podešen slijed genetičkih, okolišnih, biokemijskih i fizičkih događaja (Budday i sur. 2015). Stadiji razvoja mozga s naglaskom na razvoj kore velikog mozga su: neurogeneza i gliogeneza, migracija neurona, molekularna specijalizacija, agregiranje neurona i citoarhitektura, diferencijacija dendrita, rast aksona, sinaptogeneza i spinogeneza, stanična smrt i podrezivanje sinapsi, neurokemijsko sazrijevanje te mijelinizacija (Slika 5) (Kolb i sur. 2019, Kostović 2020).



Slika 5. Razvojni periodi pojave i intenziteta glavnih neurogenetičkih događaja u histogenezi sloja pod kortikalnom podpločom (od engl. *subplate*). Razvojno-funkcionalne faze su prikazane u gornjem dijelu dijagrama, a faze razvoja umrežavanja su u donjem dijelu dijagrama. U srednjem fetalnom i početkom kasnog fetalnog razdoblja dominiraju rast aksona, diferencijacija dendrita i sinaptogeneza. Proliferacija neurona je u ovom stadiju hipotetska, ali proliferacija glija se nastavlja. Tijekom prve godine nastavlja se rast kortikalnih puteva i razvoj intersticijskih neurona bijele tvari gyrusa. SP označava *subplate*, CP kortikalnu ploču (od engl. *cortical plate*), MZ marginalnu zonu, c-c kortiko-kortikalno (od engl. *cortico-cortical*). Preuzeto i prilagođeno iz Kostović 2020.

Neurorazvojni poremećaji (NDD, od engl. *Neurodevelopmental disorder*) obuhvaćaju veliku grupu, klinički heterogenih, kroničnih poremećaja koje karakterizira oštećenje razvoja centralnog živčanog sustava, a za posljedicu imaju narušeno funkcioniranje u domenama motoričkih vještina, kognicije, komunikacije i/ili ponašanja te su povezani s disrupcijom strogo koordiniranih događaja tijekom razvoja mozga (Gilissen i sur. 2014, Ismail i Shapiro 2019, Parenti i sur. 2020,). Upravo zbog široke definicije neurorazvojnih poremećaja u skupinu NDD-ova ubrajamo velik spektar poremećaja koji su karakterizirani funkcionalnim i bihevioralnim odstupanjima tijekom ranog razvoja. Iako se termin neurorazvojni poremećaji kao posebna dijagnostička grupa u službene medicinske klasifikacije uvodi tek prije 10 godina, u stručnoj literaturi termin se već dugi niz godina koristi za grupu poremećaja kojima je zajedničko teško oštećenje kognitivnog razvoja u ranoj dječjoj dobi i adolescenciji. Prema trenutno važećoj klasifikaciji na temelju Dijagnostičkog i statističkog priručnika za mentalne poremećaje (od engl. *Diagnostic and statistical manual of mental disorders*, 5th ed, [DSM-5]) postoji šest kategorija u okviru neurorazvojnih poremećaja koje se odnose na specifične poremećaje, i to su poremećaji intelektualnog razvoja, komunikacijski poremećaji, poremećaj iz spektra autizma, poremećaji motorike, specifični poremećaji učenja i poremećaj s deficitom pažnje i hiperaktivnošću (od engl. *Attention deficit/hyperactivity disorder* – ADHD). Širi koncept NDD-ova uključuje i cerebralnu paralizu, epilepsiju i neuropsihijatrijske poremećaje koji nastaju tijekom ranog djetinjstva ili adolescencije (Savatt i Myers, 2021). Etiološki gledano, uzroci NDD-ova mogu biti raznoliki; imunološki, nutritivni, metabolički, genetički, strukturalni, onkološki, hipoksijsko-ishemijska ozljeda, trauma, izlaganje toksinima, infekcija, starenje ili iatrogeni poremećaji (Ismail i Shapiro, 2019, Savatt i Myers 2021).

### 1.7 NEURORAZVOJNI POREMEĆAJI I PROMJENE BROJA KOPIJA DNA

Identifikacija potencijalnih genetičkih uzroka NDD-ova je ključna za razumijevanje molekularnih mehanizama odgovornih za nastanak NDD-ova i za koreliranje genotipa i fenotipa što može pomoći praćenju poremećaja i predviđanju budućih komplikacija. Različiti tipovi genetičkih promjena se povezuju s NDD-ovima; kromosomski rearanžmani, CNV, male indel, i točkaste mutacije (Parenti i sur. 2020, Savatt i Myers 2021).

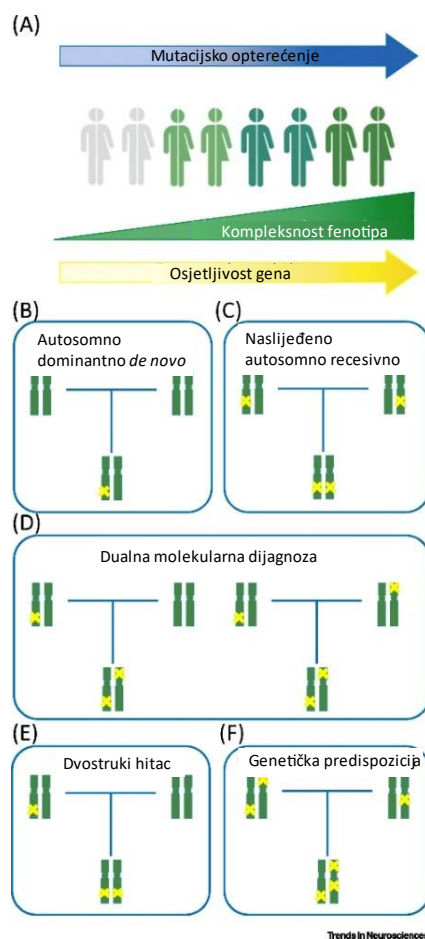
Mnogo rijetkih CNV-ova povezano je s povećanim rizikom za razvoj neurorazvojnih poremećaja (NDD-CNV). Prvotno su pronađeni u slučajevima NDD-ova koristeći metode aCGH i FISH (Chawner i sur. 2020, Dinneen i sur. 2022). NDD-CNV su primarne varijante NDD-ova i često su pronađene kod pacijenata s ASD-om i ID/DD-om koristeći kromosomske mikročipove tijekom rutinske kliničke obrade. Do 2022. najviše impliciranih lokusa CNV-ova bilo je za ID/DD (n = 70) pa za ASD (n= 15) za koji ima i najveći

zabilježeni broj pacijenata s NDD-CNV-ovima, približno 10-20% pacijenata ima barem jednu NDD-CNV varijantu (Dinneen i sur. 2022).

Neki NDD-CNV-ovi su povezani s nepotpunom penetracijom, a jedna od značajki im je i varijabilna ekspresivnost fenotipa (de la Torre-Ubieta i sur. 2016, Gudmundsson i sur. 2019, Parenti i sur. 2020). Penetracija je vjerojatnost ispoljavanja kliničkog fenotipa kada je osoba nositelj određenog genotipa. Nepotpunu penetraciju definiramo kao pojavu kod koje neke osobe s patogenim varijantama ekspimiraju očekivani klinički fenotip dok druge ne. Pleiotropija je pojava kada različite varijante istog gena izazivaju različite, potencijalno nepovezane fenotipe koji mogu biti kategorizirani i kao različite bolesti (Kingdom i Wright, 2022).

Procjene penetracije su veće za ASD i ID/DD od shizofrenije, te za veće NDD-CNV i one koje zahvaćaju lokuse netolerantne na gubitak funkcije. Širok spektar psihijatrijskih, bihevioralnih i fizičkih simptoma koji prelaze dijagnostičke granice se često bilježi kod nositelja NDD-CNV-ova. Nepotpuna penetracija i pleiotropija uočavaju se i unutar obitelji te mogu biti naslijeđeni od asimptomatskih nositelja (de la Torre-Ubieta i sur. 2016, Gudmundsson i sur. 2019, Parenti i sur. 2020).

Sve veći broj istraživanja ukazuje na to da NDD-ovi dijele genetičke mehanizme i navode da NDD-CNV-ovi imaju pleiotropni učinak na kliničke fenotipe i potencijalno utječu na konvergentne mehanizme u podlozi poremećaja (de la Torre-Ubieta i sur. 2016, Gudmundsson i sur. 2019, Parenti i sur. 2020). Kao glavni principi nastanka takvog fenotipa navode se osjetljivost gena i mutacijsko opterećenje (Slika 6) (Parenti i sur. 2020).



Slika 6. Shematski prikaz genetičkih mehanizama neurorazvojnih poremećaja, (A) mutacijsko opterećenje i stupanj osjetljivosti gena u kojima dolazi do disrupcije utječu na fenotip, (B) većina genetičkih uzroka NDD-ova uključuje *de novo* mutacije, (C) Opaža se i autosomno recesivno nasljeđivanje u kojem se nasljeđuje promijenjeni alel od neoboljelih/slabooboljelih roditelja, (D) sve veći broj pacijenata ima dualne molekularne dijagnoze, različite mutacije mogu nastati *de novo* ili biti naslijeđene od jednog ili oba roditelja, (E) pretpostavka dvostrukog hitca, (F) skup naslijeđenih uobičajenih genetičkih varijanti može povećati osjetljivost na nastanak NDD-a. Preuzeto i prilagođeno iz Parenti i sur. 2020.

Osjetljivost gena se definira kao mogućnost tolerancije disruptivnih varijanti nekog gena; što je tolerancija na mutacije niža, viša je razina osjetljivosti. Neki geni povezani s NDD-ovima su visoko osjetljivi na dozu gena (haploinsuficijentni geni) te spadaju u kategoriju visoko osjetljivih gena. Mutacije ovih gena povezane su sa značajnim rizikom za razvoj bolesti. Primjeri gena su *DEPDC5*, *CACNA1A* i *SCN8A*. Disruptivne promjene ovih gena s velikom vjerojatnošću dovode do fenotipa bolesti čak i ako nema drugih uzroka što rezultira monogenским NDD-ovima. Mutacije koje pogađaju ove gene su pod jakim negativnim selektivnim pritiskom te se u populacijskim studijama bilježi manji broj disruptivnih varijanti u usporedbi s drugim



lokusima. Ukratko, mutacije u visoko osjetljivim genima mogu se kategorizirati kao rijetke varijante povezane sa značajnim rizikom za razvoj bolesti i visoke penetracije.

Kod gena koji su manje osjetljivi na disruptivne mutacije varijante gena nisu pod negativnim selektivnim pritiskom te se često generacijski prenose u obiteljima. Kako ne dolazi do bolesti uslijed jedne disruptivne promjene kod neosjetljivih gena oni spadaju u kategoriju uobičajenih varijanti s niskim rizikom razvoja bolesti. Usprkos tome, prema novijim istraživanjima, značajni udio poligenih NDD-ova se može dovesti u vezu s uobičajenim/učestalim varijantama. U ovom slučaju fenotip bolesti rezultat je zbrojnih učinaka mutacijskih događaja i ovisi ne samo o sumi efekata pojedinih mutacija već i o fizičkim i funkcionalnim interakcijama pogođenih gena (epistazi).

Epistatske interakcije i osjetljivost doze snažno koreliraju s konceptom mutacijskog opterećenja koji govori da na penetraciju i kompleksnost fenotipa bolesti utječe niz disruptivnih događaja.

Više mutacijsko opterećenje može biti uzrokovano kombinacijom somatskih i germinativnih događaja, mehanizmom teorije dvostrukog hitca. U teoriji dvostrukog hitca naslijeđena mutacija pruža osjetljivu genetičku pozadinu te stečena somatska mutacija koja nastaje kasnije tijekom razvoja zaslužna je za izražavanje fenotipa bolesti ili za širenje skupa već prisutnih kliničkih značajki.

Primjer pleiotropnog učinka je gen *DISC1*, balansirane translokacije ovog gena povezane su sa shizofrenijom i bipolarnim poremećajem kod više članova iste obitelji. Različiti klinički ishod povezan s istom uzročnom varijantom određen je genetičkom pozadinom koja sadrži varijante specifične za bolest. Što se tiče molekularne patofiziologije, varijante povećanog rizika za razvoj NDD su pronađene u genima s ulogama u procesima regulacije sinteze proteina, epigenetičke i transkripcijske regulacije, te sinaptičke signalizacije. Homeostaza navedenih procesa može biti poremećena tijekom neurogeneze, migracije i diferencijacije neurona u prenatalnom razvoju mozga ili tijekom sazrijevanja sinapsi i pravilne emergencije inhibitorno-ekscitacijske ravnoteže u postnatalnom razvoju mozga (Parenti i sur. 2020).

## 1.8 REEVALUACIJA REZULTATA

Napretkom znanosti i razvojem tehnologije dolazimo do novih saznanja, razvijaju se novi genetički testovi i bioinformatički alati te se povećava rezolucija postojećih testova. Svakodnevno sve više razumijemo povezanost gena i neurorazvojnih poremećaja te patogenost određenih varijanti što povećava dijagnostički prinos samog testiranja. Trenutno, za samo mali broj kodirajućih gena u ljudskom genomu ustanovljena je veza s bolestima. S jačanjem referentnih baza podataka, dijeljenjem podataka o varijantama i genima kandidatima te integracijom dodatnih tehnologija poput RNA sekvenciranja, *genome-wide methylation*

*studies* i *single-cell* sekvenciranje dolazi do poboljšanja razumijevanja povezanosti gena i NDD-ova te patogenosti samih varijanti (Savatt i Myers, 2021). Nova saznanja trebala bi biti osnova za reevaluacija postojećih rezultata genetičkih testova visoke rezolucije, uključujući prije svega kromosomske mikročipove i NGS. Bolje shvaćanje povećanja količine znanja dobivamo pretraživanjem baze PubMed pojmovima *microdeletions* i *microduplications*. U periodu do 2002. pretraživanjem baze pojmom *microdeletions* dobivamo 949 rezultata dok za razdoblje od 2003.-2023. dobivamo 5741 rezultat. Za pojam *microduplications* u razdoblju do 2002. dobivamo 19 rezultata, a za razdoblje od 2003.-2023. 1379 rezultata (National Library of Medicine 2021).

Kako bi se standardizirala i olakšala reevaluacija rezultata, *American College of Medical Genetics and Genomics* izdao je smjernice za reevaluaciju varijanti, preispitivanje i reklasifikaciju varijanti koje su reportirane pacijentima u smislu nalaza genetskih testova, te za ponovnu analizu slučajeva, pregled svih varijanti eksoma ili genoma, za pronalaženje potencijalnih uzročnih varijanti zbog širenja znanja i poboljšanja bioinformatičkih alata. Smjernice za reevaluaciju podijeljene su u nekoliko kategorija:

1. opće – standardizirani protokoli za klasifikaciju varijanti, reevaluaciju i ponovnu analizu koji su periodično revidirani, pravovremeno odgovaranje na vanjske zahtjeve za reevaluaciju, održavanje baze klasificiranih varijanti unutar laboratorija, dokumentacija koja obavještava pojedince o mogućoj promjeni klasifikacije varijanti, sve klasificirane varijante i informacije o fenotipu se trebaju podnositi javno dostupnim bazama.
2. reevaluacija varijanti – prioritizacija reevaluacije varijanti nesigurne važnosti ili vjerojatno patogenih te baze podataka na temelju novih publikacije koristeći se standardiziranim ACMG protokolom
3. Ponovna analiza slučajeva – temelj ponovne analize treba biti ažurirana klinička i obiteljske anamneza koristeći poboljšane bioinformatičke alate
4. izvještavanje – prioritizacija izvještavanja rezultata koje imaju kliničku važnosti, poput varijanti nesigurne važnosti koje su reklasificirane kao patogene
5. ponovno testiranje – zbog napretka tehnologije tijekom vremena za neke pojedince potrebno je umjesto reevaluacije ponoviti testiranje (Deignan i sur. 2019).

## 2. CILJEVI ISTRAŽIVANJA

Cilj ovog rada je utvrditi postoji li promjena u interpretaciji genetičkih nalaza dobivenih pomoću kromosomskih mikročipova kod ispitanika s razvojnim zaostajanjem i intelektualnim teškoćama tijekom vremena, obzirom na nove znanstvene spoznaje.

Specifični ciljevi su:

1. Anotirati sve detektirane promjene u broju kopija DNA na novu inačicu ljudskog genoma (hg38).
2. Utvrditi hodogram reevaluacije rezultata metode CMA, s ciljem postizanja optimalnih rezultata ponovne analize.
3. Istražiti potencijalne nove kandidat gene koji su povezani s putovima ili biološkim procesima za koje se zna da su uključeni u razvoj središnjeg živčanog sustava.
4. Provesti analizu interakcije genomskih sekvenci unutar detektiranih aberacija kako bi se utvrdila potencijalna indirektna veza na gene koji su povezani s neurorazvojnim poremećajima.
5. Istražiti postoji li povezanost nekodirajućih sekvenci unutar detektiranih aberacija s neurorazvojnim poremećajima.

### 3. MATERIJALI I METODE

#### 3.1 UZORCI

U izradi diplomskog rada korišteni su podaci proizašli iz projekta MZOŠ „Molekularna citogenetika u evaluaciji mentalne retardacije nepoznate etiologije“ (šifra projekta: 108-1081870-1888) voditeljice prof.dr.sc. Lukrecije Brečević (2007.-2013.) koji je proveden na Medicinskom fakultetu Sveučilišta u Zagrebu u suradnji s Institutom za humanu genetiku Sveučilišta u Jeni, Njemačka.

Analizirani su podaci 50 ispitanika s dijagnozom razvojnog intelektualnog poremećaja (RIP-a) iz Centra za rehabilitaciju Stančić i Centra za rehabilitaciju Zagreb ili ASD-a iz Centra za rehabilitaciju Edukacijsko-rehabilitacijskog fakulteta, Kabinet za poremećaje iz autističnog centra. Za korištenje uzoraka ishodovana je potvrda Etičkog povjerenstva Medicinskog fakulteta Sveučilišta u Zagrebu te Centra za rehabilitaciju Stančić, Centra za rehabilitaciju Zagreb i Centra za rehabilitaciju Edukacijsko-rehabilitacijskog fakulteta. Iz uzoraka krvi izolirana je genomski DNA, provjerena je čistoća i koncentracija izolirane DNA apsorpcijskom metodom na Bioanalyzer sistemu (*Agilent Technologies*) uz softverski paket 2100 Expert Software (*Agilent Technologies*). Nakon provjere kvalitete DNA u agaroznom gelu provedena je analiza MLPA. Korišteni su specifični setovi MLPA proba prema sljedećem algoritmu:

1. Ženske ispitanice s kliničkom slikom sindroma Rett: pretraživanje metodom MLPA gena *MECP2* i *CDKL5* gena.
2. Muški ispitanici s RIP-om i nejasnim fenotipom: pretraživanje metodom MLPA za poznate gene *XMR/XID*.
3. Muški ispitanici s kliničkom slikom sindroma FRAX: analiza MLPA gena *FMRI* i *AFF2*
4. Ispitanici s kliničkom slikom Prader-Willi ili Angelman sindroma: pretraživanje metodom MLPA kromosomske regije PWS/AS.
5. Ispitanici s kliničkom slikom Angelman sindrom i normalnim nalazom pod (4): analiza MLPA gena *UBE3A*.
6. Ispitanici s indikacijom za specifične mikrolelecijske sindrome: primjena setova proba MLPA za mikrolelecijske sindrome.
7. Kod svih negativnih slučajeva nakon (1)-(6): analiza MLPA subtelomera.

8. Kod svih negativnih slučajeva nakon (1)-(5) i pod (7): analiza MLPA za klasične i nove mikrodelecijske sindrome.
9. Kod ispitanika s ASD-om i ispitanika s RIP-om i elementima autističnog ponašanja primjena set proba MLPA za ASD.
10. Kod negativnih slučajeva (9) primjena seta proba za kromosomsku regiju PWS/AS

Svi setovi proba MLPA detektiraju se u sustavu kapilarne elektroforeze.

Rezultati koji su inicijalno bili klasificirani kao varijante nesigurne ili sporne važnosti reevaluirani su s obzirom na nove smjernice o interpretaciji genomskih aberacija i novih znanstvenih podataka o funkcijama gena. Ukupno su ponovno analizirani rezultati 50 ispitanika.

### 3.2 BAZE PODATAKA

Mrežni alat *The University of California Santa Cruz (UCSC) Genome Browser* (<https://genome.ucsc.edu/>) služi za brzo prikazivanje traženog dijela genoma uz seriju poravnatih anotacija („tracks“). Anotacije su generirane od strane *UCSC Genome Bioinformatics Group* i vanjskih suradnika te uključuju gene i predikcije gena, rezultate sekvenciranja i mapiranja, SNP-ove, podatke o ekspresiji i regulaciji, fenotipu i varijantama te *pairwise* i *multiple-species* komparativne podatke o genomici. Baze podataka koje su u pozadini alata *Genome Browser* mogu se vidjeti, preuzeti i manipulirati pomoću mrežne aplikacije *the UCSC Table Browser* (Karolchik i sur. 2012).

Baza podataka *The Human Brain Transcriptome* (HBT, <https://hbatlas.org/>) projekt je Odjela za Neurobiologiju Yale Sveučilišta, Škole za medicinu. Javno dostupna baza sadrži podatke o transkriptomu i pridružene metapodatke za odrasli ljudski mozak i mozak u razvoju. Baza sadrži podatke o transkriptomu na razini eksona širom genoma dobivene pomoću *Affymetrix GeneChip Human Exon 1.0 ST Arrays* iz više od 1340 uzoraka tkiva uzorkovanih *post mortem* iz obje hemisfere ljudskog mozga. Uzorkovani su mozgovi od stadija embrionalnog razvoja do odraslih mozgova i reprezentativni su za muškarce i žene iz različitih etničkih skupina. Ukupno je uzorkovano 16 regija mozga: cerebralni korteks, mediodorzalna jezgra talamusa, strijatum, amigdala, hipokampus i 11 područja neokorteksa. Podatci genotipizacije širom genoma dostupni su za 2.5 milijuna markera (*Illumina Human Omni 2.5-Quad Bead Chips*) za sve uzorke.

Program *The Human Protein Atlas* (<https://www.proteinatlas.org/>) pokrenut je 2003. s ciljem mapiranja svi ljudskih proteina u stanicama, tkivima i organima koristeći integrirane *-omics* tehnologije uključujući bojanje uz pomoć antitijela, proteomiku temeljenu na masenoj spektrometriji, transkriptomiku i sistemsku biologiju. *The Human Protein Atlas* sastoji se od 12 dijelova; *Tissue, Brain, Single Cell Type, Tissue Cell Type, Pathology, Disease Blood Atlas, The Immune Cell, Blood Protein, Subcellular, Cell Line, Structure, Metabolic*. Dio *The Human Brain* pruža sveobuhvatno prostorno profiliranje mozga uključujući pregled proteinske ekspresije u mozgu sisavaca temeljen na integraciji podataka iz mozгова miševa, ljudi i svinja. Dostupni su detaljni podaci transkriptomike kombinirani s afinitetnom lokalizacijom proteina *in situ* do razine pojedine stanice. Ekspresija gena u mozgu određena je iznad *cut off* vrijednosti ( $nTPM \geq 1$ ). Na temelju toga geni su podijeljeni u kategorije:

1. izraženi u jednom tkivu: Opaženi samo u mozgu (202) (dalje u tekstu HPA1)
2. izraženi u nekim tkivima: Opaženi u mozgu i barem još jednom tkivu, ali manje od trećine tkiva (2014) (dalje u tekstu HPA2)
3. izraženi u mnogim tkivima: Opaženi u barem trećini tkiva, ali ne svima uključujući mozak (5473)

Baza SynGO (<https://www.syngoportal.org/>) javno je dostupna baza podataka i platforma za *online* analizu istraživanja sinapsi. Sadrži 1112 anotiranih gene kojima je pridružena funkcija i/ili lokalizacija u sinapsama. Geni sadržani u bazi su iznimno dugački, jako konzervirani i intolerantni na mutacije te su povezani s raznim obilježjima centralnog živčanog sustava. Anotacije se temelje na recentnoj znanstvenoj literaturi. Moguća je *online* analiza i dostupni su alati za vizualizaciju kako bi se procijenila lokacija pojedinih gena unutar sinapsi, te njihova funkcija ili izvela studija „obogaćivanja“ (Koopmans i sur. 2019).

Baza autDB javno je dostupan *web*-portal (<http://autism.mindspec.org/autdb/Welcome.do>) za neprekidno prikupljanje, ručnu anotaciju i vizualizaciju gena povezanih s poremećajima iz spektra autizma (Basu i sur. 2008). Baza predstavlja model baze podataka vođen bolešću gdje su prikupljeni geni povezani s ASD-om koji se klasificiraju ovisno o genetičkoj varijabilnosti: geni kandidati identificirani istraživanjima genetičke asocijacije, rijetke mutacije pojedinog gena i geni povezani s autizmom. Sadržaj baze proizlazi iz objavljene znanstvene literature i sadrži 1359 gena, 2042 pregledane jedinstvene reference i dodatnih nepregledanih 2296 jedinstvenih referenci.

Baza podataka SFARI *Gene* (<https://gene.sfari.org/>) pruža sveobuhvatni pregled dosadašnjih spoznaja o genetskoj podlozi poremećaja autističnog spektra. Unutar same baze, podatci su podijeljeni u module: *The Human Gene*, *Animal Models* i CNV. *The Human Gene* modul je detaljno anotiran popis gena proučavanih u kontekstu poremećaja autističnog spektra. Sadrži informacije o samim genima, relevantnu znanstvenu literaturu te opise dokaza koji povezuju gene s ASD-om. Geni su podijeljeni u nekoliko kategorija prema značajnosti. Kategorije gena su: S (sindromski), 1 (visoka vjerodostojnost), 2 (snažni kandidat) i 3 (sugestivni dokazi).

Baza podataka sastoji se od 1127 ocijenjenih gena, 2547 pregledanih i 1101 gena povezanog s ASD-om (Abrahams i sur. 2013).

Baza DECIPHER (*DatabasE of genomiC variation and Phenotype in Humans using Ensembl Resources*) je interaktivna web baza (<https://www.deciphergenomics.org/>) koja uključuje paket alata dizajniranih kao pomoć pri interpretaciji genomskih varijanti. Baza dohvaća informacije od raznolikih bioinformatičkih izvora. Varijanta pronađena kod pacijenta prikazana je u kontekstu normalne varijabilnosti i patogene varijabilnosti zabilježene na lokusu što pomaže kliničkoj interpretaciji rezultata (Firth i sur. 2009). Baza DECIPHER sadrži 45837 zapisa pacijenata, 184814 fenotipskih opažanja kod ovih pacijenata, 46800 promjena u broju kopija DNA i 12244 varijanti sekvenci.

Podaci koji su korišteni u radu su pLi, LOEUF, sHet, pHaplo, pTriplo, OMIM, Morbid, *ClinGen gene/disease*, *ClinGen Dosage Sensitivity*, *GeneReviews*, *Gene2Phenotype* i *Gene CC* s osvrtom o svakom genu.

Vjerojatnost intolerancije na gubitak gena (pLI, od engl. *Probability of Loss-of-function Intolerance*) je vjerojatnost da gen spada u kategoriju haploinsuficijentnih gena i ekstremno je intolerantan na varijantu gubitka funkcije gena. Geni visoke vrijednosti pLI ( $\geq 0.9$ ) su ekstremno intolerantni na gubitak funkcije gena dok su geni s niskom vrijednosti pLI ( $pLI \leq 0.1$ ) tolerantni na gubitak funkcije gena (Karczewski i sur. 2020). Sama vrijednost pLI temelji se na 125 748 eksoma jedinstvenih nesrodnih pojedinaca (gnomAD v2.1.1.) Metoda korištena za dobivanje vrijednosti koristi mutacijski model koji je selekcijski neutralni, temeljen na sekvenci i kontekstu kako bi se usporedio opaženi broj rijetkih varijanti po genu na očekivanu vrijednost (Samocha i sur. 2014).

LOEUF (od engl. *Loss of function Observed*) je kvantitativna mjera opaženog smanjivanja gubitka funkcije gena varijante u gnomAD u usporedbi s modelom s nula mutacija. S obzirom na to da je gnomAD uzorak populacije i možda ne predstavlja pravo smanjivanje gubitka funkcije gena, stavljena je gornja granica na ovu procjenu kako bi bila konzervativnija. Minimalna vrijednost je 0 i teoretski nema maksimalne vrijednosti. Geni s manjom vrijednosti (bliže 0) su intolerantniji na mutacije.

Selekcijski koeficijent varijanti heterozigotnog gubitka funkcije gena (sHet, od engl. *Selection coefficient of heterozygous loss-of-function variants*) je procjena učinka na fitness po genu za osobe koji imaju heterozigotne varijante gubitka funkcije gena. U populacijskoj genetici to je mjera relativnog fitnessa između dva stanja.

Predviđena vjerojatnost haploinsuficijencije (pHaplo od engl. *Predicted Probability of Haploinsufficiency*) vrijednost je predviđena iz analize velikih rijetkih CNV ustanovljenih mikročipovima kod 950278 pojedinaca. Vrijednosti pHaplo dobivene su modelom strojnog učenja treniranom na 145 karakteristika na razini gena kako bi se predvidjele vjerojatnosti gubitka cijele kopije. Vrijednost pHaplo  $\geq 0.86$  ukazuje na to da je prosječna veličina učinka delecija jednako jaka kao gubitak funkcije gena za koje se zna da su ograničeni protiv protein skraćujućih varijanti (*odd ratio*  $\geq 2.7$ ) (Karczewski i sur. 2020).

Predviđena vjerojatnost triplosenzitivnosti (pTriplo, od engl. *Predicted Probability of Triplosensitivity*) vrijednost je predviđena iz analize velikih rijetkih CNV ustanovljenih mikročipovima kod 950278 pojedinaca. Vrijednosti pTriplo dobivene su modelom strojnog učenja treniranom na 145 karakteristika na razini gena kako bi se predvidjele vjerojatnosti toga bi li dobitak cijele kopije svakog gena bio obogaćen u kohorti pojedinaca oboljelih od teških bolesti ranog početka u odnosu na opću populaciju. Vrijednost pTriplo  $\geq 0.94$  ukazuje na to da je prosječna veličina učinka duplikacija jednako jaka kao gubitak funkcije gena kod gena za koje se zna da su ograničeni nasuprot protein skraćujući varijanti (*odd ratio*  $\geq 2.7$ ) (Karczewski i sur. 2020).

*Online Mendelian Inheritance in Man* (OMIM) je *online* katalog ljudskih gena i genetičkih poremećaja Sveučilišta John Hopkins. Geni Morbid odabrani su poveznicom s fenotipom bolesti. Internacionalni izvor *GeneReviews* pruža klinički relevantne i medicinske informacije za nasljedna stanja u standardiziranom formatu koji uključuje dijagnozu, tretman i genetičko savjetovanje. Sustav *Gene2Phenotype* dizajniran je kao pomoć pri razvoju, validaciji, pregledavanju i



distribuciji velike količine setova podataka temeljenih na dokazima kako bi se mogli koristiti prilikom filtriranja dijagnostičkih varijanti. Razina pouzdanosti i poveznica na dokaz su dodijeljeni svakom unosu. *ClinGen Gene/disease validity* standardizirano pristupa određivanju kliničke validnosti svakog para gen-bolest. Klasifikacije gena su: definitivna, jaka, umjerena, ograničena, bez zabilježenih dokaza i opovrgnuta i osporena. *ClinGen Dosage Sensitivity* rezultat je prikupljanja dokaza o haploinsuficijenciji i triplosenzitivnosti gena i genomskih regija. Informacije o genima su rangirane u kategorije:

- 3, dovoljno dokaza za patogenost doze gena/regije
- 2, ima dokaza za patogenost doze gena/regije
- 1, malo dokaza za patogenost doze gena/regije
- 0, nema dokaza za patogenost doze gena/regije
- 40, dokazi ukazuju na to da gen/regija nije osjetljiva na dozu (nije vjerojatna haploinsuficijencija ili triplosenzitivnost)
- 30, gen/regija su povezani s autosomnim recesivnim stanjem

*The Gene Curation Coalition* (GenCC) evaluira validnost povezivanja gena i bolesti kako bi se razvila konzistentna terminologija za *curation* gena i procjene gena zabilježenih u povezivanju s bolesti. Parovi gen-bolest klasificirani su u kategorije: definitivni, jaki, umjereni, podržavani, ograničeni, opovrgnuti, osporeni, kod životinja ili nepoznati.

Za podatke o *matching patient variants* izvučeni su podaci o: ukupnom broju CNV, postotku *gain* i *loss*, postotku za svaku kategoriju veličine zabilježene CNV i prisutni fenotip uz CNV.

Baza ClinGen – *Clinical Genome Resource* (<https://clinicalgenome.org/>) je izvor posvećen stvaranju autoritativnog centralnog izvora koji definira kliničku važnost gena i varijanti za korištenje u preciznoj medicini i istraživanjima. Baza služi za prikupljanje fenotipskih i kliničkih informacija za varijante diljem genoma te razvoj konsenzusa za identificiranje njihove kliničke relevantnosti i pružanje takvih informacije istraživačima i kliničarima. Baza ClinGen kao izvorni cilj ima unaprijediti genomiku u kliničkoj praksi, poboljšati razumijevanje fenotipskih i funkcionalnih efekata varijanti te njihovu kliničku važnosti (Rehm i sur. 2015). U bazi su sadržani osnovni podaci o genima, *gene-disease validity*, *dosage sensitivity*, *clinical actionability*, *status and future prospects*, *resources* i poveznica na ClinVar.

Projekt *The Genotype-Tissue Expression* (GTEx, <https://gtexportal.org/home/>) cilja na pružanje izvora za istraživanje ekspresije i regulacije gena kod ljudi te poveznice s genetičkom varijabilnošću. Od genotipiziranih donora prikupljeni su i analizirani uzorci različitih tkiva kako bi se procijenila genetička varijabilnost unutar njihovih genoma (Lonsdale i sur. 2013). Najrecentniji podaci su *GTEx Analysis Release V8* (dbGaP Accession phs000424.v8.p2) koji sadrže podatke od 948 *post mortem* donora i 17382 RNA-seq uzoraka iz 54 tkivnih područja i 2 stanične linije s adekvatnom moći detekcije eQTL u 48 tkiva. U bazi GTEx ekspresija gena u tkivima iskazana je u vrijednostima transkripti na milijun (od engl. *Transcripts per Million*, TPM), izračunatih iz genskog modela izoformi kolapsiranih na pojedini gen. Podaci nisu dodatno normalizirani, kutijasti dijagrami su pokazani kao medijan i 25. i 27. percentil. Točke su prikazane kao netipične vrijednosti ako su 1.5 puta ispod ili iznad interkvartilskog raspona.

Baza *Reactome* (<https://reactome.org/>) pruža bioinformatički alat za vizualizaciju, interpretaciju i analizu informacija o biološkim putevima kako bi se poduprla temeljna i klinička istraživanja, analize genoma, modeliranje, sistemska biologija i edukacija (Gillespie i sur. 2021). U bazi je 14471 reakcija u ljudskom tijelu organizirana u 2610 bioloških puteva koji uključuju 30656 proteina i modificiranih formi proteina kodiranih s 11442 različita ljudska gena, 14153 kompleksa, 2002 malih molekula i 1113 lijekova. Ove reakcije su podržane s 36290 literaturnih navoda i projicirane su na 83933 ortologna proteina stvarajući 19417 ortolognih bioloških puteva u 14 različitih vrsta. Trenutno verzija od 30. studenog 2022., Verzija V83 ima anotacije za 4977 proteinskih varijanti i njihovih posttranslacijski modificiranih formi dobivenih od 361 proteina što je doprinijelo anotaciji 704 biološka puta i 1659 reakcija specifičnih za bolest.

Baza MGI (*Mouse Genome Informatics*, <https://www.informatics.jax.org/>) je internacionalna baza koja pruža integrirane genetičke, genomske i biološke podatke za laboratorijske mišve koji mogu pomoći proučavanju ljudskog zdravlja i bolesti. Jedan od alata je *Human – Mouse: Disease Connection* koji identificira objavljene i potencijalne mišje modele ljudskih bolesti, otkrivanje gena kandidata i istraživanje fenotipskih sličnosti između mišjeg modela i ljudskih pacijenata. Mišje mutacije, fenotipi mišjih modela za bolesti iz baze MGI integrirani su s ljudskim poveznicama gena i bolesti od *Alliance of Genome Resources* koji integrira podatke od *National Center for Biotechnology Information* (NCBI), *RGD (Rat Genome Database)*, *OMIM* i *HPO (Human Phenotype Ontology)*. Upisivanjem gena u tražilicu dobivaju se idući podaci: organizam, simbol gena, lokacija gena, genomske koordinate, povezane ljudske bolesti, abnormalni mišji

fenotip, reference u bazi MGI i broj miševa s mutacijama u tom genu. Dodatnim pretraživanjem samih bolesti dobiva se naziv bolesti, DO ID, OMIM ID, broj mišjih modela, povezani geni iz mišjih modela, povezani ljudski geni i reference koje koriste mišje modele (Blake i sur. 2020).

Baza podataka RNAcentral (<https://rnacentral.org/>) pruža integrirani pristup opsežnom i suvremenom setu nekodirajućih RNA sekvenci. Sadrži podatke o rRNA, lncRNA, modelnim organizmima, genomu, općenito i drugim RNA tipovima. U najnovijem izdanju (*release 22*) sadrži 34387991 sekvenci (Sweeney i sur. 2018).

Aplikacija Enrichr (<https://maayanlab.cloud/Enrichr/>) sadrži biblioteke novih setova gena, alternativni pristup rangiranju obogaćenih termina i različite interaktivne pristupe vizualizaciji kako bi se prikazali rezultati obogaćivanja koristeći JavaScript biblioteku, *Data Driven Documents* (D3). Sastoji se od 210 biblioteka setova gena od kojih su neke posuđene iz drugih izvora a neke su novo nastale i jedino dostupne u Enrichr-u. Biblioteke su podijeljene u 8 kategorija: *transcription, pathways, ontologies, diseases/drugs, cell types, miscellaneous, legacy* i *crowd* (Chen i sur. 2013).

Skup baza *The Universal Protein Resource* (UniProt, <https://www.uniprot.org/>) sveobuhvatan je izvor za proteinske sekvence i podatke o anotaciji. Sastoji se od više baza; *UniProt Knowledgebase* (UniProtKB), the *UniProt Reference Clusters* (UniRef), i the *UniProt Archive* (UniParc). Baza UniProtKB/Swiss-Prot sadrži visokokvalitetne ručno anotirane i neredundantne zapise proteinskih sekvenci. Podaci se analiziraju, uspoređuju i spajaju se sve dostupne sekvence za pojedini protein te se kritički pregledavaju povezani eksperimentalni i predviđeni podaci (Bateman i sur. 2020).

Projekt *Database of Genomic Variants* (<http://dgv.tcag.ca/dgv/app/home>) pruža sveobuhvatan sažetak strukturnih varijacija ljudskog genoma. Strukturne varijacije se definiraju kao genomske promjene koje uključuju segmente DNA veće od 50 pb. U bazi su prikazane samo strukturne varijacije pronađene kod kontrolnih uzoraka (MacDonald i sur. 2013).

Softverski sustav PANTHER (*Protein ANalysis THrough Evolutionary Relationships*, <http://www.pantherdb.org/>) povezuje evoluciju proteinskih sekvenci s evolucijom specifičnih proteinskih funkcija i bioloških uloga. Jezgra sustava je velika kolekcija filogenetski definiranih proteinskih obitelji i podobitelji generiranih računalnim algoritmima i pregledanih od strane biologa stručnjaka pomoću ekstenzivnih softverskih sustava za povezivanje ontoloških pojmova.

Svaka proteinska obitelj/podobitelj je prikazana filogenetskim stablom, skrivenim Markovim modelom (HMM) i poravnanjem više sekvenci (Mi i Thomas 2009, Thomas i sur. 2021).

Izvor *The SIGnaling Network Open Resource* (SIGNOR, <https://signor.uniroma2.it/>) anotira eksperimentalne dokaze o uzročnim interakcijama među proteinima i drugim entitetima biološke značajnosti (stimulusima, fenotipovima, inhibitorima enzima, kompleksima, proteinskim obiteljima). Svaki unos pokazuje eksperimentalne dokaze koji podržavaju interakciju i obogaćen je dodatnim relevantnim metapodacima. U mrežu SIGNOR 3.0 integrirano je 6750 proteina (UP000005640, *reviewed dataset*, UniProt *release 2022\_03*) i svaka interakcija je anotirana molekularnim mehanizmom koji uzrokuje efekt, pozitivnim ili negativnim na ciljni entitet (Lo Surdo i sur. 2022).

Baza *The GeneCards* (<https://www.genecards.org/>) pruža opširne informacije o svim anotiranim i predviđenim ljudskim genima. Baza automatski integrira podatke o genima iz približno 150 web izvora koji uključuju genomske, transkriptomске, proteomske, genetičke, kliničke i funkcionalne informacije. Najrecentnija verzija 5.14. sadrži 416392 unosa, 43529 odobrenih od strane HGNC-a, 21637 protein kodirajućih, 293262 RNA gena koji uključuju 131025 lncRNA, 111811 piRNA i 50426 ostalih ncRNA (Stelzer i sur. 2016).

Za naš set podataka izvučeni su idući podaci: *Entrez Gene Summary*, *GeneCards Summary* i *UniProtKB/Swiss-Prot Summary*. Baza podataka Entrez pruža integrirani pristup podacima o nukleotidnoj i proteinskoj sekvenci, informacije o genomskom i mapiranju usmjerenom na gene, 3D strukturi, PubMed MEDLINE i razno ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

### 3.3 ANOTACIJA PROMJENA BROJA KOPIJA NA NOVU INAČICU GENOMA

Kako bi se konvertirale genomske koordinate zabilježenih promjena korišten je alat *Lift Genome Annotations* baze UCSC, <https://genome.ucsc.edu/cgi-bin/hgLiftOver>. Originalne koordinate anotirane su prema inačici NCBI36/hg18, a za novi referentni ljudski genom uzeta je inačica GRCh38/hg38. Za minimalni omjer baza koje se moraju preklapati uzeta je vrijednost 0.95, uz isključivanje višestrukog anotiranja pojedinačnih regija. Za koordinate za koje uz zadani omjer preklapanja baza nije bilo moguće dobiti nove koordinate dopušten je minimalni omjer do 0.8.

Nove koordinate korištene su za anotaciju pomoću alata *Table Browser* baze UCSC, [https://genome.ucsc.edu/cgi-bin/hgTables?hgsid=1627495889\\_Lai6shxbksztUA9m979b6l56q1ax](https://genome.ucsc.edu/cgi-bin/hgTables?hgsid=1627495889_Lai6shxbksztUA9m979b6l56q1ax)

.Odabrana je grupa: *Genes and Gene Predictions*, obilježje: *NCBI RefSeq*, te tablica: *RefSeq Select and MANE (ncbiRefSeqSelect)*. Set transkripata MANE Select rezultat je suradnje pod nazivom *the Matched Annotation from NCBI and EMBL-EBI (MANE)* čiji je cilj definiranje setova transkripata i odgovarajućih proteina. Set transkripata MANE koristi se kao univerzalni standard pri izvještavanju varijanti za ljudske protein kodirajuće gene (Morales i sur. 2022).

Nakon anotacije dobiven je popis gena prema inačici NCBI36/hg18 kod kojih dolazi do promjena broja kopija DNA kod ispitivanih pacijenata – CNV lista gena.

### 3.4 ISTRAŽIVANJE POTENCIJALNIH GENA KANDIDATA

Kako bi se utvrdili potencijalni novi kandidat geni u podlozi NDD-ova prethodno dobiveni popis gena, CNV lista gena, kod kojih dolazi do promjena u broju kopija DNA korišten je kao polazište za pretraživanje baza *The Human Protein Atlas*, autDB, SFARI, SynGO i *Reactome* prema navedenim koracima:

#### 1. *The Human Protein Atlas The brain*

U prvom koraku istraženo je koji su geni s CNV liste gena izraženi isključivo ili gotovo isključivo u mozgu. Prema *HPA, The brain* bazi podataka napravljena je lista gen čija je ekspresija detektirana isključivo u mozgu i mozgu i barem još jednom tkivu, ali manje od trećine tkiva. Ova lista predstavlja gene za koje se zna da su isključivo ili skoro isključivo izraženi u mozgu, ukupno 2216 gena. Pomoću Vennovog dijagrama navedena lista uspoređena je s CNV listom gena, te su dobiveni geni čija je ekspresija detektirana samo u mozgu ili u mozgu i barem još jednom tkivu, a prisutni su kao CNV kod ispitanika s NDD-om. Usporedba setova podataka odrađena je pomoću *web* stranice <https://bioinformatics.psb.ugent.be/webtools/Venn/>.

#### 2. SynGO

CNV lista gena korišten je za pretragu baze SynGO, te su geni koji se nalaze u bazi pridodani prethodno navedenom popisu gena koji su izraženi isključivo ili gotovo isključivo u mozgu a prisutni su kao CNV kod ispitanika s NDD-om.

#### 3. SFARI i autDB

Popis gena impliciranih u etiologiju ASD-a napravljen je integriranjem liste gena iz baza SFARI i autDB. Navedena lista uspoređena je s CNV listom gena pomoću *web* stranice <https://bioinformatics.psb.ugent.be/webtools/Venn/>. Preklapajući geni pridodani su postojećem setu podataka iz prethodnog koraka.

#### 4. Reactome

Nakon izdvajanja gena koji su detektirani kao potencijalni kandidati u prethodnim koracima i bazama, ostatak kodirajućih gena na popisu CNV lista gena je korišten za pretragu *Reactome* baze. Ovim korakom dodatno su izdvojeni geni uključeni u razvoj živčanog sustava i u neuronski sustav te pridruženi postojećem setu podataka.

#### 5. UniProt

Lista protein kodirajućih gena s CNV liste gena koji nisu bili identificirani kao kandidat geni u prethodnim koracima (1-4) korištena je za pretragu baze UniProtKB/Swiss-Prot. Na temelju kriterija tkivne specifičnosti (engl. *tissue specificity*) izdvojeni su geni eksprimirani u mozgu te dodani postojećem setu podataka.

Rezultat analize kroz opisane korake je lista gena koji su detektirani kao CNV kod ispitanika s NDD-om, a specifično su vezani uz središnji živčani sustav, bilo tijekom razvoja prenatalno ili postnatalno. Na ovaj način definirana je proširena lista gena kandidata.

### 3.5 PROVJERA EKSPRESIJE IZABRANIH GENA

Za proširenu listu gena kandidata provjerena je ekspresija u mozgu prenatalno i postnatalno.

1. Kao referentna vrijednost za prenatalne ekspresije pretražena je baza HBT. Za svaki gen pogledan je graf koji pokazuje intenzitet ekspresije gena tijekom razvoja mozga, od embrionalnog razvoja do kasnog odraslog razdoblja. Na temelju izgleda grafa (pada ili rasta krivulje) i granične vrijednosti od 5 za intenzitet signala označeni su geni za koje se smatra da su bitni tijekom neurorazvoja.
2. Postnatalna ekspresija provjerena je u bazi GTEx, označeni su geni čija je ekspresija jača u tkivima živčanog sustava nego u ostalim tkivima.

Geni kojima je ekspresija povećana prenatalno što je vidljivo u bazi HBT i postnatalno, vidljivo u bazi GTE<sub>x</sub>, dodatno su označeni.

### 3.6 ODGOVARAJUĆI MODELNI ORGANIZMI

Za gene s proširene liste kandidat gena, gene koji su izdvojeni zbog potencijalne uključenosti u neurorazvoj, neurorazvojne poremećaje, funkciju sinapsi ili eksprimirani u mozgu i moždanim tkivima provjereno je postojanje odgovarajućih mišjih modela u bazi MGI. Pretražene su bihevioralne/neurološke, kraniofacijalne, embrionalne i neuralne fenotipske karakteristike za postojeće mišje modele, te su zabilježeni abnormalni fenotipovi. U obzir nisu uzimani mišji transgeni kao homolozi ljudskih gena.

### 3.7 ONTOLOGIJA GENA

Za gene koji nisu dodani na proširenu listu gena istražene su molekularne funkcije, uključenost u biološke procese i ontologiju ljudskog fenotipa. To je učinjeno unošenjem protein kodirajućih gena s CNV liste gena koji nisu dio proširene liste gena u bazu Enrichr. Nakon unošenja podataka u tražilicu, izabrana je kategorija Ontologies i izvađeni su podaci iz biblioteka *GO Biological Process 2021*, *GO Molecular Function 2021* i *Human Phenotype Ontology*. Izvađeni su podaci za top 10 bioloških procesa, molekularnih funkcija i ontologija ljudskog fenotipa na temelju p vrijednosti. Vrijednosti su poredane od najmanje do najveće p vrijednosti.

### 3.8 OSNOVNE INFORMACIJE O GENIMA I DEFINIRANJE KLINIČKE VAŽNOSTI

Za sve gene s proširene kandidat liste prikupljeni su podaci o vrijednosti p<sub>Haplo</sub> i p<sub>Triplo</sub> iz baze DECIPHER. Za gene kod kojih je kod pacijenata došlo do delecije gledala se vrijednost p<sub>Haplo</sub>, te su u daljnju analizu uzeti geni kod kojih je p<sub>Haplo</sub>  $\geq 0.86$ . Za gene kod kojih je kod pacijenata došlo do duplikacije gledala se vrijednost p<sub>Triplo</sub>  $\geq 0.94$ . Geni koji su zadovoljavali te kriterije su dodatno izdvojeni. Dodatna saznanja o samim potencijalnim genima kandidatima izdvojenim na temelju p<sub>Haplo</sub> i p<sub>Triplo</sub> izvađena su iz baze *GeneCards*.

Promjene broja kopija DNA mogu igrati važnu ulogu u patogenezi ljudskih bolesti zbog čega je detekcija i interpretacija CNV-ova klinički važna prilikom genetičkog testiranja. Poveznica između zabilježenih zdravstvenih stanja pacijenata s promjenama u genima koji čine naš set

podataka odrađena je u bazama ClinGen i DECIPHER. Za svaki gen prethodno pronađen u bazama: *The Human Protein Atlas*, autDB, SFARI, SynGO i *Reactome* izvučeni su podatci iz baza DECIPHER i ClinGen.

### 3.9 NEKODIRAJUĆE SEKVENCE UNUTAR DETEKTIRANIH ABERACIJA

Kako bi se provela analiza interakcija genomskih sekvenci unutar detektiranih aberacija i utvrdila potencijalna indirektna veza na gene koji su povezani s neurorazvojnim poremećajima za gene kojima je definirana klinička važnost provjereno je postojanje lncRNA u bazama RNAcentral i UCSC *Genome Browser*. Iz baze RNAcentral je izvučen broj zabilježenih lncRNA, njihovi nazivi i radovi povezani s neurorazvojnim poremećajima. Iz mrežnog alata UCSC *Genome Browser* izvučen je naziv zabilježenih lncRNA, *Genecode* ime, pozicija i pogledana je ekspresija u bazi GTEX.

### 3.10 DATABASE OF GENOMIC VARIANTS

Kako bi se provjerile strukturne aberacije prisutne u općoj populaciji, za gene s proširene kandidat liste pregledana je baza DGV. U bazi je postavljen izvor podataka: *Genomic Variants in Human Genome (Build GRCh38: Dec.2013, hg 38)*. U bazi se gledala alatna traka *DGV Gold Standard Variants* koja prikazuje pregledani set varijanti iz izabranog broja studija. Istraživanjima visoke rezolucije i kvalitete evaluirana je točnost i osjetljivost te su primijenjeni strogi kriteriji validacije kvalitete. Varijante na razini uzoraka koje su se preklapale su spojene u regije varijanti ako su međusobno dijelile barem 70% duljine. Dostupne su informacije o uzorcima, etnicitetu i distribuciji populacije. Početni kriteriji su da je varijanta pronađena u barem 2 različite studije i u barem 2 različita uzorka. Uzeti su podaci o tome ima li zabilježenih regija CNV, koliko ih je i koliki broj uzoraka je bio u istraživanju i u koliko je zabilježena ta CNV.

### 3.11 VIZUALIZACIJA SIGNALNIH I METABOLIČKIH PUTEVA

Za određivanje proteinskih obitelji i klasa, te molekularne funkcije i staničnih puteva u koje su uključeni proteini gena s proširene liste korištena je baza PANTHER. Podaci o molekularnoj funkciji proteina prikazani su pomoću toplinske karte kako bi se olakšala vizualizacija.



Za vizualizaciju staničnih puteva i odnosa u koje su uključeni proteini koji nastaju od potencijalnih gena kandidata lista gena unesena je u bazu SIGNOR. Kao granična vrijednost odnosa (engl. *relations score*) uzeta je vrijednost od 0.4 te nisu gledani odnosi čija je vrijednost odnosa manja od 0.4. Zasebno su pogledane kategorije interakcija: direktne, indirektna, povišenje regulacije, sniženje regulacije, vezanje i transkripcija. Podaci su prikazani toplinskom kartom kako bi se ustvrdili međusobni odnosi među proteinima i uključenost u stanične puteve te jesu li potencijalni kandidatski geni uključeni uzvodno ili nizvodno u iste.

### 3.12 LITERATURNI PREGLED

Za utvrđivanje novih spoznaja za prethodno izabrane gene korištena je baza PubMed. Za svaki gen je pretražena baza te su traženi radovi koji imaju zabilježene aberacije kod pacijenata u prikazima slučajeva ili upućuju na molekularnu funkciju proteina i biološke procese u koje su uključeni geni i njihovi produkti u posljednjih 5 godina. Isključni kriterij za odabir literature bio je povezivanje gena s neurodegenerativnim bolestima. Za gene za koje se nisu mogli pronaći nikakvi podaci u posljednjih 5 godina spuštena je granica na 10 godina. Podaci su izraženi u vidu tablice koja sadrži literaturne navode koji odgovaraju zadanim kriterijima pretrage.

## 4. REZULTATI

### 4.1 USPJEŠNOST ANOTACIJE NA NOVU INAČICU GENOMA

Od 713 polazišnih genomskih lokusa prema NCBI36/hg18 inačici za 50 pacijenata uspješno su pomoću alata *Lift Genome Annotations* baze UCSC konvertirane 662 koordinate na GRCh38/hg38 inačicu genoma. Rezultat konverzije za 51 lokus značio je *split in new* ili *deleted in new*, te ih nije bilo moguće konvertirati prema zadanim parametrima jer je alat javljao da su na novoj inačici genoma dijelovi genoma ograničeni koordinatama *split in new* ili *deleted in new*, od tih 51 koordinata njih 45 je anotirano.

Anotacija 662 novo dobivena lokusa pomoću alata *Table Browser* baze UCSC rezultirala je listom od 920 UCSC RefSeq Select and MANE (ncbiRefSeqSelect) gena (Prilog 1). Od 662 lokusa, na 136 lokusa nije anotiran nijedan gen. Usporedbom novo dobivene liste gena od 920 gena s prethodnom listom gena prema NCBI36/hg18 inačici genoma koju čini 781 gen dobiveno preklapanje gena je 56.74%. 64 lokusa koja nisu bila anotirana na inačici genoma NCBI36/hg18 na inačici genoma GRCh38/hg38 su anotirana.

### 4.2 POTENCIJALNI KANDIDATSKI GENI

Od 920 gena dobivenih anotacijom koordinata na novu inačicu genoma u bazama podataka *The Human Protein Atlas*, autDB, SFARI, SynGO, *Reactome* i UniProt pronađeno je 214 gena. U bazi *The Human Protein Atlas, detected only in brain* pronađeno je 4 gena, u bazi *The Human Protein Atlas, detected only in brain and at least one more tissue, but less than one-third of tissues* pronađeno je 63 gena, a u bazi SynGo pronađeno je 65 gena. U bazi autDB pronađeno je 70 gena, a u bazi SFARI 60 gena te se svi geni iz baze SFARI nalaze i u bazi autDB. U bazi *Reactome* pronađeno je 16 gena povezanih s neurorazvojnim sustavom i 7 gena povezanih s neuronskim sustavom, u UniProt bazi pronađeno je 15 gena. Geni pronađeni u bazama su se preklapali tako da su 3 gena pronađena u 4 baze (HPA2, SynGO, SFARI i autDB), 9 ih je pronađeno u bazama SynGO, autDB i SFARI, 1 gen je pronađen u bazama HPA2, SynGO i autDB. U dvije baze je pronađeno 7 gena (HPA2 i SynGO), 1 u HPA1 i SynGO, te 1 u HPA2 i autDB (Prilog 1). Ovaj popis gena čini popis gena kandidata povezanih s putevima ili biološkim procesima za koje se zna da su uključeni u razvoj središnjeg živčanog sustava – proširena lista gena kandidata.

#### 4.3 ONTOLOGIJA GENA

Nakon pronalazjenja 214 gena (proširena lista gena) od 920 (CNV lista gena) u bazama gena, za ostalih 706 protein kodirajućih gena istražena je molekularna funkcija i uključenost u biološke procese nastalih proteina preko baze Enrichr gledajući ontologije *Gene Ontology Biological Process 2021*, *Gene Ontology Molecular function 2021* i *Human Phenotype Ontology*. Najviše proteina uključeno je u biološke procese popravka izrezivanja nukleotida ( $p=0.0003351$ ), pozitivnu regulaciju acetilacije histona H3-K9 i regulaciju acetilacije histona H3-K9. Što se tiče molekularne funkcije najviše proteina ima aktivnost alfa-amilaze ( $p=0.0004170$ ) i amilaze te aktivnost inhibitora proteina Ser/Thr fosfataza. Najučestalije fenotipske abnormalnosti uzrokovane genima i genskim produktima su renovaskularna hipertenzija ( $p=0,0007583$ ), hiperakuzija i „jašuća“ aorta.

#### 4.4 EKSPRESIJA KANDIDATSKIH GENA I ODGOVARAJUĆI MIŠJI MODELI

Za proširenu listu gena kandidata provjerena je prenatalna ekspresija u mozgu pomoću baze HBT i zabilježena je pojačana ekspresija u mozgu za 176 gena. Geni iz baze UniProt birani su prema tome imaju li povišenu ekspresiju u mozgu i jesu li tkivno specifični za mozak.

Postnatalna ekspresija provjerena je u bazi GTEx te je zabilježena pojačana postnatalna ekspresija u mozgu u odnosu na ostala tkiva kod 80 gena. Od tih 80 gena, za njih 69 zabilježena je i povećana prenatalna ekspresija u mozgu u bazi HBT.

Nakon provjere postnatalne i prenatalne ekspresije provjereno je postojanje odgovarajućih mišjih modela u bazi MGI. Abnormalni mišji fenotip koji uključuje bihevioralni/neurološki, kraniofacijalni, embrionalni ili neuralni zabilježeni abnormalni mišji fenotip zamijećen je kod 126 gena.

#### 4.5 KLINIČKA VAŽNOST

Za sve gene kandidate iz baze DECIPHER izvađene su vrijednosti pLi, LOEUF, sHet, pHaplo, pTriplo, OMIM, Morbid, ClinGen gene/disease, ClinGen Dosage Sensitivity, GeneReviews, Gene2Phenotype i Gene CC s osvrtom o svakom genu te vrijednosti o ukupnom broju CNV, postotku *gain* i *loss*, postotku za svaku kategoriju veličine zabilježene CNV i prisutni fenotip uz CNV iz *matching patient variants* (Prilog 2). Geni u kojima CNV može imati klinički značaj s obzirom na vrijednosti pHaplo ( $pHaplo \geq 0.86$ ) i pTriplo ( $pTriplo \geq 0.94$ ) su dodatno izdvojeni. Za gene kod

kojih je zabilježena delecija u pacijenata gledana je vrijednost pHaplo, a za gene kod kojih je zabilježena amplifikacija kod pacijenata je gledana vrijednost pTriplo. Ukupno je izdvojeno 29 gena (Tablica 1).

Dodatno svi geni su provjereni u bazi ClinGen zasebno te su vrijednosti odgovarale vrijednostima iz baze DECIPHER.

Tablica 1. Popis gena izabranih kao potencijalnih gena kandidata prema vrijednostima pHaplo i pTriplo iz baze DECIPHER, crvenom bojom označene delecije, plavom bojom označene amplifikacije, OMIM označava *Online Mendelian inheritance in Man*, vrijednost koja se označava svakom unosu u bazu OMIM.

GEN	pHaplo	pTriplo	OMIM	RefSeq oznaka	Naziv gena	HGNC ID	Lokacija	Skupina gena
<i>KAT2A</i>	0,34	1	602301	NM_021078.3	lysine acetyltransferase 2A	4201	17q21.2	GCN5 related N-acetyltransferases ATAC complex SAGA complex Lysine acetyltransferases Bromodomain containing
<i>CNP</i>	0,44	0,98	123830	NM_033133.5	2',3'-cyclic nucleotide 3' phosphodiesterase	2158	17q21.2	
<i>ATP6V0A1</i>	0,51	0,98	192130	NM_001130021.3	ATPase H+ transporting V0 subunit a1	865	17q21.2	V-type ATPase subunits MicroRNA protein coding host genes
<i>NR1H2</i>	0,6	0,94	600380	NM_007121.7	nuclear receptor subfamily 1 group H member 2	7965	19q13.3	Nuclear receptor subfamily 1 group H
<i>LIMK1</i>	0,65	0,95	601329	NM_002314.4	LIM domain kinase 1	6613	7q11.23	PDZ domain containing LIMK/TESK kinase family LIM domain containing
<i>PSME3</i>	0,65	1	605129	NM_005789.4	proteasome activator subunit 3	9570	17q21.31	Proteasome
<i>YWHAH</i>	0,68	0,94	113508	NM_003405.4	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein eta	12853	22q12.3	14-3-3 phosphoserine/phosphothreonine binding proteins
<i>DPF1</i>	0,71	0,96	601670	NM_001135155.3	double PHD fingers 1	20225	19q13.2	PHD finger proteins BAF complex

<i>KCNH4</i>	0,76	0,98	604528	NM_012285.3	potassium voltage-gated channel subfamily H member 4	6253	17q21.2	Potassium voltage-gated channels PAS domain containing
<i>RAB35</i>	0,81	0,98	604199	NM_006861.7	RAB35, member RAS oncogene family	9774	12q24.23	RAB, member RAS oncogene GTPases
<i>GTF2I</i>	0,83	0,97	601679	NM_032999.4	general transcription factor Iii	4659	7q11.23	General transcription factors
<i>DHX8</i>	0,84	1	600396	NM_004941.3	DEAH-box helicase 8	2749	17q21.31	Spliceosomal P complex DEAH-box helicases Spliceosomal C complex SSU processome
<i>CABP1</i>	0,87	0,97	605563	NM_001033677.2	calcium binding protein 1	1384	12q24.31	EF-hand domain containing
<i>AP2AI</i>	0,87	1	601026	NM_130787.3	adaptor related protein complex 2 subunit alpha 1	561	19q13.33	Adaptor related protein complex 2 Clathrin/coatamer adaptor, adaptin-like, N-terminal domain containing MicroRNA protein coding host genes
<i>SCAF1</i>	0,88	0,96	617264	NM_021228.3	SR-related CTD associated factor 1	30403	19q13.3-q13.4	
<i>ARHGAP35</i>	0,95	0,99	605277	NM_004491.5	Rho GTPase activating protein 35	C:4591	19q13.32	Rho GTPase activating proteins
<i>GRIN2D</i>	0,96	1	602717	NM_000836.4	glutamate ionotropic receptor NMDA type subunit 2D	4588	19q13.33	Glutamate ionotropic receptor NMDA type subunits
<i>PRR12</i>	0,97	1	616633	NM_020719.3	proline rich 12	29217	19q13.33	
<i>MSI1</i>	0,98	0,94	603328	NM_002442.4	musashi RNA binding protein 1	7330	12q24.31	RNA binding motif containing
<i>STAT3</i>	0,98	1	102582	NM_139276.3	signal transducer and activator of transcription 3	11364	17q21.2	SH2 domain containing
<i>KDM2B</i>	0,99	0,99	609078	NM_032590.5	lysine demethylase 2B	13610	12q24.31	PHD finger proteins Lysine demethylases Zinc fingers CXXC-type MicroRNA protein coding

								host genes F-box and leucine rich repeat proteins
<i>DST</i>	1	0,94	113810	NM_001374736.1	Dystonin	1090	6p12.1	Plakins EF-hand domain containing
<i>PROX1</i>	1	0,99	601546	NM_001270616.2	prospero homeobox 1	9459	1q32.3	PROS class homeoboxes
<i>XPO1</i>	1,00	1,00	602559	NM_003400.4	exportin 1	12825	2p15	Exportins
<i>SETD1B</i>	0,94	1	611055	NM_001353345.2	SET domain containing 1B, histone lysine methyltransferase	29187	12q24.31	SET domain containing Lysine methyltransferases RNA binding motif containing
<i>CLASP2</i>	0,95	0,79	605853	NM_001365631.1	cytoplasmic linker associated protein 2	17078	3p22.3	TOG domain containing
<i>SLC17A7</i>	0,95	0,97	605208	NM_020309.4	solute carrier family 17 member 7	16704	19q13.33	Solute carrier family 17
<i>MAGI2</i>	0,98	0,96	606382	NM_012301.4	membrane associated guanylate kinase, WW and PDZ domain containing 2	18957	7q21.11	Membrane associated guanylate kinases PDZ domain containing
<i>TRIM71</i>	0,99	0,62	618570	NM_001039111.3	tripartite motif containing 71	32669	3p22.3	Ring finger proteins Tripartite motif family

#### 4.6 NEKODIRAJUĆE SEKVENCE UNUTAR DETEKTIRANIH ABERACIJA

Za 29 gena kandidata za koje je definiran klinički značaj analizirana je interakcija genomskih sekvenci unutar detektiranih aberacija provjerom postojanja postojanje lncRNA u bazama RNAcentral i UCSC *Genome Browser*. U bazi RNAcentral pronađeni su podaci za 11 gena, a u bazi UCSC *Genome Browser* samo za 4 gena od kojih su za 1 gen pronađene 2 lncRNA. Samo za jednu lncRNA, YWHAH-AS1 utvrđena je povišena ekspresija u mozgu (Tablica 2).

Tablica 2. Popis gena za koje su pronađene lncRNA u bazi RNACentral i u bazi UCSC *Genome Browser*, dodana je pozicija zapisa iz baze UCSC *Genome Browser* na kromosomu i postnatalna ekspresija zapisa iz baze GTEx.

Naziv gena	Accession number	HGNC ID	Broj zapisa u bazi RNACentral	Naziv zapisa u UCSC <i>Genome Browser</i>	Genecode trankript	Pozicija	GTEx ekspresija
<i>MAGI2</i>	NM_012301.4	HGNC:18957	182	MAGI2-AS1	ENST00000428298.1	hg38 chr7:78,939,850-78,940,895	povišena u adipoznom tkivu
				MAGI2-AS2	ENST00000411616.1	hg38 chr7:79,008,988-79,012,277	Povišena u testisima
<i>DST</i>	NM_001374736.1	HGNC:1090	19	DST-AS1	ENST00000426453.1	hg38 chr6:56,844,002-56,864,078	Povišena u testisima i
<i>YWHAH</i>	NM_003405.4	HGNC:12853	19	YWHAH-AS1	ENST00000484682.2	hg38 chr22:31,925,519-31,945,346	povišena u mozgu ( <i>Nucleus accumbens</i> , bazalni gangliji)
<i>GTF2I</i>	NM_032999.4	HGNC:4659	16	GTF2I-AS1	ENST00000669011.1	hg38 chr7:74,688,864-74,727,886	Povišeno u EBV transformiranim limfocitima
<i>KDM2B</i>	NM_032590.5	HGNC:13610	20				
<i>CABPI</i>	NM_001033677.2	HGNC:1384	17				
<i>STAT3</i>	NM_139276.3	HGNC:11364	3				
<i>NR1H2</i>	NM_007121.7	HGNC:7965	1				
<i>PRR12</i>	NM_020719.3	HGNC:29217	1				
<i>RAB35</i>	NM_006861.7	HGNC:9774	1				
<i>XPO1</i>	NM_003400.4	HGNC:12825	1				

#### 4.1 STRUKTURNE ABERACIJE PRISUTNE U OPĆOJ POPULACIJI

U bazi DGV provjerene su strukturne aberacije prisutne u općoj populaciji koje spadaju u DGV *Gold Standard Variants* za 29 gena kandidata definirane kliničke važnosti. Samo za 3 gena (*CLASP2*, *DHX8* i *MAGI2*) pronađeni su zapisi u bazi DGV od kojih su za 2 gena pronađeni samo po 1 zapis dok je za treći gen (*MAGI2*) pronađeno 10 zapisa, ali je uzorak u istraživanjima bio mali.

#### 4.2 SIGNALNI I METABOLIČKI PUTOVI

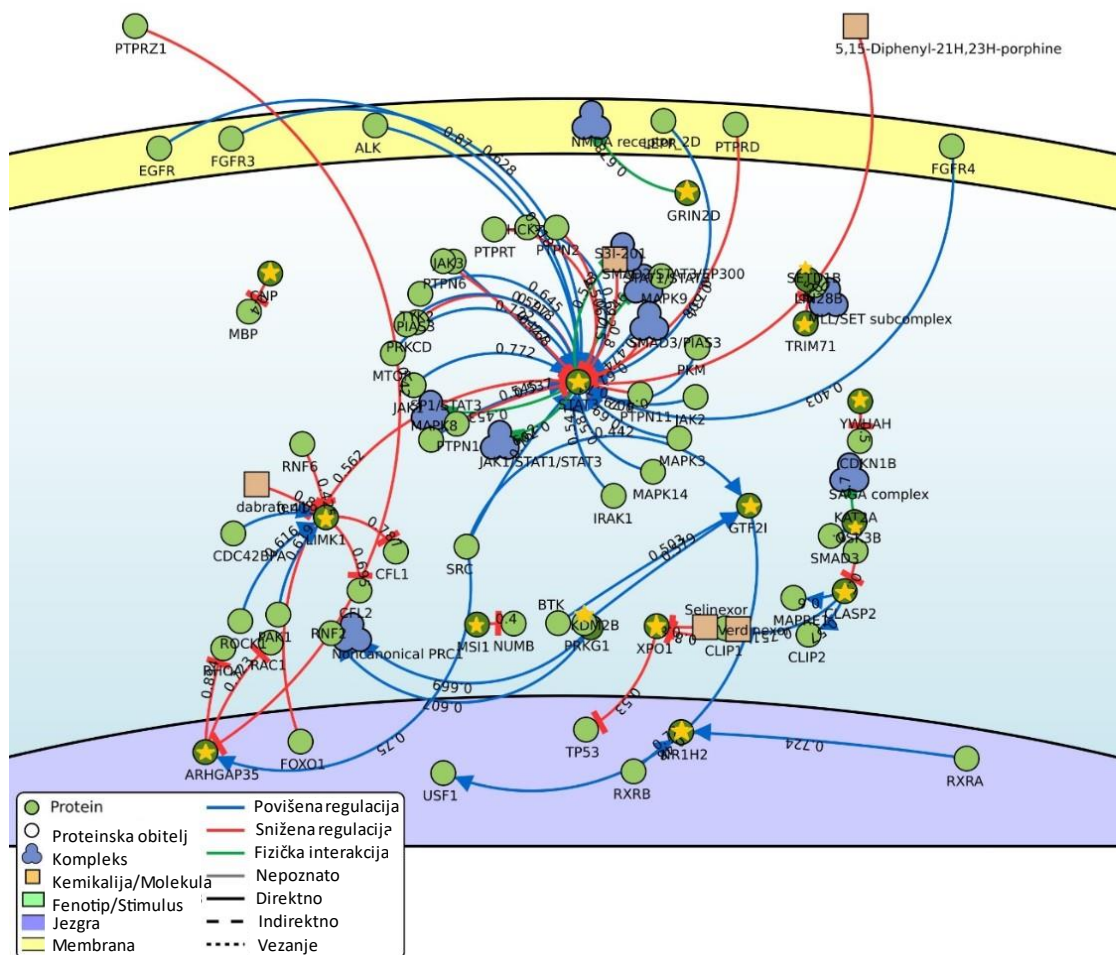
Proteinske obitelji i klase te molekularna funkcija i stanični putevi u koje su uključeni proteini nastali sintezom proteina od 29 gena kandidata dobiveni su korištenjem baze PANTHER. Za 8 gena nisu dostupni podaci (Tablica 3).



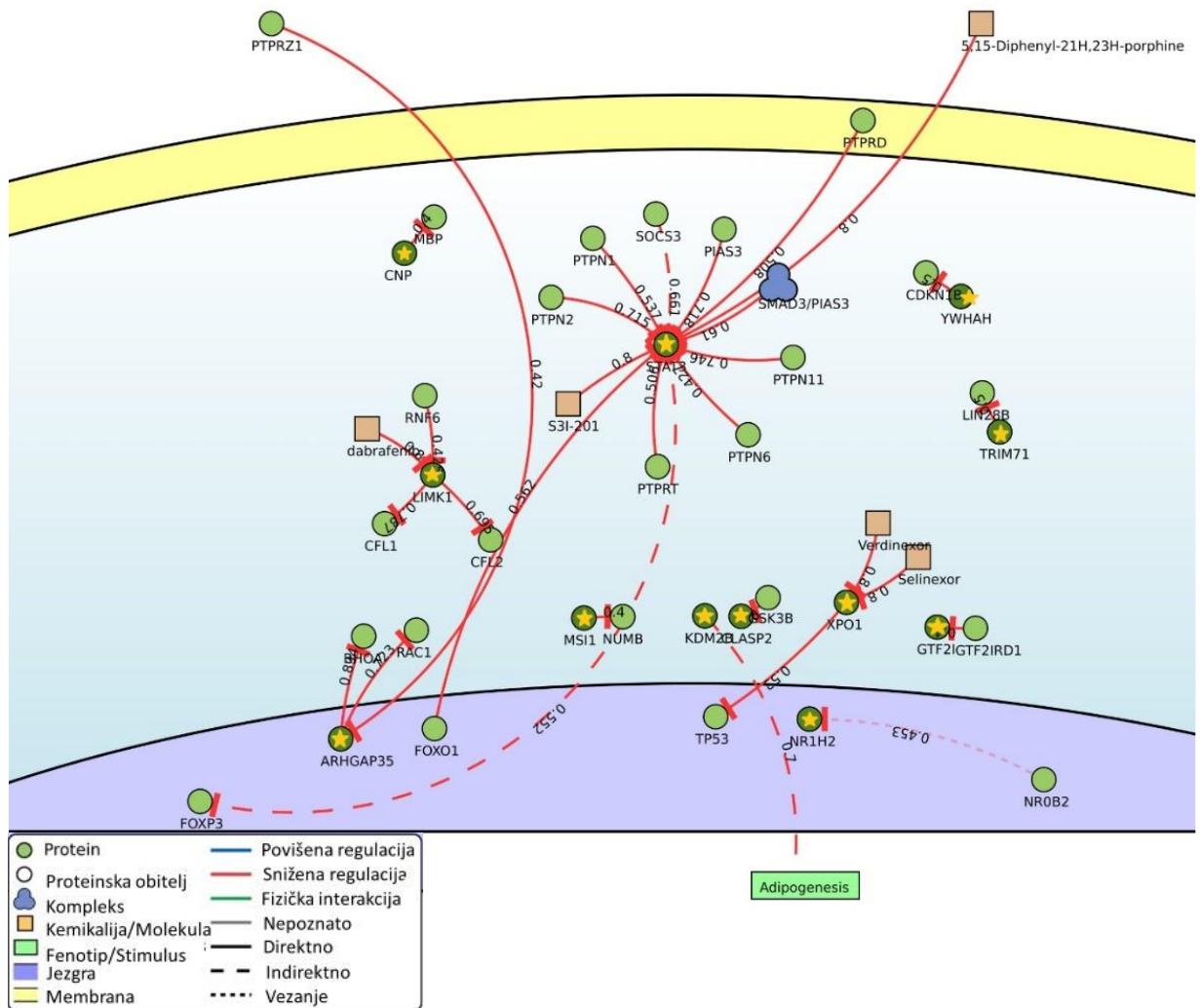
Tablica 3. Tablica uključenosti gena kandidata u biološke procese prema bazi PANTHER, zelena boja označava da je gen uključen u proces.

		<i>PSME3</i>	<i>AP2A1</i>	<i>MAGI2</i>	<i>NRIH2</i>	<i>GRIN2D</i>	<i>PROX1</i>	<i>STAT3</i>	<i>GTF2I</i>	<i>ATP6V0A1</i>	<i>DHX8</i>	<i>SCAF1</i>	<i>XPO1</i>	<i>CLASP2</i>	<i>DST</i>	<i>KDM2B</i>	<i>SETD1B</i>	<i>CNP</i>	<i>KAT2A</i>	<i>LIMK1</i>	<i>KCNH4</i>	<i>SLC17A7</i>		
Vežanje	Vežanje heterocikličkih spojeva																							
	Vežanje iona																							
	Vežanje neurotransmitera																							
	Vežanje organskih cikličkih spojeva																							
	Vežanje proteina																							
	Vežanje malih molekula																							
Katalitička aktivnost	Katalitička aktivnost djelovanja na protein																							
	Demetilazna aktivnost																							
	Hidrolazna aktivnost																							
	Transferazna aktivnost																							
Aktivnost molekularnog adaptera	Protein-makromolekula adapterska aktivnost																							
Regulator molekularne funkcije	Aktivnost regulatora enzima																							
Aktivnost molekularni provoditelj	Aktivnost signalnog receptora																							
Aktivnost transkripcijske regulacije	Aktivnost DNA-vezujućeg transkripcijskog faktora																							
	Aktivnost transkripcijskog kofaktora																							
Aktivnost prenositelja	Aktivnost transmembranskog prenositelja																							

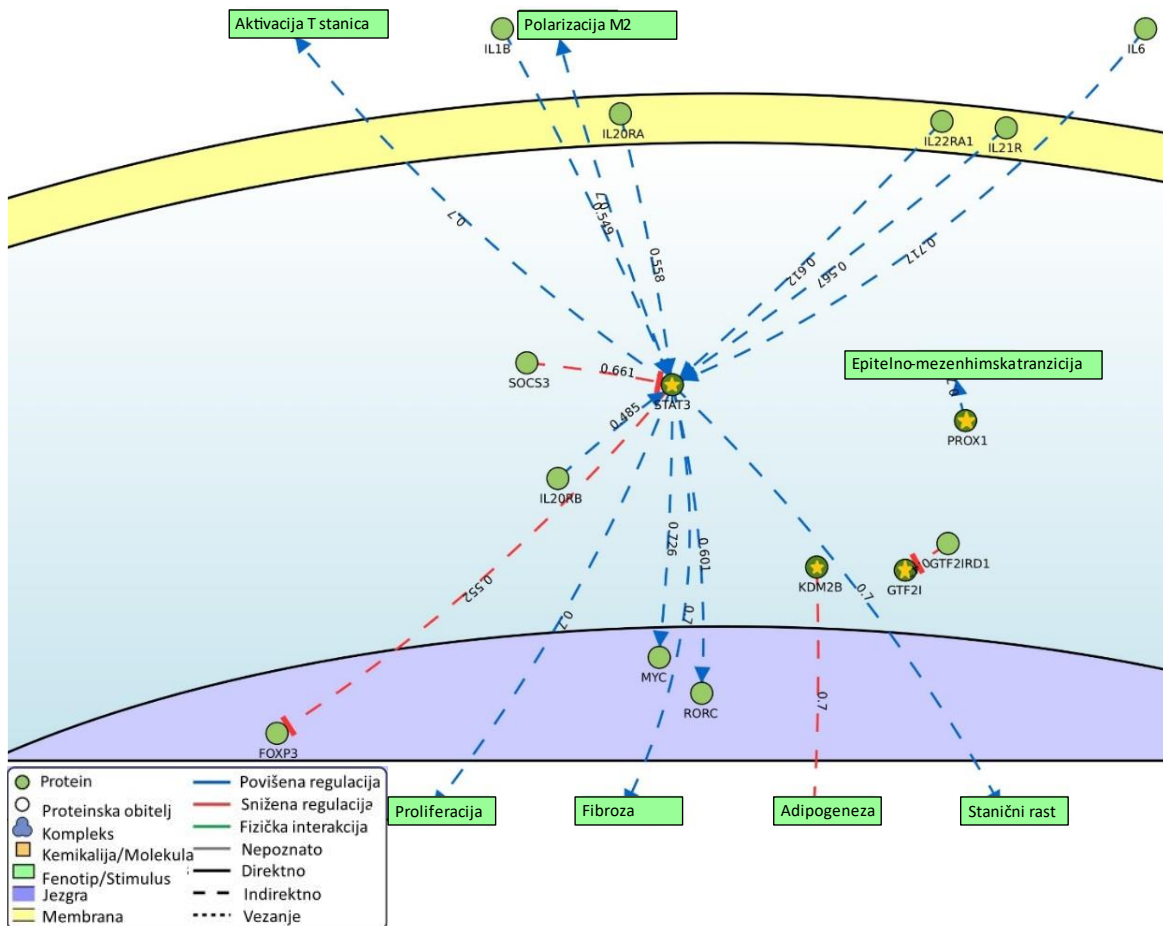
Uključenost proteina kodiranih genima kandidatima u stanične puteve i odnose vizualizirana je korištenjem baze SIGNOR. U bazi nisu pronađeni stanični putevi i odnosi za 11 proteina kodiranih genima kandidatima. Sveukupno je pronađeno 228 odnosa interakcija, nije pronađena nijedna interakcija vezanja. Interakcije su dobivene za kategorije direktnih interakcija (Slika 7), snižene ili smanjene regulacije (Slika 8), indirektnih interakcija (Slika 9), transkripcijske regulacije (Slika 10) i povišene ili povećane regulacije (Slika 11).



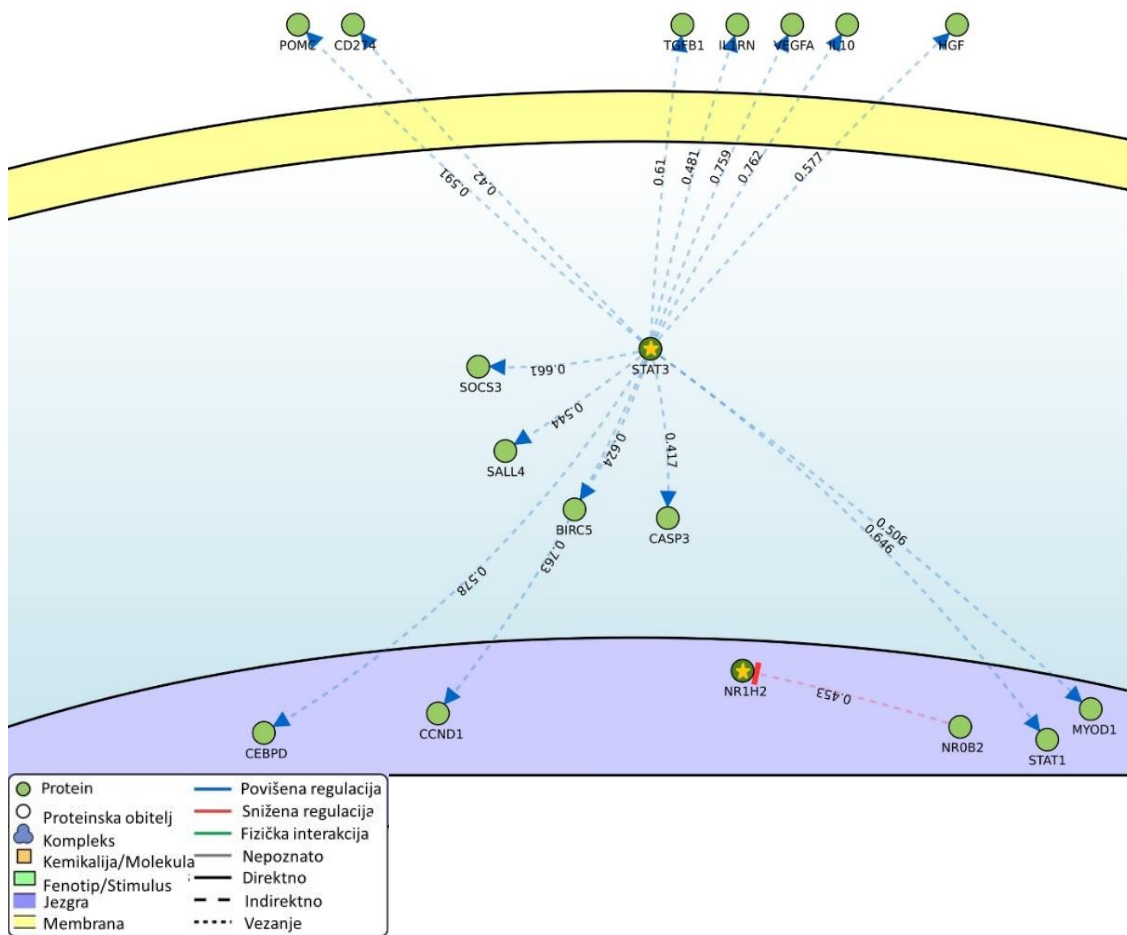
Slika 7. Prikaz direktnih interakcija i odnosa produkata gena kandidata prilagođeno iz baze SIGNOR, žutom zvjezdicom su označeni produkti gena kandidata koji su u interakciji; *ARHGAP35*, *NR1H2*, *XPO1*, *CLASP2*, *KAT2A*, *KDM2B*, *MSI1*, *LIMK1*, *STAT3*, *YHWHAH*, *SETD1B*, *TRIM71*, *GRIN2D*, *CNP*, *GTF2I*. Kod *CLASP2*, *TRIM71* i *SETD1B* kod ispitanika došlo je do amplifikacije dijela gena dok je kod svih ostalih došlo do gubitka dijela gena.



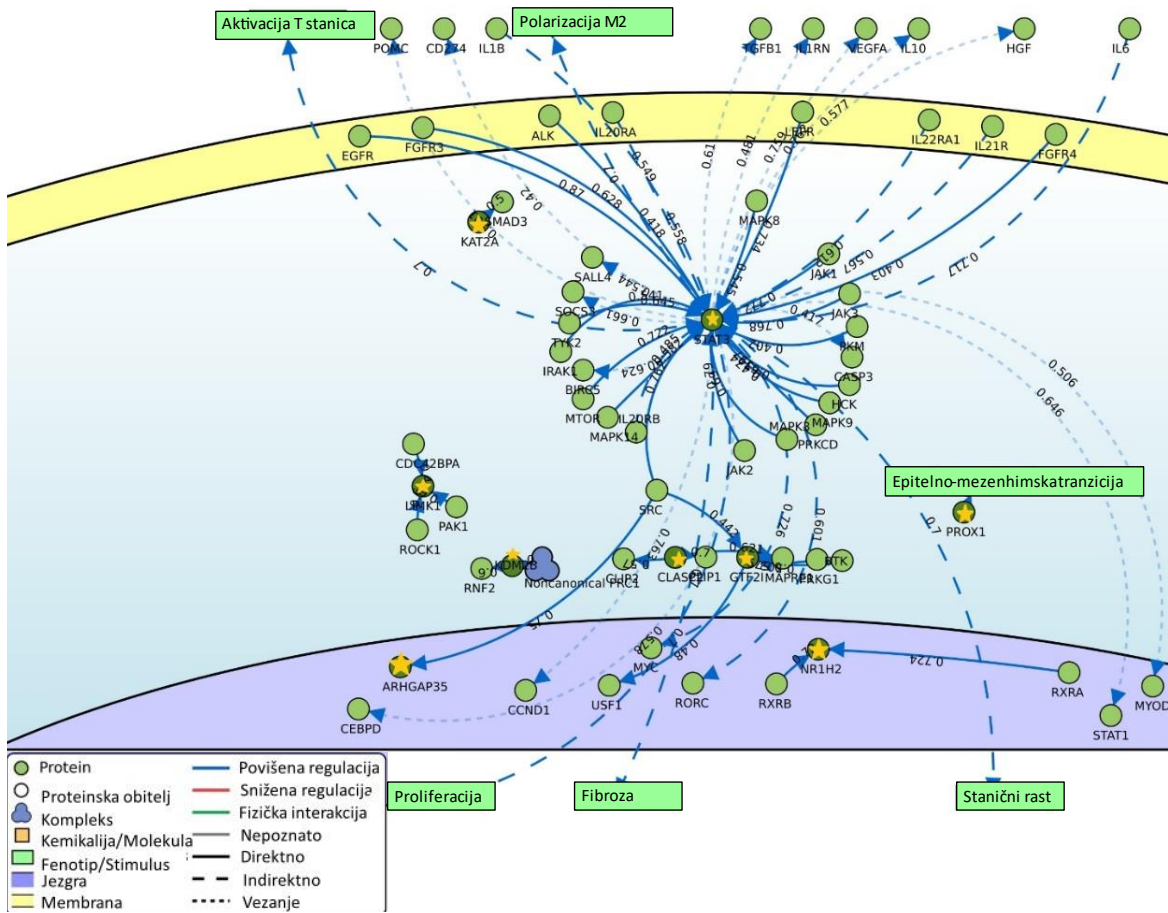
Slika 8. Prikaz snižene regulacije i odnosa produkata gena kandidata. Prilagođeno iz baze SIGNOR, žutom zvjezdicom su označeni produkti gena kandidata; *STAT3*, *LIMK1*, *YWHAH*, *TRIM71*, *GTF2I*, *NR1H2*, *KDM2B*, *MSI1*, *ARHGAP35*, *XPO1*, *CNP* i *CLASP2*.



Slika 9. Prikaz indirektnih interakcija i odnosa produkata gena kandidata. Prilagođeno iz baze SIGNOR, žutom zvjezdicom su označeni produkti gena kandidata; *GTF21*, *PROX1*, *KDM2B* i *STAT3*.



Slika 10. Prikaz transkripcijske regulacije i odnosa produkata gena kandidata. Prilagođeno iz baze SIGNOR, žutom zvjezdicom su označeni produkti gena kandidata; *STAT3* i *NR1H2*.



Slika 11. Prikaz povišene regulacije i odnosa produkata gena kandidata. Prilagođeno iz baze SIGNOR, žutom zvjezdicom su označeni produkti gena kandidata; *KAT2A*, *STAT3*, *PROX1*, *NR1H2*, *GTF2I*, *CLASP2*, *ARHGAP3*, *LIMK1* i *KDM2B*.

Za gene kandidate dodatno su provjereni točni biološki putovi u koje su uključeni njihovi produkti pomoću baze *Reactome* (Tablica 4).

Tablica 4. Geni kandidati i putovi u koje su uključeni prema bazi *Reactome*

<b>Geni</b>	<b>Biološki putovi</b>
<i>KAT2A</i>	prijenos signala; NOTCH signalizacija, gastrulacija; formacija paraksijalnog mezoderma, ekspresija gena; transkripcijska regulacija s RUNX3 i epigenetička regulacija, bolesti; NOTCH1 signalizacija u tumorima, organizacija kromatina, kardiogeneza, metabolizam proteina; deubikvitinacija
<i>CNP</i>	ekspresija gena; transkripcijska regulacija s MECP2, PRIJENOS SIGNALA; PTK6 signalizacija
<i>ATP6V0A1</i>	prijenos signala; signalizacija inzulinskim receptorom, transport i unos željeza
<i>NR1H2</i>	transport malih molekula; slaganje lipoproteina plazme, ekspresija gena, prijenos signala; signalizacija s NR1H2 i NR1H3, metabolizam lipida i proteina
<i>LIMK1</i>	signalizacija Rho GTPazama, razvoj živčanog sustava; usmjeravanje rasta aksona
<i>PSME3</i>	stanični ciklus; mitoz, replikacija DNA, transkripcijska regulacija s RUNX2, prijenos signala; NOTCH signalizacija, Hedgehog signalizacija; stanični odgovor na stres; transkripcijska regulacija s RUNX3, gastrulacija: formacija paraksijalnog mezoderma (somitogeneza)
<i>YWHAH</i>	stanični ciklus, intrinzični put apoptoze, transkripcijska regulacija s TP53, transport vezikulama
<i>DPF1</i>	transkripcijska regulacija s TP53, apoptoza, epigenetička regulacija genske ekspresije, mejoza, autofagija, prijenos signala (signalizacija s ERBB4)
<i>KCNH4</i>	neuronski sustav; kalijevi kanali
<i>RAB35</i>	metabolizam proteina; RAB geranilgeranilacija; transport vezikulama
<i>GTF2I</i>	epigenetička regulacija ekspresije gena, prijenos signala, imunostni sustav; stečeni i urođeni, metabolizam proteina, popravak DNA, mejoza, stanični odgovor na stres, metabolizam proteina, ekspresija gena, DNA popravak, bolesti
<i>DHX8</i>	metabolizam RNA
<i>CABP1</i>	senzorička percepcija; procesiranje zvuka
<i>AP2A1</i>	prijenos signala; WNT signalizacija, razvoj živčanog sustava; usmjeravanje rasta aksona, transport vezikulama,
<i>SCAF1</i>	/
<i>ARHGAP35</i>	prijenos signala; signalizacija nereceptorskim tirozin kinazama, razvoj živčanog sustava; usmjeravanje rasta aksona, prijenos signala; signalizacija Rho GTPazama
<i>GRIN2D</i>	prijenos signala; obitelj MAPK signalizacijskih kaskada, neuronski sustav; transmisija preko kemijskih sinapsi (aktivacija NMDA receptora i postsinaptički događaji), protein-protein interakcija na sinapsama
<i>PRR12</i>	/
<i>MSI1</i>	razvoj živčanog sustava; usmjeravanje rasta aksona
<i>STAT3</i>	imunostni sustav, prijenos signala, razvoj; transkripcijska regulacija pluripotentnih stem stanica, stanični odgovor na stres,
<i>KDM2B</i>	organizacija kromatina
<i>DST</i>	komunikacija između stanica; organizacija staničnih spojeva, signalizacija Rho GTPazama
<i>PROX1</i>	regulacija ekspresije gena u beta stanicama
<i>XPO1</i>	stanični ciklus, kontrolne točke staničnog ciklusa, signalizacija RhoGTPazama, signalizacija TGFB obitelji, WNT signalizacija, stanični odgovor na stres, metabolizam RNA, signalizacija obitelji MAPK
<i>SETD1B</i>	organizacija kromatina, transkripcijska regulacija s RUNX1
<i>CLASP2</i>	stanični ciklus i kontrolne točke staničnog ciklusa, signalizacija RhoGTPazama, razvoj živčanog sustava; usmjeravanje rasta aksona
<i>SLC17A7</i>	ciklus otpuštanja neurotransmitera
<i>MAG12</i>	komunikacija između stanica
<i>TRIM71</i>	stečeni imunostni sustav

#### 4.3 LITERATURNI PREGLED I OSNOVNE INFORMACIJE O GENIMA

Za sve gene s liste potencijalnih gena kandidata izvučene su osnovne informacije putem baze *GeneCards*. Za sve gene su dostupni *Entrez Gene Summary* i *GeneCards Summary*. Za sve osim PRR12 dostupan je i UniProtKB/Swiss-Prot Summary.

Prema zadanim kriterijima za proučavanje literature u bazi PubMed, literaturni navodi nisu pronađeni za 3 gena kandidata; *CNP*, *DHX8* i *AP2A1*.

#### 4.4 KANDIDATSKI GENI

Geni za koje je zabilježena klinička značajnost izdvojeni su kao geni kandidati. U tu skupinu spada 29 gena raznolikih funkcija (Tablica 3) i uključenosti u različite biološke puteve (Tablica 4). Od 29 gena, samo u 4 je došlo do amplifikacijske CNV dok je kod ostalih 25 došlo do delecijske CNV. Od tih 29 gena, 4 gena možemo pronaći u bazi HPA i pojačana im je ekspresija u mozgu i bar još jednom tkivu, ali manje od trećine tkiva. Ta 4 gena ne pokazuju povećanu prenatalnu ekspresiju u mozgu prema HBT, ali je zabilježena povećana postnatalna ekspresija u mozgu u odnosu na ostala tkiva u bazi GTEX. U bazi SynGO možemo pronaći 8 od 29 gena kandidata, s tim da se 1 gen, *GRIN2D* preklapa s bazom HPA2 te je za sve gene zabilježena povećana prenatalna ekspresija, ali ne i postnatalna ekspresija. U bazi autDB pronađeno je 8 gena od kojih se 7 nalazi i u bazi SFARI te je zabilježena povećana prenatalna ekspresija za njih 6 dok je povećana postnatalna ekspresija u mozgu zabilježena samo za 1 gen, *PRR12*. U bazi RNAcentral pronađeni su zapisi za 11 gena, a u bazi MGI je pronađeno 7 abnormalnih mišjih fenotipova povezanih s genima kandidatima. Svi geni su pronađeni u bazi UniProt. Za gene *CLASP2*, *DHX8* i *MAGI2* pronađeni su zapisi o strukturnim aberacijama prisutnim u općoj populaciji u bazi DGV (Tablica 5). Literaturni navodi koji povezuju gene *CNP*, *DHX8* i *AP2A1* s neurorazvojnim poremećajima nisu pronađeni.

Od 29 gena njih 24 zabilježeno je kod ispitanika 5, i to na chr2 (1), chr 7 (2), chr12 (5), chr 17 (7), chr19 (8) i chr 22 (1)

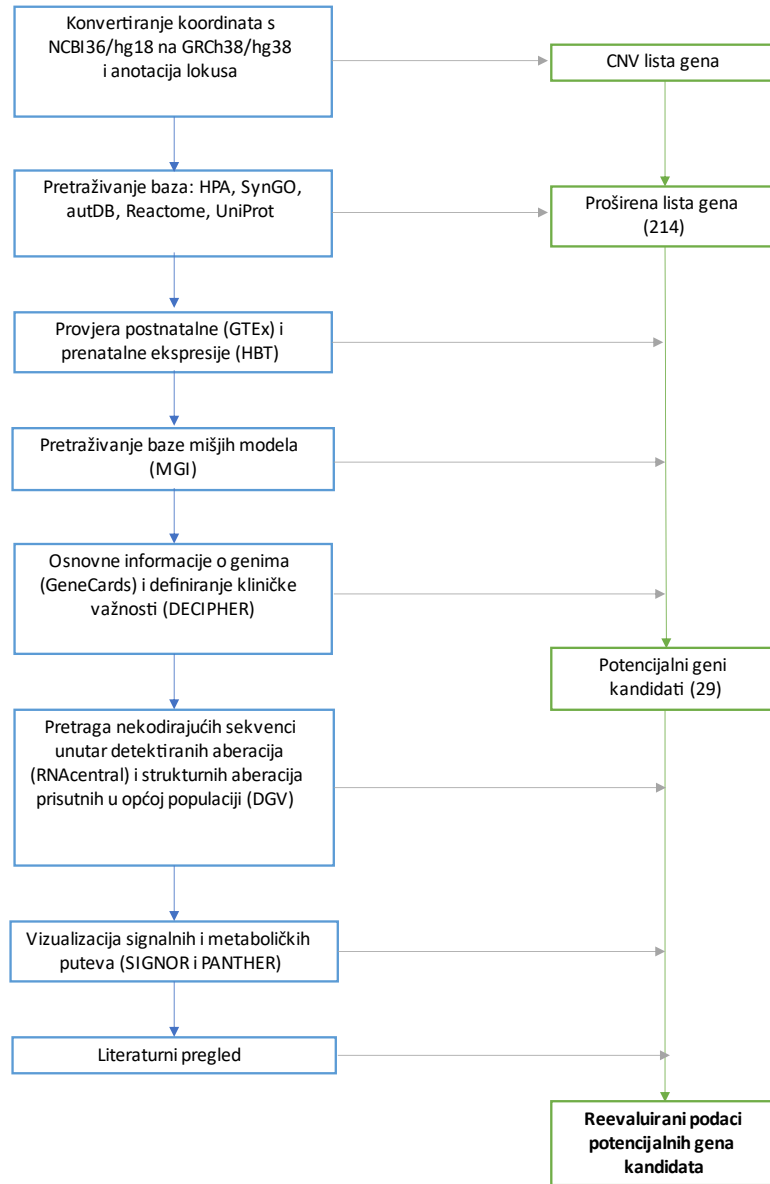


Tablica 5. Popis potencijalnih kandidatskih gena. Geni su provedeni kroz navedene baze, plus i zelena boja označavaju da su se u pojedinoj bazi pronašli podaci za gen.

Naziv gena	Accession number	Baza HPA1	Baza HPA2	SynGO	autDB	SFARI	Reactome 1	Reactome2	HBT	GTEx	RNAcentral	UniProt	MGI
<i>PROX1</i>	NM_001270616.2	-	-	-	-	-	-	-	+	-	-	+	+
<i>XPO1</i>	NM_003400.4	-	-	-	+	+	-	-	+	-	+	+	-
<i>TRIM71</i>	NM_001039111.3	-	+	-	-	-	-	-	-	+	-	+	+
<i>CLASP2</i>	NM_001365631.1	-	-	-	-	-	+	-	+	+	-	+	-
<i>DST</i>	NM_001374736.1	-	-	-	+	+	-	-	+	-	+	+	-
<i>LIMK1</i>	NM_002314.4	-	-	-	-	-	+	-	+	+	-	+	-
<i>GTF2I</i>	NM_032999.4	-	-	-	+	+	-	-	+	-	+	+	-
<i>MAGI2</i>	NM_012301.4	-	-	+	-	-	-	-	+	+	+	+	-
<i>RAB35</i>	NM_006861.7	-	-	+	-	-	-	-	+	-	+	+	-
<i>MSI1</i>	NM_002442.4	-	-	-	-	-	+	-	+	+	-	+	-
<i>CABP1</i>	NM_001033677.2	-	-	+	-	-	-	-	+	+	+	+	-
<i>KDM2B</i>	NM_032590.5	-	-	-	+	+	-	-	+	-	+	+	-
<i>SETD1B</i>	NM_001353345.2	-	-	-	+	+	-	-	-	-	-	+	+
<i>CNP</i>	NM_033133.5	-	-	-	-	-	-	-	+	+	-	+	+
<i>KAT2A</i>	NM_021078.3	-	-	-	-	-	-	-	+	-	-	+	+
<i>KCNH4</i>	NM_012285.3	-	+	-	-	-	-	-	+	+	-	+	-
<i>STAT3</i>	NM_139276.3	-	-	+	-	-	-	-	+	-	+	+	-
<i>ATP6V0A1</i>	NM_001130021.3	-	-	+	-	-	-	-	+	+	-	+	-
<i>PSME3</i>	NM_005789.4	-	-	-	-	-	+	-	+	-	-	+	-
<i>DHX8</i>	NM_004941.3	-	-	-	-	-	-	-	+	-	-	+	+
<i>DPF1</i>	NM_001135155.3	-	+	-	-	-	-	-	+	+	-	+	-
<i>ARHGAP35</i>	NM_004491.5	-	-	-	-	-	+	-	-	-	-	+	+
<i>GRIN2D</i>	NM_000836.4	-	+	+	-	-	-	-	+	+	-	+	-
<i>SLC17A7</i>	NM_020309.4	-	-	+	-	-	-	-	+	+	-	+	-
<i>PRR12</i>	NM_020719.3	-	-	-	+	+	-	-	-	+	+	+	-
<i>SCAF1</i>	NM_021228.3	-	-	-	+	+	-	-	+	-	-	+	-
<i>AP2A1</i>	NM_130787.3	-	-	-	-	-	+	+	+	-	-	+	-
<i>NR1H2</i>	NM_007121.7	-	-	-	+	-	-	-	+	-	+	+	-
<i>YWHAH</i>	NM_003405.4	-	-	+	-	-	-	-	+	+	+	+	-

#### 4.5 HODOGRAM REEVALUACIJE PODATAKA

Kako bi ponovna analiza bila optimalna, utvrđen je hodogram reevaluacije rezultata CMA (Slika 12).



Slika 12. Optimizirani hodogram reevaluacije rezultata. Od korištenih baza izostavljena je baza SFARI jer se podaci preklapaju s bazom autDB. Isto tako tijekom reevaluacije baza ClinGen je izostavljena zbog preklapanja s DECIPHER bazom.

## 5. RASPRAVA

### 5.1 ANOTACIJA ABERACIJA NA NOVU INAČICU GENOMA

Prvi cilj bio je anotirati sve detektirane promjene u broju kopija DNA na novu inačicu ljudskog genoma (hg38). Dobivene razlike u anotaciji koordinata rezultat su razlika u genomima. Prethodno dobiveni rezultati promjena broja kopija dobiveni su metodom CMA korištenjem inačice genoma NCBI36/hg18 objavljene 2006, prve inačice koja je bila u širokoj upotrebi. Reevaluacija rezultata odrađena je na inačici genoma GRCh38/hg38 objavljenoj 2013 i zadnji put ažuriranoj 2019. Preklapanje popisa gena od 56.74% objašnjava se boljom anotacijom gena u inačici genoma hg38. Inačica genoma hg38 pruža precizniji prikaz ljudskog genoma, sadrži ažuriranja koja ispravljaju greške, popunjavaju praznine i poboljšavaju kvalitetu sekvence. Bolja je pokrivenost genoma te sadrži cjelovitije sekvence centromera, telomera i ostalih zahtjevniji regija. Dodatno su ažurirane i anotacije gena i regulatornih elemenata (Guo i sur. 2017). Razlike u anotaciji rezultat su poboljšanja tehnika sekvenciranja, povećanju količine informacija, podacima o transkriptomu, boljem razumijevanju lncRNA, strukturnim varijacijama, komparativnoj genomici i integraciji više izvora podataka prilikom anotacije inačice genoma hg38 (Nassar i sur. 2022).

Osim toga, hg38 sadrži i dodatne značajke poput anotacije repetitivnih sekvenci, strukturnih varijacija i pseudogena. Zbog toga inačica hg38 pruža sveobuhvatnije razumijevanje funkcionalnih elemenata unutar genoma što omogućava točnije istraživanje genoma i analizu (Guo i sur. 2017).

Usprkos tome što hg38 ima prednosti u odnosu na prethodne inačice genoma, postoje i nedostaci u vidu 151 Mbp nepoznatih sekvenci raspršenih po genomu uključujući pericentromerne i subtelomerne regije, segmentalne duplikacije, nizove amplikona gena i rRNA od kojih su svi ključni za osnovne stanične procese. Neke od najvećih praznina su nizovi ponavljanja satelita te kratke ruke svih 5 akrocentričnih kromosoma (Miga i sur. 2014). Daljnji napredak osiguran je nasumičnim sekvenciranjem dugih očitavanja koje rezultira novom inačicom genoma T2T-CHM13. Nova inačica genoma je bez praznina i sadrži sekvence svih autosoma i X kromosoma od telomere do telomere (Nurk i sur. 2022). Za sad, za inačicu genoma T2T-CHM13 u UCSC *Genome Browser* nema još svih dostupnih podataka koji su dostupni za hg38, a relevantni su za kliničku interpretaciju genomskih aberacija (genome.ucsc.edu). Zbog navedenog T2T-CHM13 inačica nije korištena kod konverzije koordinata.

Razlike u pokrivenosti genoma, anotaciji i količini znanja o ljudskom genomu potvrđuju potrebu za reevaluacijom prethodno dobivenih rezultata kako bi bila u skladu s današnjim saznanjima.

Očekuje se da će u skorijoj budućnosti biti potrebna reevaluacija rezultata i na T2T-CHM13 inačicu kako bismo vidjeli utjecaj nove sekvence genoma na rezultate.

Delecije i amplifikacije koje zovemo promjene broja kopija DNA doprinose velikim brojem promjenama u genomu i imaju dijagnostički značaj zbog važne uloge koje mogu igrati kao uzročnici bolesti (Singh i sur. 2021). Zbog toga, važna je reevaluacija prethodno dobivenih rezultata.

## 5.2 OPTIMIZACIJA HODOGRAMA

Jedan od postavljenih ciljeva bio je utvrditi hodogram reevaluacije rezultata CMA, s ciljem postizanja optimalnih rezultata ponovne analize. Tijekom ponovne analize podataka utvrđeno je da se sadržaji baza autDB i SFARI podudaraju s tim da autDB baza sadrži sve podatke iz SFARI baze uz dodatne podatke. Zbog toga izostavljen je korak usporedbe anotiranog seta gena sa setom gena iz SFARI baze.

Izostavljanje navedenog koraka dodatno potvrđuje to što je baza SFARI vlasništvo SFARI.org i *The Simons Foundation* te je licencirana od strane *the Simons Foundation* od MindSpec-a. Baza autDB razvijena je od strane MindSpec-a i financirana od strane *The Simons Foundation*. Baze SFARI i autDB predstavljaju isti set podataka (Pereanu i sur. 2017).

Kako bi se utvrdila klinička važnost CNV-ova u genima iz našeg seta podataka korištene su baze ClinGen i DECIPHER. DECIPHER omogućava laki pristup kliničkim preporukama za procjenu patogenosti varijante korištenjem standarda dobre prakse poput smjernica od *the American College of Medical Genetics and Genomics and Association of Molecular Pathology* (ACMG/AMP) za interpretaciju varijanti te ACMG/ClinGen standarde za interpretaciju CNV-ova. Uz to, pruža poveznice na preporuke od ClinGen-a, uključujući opće preporuke i preporuke specifične za gene i bolesti (Foreman i sur. 2022, Foreman i sur. 2023). Zbog toga i zbog podudaranja podataka iz baza ClinGen i DECIPHER korak provjere kliničke važnosti u bazi ClinGen je izostavljen iz hodograma za dobivanje optimalnih rezultata.

## 5.3 PRENATALNA AKTIVNOST GENA

Od 29 gena kandidata, za njih 25 postoje podaci o ekspresiji u bazi HBT. Od tih 25 gena, njih 24 pokazuje povišenu i značajnu ekspresiju prema zadanim parametrima (intenzitet signala > 5). Ta

24 gena podijeljena su u kategorije: stabilna ekspresija, prenatalni pad ekspresije, prenatalni porast ekspresije. Stabilna ekspresija zabilježena je kod 7 gena: *CLASP2*, *GTF2I*, *STAT3*, *PSME3*, *AP2A1*, *SCAF1* i *YWHAH*. Prenatalni porast ekspresije zabilježen je kod gena *PROX1*, *DST*, *MAGI2*, *RAB35*, *CABP1*, *KDM2B*, *KAT2A*, *DPF1*, *GRIN2D*, *SLC17A7*, *CNP*, *KCNH4*. Prenatalni pad ekspresije zabilježen je kod gena: *NR1H2*, *DHX8*, *MSI1*, *LIMK1*, *XPO1*.

Geni za koje nema podataka u HBT bazi su: *TRIM71*, *SETD1B*, *ARHGAP35* i *PRR12*.

Za gene *GRIN2D*, *MAGI2*, *KDM2B*, *PROX1* i *RAB35* uočen je porast ekspresije do kasnog srednjeg fetalnog razdoblja. Za razvoj do kasnog srednjeg fetalnog razdoblja najbitniji su procesi migracije neurona, molekularne specifikacije, neuronske agregacije, diferencijacije dendrita te rast aksona, a u tom razdoblju započinju i procesi sinaptogeneze i spinogeneze te podrezivanja sinapsi i neurokemijske maturacije te mijelinizacije (Kostović 2020).

Kod gena *SLC17A7*, *CABP1*, *CNP* i *KCNH4* uočen je postnatalni porast ekspresije. Postnatalno su za razvoj mozga bitni procesi diferencijacije dendrita, neurokemijske maturacije i mijelinizacije (Kostović 2020).

Prenatalna ekspresija gena upućuje na njihovu uključenost u neurorazvojne procese. S obzirom na promjenu ekspresije ovisno o periodima razvoja mozga možemo povući korelaciju u koje su točno procese koji se događaju u tom periodu razvoja uključeni geni. Tako gledajući, geni sa stabilnom ekspresijom važni su za period od embrionalnog razvoja do kasnog odraslog doba čovjeka. Geni čija ekspresija se mijenja tijekom razvoja imaju različite uloge u različitim procesima razvoja. Stabilno eksprimirani geni uključeni su u općenitije procese poput staničnog ciklusa, razvoja, prijenosa signala, imunskog sustava, metabolizma i bolesti (Gillespie i sur. 2021).

Geni kandidati uključeni su u širok dijapazon procesa te se disrupcije nekih od njih mogu povezati s neurorazvojnim poremećajima poput neurotransmisije, senzoričkog procesiranja zvuka, organizacije kromatina, Notch signalizacije, ekspresije gena, sinapsi, otpuštanja neurotransmitera, signalizacije RhoGTP-azama, TGF $\beta$  i WNT (Michetti i sur. 2022, Mossink i sur. 2020, Nussinov i sur. 2022, Prashad i Gopal, 2020).

S obzirom na to da su *MSI*, *LIMK1* i *ARHGAP35* uključeni u jedan od ključnih procesa tijekom neurorazvoja, usmjeravanje rasta aksona, posebno ih treba izdvojiti kao gene od interesa u kojima promjene broja kopija koje dovode do disrupcije gena mogu upućivati na disrupciju procesa usmjeravanja rasta aksona i uzroka razvoja NDD-a. Dodatno treba izdvojiti gene *TRIM71*,

*SETD1B*, *ARHGAP35* i *PRR12* za koje nema podataka o njihovoj prenatalnoj ekspresiji no postoje radovi u PubMed bazi koji ih povezuju s neurorazvojem. Taj podatak dodatno daje na važnosti reevaluaciji podataka ako znamo da je baza HBT izašla 2011. i nisu dodavani recentniji podaci te bi se ovi geni mogli prethodno isključiti zbog nedovoljno informacija. Tome doprinosi i to da je gen *ARHGAP35* povezan s usmjeravanjem rasta aksona, jednim od ključnih procesa neurorazvoja.

Dodatno bi se trebala istražiti dolazi li do disrupcije samog procesa ako u genima uključenim u te procese zabilježimo promjene broja kopija kako bismo dodatno mogli povezati disrupcije samih gena s razvojem NDD-ova.

Najveći potencijal za istraživanja i tretman NDD-CNV-ova imaju tehnologije i sustavi: identificiranje mehanizama i gena pokretača, stanični inženjering i inducirane pluripotentne matične stanice, određivanje kritičnih razvojnih perioda i staničnih tipovi, korištenje životinjskih modela te translacijski i intervencijski potencijal (Buttermore i sur. 2022).

Dodatno, kako bi se otkrila povezanost promjena broja kopija u određenim genima i ljudskih genetičkih bolesti potrebni su novi sustavi modela pomoću kojih bi se otkrili različiti mehanizmi pomoću kojih promjene broja kopija u određenim genima utječu na neurorazvoj. U zadnje vrijeme nastaju modeli koji koriste ljudske pluripotentne matične stanice (hPSC, od engl. *human pluripotent stem cell*). Ovi pristupi koriste ljudske embrionalne matične stanice (hESC, od engl. *human embryonic stem cell*) i ljudske inducirane pluripotentne matične stanice (hiPSC, od engl. *human induced pluripotent stem cell*). Diferencirane hiPSC omogućuju razvoj ljudskih *in vitro* modelnih sustava koje mogu imati patologije uzrokovane genetičkom pozadinom izvorišnog materijala. Tako je omogućeno istraživanje genetičkih faktora u ljudskih organskim sustavima koji utječu na neurorazvoj (Lee i sur. 2015). Kako bi podaci dobiveni na ovaj način bili relevantni, potrebno je integrirati fenotipska, genomska, transkriptomaska i proteomska istraživanja s eksperimentalnima da bi se mogli identificirati fenotipi, molekularni putevi i potencijalni terapeutici koji bi bili efektivni za NDD-ove (Buttermore i sur. 2022).

#### 5.4 INTERAKCIJE PROTEINA

Vizualizacija interakcija proteina napravljena pomoću baze SIGNOR služila je za provjeru jesu li geni kandidati povezani i dijele li molekularne funkcije i biološke procese u koje su uključeni.

Ključni, centralni protein analize interakcije proteina kandidata je STAT3, protein s najviše interakcija. *STAT3*, signalni prijenosnik i aktivator transkripcije 3 igra važnu ulogu u obnavljanju neuralnih progenitora i diferencijaciji. Protein STAT3 odgovara na citokine i faktore rasta, fosforilira se na Tyr 105 i translocira u jezgru te aktivira transkripciju ciljanih gena. Osim toga, STAT3 promovira ekspresiju gena neovisnu o fosforilaciji (Darnell i sur. 1994, Levy i Darnell 2002). U embrionalnim neuralnim progenitorima, STAT3 neophodan je za astrocitogenezu te tijekom faze neurogeneze promovira proliferaciju stanica i negativno regulira neurogenezu (Hong i Song 2015). Dodatno je uključen u mitohondrijski metabolizam u središnjem živčanom sustavu što može igrati ulogu u preživljenju stanica i rastu aksona (Luo i sur. 2016, Su i sur. 2020). Kao takav, ističe se kao iznimno važan gen čija disrupcija promjenom broja kopija bi mogla imati razne posljedice pa tako dovesti i do razvoja neurorazvojnog poremećaja.

Unatoč pronađenom broju interakcija, nisu pronađene direktne interakcije među produktima gena kandidata. Jedine poveznice u interakciji pronađene su među *GTF2I*, *ARHGAP35* i *STAT3* koje sve povišeno regulira *SRC*. Nereceptorska tirozin kinaza Src blisko je povezana s tumorima te se spominje kao centralni medijator više signalnih puteva koji uključuju i neuropeptid (Yang i sur. 2020). Putevi povezani s citokinima, toll-like receptorima (TLR) i faktorom rasta fibroblasta (FGFR od engl. *fibroblast growth factor receptor*) česti su u NDD-ovima i tumorima te mogu djelovati preko obitelji kinaza Src i NF- $\kappa$ B (Nussinov i sur. 2023).

## 5.5 NEURORAZVOJNI POREMEĆAJI, IMUNOSNI SUSTAV I TUMORI

Imunosni sustav možemo gledati kao poveznicu između neurorazvojnih poremećaja i tumora. Prilikom razvoja embrija imunski i živčani sustav se zajedno razvijaju i imunski sustav može otpuštati različite citokine koji aktiviraju MAPK signalni put u neuralnim stanicama. U specifičnim embrionalnim staničnim tipovima mozga disregulirana signalizacija koja je rezultat germinativnih ili somatskih mutacija može dovesti do promjena u organizaciji kromatina ili dostupnosti gena i time promijeniti razinu ekspresije gena važnih za neurorazvoj te dovesti do nastanka NDD-a (Nussinov i sur. 2022). Osim signalnog puta MAPK, disfunkcija živčanog sustava može biti uzrokovana i disregulacijom u signalnom putu PI3K/PDK1/AKT/mTOR (Nussinov i sur. 2023).

Disregulirani signalni putevi PI3K/PDK1/AKT/mTOR i MAPK dovode do disregulacije u staničnom ciklusu i ometanja normalne stanične proliferacije i diferencijacije koja može dovesti do razvoja NDD-ova ili tumora (Torii i sur. 2006). Razlika tumora i NDD-ova je to što se NDD-ovi povezuju s ekspresijom promijenjenih proteina koji su kodirani genima koji sadrže germinativne mutacije ili mutacije koje se pojavljuju rano tijekom embrionalnog razvoja ili delecijama gena koji kodiraju ključne proteina. S druge strane, mutacije koje su povezane s tumorima često su sporadične i javljaju se tijekom života (Nussinov i sur. 2023).

Uključenost pronađenih potencijalnih gena u stanični ciklus, imunosni sustav i razvoj čije je ispravno funkcioniranje bitno za neurorazvoj daje dodatne dokaze da promjene broja kopija zabilježene u tim genima koje dovode do disrupcije gena mogu dovesti do razvoja NDD-ova te je potrebno dodatno istražiti utjecaj disrupcije tih gena na razvoj NDD-ova.

## 5.6 DUGE NEKODIRAJUĆE MOLEKULE RNA

Razvoj ljudskog mozga je visoko reguliran proces za koji je bitno u kojem trenutku i gdje dolazi do ekspresije gena. Dinamička, prostorno-vremenska ekspresija gena orkestrira sve faze neurorazvoja. Greške u regulaciji ovih procesa mogu dovesti do NDD-ova. Jedan od važnih regulatornih elemenata ekspresije gena su i Duge nekodirajuće RNA (lncRNA, od engl. *long non-coding* RNA). Duge nekodirajuće RNA predstavljaju najveću i najraznolikiju skupinu nekodirajućih RNA te ih definiramo kao RNA dužine od 200 nt koje se ne transliraju u funkcionalne proteina (Fang i sur. 2017, Uszczynska-Ratajczak i sur. 2018). LncRNA reguliraju ekspresiju gena pomoću više mehanizama i to *cis*, na mjestu transkripcije i *trans* (D'haene i Vergult 2020). Profiliranje ekspresije od strane *GENCODE Consortium* pokazuje da je većina lncRNA tkivno-specifična te da se najveća skupina eksprimira u mozgu (Derrien i sur. 2012). Kod pacijenata s ASD-om utvrđena je diferencijalna ekspresija desetaka lncRNA u *post mortem* uzorcima prefrontalnog korteksa i malog mozga za razliku od kontrole (Ziats i Rennert, 2012). Primijećen je veći broj disreguliranih ncRNA u tkivima pacijenata s ASD-om u putevima povezanim s imunosnim sustavom, biogenezi proteina za transport u sinaptičkim vezikulama i konfiguraciji kromatina što uz ostale dokaze predstavlja lncRNA kao pretpostavljene sudionike u razvoju ASD-a (Ghafouri-Fard i sur. 2021).



Za 11 od 29 potencijalnih gena kandidata utvrđeni su zapisi o lncRNA, i s tim da je jedan od zapisa *YWHAH-ASI* čija je ekspresija zabilježena u mozgu. Kako položaj tih zapisa na genomu odgovara položaju gena kandidata, u tim zapisima zabilježena je promjena broja kopije DNA. Kako je već otkriveno da su lncRNA uključene u različite biološke procese tijekom razvoja, promjene broja kopija mogu dovesti do njihove disregulacije i razvoja NDD-ova.

Za lncRNA poznate su različite funkcije; epigenetičke modifikacije, transkripcijske modifikacije i post-transkripcijske modifikacije. Čak je regulacija pomoću lncRNA potvrđena uzorcima tkiva i stanica ljudskog i mišjeg podrijetla, no točna funkcija i mehanizam djelovanja potvrđen je ustvari za mali broj lncRNA (Shi i sur. 2017). Zbog toga, manjka informacija te istraživanja koja povezuju disregulaciju lncRNA s razvojem NDD-ova možemo reći da je ovo vrlo bitno polje za buduća istraživanja, pogotovo u vidu CNV-ova koji se većinski rade kao uzrok disrupcije samih gena.

## 6. ZAKLJUČAK

Rezultati koji su inicijalno bili klasificirani kao varijante nesigurne ili sporne važnosti reevaluirani su s obzirom na nove smjernice o interpretaciji genomskih aberacija i novih znanstvenih podataka o funkcijama gena. Od 713 genomskih lokusa dobivena je sužena lista od 29 potencijalnih gena kandidata koji su od kliničke važnosti za razvoj neurorazvojnih poremećaja.

Pritom je optimiziran hodogram reevaluacije podataka CMA za efikasno dobivanje optimalnih rezultata. Utvrđena je prenatalna i postnatalna ekspresija gena u mozgu, procesi u koje su geni uključeni te interakcije proteina.

Od 29 gena kandidata dodatno se ističu geni *STAT3*, *GTF2I* i *ARHGAP35* koji su povezani interakcijom s kinazom Src. Dodatno treba izdvojiti i gene *TRIM71*, *SETD1B*, *ARHGAP35* i *PRR12* za koje nema podataka o njihovoj prenatalnoj ekspresiji dok se u recentnijoj literaturi i bazama povezuju s neurorazvojem. Geni *MSH1*, *AP2A1* i *LIMK1* uključeni su u usmjeravanje rasta aksona. Bitno je i izdvojiti gene *CNP*, *DHX8* i *AP2A1* za koje nema recentnije literature koja ih povezuje s neurorazvojem.

Osim samih gena, u reevaluaciji je potrebno se osvrnuti i na lncRNA koje mogu igrati važnu ulogu u neurorazvoju te čija disrupcija može dovesti do razvoja NDD-ova, od zabilježenih lncRNA ističe se *YWHAH-AS1* s povišenom razinom ekspresije u mozgu. Ovi rezultat dodatni su dokazi za važnost reevaluacije rezultata u skladu s napretkom tehnologije i povećanju količine informacija.

## 7. LITERATURA

- Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S. and Packer, A. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Molecular Autism*, 4(1), 36.
- Barak, B., Zhang, Z., Liu, Y., Nir, A., Trangle, S.S., Ennis, M., Levandowski, K.M., Wang, D., Quast, K., Boulting, G.L., Li, Y., Bayarsaihan, D., He, Z. and Feng, G. (2019). Neuronal deletion of Gtf2i, associated with Williams syndrome, causes behavioral and myelin alterations rescuable by a remyelinating drug. *Nature Neuroscience*, 22(5), 700–708.
- Basu, S.N., Kollu, R. and Banerjee-Basu, S. (2008). AutDB: a gene reference resource for autism research. *Nucleic Acids Research*, 37(suppl\_1), D832–D836
- Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Da Silva, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Castro, L.G. and Garmiri, P. (2020). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1), D480–D489.
- Blake, J.A., Baldarelli, R., Kadin, J.A., Richardson, J.E., Smith, C. and Bult, C.J. (2020). Mouse Genome Database (MGD): Knowledgebase for mouse–human comparative biology. *Nucleic Acids Research*, 49(D1), D981–D987.
- Budday, S., Steinmann, P. and Kuhl, E. (2015). Physical biology of human brain development. *Frontiers in Cellular Neuroscience*, 9.
- Buttermore, E., Chamberlain, S., Cody, J., Costain, G., Dang, L., DeWoody, A., DeWoody, Y., Dies, K., Eichler, E., Girirajan, S., Gramm, M., Halladay, A., Lal, D., Lalli, M., Levy, T., Logsdon, G., Lowenstein, D., Mefford, H., Mulle, J. and Muotri, A. (2022). Neurodevelopmental copy-number variants: A roadmap to improving outcomes by uniting patient advocates, researchers, and clinicians for collective impact. *The American Journal of Human Genetics*, 109(8), 1353–1365.

Cahan, P., Li, Y., Izumi, M. and Graubert, T.A. (2009). The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells. *Nature Genetics*, 41(4), 430–437.

Chawner, S.J.R.A., Owen, M.J., Holmans, P., Raymond, F.L., Skuse, D., Hall, J. and van den Bree, M.B.M. (2019). Genotype–phenotype associations in children with copy number variants associated with high neuropsychiatric risk in the UK (IMAGINE-ID): a case-control cohort study. *The Lancet Psychiatry*, 6(6), 493–505.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G., Clark, N.R. and Ma’ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, 14(1), 128.

D’haene, E. and Vergult, S. (2020). Interpreting the impact of noncoding structural variation in neurodevelopmental disorders. *Genetics in Medicine*, 23(1), 34–46.

Darnell, J., Kerr, I. and Stark, G. (1994). Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science*, 264(5164), 1415–1421.

de la Torre-Ubieta, L., Won, H., Stein, J.L. and Geschwind, D.H. (2016). Advancing the understanding of autism disease mechanisms through genetics. *Nature Medicine*, 22(4), 345–361.

Deignan, J.L., Chung, W.K., Kearney, H.M., Monaghan, K.G., Rehder, C.W. and Chao, E.C. (2019). Points to consider in the reevaluation and reanalysis of genomic test results: a statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, 21(6), 1267–1270.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J.B., Lipovich, L., Gonzalez, J.M. and Thomas, M. (2012). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*, 22(9), 1775–1789.

Dinneen, T.J., Ghráiligh, F.N., Walsh, R., Lopez, L.M. and Gallagher, L. (2022). How does genetic variation modify ND-CNV phenotypes? *Trends in Genetics*, 38(2), 140–151.

Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., Zhao, L., Li, X., Teng, X., Sun, X., Sun, L., Zhang, M.Q., Chen, R. and Zhao, Y. (2017). NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Research*, 46(D1), D308–D314.

Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Vooren, S.V., Moreau, Y., Pettett, R.M. and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *The American Journal of Human Genetics*, 84(4), 524–533.

Foreman, J., Brent, S., Perrett, D., Bevan, A., Hunt, S.E., Cunningham, F., Hurles, M.E. and Firth, H.V. (2022). DECIPHER: Supporting the interpretation and sharing of rare disease phenotype-linked variant data to advance diagnosis and research. *Human Mutation*, 43(6).

Foreman, J., Perrett, D., Mazaika, E., Hunt, S.E., Ware, J.S. and Firth, H.V. (2023). DECIPHER: Improving Genetic Diagnosis Through Dynamic Integration of Genomic and Clinical Data. *The Annual Review of Genomics and Human Genetics*, 24(1).

genome.ucsc.edu. (n.d.). *UCSC Genome Browser Home*. <https://genome.ucsc.edu/index.html>. (pristupljeno 22.11.2022.)

Ghafouri-Fard, S., Noroozi, R., Brand, S., Hussen, B.M., Eghtedarian, R., Taheri, M. and Ebrahimzadeh, K. (2021). Emerging Role of Non-coding RNAs in Autism Spectrum Disorder. *Journal of Molecular Neuroscience*, 72(2), 201–216.

Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W.M., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., Leach, R., Klein, R., Tearle, R., Bo, T., Pfundt, R., Yntema, H.G., de Vries, B.B.A., Kleefstra, T., Brunner, H.G. and Vissers, L.E.L.M. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature*, 511(7509), 344–347.

Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., Deng, C., Varusai, T., Ragueneau, E., Haider, Y., May, B., Shamovsky, V., Weiser, J., Brunson, T., Sanati, N. and Beckman, L. (2021). The reactome pathway knowledgebase 2022. *Nucleic Acids Research*, 50(D1), D687–D692.

Gudmundsson, O.O., Walters, G.B., Ingason, A., Johansson, S., Zayats, T., Athanasiu, L., Sonderby, I.E., Gustafsson, O., Nawaz, M.S., Jonsson, G.F., Jonsson, L., Knappskog, P.-M., Ingvarsdottir, E., Davidsdottir, K., Djurovic, S., Knudsen, G.P.S., Askeland, R.B., Haraldsdottir, G.S., Baldursson, G. and Magnusson, P. (2019). Attention-deficit hyperactivity disorder shares copy number variant risk with schizophrenia and autism spectrum disorder. *Translational Psychiatry*, 9(1), 1–9.

Guo, Y., Dai, Y., Yu, H., Zhao, S., Samuels, D.C. and Shyr, Y. (2017). Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. *Genomics*, 109(2), 83–90.

health.ec.europa.eu. (n.d.). *Rare diseases*. [https://health.ec.europa.eu/non-communicable-diseases/expert-group-public-health/rare-diseases\\_en](https://health.ec.europa.eu/non-communicable-diseases/expert-group-public-health/rare-diseases_en) (pristupljeno 01.03.2023.).

Heckel, T., Singh, A., Gschwind, A., Reymond, A. and Certa, U. (2015). *Chapter 4 - Genetic Variations in the Macaca fascicularis Genome Related to Biomedical Research*. ScienceDirect. (Pristupljeno 08. svibnja 2023).

Hehir-Kwa, J., Pfundt, R., Veltman, J. and de Leeuw, N. (2013). Pathogenic or not? Assessing the clinical relevance of copy number variants. *Clinical Genetics*, 84(5), 415–421.

Henderson, L.B., Applegate, C.D., Wohler, E., Sheridan, M.B., Hoover-Fong, J. and Batista, D.A.S. (2014). The impact of chromosomal microarray on clinical management: a retrospective analysis. *Genetics in Medicine*, 16(9), 657–664.

Henrichsen, C.N., Chaignat, E. and Reymond, A. (2009). Copy number variants, diseases and gene expression. *Human Molecular Genetics*, 18(R1), R1–R8.

Highsmith Jr., W.E. ed., (2013). *Molecular Diagnostics: 12 Tests That Changed Everything*. 1st ed. Humana New York, NY, 257.

Hong, S. and Song, M.-R. . (2015). Signal Transducer and Activator of Transcription-3 Maintains the Stemness of Radial Glia at Mid-Neurogenesis. *Journal of Neuroscience*, 35(3), 1011–1023.

Ismail, F.Y. and Shapiro, B.K. (2019). What are neurodevelopmental disorders? *Current Opinion in Neurology*, 32(4), 611–616.

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A. and Walters, R.K. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809), 434–443.

Karolchik, D., Hinrichs, A.S. and Kent, W.J. (2012). The UCSC Genome Browser. *Current Protocols in Bioinformatics*, 40(1).

Kingdom, R. and Wright, C.F. (2022). Incomplete Penetrance and Variable Expressivity: From Clinical Studies to Population Cohorts. *Frontiers in Genetics*, 13.

Kolb, B., Whishaw, I.Q. and G Campbell Teskey (2019). *An introduction to brain and behavior*. 6th ed. Vancouver, B.C.: Worth Publishers, 688.

Koopmans, F., van Nierop, P., Andres-Alonso, M., Byrnes, A., Cijssouw, T., Coba, M.P., Cornelisse, L.N., Farrell, R.J., Goldschmidt, H.L., Howrigan, D.P., Hussain, N.K., Imig, C., de Jong, A.P.H., Jung, H., Kohansalnodehi, M., Kramarz, B., Lipstein, N., Lovering, R.C., MacGillavry, H. and Mariano, V. (2019). SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron*, 103(2), 217-234.

Kostović, I. (2020). The enigmatic fetal subplate compartment forms an early tangential cortical nexus and provides the framework for construction of cortical connectivity. *Progress in Neurobiology*, 194, 101883.

Ku, C.S., Loy, E.Y., Salim, A., Pawitan, Y. and Chia, K.S. (2010). The discovery of human genetic variations and their use as disease markers: past, present and future. *Journal of Human Genetics*, 55(7), 403–415.

Lee, C.-T., Freed, W.J. and Mash, D.C. (2015). CNVs in neurodevelopmental disorders. *Oncotarget*, 6(21), 18238–18239.

Levy, D.E. and Darnell, J.E. (2002). STATs: transcriptional control and biological impact. *Nature Reviews Molecular Cell Biology*, 3(9), 651–662.

Lo Surdo, P., Iannuccelli, M., Contino, S., Castagnoli, L., Licata, L., Cesareni, G. and Perfetto, L. (2022). SIGNOR 3.0, the SIGnaling network open resource 3.0: 2022 update. *Nucleic Acids Research*.

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J. and Fleming, J. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, 45(6), 580–585.

Luo, X., Ribeiro, M., Bray, E.R., Lee, D.-H., Yungher, B.J., Mehta, S.T., Thakor, K.A., Diaz, F., Lee, J.K., Moraes, C.T., Bixby, J.L., Lemmon, V.P. and Park, K.K. (2016). Enhanced Transcriptional Activity and Mitochondrial Localization of STAT3 Co-induce Axon Regrowth in the Adult Central Nervous System. *Cell Reports*, 15(2), 398–410.

MacDonald, J.R., Ziman, R., Yuen, R.K.C., Feuk, L. and Scherer, S.W. (2013). The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Research*, 42(D1), D986–D992.

Mi, H. and Thomas, P. (2009). PANTHER Pathway: an ontology-based pathway database coupled with data analysis tools. *Methods in molecular biology (Clifton, N.J.)*, 563, 123.

Michetti, C., Falace, A., Benfenati, F. and Fassio, A. (2022). Synaptic genes and neurodevelopmental disorders: From molecular mechanisms to developmental strategies of behavioral testing. *Neurobiology of Disease*, 173, 105856.

Miga, K.H., Newton, Y., Jain, M., Altemose, N., Willard, H.F. and Kent, W.J. (2014). Centromere reference models for human chromosomes X and Y satellite arrays. *Genome Research*, 24(4), 697–707.

Morales, J., Pujar, S., Loveland, J.E., Astashyn, A., Bennett, R., Berry, A., Cox, E., Davidson, C., Ermolaeva, O., Farrell, C.M., Fatima, R., Gil, L., Goldfarb, T., Gonzalez, J.M., Haddad, D., Hardy,



M., Hunt, T., Jackson, J., Joardar, V.S. and Kay, M. (2022). A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature*, 604(7905), 310–315.

Mossink, B., Negwer, M., Schubert, D. and Nadif Kasri, N. (2020). The emerging role of chromatin remodelers in neurodevelopmental disorders: a developmental perspective. *Cellular and Molecular Life Sciences*, 78(6), 2517–2563.

Nassar, L.R., Barber, G.P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J.N., Hinrichs, A., Lee, B., Lee, C., Muthuraman, P., Nguy, B., Pereira, T., Nejad, P., Perez, G., Raney, B., Schmelter, D., Speir, M. and Wick, B. (2022). The UCSC Genome Browser database: 2023 update. *Nucleic Acids Research*, 51(D1), D1188–D1195.

National Library of Medicine (2021). *PubMed Labs*. PubMed Labs. <https://pubmed.ncbi.nlm.nih.gov/>. (pristupljeno 13.04.2023.)

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, 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. and Caldas, G.V. (2022). The complete sequence of a human genome. *Science*, 376(6588), 44–53.

Nussbaum, R.L., Mcinnes, R.R., Williard, H.F. and Ada Hamosh (2007). *Thompson & Thompson Genetics in medicine*. Saunders, 585.

Nussinov, R., Tsai, C.-J. and Jang, H. (2022). Neurodevelopmental disorders, immunity, and cancer are connected. *iScience*, 25(6), 104492.

Nussinov, R., Yavuz, B.R., Arici, M.K., Demirel, H.C., Zhang, M., Liu, Y., Tsai, C.-J., Jang, H. and Tuncbag, N. (2023). Neurodevelopmental disorders, like cancer, are connected to impaired chromatin remodelers, PI3K/mTOR, and PAK1-regulated MAPK. *Biophysical Reviews*, 15(2), 163–181.

Parenti, I., Rabaneda, L.G., Schoen, H. and Novarino, G. (2020). Neurodevelopmental Disorders: From Genetics to Functional Pathways. *Trends in Neurosciences*, 43(8), 608–621.

Pereanu, W., Larsen, E.C., Das, I., Estévez, M.A., Sarkar, A.A., Spring-Pearson, S., Kollu, R., Basu, S.N. and Banerjee-Basu, S. (2017). AutDB: a platform to decode the genetic architecture of autism. *Nucleic Acids Research*, 46(D1), D1049–D1054.

Pinkel, D., Se Graves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Dairkee, S.H., Ljung, B.M., Gray, J.W. and Albertson, D.G. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genetics*, 20(2), 207–211.

Pös, O., Radvanszky, J., Buglyó, G., Pös, Z., Rusnakova, D., Nagy, B. and Szemes, T. (2021). DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. *Biomedical Journal*, 44(5), 548–559.

Prashad, S. and Gopal, P.P. (2020). RNA-binding proteins in neurological development and disease. *RNA Biology*, 1–16.

Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L., Plon, S.E., Ramos, E.M., Sherry, S.T. and Watson, M.S. (2015). ClinGen — The Clinical Genome Resource. *New England Journal of Medicine*, 372(23), 2235–2242.

Riggs, E.R., Andersen, E.F., Cherry, A.M., Kantarci, S., Kearney, H., Patel, A., Raca, G., Ritter, D.I., South, S.T., Thorland, E.C., Pineda-Alvarez, D., Aradhya, S. and Martin, C.L. (2019). Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genetics in Medicine*, (22).

Rinčić, M. (2014). *Molekularna citogenetika nekih neurorazvojnih poremećaja*. Doktorski rad.

Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnström, K., Mallick, S., Kirby, A., Wall, D.P., MacArthur, D.G., Gabriel, S.B., DePristo, M., Purcell, S.M., Palotie, A., Boerwinkle, E., Buxbaum, J.D., Cook, E.H. and Gibbs, R.A. (2014). A framework for the interpretation of de novo mutation in human disease. *Nature Genetics*, 46(9), 944–950.

Savatt, J.M. and Myers, S.M. (2021). Genetic Testing in Neurodevelopmental Disorders. *Frontiers in Pediatrics*, 9.

Schouten, J.P. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Research*, 30(12).

Shi, C., Zhang, L. and Qin, C. (2017). Long non-coding RNAs in brain development, synaptic biology, and Alzheimer's disease. *Brain Research Bulletin*, 132, 160–169.

Singh, A.K., Olsen, M.F., Lavik, L.A.S., Vold, T., Drabløs, F. and Sjørusen, W. (2021). Detecting copy number variation in next generation sequencing data from diagnostic gene panels. *BMC Medical Genomics*, 14(1).

Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Döhner, H., Cremer, T. and Lichter, P. (1997). Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes, Chromosomes & Cancer*, 20(4), 399–407.

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M. and Lancet, D. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current Protocols in Bioinformatics*, 54(1), 1.30.1–1.30.33.

Stuppia, L., Antonucci, I., Palka, G. and Gatta, V. (2012). Use of the MLPA Assay in the Molecular Diagnosis of Gene Copy Number Alterations in Human Genetic Diseases. *International Journal of Molecular Sciences*, 13(3), 3245–3276.

Su, Y., Zhang, W., Patro, K., Zhao, J., Mu, T., Ma, Z., Xu, J., Hon, K., Yi, C. and Zhou, Y. (2020). STAT3 Regulates Mouse Neural Progenitor Proliferation and Differentiation by Promoting Mitochondrial Metabolism. *Front. Cell Dev. Biol.*, 8.

Sweeney, B.A., Petrov, A.I., Burkov, B., Finn, R.D., Bateman, A., Szymanski, M., Karłowski, W.M., Gorodkin, J., Seemann, S.E., Cannone, J.J., Gutell, R.R., Fey, P., Basu, S., Kay, S., Cochrane, G., Billis, K., Emmert, D., Marygold, S.J., Huntley, R.P. and Lovering, R.C. (2018).

RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Research*, 47(D1), D221–D229.

Talseth-Palmer, B.A. and Scott, R.J. (2011). Genetic Variation and its Role in Malignancy. *International Journal of Biomedical Science : IJBS*, 7(3), 158–171.

Thomas, P.D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L. and Mi, H. (2021). PANTHER : Making genome-scale phylogenetics accessible to all. *Protein Science*, 31(1), 8–22.

Torii, S., Yamamoto, T., Tsuchiya, Y. and Nishida, E. (2006). ERK MAP kinase in G1 cell cycle progression and cancer. *Cancer Science*, 97(8), 697–702.

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R. and Johnson, R. (2018). Towards a complete map of the human long non-coding RNA transcriptome. *Nature Reviews Genetics*, 19(9), 535–548.

Wetzel, A.S. and Darbro, B.W. (2022). A comprehensive list of human microdeletion and microduplication syndromes. *BMC Genomic Data*, 23(1).

www.ncbi.nlm.nih.gov. (n.d.). *Entrez Molecular Sequence Database System*. <https://www.ncbi.nlm.nih.gov/Web/Search/entrezfs.html#:~:text=Entrez%20is%20a%20molecul ar%20biology> (pristupljeno 23. svibnja 2023).

Yang, H., Wang, L., Zang, C., Wang, Y., Shang, J., Zhang, Z., Liu, H., Bao, X., Wang, X. and Zhang, D. (2020). Src Inhibition Attenuates Neuroinflammation and Protects Dopaminergic Neurons in Parkinson's Disease Models. *Frontiers in Neuroscience*, 14.

Ziats, M.N. and Rennert, O.M. (2012). Aberrant Expression of Long Noncoding RNAs in Autistic Brain. *Journal of Molecular Neuroscience*, 49(3), 589–593.

## ŽIVOTOPIS

Rođena sam u Đakovu. Nakon osnovne škole upisala sam Gimnaziju Antuna Gustava Matoša u Đakovu. Nakon završenog srednjoškolskog obrazovanja s odličnim uspjehom, 2017. upisala sam preddiplomski studij Molekularna biologija na Prirodoslovno-matematičkom fakultetu Sveučilišta u Zagrebu. Tijekom preddiplomskog studija bila sam demonstrator na kolegiju Zoologija, sudjelovala u manifestaciji „Noć Biologije“, volontirala na Simpoziju studenata bioloških usmjerenja i bila pasivni sudionik na raznim simpozijima. Nakon završenog preddiplomskog studija s velikom pohvalom 2020. nastavila sam školovanje na Prirodoslovno-matematičkom fakultetu Sveučilišta u Zagrebu upisujući diplomski studij Molekularna biologija. Tijekom studija volontirala sam na projektu „Dopaminska regulacija koptitivnog ponašanja guštera *Podarcis sicula* i *Podarcis melisellensis*“ gdje sam odradila Laboratorijsku stručnu praksu i stekla znanje rada u laboratoriju. Redoviti sam član Udruge studenata biologije – BIUS na čijim projektima poput Velikog terena i Simpozija studenata bioloških usmjerenja sam sudjelovala te sam u godini 2022. bila potpredsjednica Udruge i sudjelovala u njihovoj organizaciji. Dodatno, predstavljala sam rad fakulteta na Smotri Sveučilišta i sudjelovala sam u organizaciji WISE-a, Dana karijera PMF-a. Diplomski rad izradila sam u Laboratoriju za medicinsku genetiku, Hrvatskog instituta za istraživanje mozga Medicinskog fakulteta Sveučilišta u Zagrebu pod vodstvom dr.sc. Martine Rinčić.

## PRILOZI

Prilog 1 Proširena lista gena. Plus označava prisutnost gena u bazi

Naziv gena	Accession number	Baza HPA1	Baza HPA2	SynGO	autDB	SFARI	Reactome1	Reactome2	HBT	GTEx	RNAcentral	UniProt	MGI
CACNG8	NM_031895.6	+	-	+	-	-	-	-	+	+		+	+
HCRT	NM_001524.1	+	-	-	-	-	-	-	+	+		+	+
PTH2	NM_178449.4	+	-	-	-	-	-	-	+	+		+	
SCART1	NM_00139605 0.1	+	-	-	-	-	-	-	-			+	-
AANAT	NM_00116657 9.2	-	+	-	-	-	-	-	+			+	+
ADAM20	NM_003814.5	-	+	-	-	-	-	-	+		+	+	+
ATP2B3	NM_00138836 2.1	-	+	+	-	-	-	-	+			+	+
B3GNT4	NM_00133049 2.2	-	+	-	-	-	-	-	+	+		+	
CATSPERG	NM_00133049 6.2	-	+	-	-	-	-	-	+	+	+	+	
CCDC62	NM_201435.5	-	+	-	-	-	-	-	+			+	
CD160	NM_007053.4	-	+	-	-	-	-	-	+			+	
CDH12	NM_00136410 5.2	-	+	-	-	-	-	-	+	+	+	+	
CFAP97D1	NM_00135340 0.2	-	+	-	-	-	-	-	-			+	+

CLEC18C	NM_173619.4	-	+	-	-	-	-	-	-	-	+	+	
CNTD1	NM_00133022 2.2	-	+	-	-	-	-	-	+			+	
CNTN5	NM_00124327 1.2	-	+	+	+	+	-	-	+	+		+	
CNTNAP2	NM_014141.6	-	+	+	+	+	-	-	+	+	+	+	+
CRHR2	NM_00120248 3.2	-	+		+	-	-	-	+	+		+	+
CRYBB2	NM_000496.3	-	+	-	-	-	-	-	+			+	+
DHDH	NM_014475.4	-	+	-	-	-	-	-	+			+	
DPF1	NM_00128997 8.2	-	+	-	-	-	-	-	+	+		+	+
DUSP9	NM_001395.4	-	+	-	-	-	-	-	+			+	+
FOXJ1	NM_001454.4	-	+	-	-	-	-	-	+			+	+
FZD9	NM_003508.3	-	+	+	-	-	-	-	+	+		+	+
GADL1	NM_207359.3	-	+	-	-	-	-	-	+			+	
GAST	NM_000805.5	-	+	-	-	-	-	-	+	+		+	
GGN	NM_152657.4	-	+	-	-	-	-	-	+		+	+	
GPR3	NM_005281.4	-	+	-	-	-	-	-	+	+		+	+
GRIN2D	NM_000836.4	-	+	+	-	-	-	-	+	+		+	+

H2BW2	NM_00138846 4.1	-	+	-	-	-	-	-	-	-	-	+	-
HEATR4	NM_00122048 4.1	-	+	-	-	-	-	-	+	+	+	+	-
HSPB9	NM_033194.3	-	+	-	-	-	-	-	+	-	-	+	+
IFNE	NM_176891.5	-	+	-	-	-	-	-	+	-	-	+	-
KASH5	NM_144688.5	-	+	-	-	-	-	-	-	-	-	+	+
KCNH4	NM_012285.3	-	+	-	-	-	-	-	+	+	-	+	-
KIF18B	NM_00126457 3.2	-	+	-	-	-	-	-	-	-	+	+	-
LIN28A	NM_024674.6	-	+	-	-	-	-	-	-	-	+	+	+
LRGUK	NM_00140151 9.1	-	+	-	-	-	-	-	+	-	-	+	-
LYG2	NM_175735.4	-	+	-	-	-	-	-	+	+	-	+	-
MALRD1	NM_00114230 8.3	-	+	-	-	-	-	-	-	-	-	+	-
MEIOC	NM_00114508 0.3	-	+	-	-	-	-	-	-	+	-	+	-
MGAT5B	NM_144677.3	-	+	-	-	-	-	-	+	+	-	+	-
MORN3	NM_173855.5	-	+	-	-	-	-	-	+	+	-	+	-
NPAS1	NM_002517.4	-	+	-	-	-	-	-	+	+	-	+	-
NPBWR1	NM_005285.5	-	+	-	-	-	-	-	+	-	-	+	+



OPRD1	NM_000911.4	-	+	+	-	-	-	-	+	+		+	+
OVOL3	NM_00130275 7.2	-	+	-	-	-	-	-	-	+		+	-
PLGLB2	NM_002665.4	-	+	-	-	-	-	-	-			+	-
PRCD	NR_033357.2	-	+	-	-	-	-	-	-	+	+	+	+
PTPRT	NM_133170.4	-	+	+	+	+	-	-	+	+	+	+	
RIBC2	NM_015653.5	-	+	-	-	-	-	-	+			+	+
RIMBP3	NM_015672.2	-	+	+	-	-	-	-	-			+	-
SGCZ	NM_139167.4	-	+	-	-	-	-	-	+	+		+	
SLC4A1	NM_000342.4	-	+	-	-	-	-	-	+		+	+	
SPRED3	NM_00139433 8.1	-	+	-	-	-	-	-	-	+		+	+
STRC	NM_153700.2	-	+	-	-	-	-	-	+	+		+	+
SYNGR4	NM_012451.4	-	+	+	-	-	-	-	+	+		+	
SYT3	NM_032298.3	-	+	+	+	-	-	-	+	+	+	+	
TKTL1	NM_00114593 3.2	-	+	-	-	-	-	-	+			+	
TMEM156	NM_024943.3	-	+	-	-	-	-	-	-			+	-
TMEM191B	NM_00124231 3.1	-	+	-	-	-	-	-	+	+		+	

TRIM71	NM_00103911 1.3	-	+	-	-	-	-	-	-	+		+	+
TULP2	NM_003323.3	-	+	-	-	-	-	-	+			+	
VSTM2A	NM_00131784 3.2	-	+	-	-	-	-	-	+	+	+	+	
WNT7A	NM_004625.4	-	+	+	-	-	-	-	+	+		+	+
ZNF114	NM_00136981 1.1	-	+	-	-	-	-	-	+		+	+	
ZNF683	NM_00111475 9.3	-	+	-	-	-	-	-	+			+	
ACTN4	NM_00132203 3.2	-	-	-	+	+	-	-	+		+	+	+
ADAM11	NM_00131893 3.2	-	-	-	-	-	-	-	+	+		+	+
AGBL4	NM_00132357 4.2	-	-	-	+	+	-	-	+	+	+	+	+
AHDC1	NM_00102988 2.3	-	-	-	+	+	-	-	+			+	
AP2A1	NM_130787.3	-	-	-	-	-	+	+	+			+	+
AP2S1	NM_00130107 6.3	-	-	+	+	+	-	-	+			+	+
APOD	NM_001647.4	-	-	-	-	-	-	-	+			+	+
AQP1	NM_00132987 2.2	-	-	-	-	-	-	-	+			+	+
ARHGAP35	NM_004491.5	-	-	-	-	-	+	-	-			+	+
ATP6AP1	NM_001183.6	-	-	+	-	-	-	-	+	+	+	+	+
ATP6V0A1	NM_00137854 9.1	-	-	+	-	-	-	-	+	+		+	+

BAX	NM_00129143 1.2	-	-	-	-	-	-	-	+			+	+
BICDL1	NM_00136788 6.1	-	-	-	+	+	-	-	-	+		+	+
BICRA	NM_015711.3	-	-	-	+	+	-	-	-		+	+	-
C1QL1	NM_006688.5	-	-	+	-	-	-	-	+	+		+	+
CABP1	NM_00103367 7.2	-	-	+	-	-	-	-	+	+	+	+	+
CADM2	NM_00138196 4.1	-	-	-	+	+	-	-	+	+	+	+	+
CAMKK2	NM_00127048 6.1	-	-	-	-	-	-	+	+	+		+	+
CCT4	NM_006430.4	-	-	-	+	+	-	-	+		+	+	
CHRNA9	NM_017581.4	-	-	+	-	-	-	-	+			+	+
CLASP2	NM_00140041 3.1	-	-	-	-	-	+	-	+	+		+	
CLIP2	NM_032421.3	-	-	-	+	+	-	-	+	+		+	+
CNKSR2	NM_00133077 1.2	-	-	+	+	+	-	-	+	+	+	+	+
CNP	NM_033133.5	-	-	-	-	-	-	-	+	+		+	+
CNTNAP1	NM_003632.3	-	-	+	-	-	-	-	+	+	+	+	+
CPT1C	NM_00137848 5.1	-	-	+	-	-	-	-	+	+		+	+
CTNNA3	NM_00112738 4.3	-	-	-	+	+	-	-	+	+	+	+	+
CYFIP1	NM_00103302 8.3	-	-	+	+	+	-	-	+			+	+

CYTH2	NM_004228.7	-	-	+	-	-	-	-	+			+	+
DENR	NM_003677.5	-	-	-	+	+	-	-	+			+	+
DEPDC5	NM_00136990 2.1	-	-	-	+	+	-	-	+			+	+
DHX8	NM_00132221 7.2	-	-	-	-	-	-	-	+			+	+
DPP6	NM_00136449 8.2	-	-	-	+	+	-	-	+	+	+	+	+
DST	NM_015548.5	-	-	-	+	+	-	-	+		+	+	+
DUSP22	NM_020185.6	-	-	-	+	-	-	-	+		+	+	
EGFR	NM_201283.2	-	-	-	-	-	+	-	+		+	+	+
EPB41	NM_00137602 8.1	-	-	-	-	-	-	+	+		+	+	
FAM98C	NM_174905.4	-	-	-	+	+	-	-	+	+		+	
FLNA	NM_00111055 6.2	-	-	+	-	-	-	-	+		+	+	+
FRG1	NM_004477.3	-	-	-	+	+	-	-	+		+	+	
FXN	NM_000144.5	-	-	-	+	+	-	-	+			+	+
GDI1	NM_001493.3	-	-	+	-	-	-	-	+	+		+	+
GFAP	NM_002055.5	-	-	-	+	+	-	-	+	+		+	+
GJC1	NM_005497.4	-	-	-	-	-	-	+	+			+	+

GRB2	NM_002086.5	-	-	-	-	-	+	-	+		+	+	+
GRN	NM_002087.4	-	-	+	-	-	-	-	+			+	+
GTF2I	NM_00128080 0.2	-	-	-	+	+	-	-	+		+	+	+
HAP1	NM_00136745 9.1	-	-	+	-	-	-	-	+	+		+	+
HCFC1	NM_005334.3	-	-	-	+	+	-	-	+		+	+	+
HIP1R	NM_00130309 9.2	-	-	+	-	-	-	-	+			+	+
HLA-A	NM_002116.8	-	-	-	+	+	-	-	-		+	+	+
HLA-DRB1	NM_002124.4	-	-	-	+	+	-	-	+			+	
IMMP2L	NM_032549.4	-	-	-	+	+	-	-	+		+	+	+
ITGA2B	NM_000419.5	-	-	-	-	-	+	-	+			+	+
ITGB4	NM_00132112 3.2	-	-	+	-	-	-	-	+			+	+
KANSL1	NM_00119346 6.2	-	-	-	+	+	-	-	-		+	+	-
KAT2A	NM_00137622 7.1	-	-	-	-	-	-	-	+			+	+
KCNA7	NM_031886.3	-	-	-	-	-	-	+	+			+	
KCNC3	NM_00137230 5.1	-	-	+	-	-	-	-	+	+		+	+
KCNK6	NM_004823.3	-	-	-	-	-	-	+	+			+	

KCNJ14	NM_013348.4	-	-	-	-	-	-	-	+	+		+	+	
KDM2B	NM_00100536 6.2	-	-	-	+	+	-	-	-	+		+	+	+
KPTN	NM_007059.4	-	-	+	+	+	-	-	-	+	+	+	+	+
L1CAM	NM_024003.3	-	-	+	-	-	-	-	-	+	+	+	+	+
LIMK1	NM_002314.4	-	-	-	-	-	+	-	-	+	+		+	+
LIMK2	NM_005569.4	-	-	-	-	-	+	-	-	+			+	+
LIN7B	NM_00130841 9.2	-	-	-	+	+	-	-	-	+	+		+	
LMTK3	NM_00138848 5.1	-	-	-	+	+	-	-	-	-	+		+	+
LRRC4B	NM_00134856 8.1	-	-	+	-	-	-	-	-	-	+	+	+	+
MAGI2	NM_012301.4	-	-	+	-	-	-	-	-	+	+	+	+	+
MAP4K1	NM_007181.6	-	-	-	+	+	-	-	-	+		+	+	
MECP2	NM_00136939 4.2	-	-	+	+	+	-	-	-	+			+	+
MED25	NM_030973.4	-	-	-	-	-	-	-	-	+			+	+
MPP2	NM_00127837 6.3	-	-	+	-	-	-	-	-	+	+		+	
MSI1	NM_002442.4	-	-	-	-	-	+	-	-	+	+		+	+
MSR1	NM_00136374 4.1	-	-	-	+	+	-	-	-	+			+	

MUC4	NM_018406.7	-	-	-	+	+	-	-	+			+	
MYH11	NM_00104011 3.2	-	-	-	-	-	+	-	+		+	+	
MYH14	NM_024729.4	-	-	-	-	-	+	-	+		+	+	+
NAGS	NM_153006.3	-	-	-	-	-	-	-	+			+	+
NAPA	NM_003827.4	-	-	+	-	-	-	-	+		+	+	+
NIPA1	NM_00114227 5.1	-	-	-	+	+	-	-	+	+	+	+	
NIPA2	NM_00100889 4.3	-	-	-	+	+	-	-	+			+	
NR1D2	NM_005126.5	-	-	-	-	-	-	-	+			+	+
NR1H2	NM_00125664 7.3	-	-	-	+	-	-	-	+		+	+	+
NSF	NM_006178.4	-	-	+	-	-	-	-	+	+		+	+
NUP155	NM_004298.4	-	-	-	+	+	-	-	+			+	+
ODAD4	NM_00135031 9.2	-	-	-	+	-	-	-	-			+	+
OR2T10	NM_00100469 3.2	-	-	-	+	+	-	-	-			+	-
P2RX4	NM_00126139 7.2	-	-	-	+	-	-	-	+	+		+	+
P2RX7	NM_002562.6	-	-	+	-	-	-	-	+	+	+	+	+
PDZD4	NM_00130351 5.2	-	-	-	+	-	-	-	+	+		+	
PLEKHA5	NM_00114382 1.3	-	-	+	-	-	-	-	+			+	+

PLXNA3	NM_017514.5	-	-	-	+	+	-	-	+			+	+
PLXNB3	NM_005393.3	-	-	-	-	-	+	-	+		+	+	
POMK	NM_00127797 1.2	-	-	-	-	-	-	-	-			+	+
PPFIA3	NM_003660.4	-	-	+	-	-	-	-	+	+		+	+
PRKN	NM_004562.3	-	-	+	+	+	-	-	-			+	+
PRODH	NM_016335.6	-	-	-	+	+	-	-	+			+	+
PROX1	NM_00127061 6.2	-	-	-	-	-	-	-	+			+	+
PRR12	NM_020719.3	-	-	-	+	+	-	-	-	+	+	+	-
PRR14L	NM_173566.3	-	-	-	+	+	-	-	-			+	-
PSMD8	NM_002812.5	-	-	-	-	-	+	-	+			+	
PSMD9	NM_00126140 0.3	-	-	-	-	-	+	-	+			+	
PSME3	NM_005789.4	-	-	-	-	-	+	-	+			+	
RAB35	NM_006861.7	-	-	+	-	-	-	-	+		+	+	+
RAB5C	NM_201434.3	-	-	+	-	-	-	-	+		+	+	+
RB1CC1	NM_014781.5	-	-	-	+	-	-	-	+	+	+	+	
RNF10	NM_014868.5	-	-	+	-	-	-	-	+		+	+	+
ROBO2	NM_00112892 9.3	-	-	-	+	+	-	-	-	+	+	+	+



RPL10	NM_00125658 0.2	-	-	+	+	+	-	-	+		+	+	
RPL13A	NM_00127049 1.2	-	-	+	-	-	-	-	-			+	+
RPL15	NM_00125337 9.2	-	-	+	-	-	-	-	+			+	
RPL18	NM_000979.4	-	-	+	-	-	-	-	+			+	
RPL27	NM_00134992 1.2	-	-	+	-	-	-	-	-			+	-
RPL9	NM_00102492 1.4	-	-	+	-	-	-	-	+			+	
RPLP0	NM_053275.4	-	-	+	-	-	-	-	+			+	
RPS11	NM_001015.5	-	-	+	-	-	-	-	-			+	-
RPS6KA1	NM_002953.4	-	-	+	-	-	-	-	+			+	+
RRAS	NM_006270.5	-	-	-	-	-	+	-	+			+	+
SAE1	NM_00114571 3.2	-	-	-	+	+	-	-	+			+	
SCAF1	NM_021228.3	-	-	-	+	+	-	-	+		+	+	
SCFD2	NM_152540.4	-	-	-	+	+	-	-	+			+	
SETD1B	NM_00135334 5.2	-	-	-	+	+	-	-	-			+	+
SLC17A7	NM_020309.4	-	-	+	-	-	-	-	+	+		+	+
SLC25A39	NM_00114378 0.3	-	-	-	+	+	-	-	+			+	

SLC2A3	NM_006931.3	-	-	+	-	-	-	-	+			+	+
SLC30A5	NM_00125196 9.2	-	-	-	+	-	-	-	+			+	+
SLC6A8	NM_00114280 5.2	-	-	+	+	+	-	-	+			+	+
SLC8A2	NM_015063.3	-	-	+	-	-	-	-	-	+	+	+	+
SLCO1B3	NM_019844.4	-	-	-	+	+	-	-	+			+	
SMN1	NM_022874.2	-	-	-	-	-	-	-	-			+	+
STAT3	NM_00138499 0.1	-	-	+	-	-	-	-	+		+	+	+
STX12	NM_177424.3	-	-	+	-	-	-	-	+		+	+	
STX1A	NM_004603.4	-	-	+	+	+	-	-	+	+		+	+
SYCE1	NM_00114376 3.2	-	-	-	+	+	-	-	+	+	+	+	+
TEN1	NM_00111332 4.3	-	-	+	-	-	-	-	-		+	+	+
TGFBR2	NM_00102484 7.2	-	-	-	-	-	-	-	+			+	+
TUBGCP5	NM_00135437 2.2	-	-	-	+	+	-	-	+			+	
UBE2K	NM_00131264 7.2	-	-	-	-	-	-	-	+	+		+	+
VAT1	NM_006373.4	-	-	+	-	-	-	-	+			+	
VRK3	NM_016440.4	-	-	-	+	-	-	-	+			+	+
WASF2	NM_006990.5	-	-	+	-	-	-	-	+			+	+

WDR62	NM_173636.5	-	-	-	-	-	-	-	+			+	+
XPO1	NM_003400.4	-	-	-	+	+	-	-	+		+	+	
YWHAH	NM_003405.4	-	-	+	-	-	-	-	+	+	+	+	
ZC3H4	NM_015168.2	-	-	-	+	+	-	-	-			+	+

Prilog 2 Popis top 10 bioloških procesa u koje su uključeni geni i genski produkti koji nisu pronađeni u ostalim bazama prema bazi Enrichr u ontologiji GO Biological process 2021 prema p vrijednosti

INDEX	NAME	P-VALUE	ADJUSTED P-VALUE	ODDS RATIO	COMBINED SCORE
1	nucleotide-excision repair (GO:0006289)	0.0003351	0.2500	3.56	28.52
2	positive regulation of histone H3-K9 acetylation (GO:2000617)	0.0004170	0.2500	41.10	319.88
3	regulation of histone H3-K9 acetylation (GO:2000615)	0.0004189	0.2500	15.68	121.93
4	microtubule cytoskeleton organization involved in mitosis (GO:1902850)	0.0006111	0.2500	3.12	45283
5	regulation of ATP metabolic process (GO:1903578)	0.0007428	0.2500	4.59	33.06
6	defense response to Gram- positive bacterium (GO:0050830)	0.0007754	0.2500	4.07	45289
7	negative regulation of locomotion (GO:0040013)	0.0008577	0.2500	12.19	86.09
8	ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway (GO:0043162)	0.0008685	0.2500	8.08	56.93
9	transcription-coupled nucleotide-excision repair (GO:0006283)	0.001054	0.2500	3.87	26.56
10	positive regulation of histone acetylation (GO:0035066)	0.001078	0.2500	7.63	52.11

Prilog 3 Popis top 10 molekularnih funkcija gena i genskih produkata koji nisu pronađeni u bazama podataka prema bazi Enrichr u ontologiji GO Molecular Function 2021 prema p vrijednosti

INDEX	NAME	P-VALUE	ADJUSTED P-VALUE	ODDS RATIO	COMBINED SCORE
1	alpha-amylase activity (GO:0004556)	0.0004170	0.06855	41.10	319.88
2	amylase activity (GO:0016160)	0.0004170	0.06855	41.10	319.88
3	protein serine/threonine phosphatase inhibitor activity (GO:0004865)	0.0004189	0.06855	15.68	121.93
4	CCR6 chemokine receptor binding (GO:0031731)	0.001384	0.1520	20.55	135.27
5	G protein-coupled photoreceptor activity (GO:0008020)	0.001548	0.1520	9.97	64.54
6	oxidoreductase activity, acting on the CH-NH2 group of donors, oxygen as acceptor (GO:0016641)	0.003150	0.2577	13.70	78.91
7	purine ribonucleoside triphosphate binding (GO:0035639)	0.007281	0.4390	1.73	18841
8	RNA polymerase II cis- regulatory region sequence- specific DNA binding (GO:0000978)	0.009322	0.4390	1.43	25720
9	cis-regulatory region sequence- specific DNA binding (GO:0000987)	0.009322	0.4390	1.43	25720
10	ribose phosphate diphosphokinase activity (GO:0004749)	0.01162	0.4390	18.24	81.26

Prilog 4 Popis top 10 zabilježenih fenotipskih abnormalnosti za gene i genske produkte koji nisu pronađeni u ostalim bazama prema bazi Enrichr, ontologiji Human Phenotype Ontology

INDEX	NAME	P-VALUE	ADJUSTED P-VALUE	ODDS RATIO	COMBINED SCORE
1	Renovascular hypertension (HP:0100817)	0,0007583	0.0005710	33.02	465.29
2	Hyperacusis (HP:0010780)	0.00001183	0.002227	27.48	311.71
3	Overriding aorta (HP:0002623)	0.00001183	0.002227	27.48	311.71
4	Elfin facies (HP:0004428)	0.00001183	0.002227	27.48	311.71
5	Insomnia (HP:0100785)	0.00003505	0.003770	19.62	201.31
6	Renal duplication (HP:0000075)	0.00003505	0.003770	19.62	201.31
7	Flat cornea (HP:0007720)	0.00003505	0.003770	19.62	201.31
8	Abnormality of the gastric mucosa (HP:0004295)	0.0001041	0.009799	7.71	70.68
9	Abnormality of the diencephalon (HP:0010662)	0.0001216	0.01013	13.73	123.81
10	Spina bifida occulta (HP:0003298)	0.0001345	0.01013	6.12	54.57

