Klasifikacija statusa bolesti putem strojnog učenja na podacima RNA sekvenciranja pojedinačnih stanica

Huzjak, Klara

SVEUČILIŠTE U RIJECI
FAKULTET BIOTEHNOLOGIJE I RAZVOJA LIJEKOVA
Diplomski sveučilišni studij
Biotehnologija u medicini

Klara Huzjak

Klasifikacija statusa bolesti putem strojnog učenja na podacima RNA sekvenciranja pojedinačnih stanica

Diplomski rad

Rijeka, 2024. godine

SVEUČILIŠTE U RIJECI
FAKULTET BIOTEHNOLOGIJE I RAZVOJA LIJEKOVA
Diplomski sveučilišni studij
Biotehnologija u medicini

Klara Huzjak

Klasifikacija statusa bolesti putem strojnog učenja na podacima RNA sekvenciranja pojedinačnih stanica

Diplomski rad

Rijeka, 2024. godine

Mentor rada: doc.dr.sc. Katarina Kapuralin
Komentor rada: doc.dr.sc. Mario Lovrić

UNIVERSITY OF RIJEKA
FACULTY OF BIOTECHNOLOGY AND DRUG DEVELOPMENT
Graduate university study
Biotechnology in medicine

Klara Huzjak

Classification of disease status using machine learning on single cell RNA-sequencing data

Master thesis

Rijeka, 2024.

Mentor: doc.dr.sc. Katarina Kapuralin
Co-mentor: doc.dr.sc. Mario Lovrić

Diplomski rad obranjen je dana 25.9. 2024.

Pred povjerenstvom:

1. _____
2. _____
3. _____

Rad ima 81 stanicu, 33 slike, 4 tablice i 89 literaturnih navoda.

SAŽETAK

Alzheimerova bolest karakterizirana kao stanje progresivnog gubitka kognitivnih funkcija, uključujući pamćenje, orijentaciju i prosuđivanje, predstavlja vodeći oblik demencije u svijetu. Iako točan uzrok bolesti još nije potpuno razjašnjen, ključni patološki mehanizmi bolesti uključuju nakupljanje izvanstaničnih beta-amiloidnih plakova i hiperfosforiliranog Tau proteina unutar živčanih stanica, što dovodi do neurofibrilarnih čvorova sa posljedicom atrofije mozga i odumiranjem živčanih stanica te gubitkom neurona i sinapsi. Cilj ovog diplomskog rada bio je istražiti potrencijalne genetske markere i klasificirati status Alzheimerove bolesti pomoću strojnog učenja, korištenjem podataka dobivenih RNA sekvenciranjem pojedinačnih stanica (scRNA-seq). Visoko-propusno RNA sekvenciranje na pojedinačnim stanicama revolucionarna je tehnika koja je omogućila analizu cijelokupnog transkripcijskog profila pojedinačnih stanica u velikom obujmu, što je ključna prednost u proučavanju heterogenosti stanica u kompleksnim tkivima poput mozga. Za potrebe rada podatci RNA sekvenciranja pojedinačnih stanica entorinalnog korteksa i superiornog frontalnog girusa mozga *post mortem* (Braak stadiji 0, 2 i 6) bili su prikupljeni i prethodno obrađeni korištenjem računalnog alata Seurat. Uporabom alata strojnog učenja kao što su modeli logističke regresije i nasumičnih šuma, analizirali smo podatke kako bi identificirali ključne gene i signalne puteve povezane s progresijom bolesti. Analizom podataka identificirali smo nekoliko gena kao potencijalne markere ovisno o stadiju bolesti poput UTP8 gena za ranij stadij ili HNRNPH1, RPL10A i CLOCK gene za kasniji stadij bolesti.

Ključne riječi: Alzheimerova bolest, Braakovi stadiji, entorinalni korteks, scRNA_seq, Seurat, strojno učenje

SUMMARY

Alzheimer’s disease, characterized by the progressive loss of cognitive functions, including memory, orientation, and judgment, represents the leading form of dementia worldwide.

Although the exact cause of the disease remains unclear, key pathological mechanisms include the accumulation of extracellular beta-amyloid plaques and hyperphosphorylated Tau protein within neurons, leading to neurofibrillary tangles, brain atrophy, neuronal death, and the loss of neurons and synapses. The aim of this thesis was to investigate potential genetic markers and classify Alzheimer’s disease status using machine learning, based on data obtained from single-cell RNA sequencing (scRNA-seq). High-throughput single-cell RNA sequencing is a revolutionary technique that has enabled the large-scale analysis of the entire transcriptional profile of individual cells, which is a key advantage in studying the heterogeneity of cells in complex tissues such as the brain. For the purposes of this study, RNA sequencing data from single cells of the entorhinal cortex and superior frontal gyrus of *post-mortem* brains (Braak stages 0, 2, and 6) were collected and preprocessed using the computational tool Seurat. Using machine learning tools such as logistic regression models and random forests, we analyzed the data to identify key genes and signaling pathways associated with disease progression. Our analysis identified several genes as potential markers depending on the disease stage, such as the UTP8 gene for earlier stages and the HNRNPH1, RPL10A, and CLOCK genes for advanced stages.

Key words: Alzheimer's disease, Braak stages, entorhinal cortex, scRNA_seq, Seurat, machine learning

POJMOVNIK SKRAĆENICA

AB – Alzheimerova bolest

Aβ - beta-amiloid

APP – amiloidni prekursorski protein

cDNA - komplementarna DNA

CI – Interval pouzdanosti (eng. *Confidence Interval*)

DROP-Seq – sekvenciranje temeljno na kapljicama (*eng. Droplet sequencing*)

DEG - Diferencijalno eksprimirani geni

EC - Entorinalni korteks (*eng. Enthorinal Cortex*)

GEM - gelirana kuglica u emulziji (*eng. GEM - Gel Bead-In-Emulsion*)

IWG – Međunarodna radna skupina (*eng. International Working Group*)

LR - Logistička regresija

MAP – Potein povezan s mikrotubulima (*eng. Microtubule associated protein*)

NFT – neurofibrilni čvorovi (*eng. Neurofibrillary tangles*)

NIA – Nacionalni institut za starenje (*eng. National Institute on Aging*)

NIA – AA – Nacionalni institut za starenje – Udruga za Alzheimerovu bolest (eng. *National Institute on Aging and Alzheimer's Association*)

NGS - Sekvenciranje nove generacije (*eng. New Generation Sequencing*)

PCA - Analiza glavnih komponenata (*eng. Principal Component Analysis*)

RF – Model nasumičnih šuma (*eng. Random Forest*)

ROC-AUC – Područje ispod krivulje radne karakteristike prijamnika (eng. *Receiver Operating Characteristic - Area Under the Curve*)

SFG - Superiorni frontalni girus (*eng. Superior Frontal Gyrus*)

Sc RNA-seq - RNA sekvenciranje pojedinačnih stanica (*eng. Single cell RNA Sequencing*)

STAMPs - jednostanični transkriptomi pripojeni na mikročestice (*eng. Single-cell Transcriptomes Attached to Microparticles*)

tSNE - t-distribuirano stohaističko susjedno ugrađivanje (*eng. t-distributed Stochastic Neighbor Embedding*)

UMAP - uniformna aproksimacija i projekcija mnogostrukosti (*eng. Uniform Manifold Approximation and Projection*)

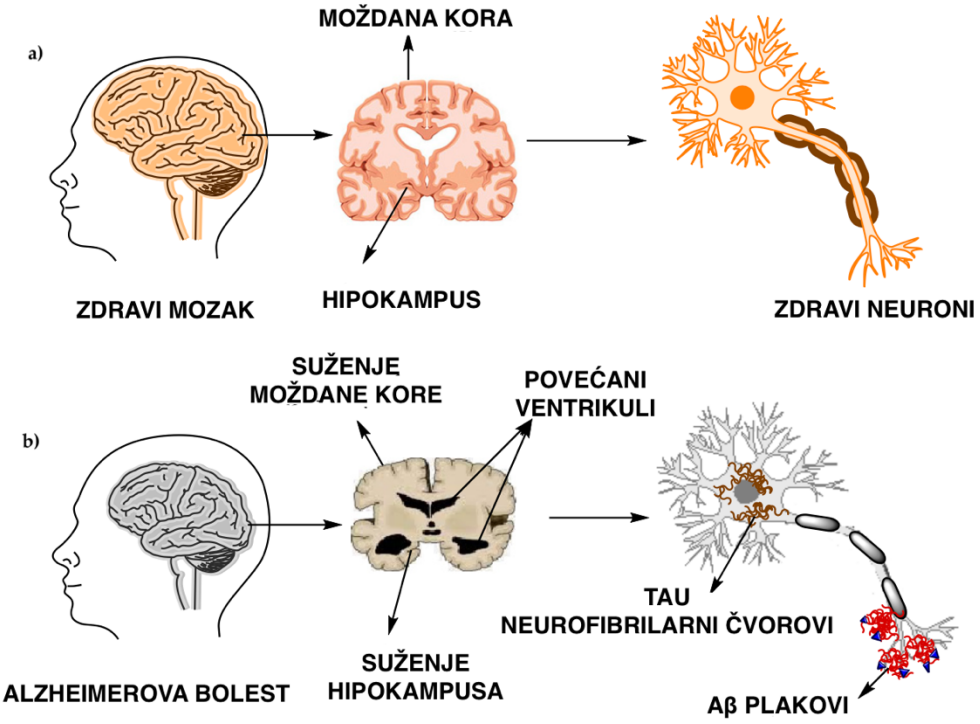
SADRŽAJ

1. UVOD

Demencija je zajednički naziv za progresivni i trajni gubitak intelektualnih sposobnosti poput poteškoća s pamćenjem, kognitivnih smetnji shvaćanja, prosuđivanja i mišljenja, kao i dezorijentacije u vremenu i prostoru [1].
Alzheimerova je bolest progresivna neurodegenerativna bolest i vodeći uzrok demencije u Svijetu, koja je odgovorna za 60-80% svih dijagnosticiranih slučajeva [2].
Bolest je prvi opisao njemački liječnik Alois Alzheimer koji ju je 1905. godine okarakterizirao kao presenilnu demenciju (*eng. Dementia in the presenium*) na temelju kliničko - patološkog istraživanja na 51. godišnjoj pacijentici Auguste Deter.
Auguste Deter opisana je kao zaboravljiva, dezorijentirana, sumnjičava i anksiozna žena sa dodatnim auditornim halucinacijama i uznapredovalim gubitkom pamćenja. Nakon smrti pacijentice 1906. godine provedena je obdukcija i biopsija tkiva mozga i leđne moždine, na temelju koje je Alzheimer opisao otkrivene promjene kao milijarna žarišta izvanstaničnih struktura, danas poznatnih kao plakovi, te unutarstanične snopove, odnosno neurofibrilarne čvorove u korteksu mozga pacijentice [3,4].
Klasifikacija Alzheimerove bolesti dijeli se na sporadični i genetski tip. Sporadični oblik čini oko 95% slučajeva ove bolesti, sa početkom manifestacije simptoma nakon 65. godine života. Genetski tip prisutan je u oko 5% slučajeva, a karakteriziran je ranijim početkom razvoja bolesti (prije 65. godine života) i uzrokovani je mutacijama gena APP, PSEN1 I PSEN2 [2,8].
Prema Svjetskoj Zdravstvenoj Organizaciji (WHO*, eng. World Health Organization)* u svijetu trenutno živi 55,2 milijuna ljudi s demencijom, a očekuje se da će taj broj porasti na 78 milijuna do 2030. i 139 milijuna do 2050. godine zbog globalnog starenja populacije [5].

1.1 NEUROPATOLOGIJA ALZHEIMEROVE BOLESTI

Alzheimerova se bolest očituje postupnim i progresivnim gubitkom pamćenja te kognitivnim poremećajima. Gubitak neuroloških funkcija kod pacijenata većinom je uzrokovan nakupljanjem toksičnih proteinskih fragmenata u živčanom sustavu, odnosno taloženjem izvanstaničnih naslaga beta-amiloida (Aβ), te unutarstaničnim nakupljanjem Tau proteina koji je odgovoran za održavanje stabilnosti mikrotubula [8] (slika 1).



Slika 1 Shematski prikaz mozga i neurona u (a) zdravom mozgu i u (b) Alzheimerovoj bolesti (Slika je preuzeta i prevedena sa Breijyeh Z, Karaman R. Comprehensive Review on Alzheimer's Disease: Causes and Treatment. Molecules. 2020;25(24):5789. Published 2020 Dec 8. doi:10.3390/molecules25245789)

1.1.1 AMILOIDNA PATOLOGIJA

Amiloidni plakovi su izvanstanične naslage netopljivih monomera amiloida β (Aβ) s različitim morfološkim obilježjima.
Agregacijom Aβ fragmenata koji nastaju proteolitičkim cijepanjem transmembranskog amiloidnog prekursorskog proteina (APP-a) pomoću beta i gamma sekretaza, stvaraju se peptidi koji su razlog formiranja senilnih plakova [6,7]. APP je membranski glikoprotein tipa 1 koji je važan za razvoj neurona, signalizaciju i unutarstanični transport. Beta sekretaza cijepa APP u izvanstaničnom prostoru, dok gamma sekretaza cijepa preostale fragmente, stvarajući Aβ peptide [9, 11].

Amiloidna hipoteza predlaže nakupljanje Aβ kao glavni uzrok Alzheimerove bolesti, sugerirajući da pogrešno savijanje Aβ proteina u plakovima, zajedno s unutarstaničnim taloženjem Tau proteina u neurofibrilarnim čvorovima uzrokuje gubitak pamćenja, konfuznost i kognitivni pad [11].
Aβ peptidi imaju intrinzičnu sklonost agregaciji u oligomere, protofibrile ili zrele fibrile koji se talože u plakovima. Aβ40 i Aβ42 glavne su komponente tih plakova, pri čemu povećana razina Aβ42, koja brže formira plakove, dominira u patologiji bolesti [7].

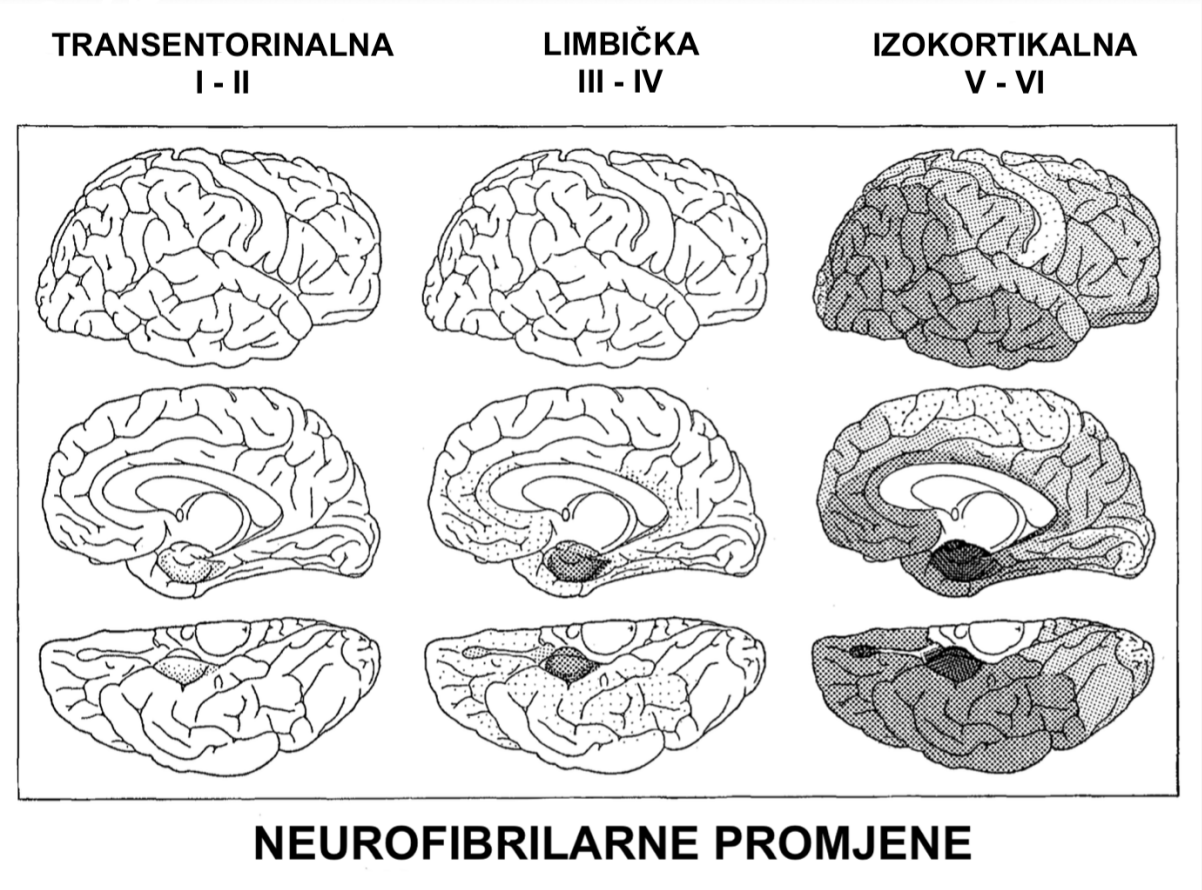
Širenje amiloidnih plakova izvan neurona odvija se u pet faza. Prva faza uključuje rane naslage u neokorteksu, dok se u drugoj fazi plakovi pojavljuju u limbičkim regijama poput entorinalnog korteksa, amigdale i cingularnog girusa. Treća faza obuhvaća širenje u subkortikalna područja kao što su bazalni gangliji i talamus. Kasniji stadiji, četvrta i peta faza, zahvaćaju moždano deblo, produljenu moždinu, te na kraju koru malog mozga. Demencija je u pacijenata povezana s četvrtom i petom fazom, dok su prve tri faze asimptomatske [10, 12].

1.1.2 TAU PATOLOGIJA

Neurofibrilarni čvorovi (NFT, *eng. Neurofibrillary tangles*) posljedica su nakupljanja hiperfosforiliranog Tau proteina u citoplazmi neurona [6].
Tau je protein koji pripada obitelji *MAP* (*eng.– microtubule-associated proteins*) odgovornih za regulaciju i povezivanje mikrotubula unutar neurona [13]. Nakupine neurofibrilarih čvorova drugi su glavni uzrok patologije Alzheimerove bolesti [10].

Tau je odgovoran za stabilizaciju i sastavljanje mikrotubula te pospješuje aksonski transport i strukturu dendrita, a njegova aktivnost regulirana je posttranslacijskim modifikacijama, uključujući fosforilaciju na više mjesta. U Alzheimerovoj bolesti hiperfosforilacija Tau proteina dovodi do gubitka njegovih funkcija, što rezultira problemima u sastavljanju mikrotubula, aksonskom transportu i strukturi dendrita, gubitkom sinapsi, apoptozom neurona i posljedično, demencijom [14].

Patologija neurofibrilarnih čvorova u Alzheimerovoj bolesti klasificirana je metodom poznatom kao Braakovi stadiji. Prema patologiji zahvaćenosti mozga, bolest možemo podjeliti na šest stadija, u tri regije mozga [10]. Prvi znakovi pojave neurofibrilarnih čvorova nalaze se u transentorinalnoj regiji hipokampalne formacije, što označava najraniji stadij bolesti, B1 (stadiji I, II).Daljnje napredovanje bolesti dovodi do širenja neurofibrilarnih čvorova u druge neokortikalne regije zahvaćajući i limbički sustav, što označuje srednju fazu bolesti, B2(stadiji III, IV) [10,15].
U kasnijim fazama, promjene postaju izrazito intenzivne u hipokampalnoj formaciji i šire se na sekundarna asocijacijska područja i na kraju, na primarna kortikalna područja. Ove kasne faze bolesti karakterizira ozbiljno oštećenje neokorteksa i označava treću fazu bolesti, B3 (stadiji V I VI) (Slika 2) [15].



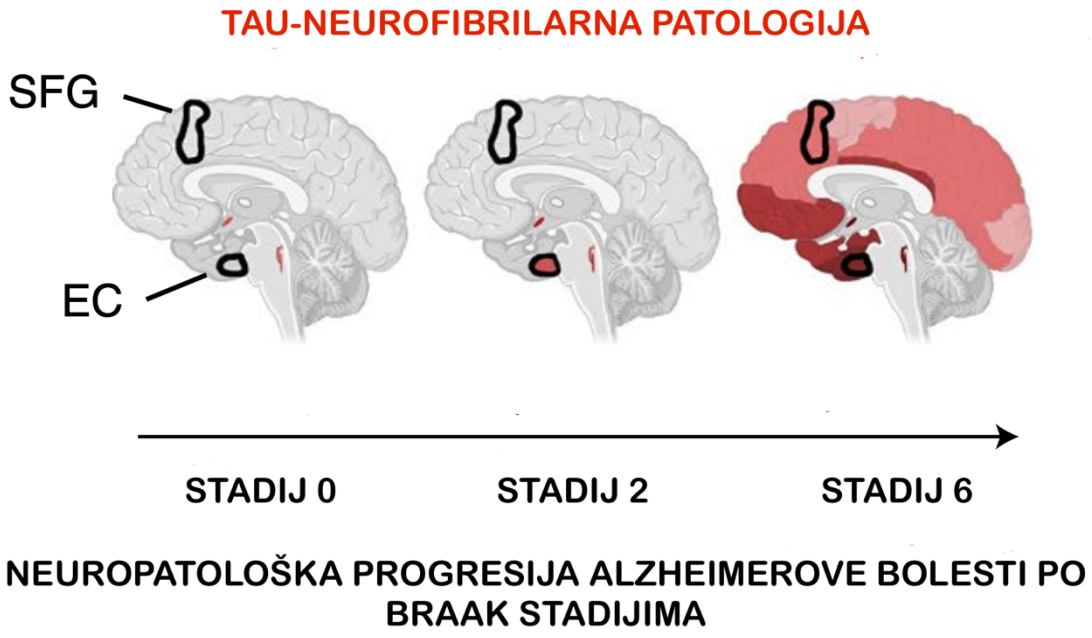
Slika 2 Shematski prikaz šest stadija neurofibrilarnih promjena u pogođenim regijama prema Braakovim stadijima (slika preuzeta i prilagođena sa: Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol. 1991;82(4):239-59. doi: 10.1007/BF00308809. PMID: 1759558.)

Progresija patologije prema Braakovim stadijima (Slika 3) pokazuje određenu korelaciju s kliničkim simptomima, pri čemu su kasniji stadiji povezani s izraženijim oblicima demencije, dok su rani stadiji uglavnom prisutni kod klinički asimptomatskih osoba [10].

Entorinalni korteks poznat kao Brodmannovo područje 28, jedna je od prvih regija zahvaćena neurodegeneracijom u Alzheimerovoj bolesti.
Dio je parahipokampalnog girusa smještenog u medijalnom temporalnom režnju, a glavnu ulogu drži u kognitivnim procesima i prostornoj orijentaciji.
U Alzheimerovoj bolesti EC mjesto je stvaranja unutarstaničnih neurofibrilarnih čvorova i glavna regija u kojoj se nakupljaju izvanstanični amiloidni plakovi što rezultira ometanjem normalne stanične funkcije i dovodi do odumiranja neurona.

Ove promjene dovode do patološke slike u EC, pridonoseći ranom gubitku pamćenja i kognitivnom padu karakterističnom za Alzheimerovu bolest [61].

Superiorni frontalni girus (SFG) smješten je u gornjem dijelu prefrontalnog korteksa i dio je frontalnog režnja [62]. Lijevi, dominantni SFG ključna je komponenta u neuronskoj mreži radne memorije kao i prostorne orijentacije, dok je desni, nedominantni SFG uključen u kontrolu impulsa tako što svojom aktivacijom modulira inhibicijsku kontrolu i motoričku hitnost [63].
SFG je pogođen u kasnijim stadijima progresije bolesti, a funkcije SFG također su oštećene kao posljedica nakupljanja izvanstaničnih plakova i neurofibrilrnih čvorova sa atrofijom te regije zbog gubitka neurona i sinapsi [63].



Slika 3 Shematski prikaz Tau patologije Alzheimerove bolesti po Braakovim stadijima sa pogođenim regijama mozga: EC- entorinalni korteks, SFG-superiorni frontalni girus. (slika preuzeta i prevedena sa Leng, Kun et al. "Molecular characterization of selectively vulnerable neurons in Alzheimer's disease." Nature neuroscience vol. 24,2 (2021): 276-287. doi:10.1038/s41593-020-00764-7)

1.1.3 DIJAGNOSTIČKI KRITERIJI ALZHEIMEROVE BOLESTI

Dijagnostički kriteriji i klasifikacija statusa Alzheimerove bolesti su se mijenjali od sredine 20. stoljeća. Godine 1984. Nacionalni institut za neurološke i komunikacijske poremećaje i moždani udar (*NINCDS, eng. National Institute of Neurological and Communicative Disorders and Stroke*) i Udruga za Alzheimerovu bolest i srodne poremećaje (*ADRDA, eng. Alzheimer's Disease and Related Disorders Association*) definirali su Alzheimerovu bolest kao dementni poremećaj srednje ili kasne životne dobi i klasificirali dijagnozu bolesti u tri stupnja: vjerojatno, moguće i definitivno [6,7].

Značajni se napredak u dijagnostici dogodio uvođenjem IWG (*eng. Internacional Working Group*) kriterija 2007. godine, koji su redefinirali razumijevanje Alzheimerove bolesti uvodeći pojmove pretklinička i prodormalna (stanje prije demencije) faza Alzheimerove bolesti.
Pretklinička faza označava asimptomatsko razdoblje u kojem se razvijaju neuropatološke promjene, dok prodormalna uključuje bolesnike s umjerenim kognitivnim poremećajima (MCI, *eng. Mild Cognitive Impairment*) povezanih s Alzheimerovom bolešću, a označava simptomatsku fazu prije demencije [16]. IWG kriteriji identificirali su bolest u njezinoj prodormalnoj fazi s dvije razine sigurnosti:vjerojatna i definitivna bolest i klasificirali dijagnostičke i prognostičke markere [7].

Nacionalni institut za starenje – Udruga za Alzheimerovu bolest (NIA-AA , eng. *National Institute on Aging—Alzheimer's Association*) dalje je unaprijedio dijagnostičke kriterije, odvajajući patofiziološki proces Alzheimerove bolesti (eng. AD-P) od kliničke faze bolesti (eng. AD-C). Na temelju toga predložena je nova terminologija dijagnostičkih kriterija, podijeljena u tri faze: pretklinička faza, faza umjerenog kognitivnog poremećaja uzrokovanog patofiziološkim procesom Alzheimerove bolesti, i demencija uzrokovana patofiziološkim procesom Alzheimerove bolesti [17].

Biomarkeri su postali ključni alat u dijagnostici Alzheimerove bolesti, a podijeljeni su u dvije glavne skupine: biomarkeri taloženja beta-amiloida (Aβ u cerebrospinalnoj tekućini i amiloid-PET), te biomarkeri neuronske ozljede (tau, fosforilirani tau u cerebrospinalnoj tekućini, FDG-PET, MRI) [17]. Prisutnost ovih biomarkera određuje vjerojatnost patološkog procesa AB, klasificirajući ih u tri kategorije: visoka vjerojatnost (pozitivna oba biomarkera), srednja vjerojatnost (pozitivan jedan biomarker) i mala vjerojatnost (oba biomarkera negativna) [17].

Najnoviji NIA-AA kriteriji uvode A/T/N sustav za klasifikaciju biomarkera: A (amiloidna patologija), T (tau patologija) i N (neurodegeneracija) [17]. Takav sustav omogućuje precizniju dijagnozu i razumijevanje same bolesti, pri čemu je prisutnost A (amiloid) potrebna, ali ne i dovoljna za dijagnozu Alzheimerove bolesti, dok kombinacija A (amiloid) i T (tau) odgovara neuropatološkoj dijagnozi Alzheimera. N (neurodegeneracija) ukazuje više na stupanj ozbiljnosti bolesti, nego na samu prisutnost bolesti [17, 25]. Također, razvoj novih biomarkera, poput tau-PET skeniranja, obećava bolju diferencijaciju i precizniju prognozu Alzheimerove bolesti [26].

1.1.4 GENETIKA ALZHEIMEROVE BOLESTI

Ključni geni povezani s Alzheimerovom bolešću uključuju gen za amiloidni prekursorski protein (APP) koji je smješten na kromosomu 21, te gene za Presenilin-1 (PSEN-1 gen) i Presenilin-2 (PSEN-2 gen). Ovi geni su izravno uključeni u patogenezu Alzheimerove bolesti, posebno u ranoj fazi bolesti [18]. Mutacije u ovim genima dovode do abnormalne proizvodnje beta-amiloida, glavne komponente senilnih plakova [18].

Još jedan ključan gen u razvoju Alzheimerove bolesti je ApoE gen, smješten na kromosomu 19 koji kodira za apolipoprotein E. Apolipoprotein E je glikoprotein visoko eksprimiran u stanicama jetre i astrocitima mozga [19]. ApoE protein igra ključnu ulogu u metabolizmu lipida, služeći kao ligand za endocitozu posredovanu receptorom za lipoproteinske čestice, poput kolesterola. Taj je mehanizam ključan za proizvodnju mijelina i održavanje normalne funkcije mozga [19].

ApoE gen ima tri glavne izoforme: ApoE2, ApoE3 i ApoE4, koje su posljedica polimorfizama jednog nukleotida (*eng. Single Nucleotide Polymorphism - SNP*) [20]. Alel ApoEε4 prepoznat je kao najsnažniji genetski čimbenik rizika za razvoj Alzheimerove bolesti, povezan s povišenim taloženjem amiloid-beta peptida kao senilnog plaka i pojavom cerebralne amiloidne angiopatije (CAA), koja je također specifični marker Alzheimerove bolesti [19]. Nasuprot tome, aleli ApoEε2 i ApoEε3 povezani su s nižim rizikom od razvoja bolesti, time da ApoEε2 pokazuje određeni zaštitni učinak [19].
Također je dokazano da je alel ApoEε4 povezan s vaskularnim oštećenjem unutar mozga, što dodatno doprinosi patogenezi Alzheimerove bolesti kroz mehanizme vaskularne disfunkcije i upalnih procesa [19, 20].

1.1.5 LIJEČENJE ALZHEIMEROVE BOLESTI

Alzheimerova je bolest sve veći problem globalno, s otprilike 24 milijuna oboljelih, a procjenjuje se da bi se broj slučajeva oboljelih od demencije mogao učetverostručiti do 2050. godine [6].
Trenutno dostupna farmakološka terapija za liječenje Alzheimerove bolesti uključuje dizajn molekula usmjerenih na osnovnu biologiju bolesti, s ciljem ublažavanja njezine progresije pomoću simptomatskih kognitivnih pojačivača [8]. Liječenje uključuje primjenu inhibitora acetilkolinesteraze (donepezil, galantamin i rivastigmin) i antagonista N-metil-D-aspartat (NMDA) receptora (memantin), s ciljem povećanja kolinergičke neurotransmisije i aktivacije NMDA receptora [21].

Inhibitori kolinesteraze djeluju spriječavajući razgradnju acetilkolina (ACh), neurotransmitera ključnog za pamćenje i kognitivne funkcije, čija se razina smanjuje kod oboljelih od Alzheimerove bolesti zbog odumiranja stanica koje proizvode acetilkolin [22]. Ovi lijekovi inhibiraju enzime kolinesteraze, što dovodi do povećanja razine ACh i poboljšanja kolinergijskog prijenosa u mozgu. Antagonisti NMDA receptora, s druge strane, djeluju blokirajući prekomjernu aktivaciju NMDA receptora, koja može uzrokovati povećan ulazak kalcija i dovesti do oštećenja neurona [22, 23].
Iako ovi lijekovi pružaju simptomatsko olakšanje, oni ne zaustavljaju napredovanje bolesti. Pokazalo se kako najbolji terapijski učinak imaju inhibitori kolinesteraze u kombinaciji sa antagonistima NMDA receptorima [22,23].

Trenutna nova istraživanja za Alzheimerovu bolest usmjerena su na promjenu tijeka bolesti ciljajući na specifične puteve poput patologije puta amiloida-beta [24]. Terapije monoklonalnim protutijelima, kao što su solanezumab i bapinezumab prošla su klinička ispitivanja, ali nažalost nisu pokazala učinkovitost u završnim fazama istraživanja [24].

1.2 MODERNE OMICS ANALIZE – SINGLE CELL RNA-SEQ

RNA sekvenciranje pojedinačnih stanica *(*eng*. Single cell RNA Sequencing - scRNA-seq*) je tehnika koja je unaprijedila razumijevanje transkriptoma stanica različitih organizama, omogućavajući analizu ekspresije gena na razini pojedinačnih stanica i otkrivajući staničnu heterogenost te rijetke stanične populacije koje mogu biti ključne u razvoju bolesti [27, 28]. ScRNA-seq prvi je puta opisana 2009. godine sekvenciranjem transkriptoma jedne blastomere i oocite [28], dok je revolucionarna tehnika visokopropusnog RNA sekvenciranja pojedinačnih stanica, koja omogućuje istovremeno paralelno sekvenciranje do desetaka tisuća stanica prvi puta opisana 2015. godine [32].

Tehnika sekvenciranja RNA pojedinačnih stanica sastoji se od šest ključnih koraka: izolacije i hvatanja pojedinačnih stanica, lize stanica, reverzne transkripcije (pretvorbe RNA u cDNA), barkodiranja i amplifikacije komplementarne DNA (cDNA) te na kraju pripreme cDNA biblioteke [29].

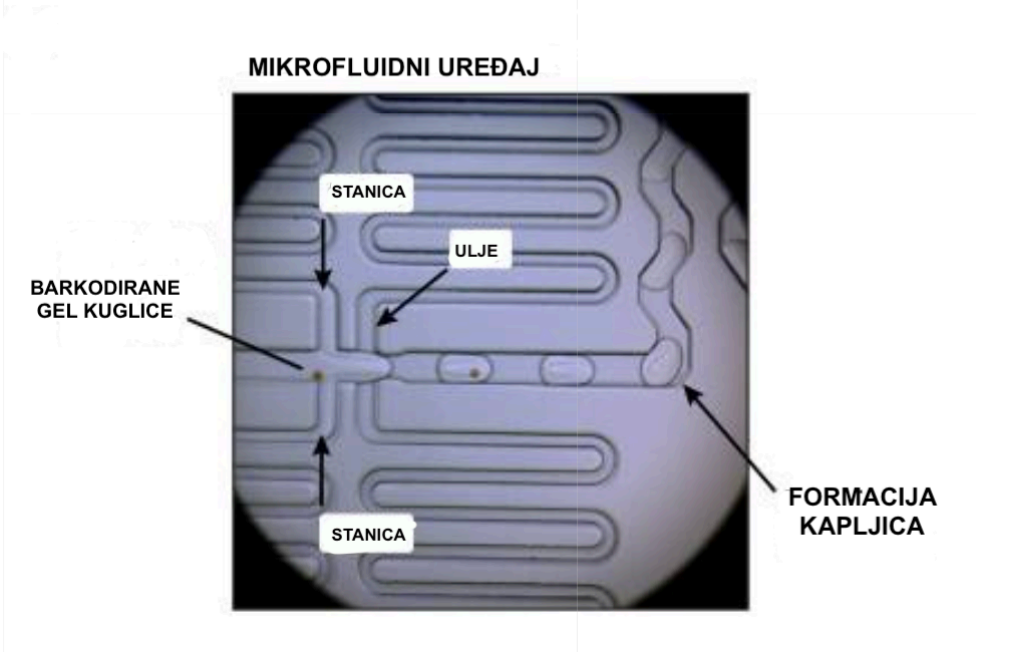
Danas postoji više različitih tehnika sekvenciranja RNA iz pojedinačnih stanica (*eng.scRNA-seq*) koje omogućuju detaljnu analizu ekspresije gena na razini pojedinačnih stanica ovisno o potrebi istraživanja [30].

Također, postoje i značajni izazovi u paralelnoj obradi tisuća stanica, kao i u preciznom rukovanju s malim uzorcima stanica kako bi se osiguralo da svaka stanica bude precizno izmjerena [33].
Točnost mjerenja scRNA-seq može jako ovisiti o učinkovitosti enzimskih koraka poput reverzne transkripcije i amplifikacije komplementarne DNA, pri čemu niska učinkovitost može rezultirati lažnim ili nedostatnim rezultatima [34]. Također, amplifikacija nižih koncentracija RNA može izazvati pogreške i unijeti dodatne varijacije u rezultate, što dodatno otežava preciznu kvantifikaciju gena [34].

Mikrofulidika, posebice metoda koja koristi mikrofluidne kapljice značajno je unaprijedila scRNA-seq tehniku i izazove koje nosi omogućujući enkapsulaciju tisuća pojedinačnih stanica u odvojenim nanolitarskim kapljicama i time poboljšala prinos komplementarne DNA i smanjenje tehničkih varijabilnosti, kao i povećavanja točnosti mjerenja velikog broja malih uzoraka [33].

1.2.1 DROP – SEQ

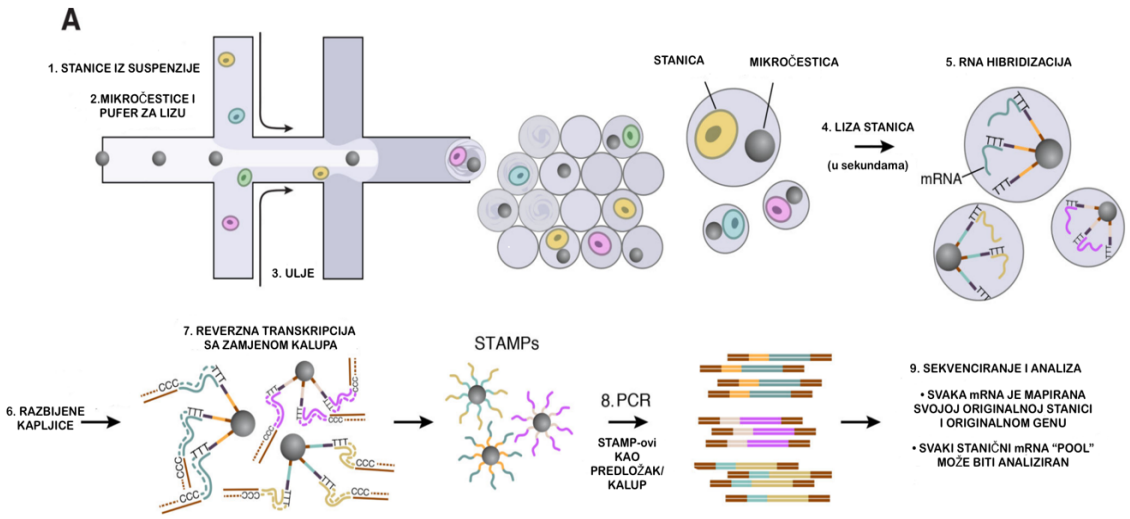
Visokopropusno RNA sekvenciranje pojedinačnih stanica temeljeno na kapljicama, (*eng. high-throughput droplet based single cell RNA-seq*) izuzetno je značajna platforma molekularne biologije koja omogućuje analizu desetaka tisuća pojedinačnih stanica istovremeno, pružajući uvid u kompletnu molekularnu strukturu pojedinačnih stanica unutar tkiva [32]. Drop-seq je metoda jednostaničnog sekvenciranja koja koristi mikrofluidni uređaj za razdvajanje kapljica koje sadrže pojedinačne stanice, pufer za lizu i mikrozrnca prekrivena početnicama s barkodom [38] (Slika 4).



Slika 4 Slikovni prikaz mikrofulidnog uređaja za Drop-seq analizu (preuzeto i uređeno sa Macosko, Evan Z et al. "Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets." Cell vol. 161,5 (2015): 1202-1214. doi:10.1016/j.cell.2015.05.002)

Dropleti su nanolitarske kapljice stvorene preciznim kombiniranjem vodenih i uljnih tokova u mikrofluidnom uređaju, a takav mikrofluidni sustav se koristi za inkapsulaciju pojedinačnih stanica zajedno s barkodiranim oligonukleotidima u kapljicama, što omogućuje masovnu analizu transkriptoma stanica u velikim količinama [32].

Postupak Drop-seq tehnike se sastoji od pripreme suspenzije pojedinačnih stanica iz tkiva, zatim se svaka stanica zajedno s jedinstveno barkodiranim mikročesticama (perlama) enkapsulira u nanolitarske kapljice. Nakon što su stanice izolirane u kapljicama, dolazi do lize stanica. U ovom procesu, mRNA svake stanice se hvata na pripadajuću mikročesticu, stvarajući STAMP-ove (*eng. Single-cell Transcriptomes Attached to Microparticles*). U daljnjem koraku, STAMP-ovi se reverzno transkribiraju, amplificiraju i potom sekvenciraju u jednoj reakciji. Na kraju, barkodovi koji su dio STAMP-a, se koriste za određivanje podrijetla svake mRNA, te je točno poznato iz koje je stanice svaki transkript potekao [32, 33] (slika 5).



Slika 5 Shematski prikaz barkodiranja Drop-seq tehnnike (slika preuzeta i prilagođena sa Macosko, Evan Z et al. "Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets." Cell vol. 161,5 (2015): 1202-1214. doi:10.1016/j.cell.2015.05.002)

1.2.2 10X GENOMICS CHROMIUM

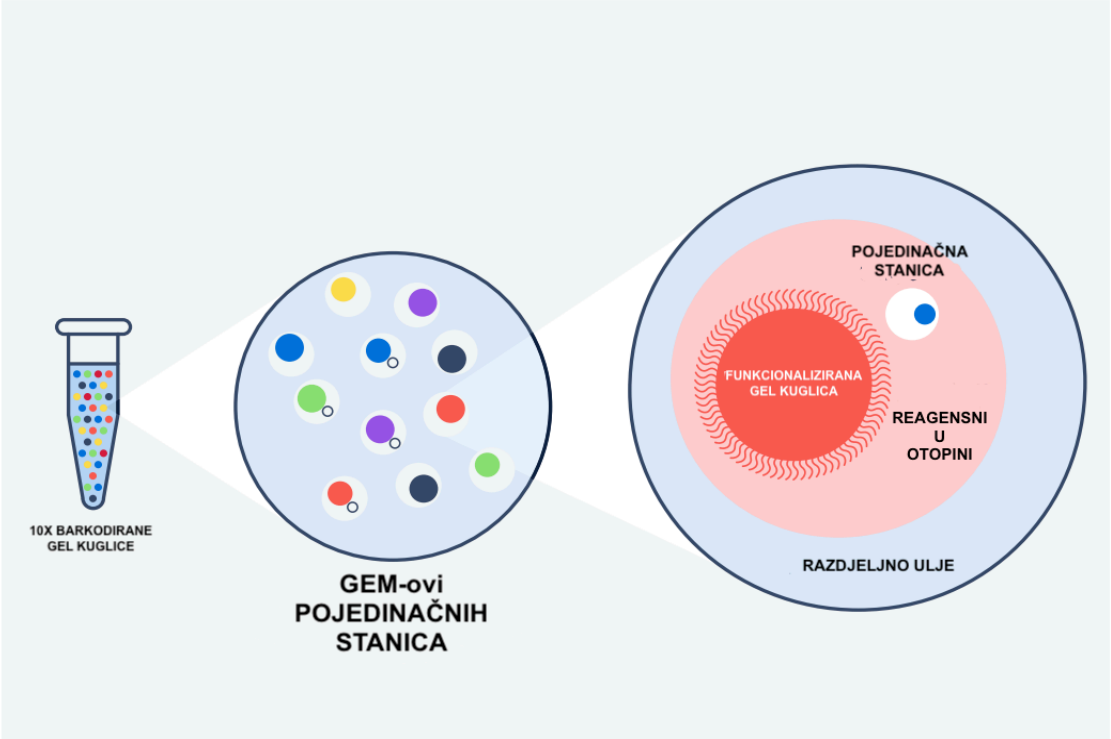
10x Genomics Chromium naziv je automatizirane, komercijalne platforme temeljene na mikrofluidnom sustavu, koja omogućuje pripremu knjižica pojedinačnih stanica za visokopropusno RNA sekvenciranje (slika 6).
Ova platforma omogućuje integriranu analizu pojedinačnih stanica u velikom obujmu koristeći tehnologiju geliranih kuglica u emulziji, GEM-X (*eng. GEM - Gel Bead-In-Emulsion*) [39].
Slično kao i Drop-seq, Chromium sustav koristi metodu temeljenu na kapljicama za enkapsuliranje pojedinačnih stanica zajedno s barkodovima označenih oligonukleotidima unutar kapljica koje sadrže reagense potrebne za prepisivanje RNA u cDNA. Proces započinje unosom stanica u sustav, gdje se svaka stanica kombinira s kapljicom koja sadrži specifičan barkod. Barkod omogućuje kasniju identifikaciju molekula RNA koje potječu iz iste stanice nakon sekvenciranja [37, 41].

Za razliku od Drop-seq metode, 10X Genomics Chromium sustav primjenjuje napredne metode mikrofluidike za precizno usmjeravanje stanica u kapljice, što poboljšava učinkovitost i smanjuje varijabilnost u broju molekula koje se mogu analizirati po stanici [41].

Slika 6 Chromium X uređaj (slika preuzta sa: https://www.10xgenomics.com/instruments/chromium-x-series)

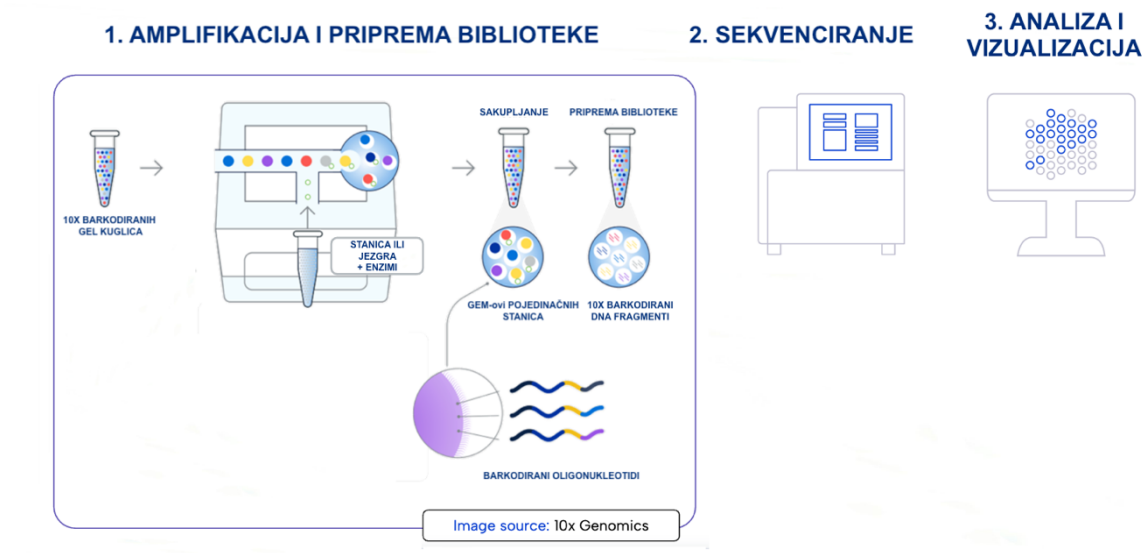
Postupak se sastoji od 4 glavna koraka formiranja kapljica:

1. Suspenzija pojedinačnih stanica ili jezgri pomiješana s reagensima optimiziranim za odabrani test, što uključuje pufer za lizu, oligonukleotide, gel kuglice (*eng. Gel Beads*) i enzime, učitava se na mikrofluidički čip.
2. Unutar Chromium instrumenta, stanice točno određene koncentracije kreću se kroz kanale kako bi se generirali desetci tisuća GEM-ova (*eng. Gel Bead-in-Emulsions*). GEM-ovi se stvaraju u Next GEM mikrofluidičkim čipovima (Slika 7). Nakon formiranja kapljica, mRNA molekule se obrnutno transkribiraju u komplementarnu DNA, cDNA, koja se zatim amplificira i priprema za sekvenciranje. Sve cDNA iz jedne stanice, koje potječu iz jedne gelirane kuglice, imat će isti 10x crtični kod, što omogućuje mapiranje svakog prijepisa natrag na njegovu stanicu podrijetla [41].
3. Svaki GEM djeluje kao pojedinačna reakcijska kapljica u kojoj se gel kuglice otapaju, a molekule od interesa iz svake stanice hvataju se i označavaju barkodom.
4. Nakon barkodiranja, svi fragmenti iz iste stanice ili jezgre dijele zajednički 10x barkod. Barkodirani fragmenti nekoliko tisuća stanica zajedno se skupljaju u daljnje reakcije kako bi se stvorile biblioteke kompatibilne sa uređajem za sekvenciranjem [41].



Slika 7 Shematski prikaz sastava 10X barokodiranih Gel Beads-a (slika preuzeta i dorađena sa: https://pages.10xgenomics.com/rs/446-PBO-704/images/10x_LIT000025_Brochure_Chromium-products_Letter_Digital.pdf

Sekvenciranje se provodi na NGS (*eng. New Generation Sequencing*) uređajima, a dobiveni podaci se analiziraju za identifikaciju transkripcijskih profila pojedinačnih stanica [40, 41] (slika 8).



Slika 8 Shematski prikaz 10X Genomics Chromium postupka sc-RNA-seq analize (slika preuzeta i dorađena sa: https://app.hubspot.com/documents/4935197/view/728088454?accessId=d75f75)

Platforma Chromium standardizira tijek rada scRNA-seq, čineći pouzdane i reproducibilne uvide u pojedinačne stanice, a Chromium Single Cell testovi podržani su instrumentima koji automatiziraju najvažniji korak u bilo kojem scRNA-seq eksperimentu — dijeljenje stanica i barkodiranje [39-41].

1.3 SEURAT – ANALIZA SINGLE CELL RNA-SEQ PODATAKA

Seurat je računalni softver dizajniran za analizu i vizualizaciju podataka RNA sekvenciranja na razini pojedinačnih stanica. RNA sekvenciranje pojedinačnih stanica, zajedno s moćnim bioinformatičkim analizama pomogli su u razumijevanju raznolikosti vrsta stanica i njihove genske ekspresije unutar heterogenog tkiva [42].

Seurat je paket R programskog jezika, razvijen od strane laboratorija prof. Rahul Satija, a nazvan je po umjetniku Georgesu Seuratu, začetniku slikarske tehnike poentilizma povlačeći analogiju između ovog umjetničkog pristupa i složenog prostornog uređenja opaženog u pojedinačnim stanicama [42].
Trenutno postoji pet verzija Seurata, a Seurat v5 najnovija je verzija koja uz postojeće analize podataka nudi i analizu prostorne transkriptomike i mogućnosti integracije podataka različitih OMICS tehnologija [46].

Upute (*eng. Vignettes*) u Seuratu su detaljni tutorijali koji pokazuju kako vršiti specifične analize korištenjem Seurat paketa.
Trenutno dostupne upute u Seuratu jesu: Upute za osnovnu analizu u kojem je objašnjeno kako podatke obraditi, klasterirati i vizualizirati. SCTransfrom tutorijal koji služi kao vodič o korištenju SCTransforma za poboljšanu normalizaciju i stabilizaciju varijance u scRNA-seq podacima, Upute za analizu prostornih skupova podataka, upute za integracije skupova podataka za usporedbe različitih stanja u više skupova podataka o pojedinačnim stanicama i upute za analizu diferencijalne ekspresije gena koji sadrži upute kako identificirati diferencijalno eksprimirane gene između različitih staničnih populacija ili biološko-patoloških stanja.

Seurat nudi niz funkcija, od normalizacije podataka, redukacije dimenzionalnosti, klasteriranja, integracije podataka pa do izvođenja statističkih testova poput analize diferencijalno eksprimiranih gena [60].

Prvi korak u analizi podataka RNA sekvenciranja pojedinačnih stanica pomoću Seurat paketa je normalizacija scRNA-seq podataka kako bi se dobili podaci o biološki značajnim stanicama. Zatim slijedi identifikacija varijabilnih gena, gdje Seurat identificira gene koji pokazuju najviše varijacija među stanicama. Nakon toga, kako su podaci scRNA-seq vrlo visokodimenzionalni jer svaka stanica i gen predstavljaju svoju dimenziju, za pojednostavljenje analize, Seurat smanjuje dimenzionalnost podataka koristeći analize kao što su: analiza glavnih komponenata, PCA (*eng. Principal Component Analysis*), uniformna aproksimacija i projekcija mnogostrukosti, UMAP (eng. *Uniform Manifold Approximation and Projection*) i Ugradnja pomoću t-distribuiranog stohaičkog susjeda, *t-SNE* (eng. *t-distributed Stochastic Neighbor Embedding*) [42].

Analiza glavnih komponenti (*eng.PCA*) je statistička tehnika koja se koristi za smanjenje dimenzionalnosti scRNA-seq podataka uz očuvanje što je moguće veće varijance u podacima, transformirajući izvorne podatke o ekspresiji gena u manji skup novih varijabli koje se nazivaju glavne komponente [43, 44]. U Seuratu, PCA pomaže identificirati najznačajnije obrasce u podacima o genskoj eksresiji, koji se zatim koriste za razlikovanje između različitih vrsta stanica [42].

Nakon što se podaci reduciraju, Seurat grupira stanice u skupine, odnosno klastere (eng.*clusters*) na temelju njihovih profila genske ekspresije. Seurat primjenjuje pristup grupiranja temeljen na dijagramima, koristeći algoritme k-sredine, Louvain ili SLM . Ti se klasteri zatim analiziraju za utvrđivanje vrsta stanica koje predstavljaju [42].

Seurat nudi različite tehnike redukcije nelinearne dimenzije, kao što su tSNE i UMAP, za vizualizaciju i istraživanje klastera. Cilj ovih algoritama je razumjeti temeljne strukture skupa podataka, kako bi se slične stanice smjestile zajedno u niskodimenzionalni prostor [60].

Ugradanja pomoću t-distribuiranog stohaičkog susjeda (*eng. t-SNE*), nelinearna je tehnika ugrađivanja za vizualizaciju podataka pojedinačnih stanica. Ovim algoritmom kreiraju se 2D ili 3D dijagrami u kojima svaka točka predstavlja stanicu, a udaljenost između točaka odražava koliko su stanice slične u smislu ekspresije gena [43, 44].

Seurat također omogućuje daljnju analizu, kao što je identificiranje genskih markera za svaki klaster, gena koji su specifično izraženi u određenom tipu stanica ili klasteru i pružaju uvid u samu biologiju stanice i njezine funkcije [60].

1.4 STROJNO UČENJE I PREDIKTIVNI MODELI ZA ANALIZU MULTIDIMENZIONALNIH PODATAKA SINGLE CELL RNA-SEQ

Strojno učenje je grana umjetne inteligencije koja se bavi razvojem algoritama i statističkih modela koje računala koriste za učenje iz podataka i donošenje odluka bez eksplicitnog programiranja. Osnovna ideja je da računala mogu prepoznati obrasce i prilagoditi donošenje odluka na temelju prethodnih iskustava. Postoje četiri glavne kategorije strojnog učenja: nadzirano učenje, nenadzirano učenje, polunadzirano učenje i učenje pojačanjem (Slika 9) [47].

Nadzirano učenje uključuje rad s označenim podacima (krajnja točka, ciljna varijabla), pri čemu je algoritmu unaprijed poznata ispravna izlazna vrijednost za svaku ulaznu vrijednost. Unutar nadziranog učenja postoje dva glavna tipa problema: klasifikacija i regresija [47].

Klasifikacija je proces predviđanja diskretne kategorije ili klase za dani ulazni podatak (npr. pozitivni, negativni ishod). Algoritmi klasifikacije koriste se kada je izlazna varijabla kategorizirana u određene klase. Primjeri uključuju prepoznavanje rukopisa (koje slovo ili broj je napisano), dijagnozu bolesti (ima li pacijent određenu bolest ili ne) i filtriranje neželjene pošte (je li e-mail spam ili ne). Uobičajni algoritmi za klasifikaciju uključuju logističku regresiju, stabla odluke, algoritam nasumičnih šuma (*eng. Random Forest*), k-algoritam najbližih susjeda (k-NN) i stroj vektornih potpora SVM (*eng. Support Vector Machine*) [47].

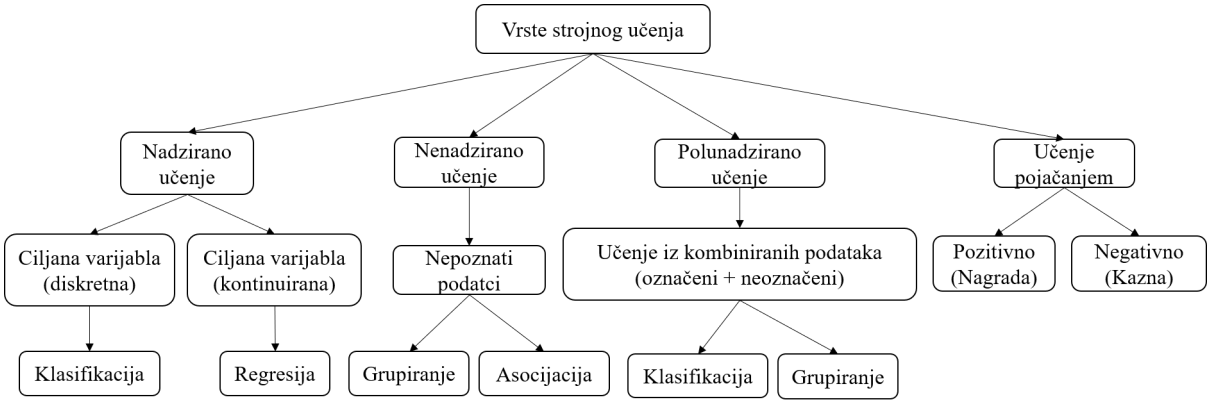
Regresija je proces predviđanja kontinuirane, kvantitativne izlazne varijable. Algoritmi regresije koriste se kada je izlazna varijabla brojčana. Primjeri uključuju predviđanje cijena nekretnina na temelju karakteristika kuće, procjenu količine padalina koja će pasti u određenom razdoblju i predviđanje prodaje proizvoda na temelju marketinških strategija [47]. Algoritmi regresije uključuju linearnu regresiju, polinomnu regresiju, Ridge regresiju i Lasso regresiju.

Nenadzirano učenje radi s neoznačenim podacima i cilj mu je otkriti skrivene obrasce ili strukture u podacima koji se ne vide ako je veliki broj dimenzija podataka (varijabli). Ključne tehnike uključuju grupiranje (eng. *clustering*), gdje se podaci grupiraju u klastere na temelju sličnosti, i reduciranje dimenzionalnosti, koje pojednostavljuje složene podatke u manje dimenzionalni oblik [47].

Polunadzirano učenje kombinira elemente nadziranog i nenadziranog učenja, jer koristi označene i neoznačene podatke. Ova je metoda korisna kada postoji mali broj označenih podataka i veliki broj neoznačenih podataka. Cilj je postići bolje rezultate predikcije nego što bi se postiglo korištenjem samo označenih podataka. Primjene uključuju strojno prevođenje, otkrivanje prijevara, označavanje podataka i klasifikaciju teksta [47].

Učenje pojačanjem omogućuje softverskim agentima i strojevima da automatski ocjenjuju optimalno ponašanje u određenom kontekstu kako bi poboljšali svoju učinkovitost. Ova metoda temelji se na nagradama i kaznama, s ciljem maksimiziranja ukupne nagrade ili minimiziranja rizika. Koristi se za obuku modela umjetne inteligencije AI (*eng. Artificial Inteligence*), tehnologije koja simulira ljudsku inteligenciju u složenim sustavima poput robotike, autonomne vožnje, proizvodnje i logistike, ali nije prikladan za rješavanje osnovnih ili jednostavnih problema [47].

Svaka od ovih tehnika igra važnu ulogu u izgradnji učinkovitih modela za različite primjene, ovisno o prirodi podataka i željenom ishodu.



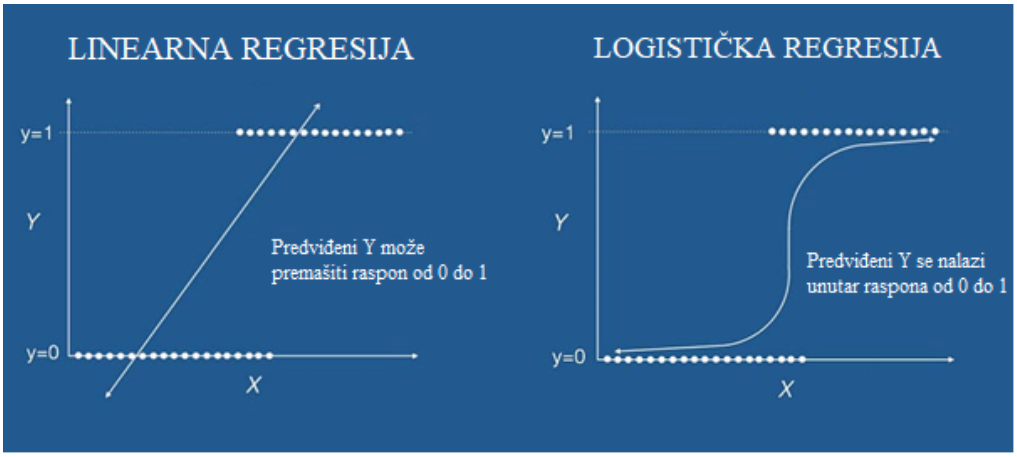
Slika 9 Kategorije strojnog učenja (slika preuzeta i dorađena sa: Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x)

Strojno učenje transformira mnoge industrije, omogućujući pametnije sustave koji mogu poboljšati učinkovitost, smanjiti troškove i pružiti bolje korisničko iskustvo. Svake godine, strojno učenje se širi na brojna istraživačka područja kao što su bioinformatika [48], biokemija [49], meteorologija [50], medicina [51], ekonomske znanosti [52], kemoinformatika [53], robotičke znanosti [54] i klimatologija [55].

1.4.1 LOGISTIČKA REGRESIJA

Logistička regresija je popularna metoda strojnog učenja koja se koristi za binarnu klasifikaciju, što znači da pomaže u predviđanju jedne od dvije moguće klase. Iako se zove regresija, ova metoda se prvenstveno koristi za rješavanje problema klasifikacije. Temelji se na logit funkciji, poznatoj i kao sigmoidna funkcija, koja transformira linearne kombinacije ulaznih varijabli u vjerojatnosti (Slika 10) [56]. Logistička se regresija na primjer, može koristiti za predviđanje hoće li neki pacijent imati određenu bolest (da ili ne) na temelju različitih medicinskih pokazatelja. Prednost ove metode je jednostavnost implementacije i interpretacije rezultata, kao i efikasnost u slučaju kada su odnosi između varijabli linearni.

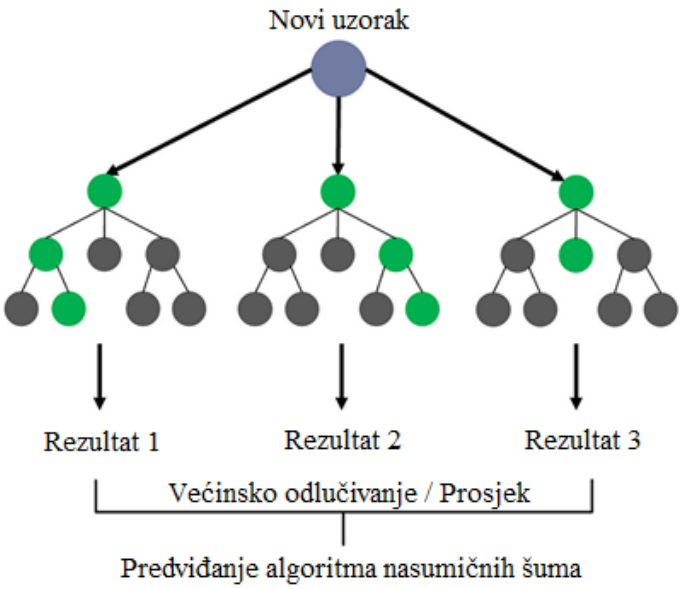
Međutim, logistička regresija može se pretjerano prilagoditi (eng. *overfit*) visokodimenzionalnim skupovima podataka, iako se to može ublažiti primjenom regularizacijskih tehnika (L1 i L2). Iako je logistička regresija nazvana regresija, češće se koristi za klasifikacijske probleme. Glavni nedostatak ove metode je pretpostavka linearne povezanosti između nezavisnih varijabli i logaritamskih omjera (eng. *log-odds*), što može ograničiti njezinu primjenu u složenijim slučajevima [57].



Slika 10 Slikovni prikaz grafova linearne i logističke regresije. (slika preuzeta i dorađena sa: https://towardsdatascience.com/introduction-to-logistic-regression-66248243c148)

1.4.2 ALGORITAM NASUMIČNIH ŠUMA

Algoritam nasumičnih šuma (*eng. Random Forest*) je napredna tehnika strojnog učenja koja se koristi za klasifikaciju i regresiju. Ova metoda se temelji na ansamblu modela kombinirajući rezultate više stabala odluke kako bi se poboljšala ukupna točnost predikcija [58]. Svako stablo u šumi trenira se na različitim podskupovima podataka koristeći metodu *bootstrap* agregacije (*eng. bagging*) i slučajni odabir značajki, čime se smanjuje problem preprilagodbe (*eng. overfitting*) i povećava robusnost modela. Algoritam nasumičnih šuma koristi "paralelno ansambliranje" (Slika 11), gdje se više stabala trenira paralelno na različitim podskupovima podataka, a konačni rezultat dobiva se glasanjem većine (za klasifikaciju) ili izračunavanjem prosjeka (za regresiju). Ova tehnika je posebno učinkovita u radu s velikim količinama podataka i raznovrsnim vrstama ulaznih varijabli, te se može primjeniti na razne zadatke poput klasifikacije slika, dijagnostike bolesti ili predikcije tržišnih trendova. Prednosti ovog algoritma uključuju njegovu sposobnost da se nosi s nedostajućim podacima, visoku točnost predikcija, i bolju kontrola nad preprilagodbom u usporedbi s modelima temeljenim na pojedinačnim stabalima odluke [59].



Slika 11 Predviđanje algoritma nasumičnih šuma.(Slika preuzeta i dorađena sa:https://medium.com/@roiyeho/random-forests-98892261dc49)

1. CILJ RADA

Hipoteza ovog rada jest da nam podatci RNA sekvenciranja pojedinačnih stanica, u kombinaciji sa strojnim učenjem, omogućuju identifikaciju ključnih gena i signalnih puteva povezanih s progresijom Alzheimerove bolesti, te mogu doprinjeti preciznijoj klasifikaciji bolesti i otkrivanju novih genetskih markera.

Glavni cilj ovog diplomskog rada je istražiti mogućnost klasifikacije statusa Alzheimerove bolesti korištenjem Seurat paketa i tehnika strojnog učenja primjenjenih na podatcima RNA sekvenciranja pojedinačnih stanica.

Glavni cilj uključuje ostvarivanje specifičnih ciljeva:

1. Bioinformatička obrada setova podataka sekvenciranja pojedinačnih stanica iz entorinalnog korteksa i superiornog frontalnog girusa ljudskog mozga *post mortem* različitih stadija Alzheimerove bolesti.
2. Analiza diferencijalne ekspresije gena za identifikaciju statistički značajnih diferencijalno eksprimiranih gena, bez obzira na tip stanica kroz različite stadije Alzheimerove bolesti prema Braak klasifikaciji statusa bolesti.
3. Razvoj modela strojnog učenja za klasifikaciju statusa bolesti na temelju identificiranih diferencijalno eksprimiranih gena.
4. Istraživanje potencijala identificiranih diferencijalno eksprimiranih gena kao novih biomarkera za rano otkrivanje i praćenje progresije Alzheimerove bolesti.

2. MATERIJALI I METODE

2.1. MATERIJALI

2.1.1. CELLXGENE BAZA PODATAKA

CELLXGENE [88] je internetski pretraživač baza podataka sekvenciranja pojedinačnih stanica sa kojeg su skinuti podaci RNA sekvenciranja pojedinačnih stanica dobivenih pomoću platforme 10X Genomics Chromium. Analizirali smo stanice mozga pacijenata koji su bolovali od Alzheimerove bolesti *post mortem*: 42,528 tisuća stanica entorinalnog korteksa i 63,608 tisuća stanica superiornog frontalnog girusa. Ovi podaci objavljeni su u radu "*Molecular characterization of selectively vulnerable neurons in Alzheimer's disease*" (Leng K et al., Nature Neuroscience, 2021).

2.1.2. TKIVO LJUDSKOG MOZGA *post mortem*

Leng K. i suradnici ekstahirali su jezgre stanica *post mortem* moždanog tkiva u području entorinalnog korteksta na razini srednjeg unkusa i iz gornjeg frontalnog girusa (SFG) na razini prednje komisure (Brodmannovo područje 8). Ukupno je kohorta uključivala 10 osoba muškog spola, genetskog profila APOE ε3/ε3, podjeljenih u 3 skupine, odnosno Braakove stadije 0 (skupina 1-3), 2 (skupina 4-7) i 6 (skupina 8-10) [65].
Knjižice pojedinačnih stanica dobivene su pomoću Chromium Single Cell 3' Reagent Kits v2 tvrtke 10X Genomics [65]. Nakon izolacije i obrnute transkripcije, cDNA je sintetizirana i barkodirana u kapljicama, omogućujući jedinstvenu identifikaciju svake jezgre. cDNA je zatim umnožena lančanom reakcijom polimeraze (PCR), a biblioteke za sekvenciranje su pripremljene fragmentiranjem cDNA i dodavanjem adaptera za skevenciranje. Sekvenciranje visoke propusnosti (eng. *New Generation Sequencing*) provedeno je na platformi Ilumina NovaSeq [65].

2.2. METODE

2.2.1. SEURAT

2.2.1.1. ANALIZA PODATAKA

Za analizu podataka korišten je Seurat paket [60]. Podaci su preuzeti iz CELLXGENE baze podataka, te učitani u programskom jeziku R i obrađeni kao Seurat objekt.
Nakon sekvenciranja podaci su usklađeni i grupirani za identifikaciju različitih tipova stanica i staničnih subpopulacija. Unakrsnim usklađivanjem osiguralo se smanjenje tehničkih varijablinosti i osiguralo da grupiranje odražava biološke razlike, a ne artefakte.

Iduće je provedena klasifikacija tipova stanica na temelju ekspresije gena-markera, u stanične podtipove, kao što su ekscitatorni neuroni,inhibicijski neuroni, astrociti, oligodendrociti, mikroglija i endotelne stanice [65].
Podaci su zatim klasificirani i identificirani prema Braak stadijima 0 i 6, pri čemu su podaci iz Braak stadija 0 korišteni kao kontrolna skupina, a podaci iz Braak stadij 6 kao eksperimentalna skupina.
Prvo je analizirana regija entorinalnog korteksa, a nakon toga regija superiornog frontalnog girusa.

Nakon normalizacije podataka, korištena je metoda redukcije ne linearne dimenzije - t-distribuirano stohaističko susjedno ugrađivanje (t-SNE, eng. *t-distributed Stochastic Neighbor Embedding*) za vizualizaciju visoko-dimenzionalnih podataka u dvodimenzionalnom prostoru. t-SNE je korišten kao prikaz distribucije stanica, omogućujući usporedbu između kontrolne i eksperimentalne (Braak stadij 6) skupine.

t-SNE je nelinearna tehnika redukcije dimenzionalnosti koja je pogodna za analizu velikih i složenih skupova podataka, kao što su podaci sekvenciranja RNA pojedinačnih stanica. Ova metoda izrađuje model sličnosti među podatcima u visoko-dimenzionalnom prostoru i zatim ih projicira u nisko dimenzionalni prostor na način da slični uzorci ostanu blizu jedan drugome, dok različite uzorke međusobno udaljava.

3.2.1.2 IDENTIFIKACIJA DIFERENCIJALNO EKSPRIMIRANIH GENA (DEG)

Prvi korak za analizu diferencijalno eksprimiranih gena bio je podjeliti stanice Braak stadija 0 i Braak stadija 6 unutar odabranih regija mozga, te nakon toga uz pomoć funkcija *subset* i *merge* spojiti podskupove za izravnu usporedbu stadija.

Za identifikaciju (DEG) između Braak stadija 0 i Braak stadija 6 korištena je funkcija *FindMarkers* iz Seurat paketa [60]. Funkcija uspoređuje ekspresiju gena između dviju skupina stanica.
Nakon toga su rezultati filtrirani prema sljedećim kriterijima:

- Abs(avg_log2FoldChange) > 3
- p_val_adjusted < 0.001

Log2 fold change – logaritamska promjena genske ekspresije baze 2 (log2FC) je statistička mjera koja opisuje omjer promjene ekspresije gena između dviju različitih skupina stanica ili uvjeta, odnosno govori kolika je razlika između dva uzorka.
avg_log2FC – prosječna log2 promjena ekspresije predstavlja prosječnu log2 promjenu ekspresije gena između dvije skupine stanica koje uspoređujemo. Pozitivne vrijednosti (*eng. Upregulated*) pokazuju da je ekspresija gena povećana, dok negativne (*eng. Downregulated*) vrijednosti pokazuju da je ekspresija gena snižena.
abs(avg_log2FC) – apsolutna vrijednost prosječne log2 promjene ekspresije uklanja znak (pozitivan ili negativan) iz log2 promjene, ostavljajući samo veličinu promjene, omogućujući fokusiranje na gene koji pokazuju najveće promjene u ekspresiji, bilo da je ta promjena povećanje ili smanjenje ekspresije.

Prag *abs(avg_log2FC)> 3* izabran je kako bi se identificirali geni s vrlo značajnim promjenama u ekspresiji, čime se osigurava da su identificirani geni biološki zaista značajni i relevantni za daljnju analizu.
P-vrijednost je vjerojatnost da se opažena razlika u ekspresiji gena između dvije skupine pojavila slučajno, pod pretpostavkom da ne postoji stvarna razlika između tih skupina. Rezultat je statistički značajniji što je p-vrijednost niža.
Prilagođena p-vrijednost (p-value adjusted) je statistička mjera koja je prošla kroz proces korekcije p-vrijednosti za smanjenje rizika lažno pozitivnih rezultata te se koristi u analizi diferencijalne ekspresije gena kako bi se procijenila značajnost rezultata uzimajući u obzir problem višestrukih usporedbi. Korištenje *praga p_val_adj < 0.001* omogućuje identifikaciju gena čija je statistička značajnost vrlo visoka, što dodatno osigurava pouzdanost rezultata.

Nakon analize podataka za EC regiju, postupak je ponovljen na drugom setu podataka, odnosno SFG regiji.

Nadalje, provedena je usporedba DEG između EC i SFG kako bismo identificirali gene koji se obično različito izražavaju u obje regije.

3.2.1.3 VIZUALIZACIJA REZULTATA

Najznačajniji diferencijalno eksprimirani geni sortirani su prema prilagođenoj p-vrijednosti i apsolutnoj log2 promjeni ekspresije (p-val _adj. <0.001, abs(avg_log2FC)> 3) i kasnije vizualizirani koristeći Seurat paket. Vizualizacije omogućuju grafički prikaz ekspresije gena u različitim Braakovim stadijima, te pružaju uvid u njihovu moguću ulogu u patogenezi Alzheimerove bolesti.
Korištene metode vizualizacija iz paketa Seurat su:*Dim plot*, *Dot Plot* i *Feature Plot*.

DimPlot je funkcija iz Seurat paketa, koja koristi t-SNE metodu za smanjenje dimenzionalnosti i prikaz stanica u dvodimenzionalnom prostoru, a plot je korišten za prikaz raspodjele stanica iz entorinalnog korteksa prema Braakovim stadijima (Braak 0 i 6).

Dot Plot (točkasti dijagram) je metoda koja prikazuje ekspresiju odabranih gena u različitim skupinama stanica, kombinirajući informacije o intenzitetu ekspresije i postotku stanica koje eksprimiraju određeni gen. Svaki krug (*eng. dot*) predstavlja ekspresiju jednog gena u određenom klasteru stanica. Veličina kruga pokazuje postotak stanica u klasteru koje eksprimiraju taj gen, dok nijansa boje označava razinu ekspresije (svjetlija boja znači nižu ekspresiju, a tamnija boja višu ekspresiju).

Feature Plot je vizualizacija koja prikazuje ekspresiju odabranih gena u dvodimenzionalnom prostoru, koristeći t-SNE redukciju dimenzionalnosti. Svaki graf prikazuje pojedinačnu stanicu, obojanu prema razini ekspresije određenog gena.

3.2.2 STROJNO UČENJE (MACHINE LEARNING)

3.2.2.1 PRIKUPLANJE I PRETHODNA OBRADA PODATAKA

Nakon analize DEG uz pomoć Seurata, izveli smo klasifikaciju gena korištenjem strojnog učenja.

Normalizirana matrica ekspresije izdvojena je iz Seurat objekta, transponirana i pretvorena u podatkovni okvir za daljnju analizu. Metapodaci, uključujući Braakovu fazu i informacije o tipu stanica, također su izdvojeni i spojeni s transponiranim podatkovnim okvirom.

Kako bi se smanjila buka i složenost obrade, geni su filtrirani prema specifičnim kriterijima ekspresije:

- Geni s više od 99% ili manje od 1% ekspresije u cijelom skupu podataka uklonjeni su kako bi se isključili oni koji vjerojatno nisu povezani s bolešću.
- Geni koji pokazuju više od 95% ekspresije u bilo kojem tipu stanica ili pojedincu isključeni su jer su smatrani previše specifičnima.
- Geni s manje od 5% ekspresije kod pacijenata uklonjeni su jer su smatrani nepovezanima s bolešću.
- Geni s visokom korelacijom (Pearsonov koeficijent korelacije > 0,80) s drugim genima također su isključeni kako bi se izbjegla redundantnost.

Za analizu su odabrani samo uzorci u Braakovoj fazi 0 (zdrave kontrole) i Braakovoj fazi 6 (uznapredovali oblik Alzheimerove bolesti).

3.2.2.2 LOGISTIČKA REGRESIJA

Za klasifikaciju uzoraka u kategorije "Zdravi" i "Bolesni" korištena je logistička regresija temeljena na podacima ekspresije gena. Primjenjena su dva pristupa: model predikcije bolesnih i zdravih stanica te klasičan regresijski model putem funkcije (lme4 paket) za dobivanje p-vrijednosti.

Ime4 je paket iz programskog jezika R i koristi se za prilagođavanje linearnih modela koji uključuju fiksne (ekspresija gena) i nasumične učinke (varijabilnosti među uzorcima). Ovdje se koristi kod regresijskog modela za procjenu genskih koeficijenata i p-vrijednosti.

Skup podataka podijeljen je na trening (80%) i testni (20%) skup pomoću stratificiranog uzorkovanja.
Proces treniranja proveden je uz pomoć peterostruke unakrsne validacije, optimizirane za površinu ispod krivulje karakteristike prijemnika (ROC-AUC) pomoću paketa Caret u R-u. Učinkovitost modela procijenjena je pomoću pokazatelja uključujući točnost, osjetljivost, specifičnost i ROC-AUC.

Logistički regresijski modeli prilagođeni su za svaki gen pojedinačno, pri čemu je Braakova faza upotrebljavana kao zavisna varijabla. Izračunate su p-vrijednosti za koeficijente gena, a Bonferronijeva metoda primjenjena je za korekcije višestrukog testiranja.

3.2.2.3 ALGORITAM NASUMIČNIH ŠUMA

Za klasifikator nasumičnih šuma (RF, eng. *Random Forest*) korišten je paket SKLearn iz programskog jezika Python koji pruža alate za nadzirano učenje, nenadzirano učenje i evaluacijske modele, kao i alate za skaliranje podataka. RF je odabran zbog svoje sposobnosti da učinkovito obrađuje složene, visokodimenzionalne podatke, te zbog otpornosti na prekomjerno prilagođavanje modela. Podaci su podijeljeni na trening (80%) i testni (20%) skup uz stratifikaciju. Optimizacija hiperparametara provedena je uporabom *RandomizedSearchCV* s peterostrukom unakrsnom validacijom. F1 mjera korištena je kao evaluacijski parametar za uravnoteženje preciznosti i odziva. Nakon optimizacije, učinkovitost RF klasifikatora procijenjena je na testnom skupu korištenjem standardnih parametara, uključujući matricu konfuzije, izvještaj o klasifikaciji i ROC-AUC. Informacije o važnosti pojedinih značajki dobivene su pomoću treniranog modela, a potom su uspoređene rezultatima logističke regresije, kako bi se identificirali geni koji se dosljedno ističu po važnosti.

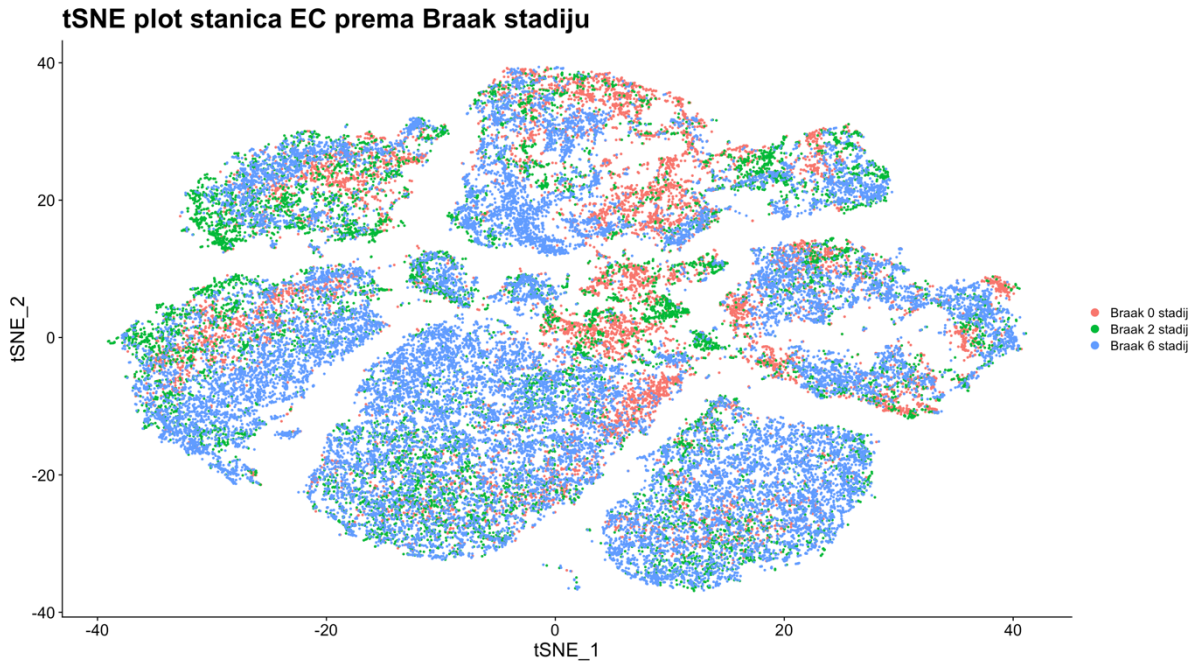
3. REZULTATI

3.1. SEURAT

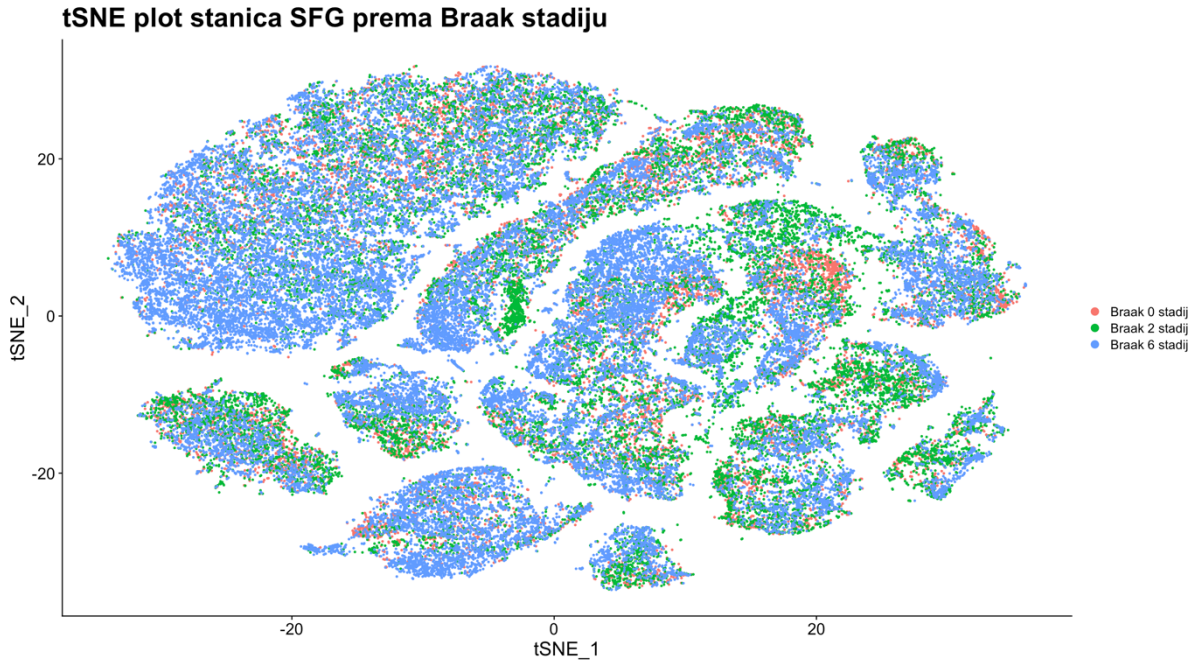
3.1.1. tSNE VIZUALIZACIJA STANICA I STANIČNIH TIPOVA

Prvo dio istraživačkog djela bio je okarakterizirati stanični sastav dviju moždanih regija – entorinalnog korteksa i superiornog frontalnog girusa u različitim fazama Braak stadija koji se koristi za klasifikaciju Alzheimerove bolesti. Kako bi istražili prostorni odnos i tendenciju grupiranja stanica u klastere na temelju Braak stadija izradili smo tSNE dijagram koji ilustrira stanice EC i SFG regija u dvodimenzionalnom prostoru (slika 12, 13).

Stanice su označene bojama na temelju njihove klasifikacije različitih Braak stadija: Braak 0 – crveno (zdravo), Braak 2 – zeleno (rani stadij AB), Braak 6 – plavo (uznapredovali stadij AB).
Prvo smo napravili tSNE na stanicama entorinalnog korteksa (EC, slika 12), a zatim na stanicama superiornog frontalnog girusa (SFG, slika 13).



Slika 12 Grafički prikaz raspodjele stanica entorinalnog korteksa kroz različite faze Braak stadija

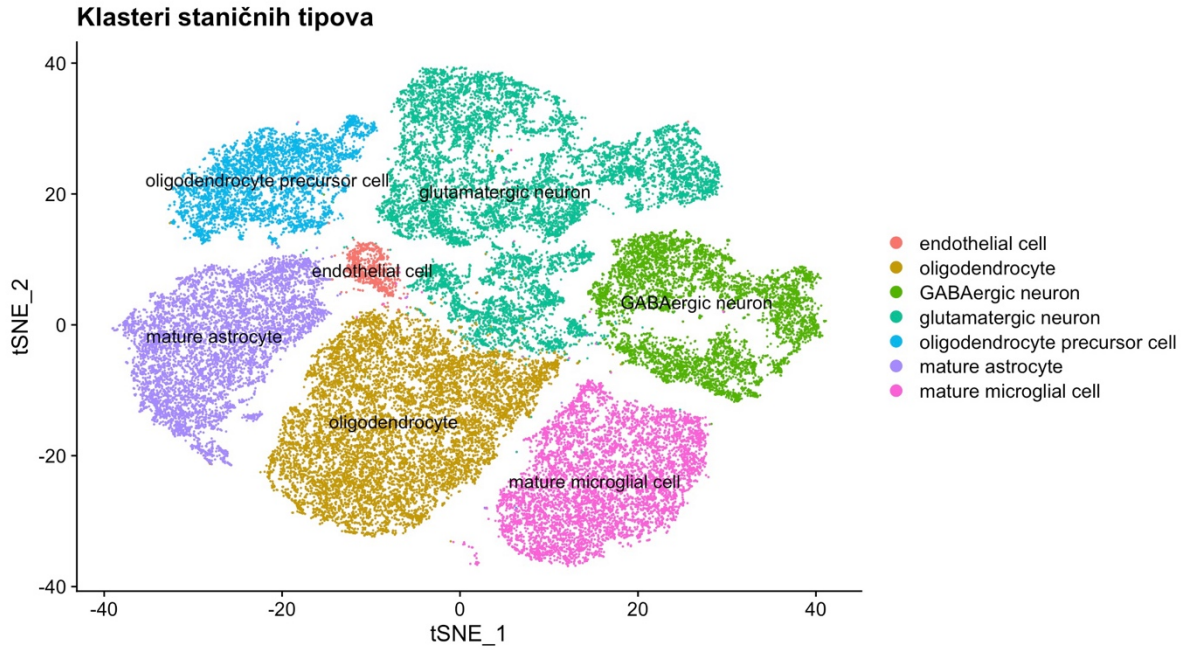


Slika 13 Grafički prikaz raspodjele stanica superiornog frontalnog girusa kroz različite faze Braak stadija. Dijagram je napravljen u programskom jeziku R pomoću Seurat paketa.

Stanice iz ranijih stadija (Braak 0 i 2) pokazuju kompaktnije grupiranje, što ukazuje na manju varijabilnost stanica, dok stanice u Braakovom stadiju 6 koji predstavlja najuznapredovaliji stadij patologije Alzheimerove bolesti, pokazuju šire rasprostranjenje, što potencijalno ukazuje na povećanu heterogenost staničnih populacija kako bolest napreduje.

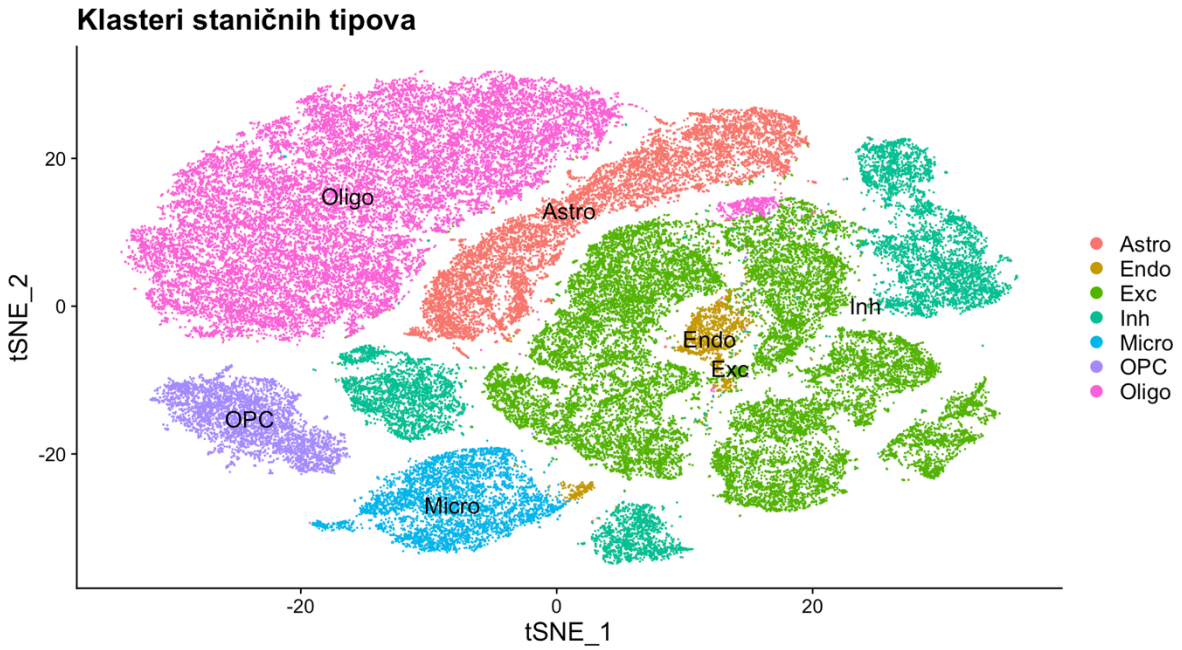
Nakon toga, generirali smo graf koji ilustrira različite klastere stanica unutar EC i SFG regija mozga pri čemu svaki klaster predstavlja jedan određeni tip stanica (slika 14, 15).
Jasne granice između klastera ukazuju na različite obrasce ekspresije gena karakteristične za svaki tip stanica, ali specifične za regiju mozga u kojoj se nalaze.

EC regija, dio medijalnog temporalnog režnja važnog za pamćenje i orijentaciju u prostoru pretežito sadrži zrele astrocite, zrelu mikrogliju i neurone – ekscitatorne (Glutamatergičke) i inhibitorne (GABAnergične) koji su važni za sinaptičku podršku i regulaciju neurotransmitera u područjima mozga povezanih s pamćenjem (slika 14).



Slika 14. Klasteri tipova stanica entorinalnog korteksa

SFG regija, koja je dio frontalnog režnja odgovornog za kognitivne i motoričke funkcije, te radnu memoriju pokazuje veći udio oligodendrocita, (stanica koje proizvode mijelin), prekursorskih stanica oligodendrocita, astrocita i pretežito ekscitatornih neurona što je u skladu s potrebom za učinkovitom komunikacijom i brzim prijenosom aksonskog potencijala u frontalnom režnju (slika 15).



Slika 15. Klasteri tipova stanica superiornog fronalnog girusa

3.1.2. ANALIZA DIFERENCIJALNO EKSPRIMIRANIH GENA

1. REZULTATI ANALIZE DEG ENTORINALNOG KORTEKSA:

   Identificirali smo 1939 značajno diferencijalno eksprimiranih gena između Braak stadija 0 i 6 u entorinalnom korteksu.
   Geni su filtrirani korištenjem strogih kriterija (avg_log2FC > 3 i p_val_adj < 0,001), osiguravajući da se u obzir uzmu samo najznačajnije promjene.

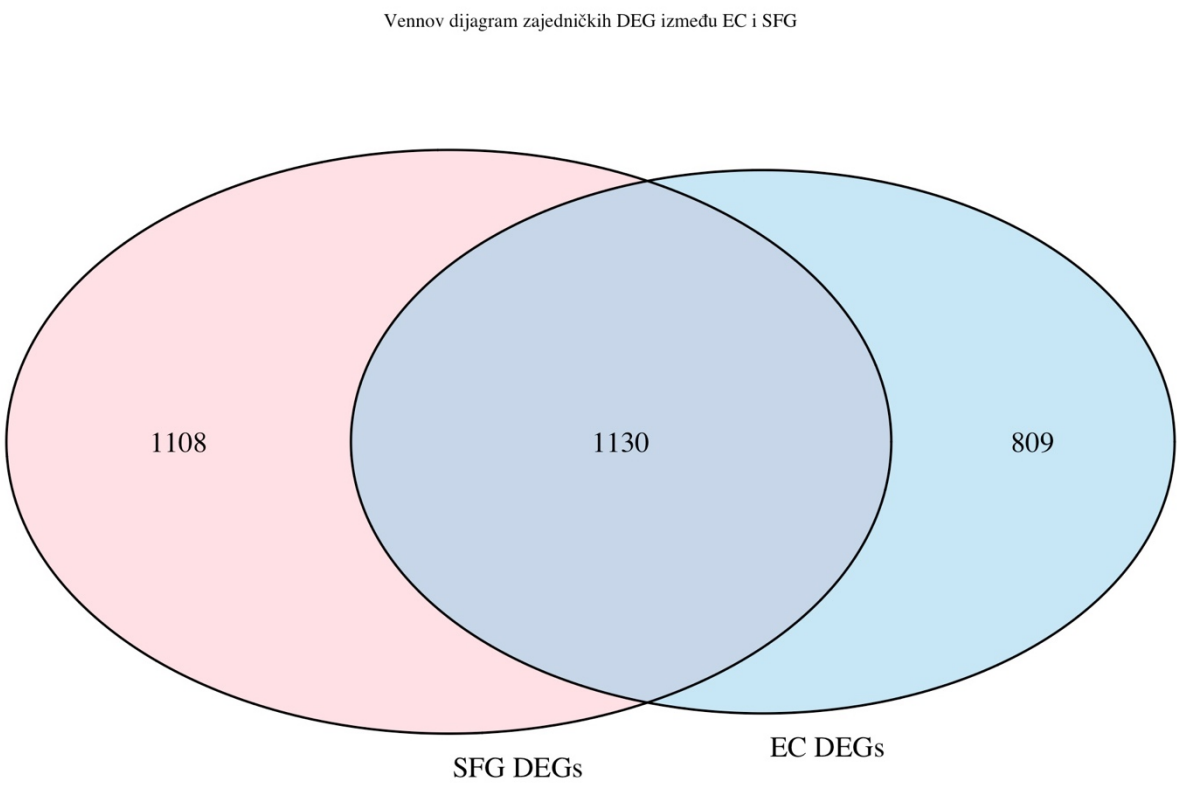
2. REZULTATI ANALIZE DEG SUPERIORNOG FRONTALNOG GIRUSA

   Slično, utvrdili smo je da je 2238 gena značajno diferencijalno eksprimirano između Braak stadija 0 i Braak stadija 6 u superiornom frontalnom girusu.
   Za prag značajnosti također su korišteni isti strogi kriteriji.

3. REZULTATI ANALIZE ZAJEDNIČKIH DEG DVIJU REGIJA

   Ukupno 1130 zajedničkih diferencijalno eksprimiranih gena je identificirano u EC i SFG.

Za vizualizaciju dobivenih rezultata napravili smo Vennov dijagram, (slika 16) koji obuhvaća DEG samo SFG regije (2230 - 1130), DEG samo EC regije (1939 -1130) i presjek zajedničkih gena u obje regije (1130).



Slika 16 Vennov dijagram zajedničkih diferencijalno eksprimiranih gena EC i SFG regije.

3.2. STROJNO UČENJE

Drugi dio istraživanja, nakon analize diferencijalno eksprimiranih gena uz pomoć Seurat paketa, bio je pomoću tehnika strojnog učenja preciznije klasificirati diferencijalno eksprimirane gene za pouzdanije rezultate.

3.2.1 DISTRIBUCIJA PODATAKA I POČETNA OBRADA

Analizirali smo dvije skupine podataka: uzorci triju zdravih osoba (EC1-EC3 za skup podataka 1 i SFG1-SFG3 za skup podataka 2) i triju osoba dijagnosticiranih s Alzheimerovom bolešću (EC8-EC10 za skup podataka 1 i ID - SFG8-SFG10 za skup podataka 2).
Distribucija vrsta stanica unutar ovih skupova podataka sažeta je u sljedećoj tablici:

Tablica 1. Tablični prikaz distribucije tipova stanica unutar dva skupa podataka

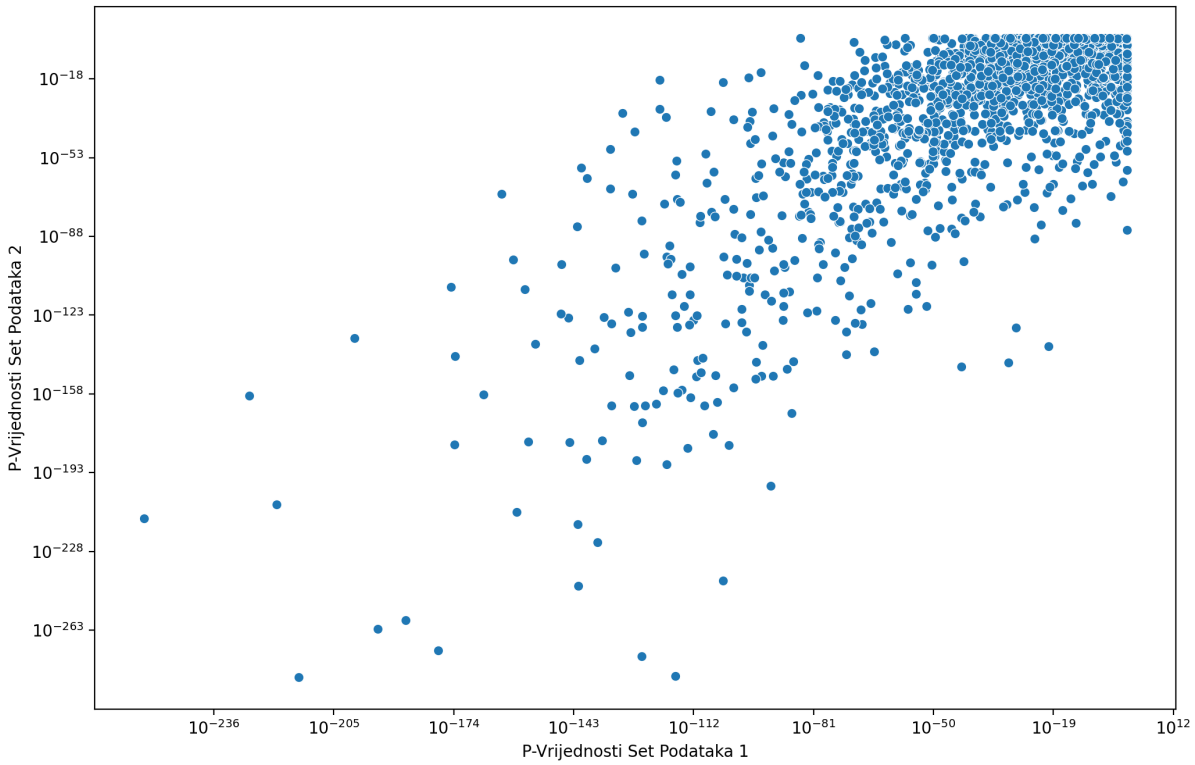
Tip stanica	Broj stanica (Skup podataka 1)	Broj stanica (Skup podataka 2)
GABAergički neuron	4536	5502
Endotelna stanica	406	776
Glutamatergični neuron	8123	13173
Zreli astrocit (Skup 1) / Astrocit (Skup 2)	4416	5393
Zrela mikroglijalna stanica (skup 1) / mikroglijalna stanica (skup 2)	4746	3506
Oligodendrociti	7590	13593
Prekursorska stanica oligodendrocita	2322	2071

Početni korak filtriranja gena implementiran je kako bi se uklonili geni koji su pokazivali specifične obrasce ekspresije za pojedine osobe i nisu bili povezani s patološkim stanjem bolesti. Nakon filtriranja, skup podataka 1 (EC regija) obuhvaćao je 3,579 gena, dok je skup podataka 2 (SFG regija) obuhvaćao 3,998 gena.
Identificirano je 2,981 zajedničkih gena između ova dva skupa.

4.2.1 REZULTATI LOGISTIČKE REGRESIJE

Za analizu smo koristili dva pristupa logističkoj regresiji: klasični model koji nam je dao informaciju o p-vrijednosti i predikcijski pomoću kojeg smo detektirali značajne gene.

Prvo smo primijenili klasičnu logističku regresiju na skupu podataka 1 (EC regija), koristeći paket lme4 za izračun p-vrijednosti.
Analiza je identificirala 2,763 gena kao značajne ($p < 0.05$), što predstavlja 77% ukupnih gena iz skupa podataka 1. Sličan pristup primijenjen je na skupu podataka 2 (SFG regija), gdje je 1,594 gena pokazalo značajnu povezanost ($p < 0.05$).
Slika 17 pokazuje distribuciju ovih podataka.



Slika 17 Prikaz distribucije p vrijednosti između skupova podataka 1 i 2. P vrijednosti su dobivene kao rezultat logističke regresije sa lme4 paketom u R-u.

Model logističke regresije evaluiran je korištenjem peterostruke unakrsne validacije, što je omogućilo procjenu diskriminacijske sposobnosti između zdravih i bolesnih uzoraka.

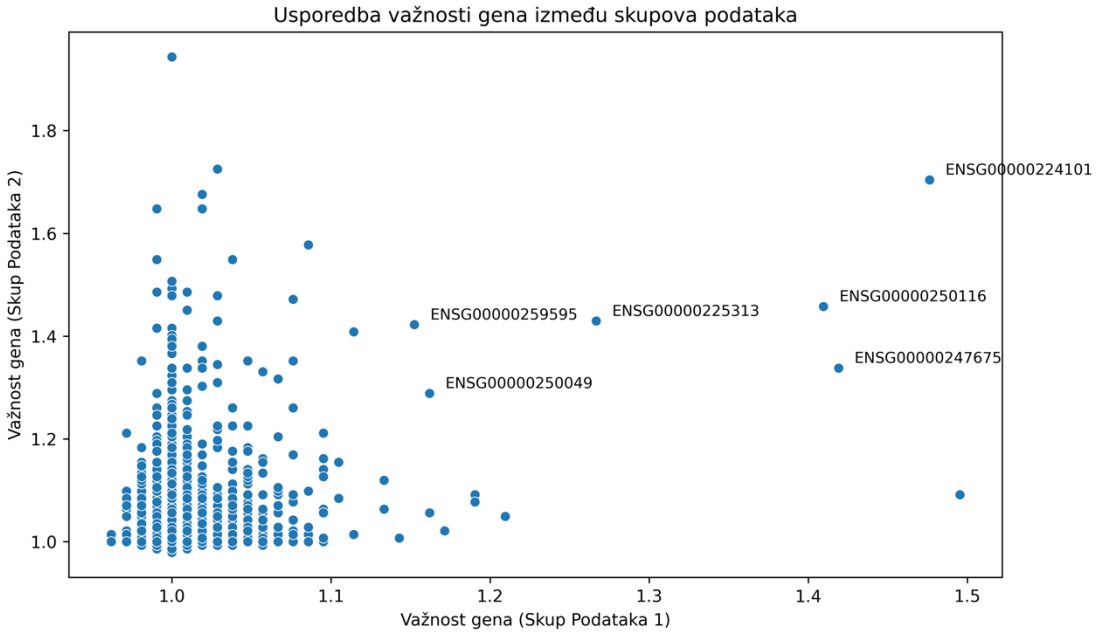
ROC-AUC, područje ispod krivulje radne karakteristike prijamnika (eng. *Receiver Operating Characteristic - Area Under the Curve)* metoda je za mjerenje izvedbe klasifikacijskog modela, analizirajući odnos između osjetljivosti (stvarna pozitivna stopa) i specifičnosti (lažna pozitivna stopa) za različite klasifikacijske pragove.
AUC vrijednost, koja se kreće od 0 do 1, predstavlja numeričku procjenu kvalitete modela, pri čemu vrijednost 1 označava savršenu klasifikaciju, dok vrijednost 0.5 označava nasumično pogađanje [89].

Na Skupu podataka 1, model je postigao prosječan ROC-AUC od 0.932 (93,2%). Na testnom skupu, model je pokazao visoku točnost od 0.984, intervala pouzdanosti (CI, eng. *Confidence Interval*) 95% (CI: 0.980–0.987), s osjetljivošću od 0.973 i specifičnošću od 0.988.
CI 0,980–0,987 pokazuje da je stvarna točnost modela vrlo vjerojatno između 98,0% i 98,7% s 95% pouzdanosti.
Slični rezultati postignuti su s modelom primijenjenim na skupu podataka 2, gdje je točnost iznosila 0.982, intevala pouzdanosti 95% (CI: 0.9792–0.9848), uz osjetljivost od 0.970 i specifičnost od 0.987.

Ovi parametri označavaju da je model vrlo dobar u klasifikaciji dvaju skupova sa uskim rasponom intervala pouzdanosti što ukazuje na stabilnu i pouzdanu točnost modela.

Za identifikaciju važnih gena iz modela, korišten je im4 paket, koji koristi permutaciju za procjenu utjecaja svake varijable na prediktivnost modela. Kako bi se izbjegla pristranost jednog modela, provedena je usporedba važnih varijabli između modela razvijenog na skupu podataka 1 i modela

razvijenog na skupu podataka 2 (Slika 18) gdje je uviđeno da se šest gena najviše istaknulo, te oni potencijalno mogu biti i značajni.



Slika 18 Grafički prikaz usporedbe utjecaja gena na model između skupa podataka 1 i 2. Scatterplot vizualizira važnost gena, prikazujući kako utjecaj svakog gena u modelu 1 korelira s njegovom važnosti u modelu 2.

4.2.2 REZULTATI ALGORITMA NASUMIČNIH ŠUMA

Kako bismo primijenili klasifikator modela nasumičnih šuma (eng. *Random Forest*) i procijenili njegova svojstva u klasifikaciji, prvo smo morali odabrati odgovarajuće hiperparametre. Za optimizaciju hiperparametara ispitano je nekoliko opcija, a najbolji rezultati postignuti su s sljedećim postavkama: broj stabala u šumi (n_estimators) postavljen je na 200, maksimalna dubina stabala (max_depth) na 10, maksimalan broj značajki razmatranih pri podjeli (max_features) na 'sqrt', minimalan broj uzoraka potreban za podjelu čvora (min_samples_split) na 10, minimalan broj uzoraka u listu (min_samples_leaf) na 2, te težina klase (class_weight) na 'balanced'.

Za skup podataka 1, model je postigao točnost unakrsne provjere od 0.9029 (90,29%).
Točnost modela konfuzijske matrice bila je 0,87, što označava 87% svih predviđanja (zdravih i bolesnih).
Preciznost za kontrolnu skupinu iznosila je 0,71, što ukazuje da je 71% uzoraka klasificiranih kao zdravo zapravo bilo zdravo, a za eksperimentalnu skupinu 0,98, što pokazuje da je 98% uzoraka klasificiranih kao bolesno zapravo bilo bolesno.
Osjetljivost ili odziv je za kontrolnu skupinu iznosio 0,96, što pokazuje da je model ispravno identificirao 96% svih pravih zdravih uzoraka, te eksperimentalnu 0,83, što pokazuje da je 83% svih pravih bolesnih uzoraka točno identificirano.
Za procjenu izvedbe klasifikacijskog modela koristili smo F1 parametar (balansira preciznost i osjetljivost) koji je iznosio 0.82 za zdrave i 0.90 za bolesne uzorke.
Slična izvedba zabilježena je u Skupu podataka 2, gdje je model postigao točnost unakrsne validacije od 0.8673 što pokazuje dosljednu izvedbu u različitim skupovima podataka, pojačavajući pouzdanost modela.

4.3 ANALIZA I VIZUALIZACIJA REZULTATA

Nakon rezultata dobivenih logističkom regresijom i modelom nasumične šume (eng. *Random Forest*) napravili smo presjek gena i dobili rezultat za 126 gena za koja su se tri od četiri modela složila da su bitni.

U programskom jeziku R smo koristili paket *biomaRt* za spajanje pripadajućih HGNC simbola gena na temelju liste Ensembl identifikatora gena (eng. *Ensembl Gene ID's*), sprecifično korištenjem baze podataka ljudskog genoma (eng. *hsapiens_gene_ensembl*).
Na temelju dostupnih HGNC simbola iz početnog skupa podataka, pomoću funkcije *getBM* pronašli smo više različitih tipova gena, te poslije spajanja podataka, filtrirani su samo oni geni koji su klasificirani kao geni koji kodiraju za protein. Tih 87 gena izdvojeno je za daljnju analizu.

Za daljnju analizu dobivenih gena koristili smo dostupnu literaturu, te baze podataka: *NIH- National Institutes of health* bazu podataka *NCBI Gene*, za informacije o ekspresiji i ulozi gena u organizmu; te *DisGeNET* bazu podataka u kojoj su objedinjene informacije iz znanstvene literature o povezanosti ljudskih gena i bolesti.

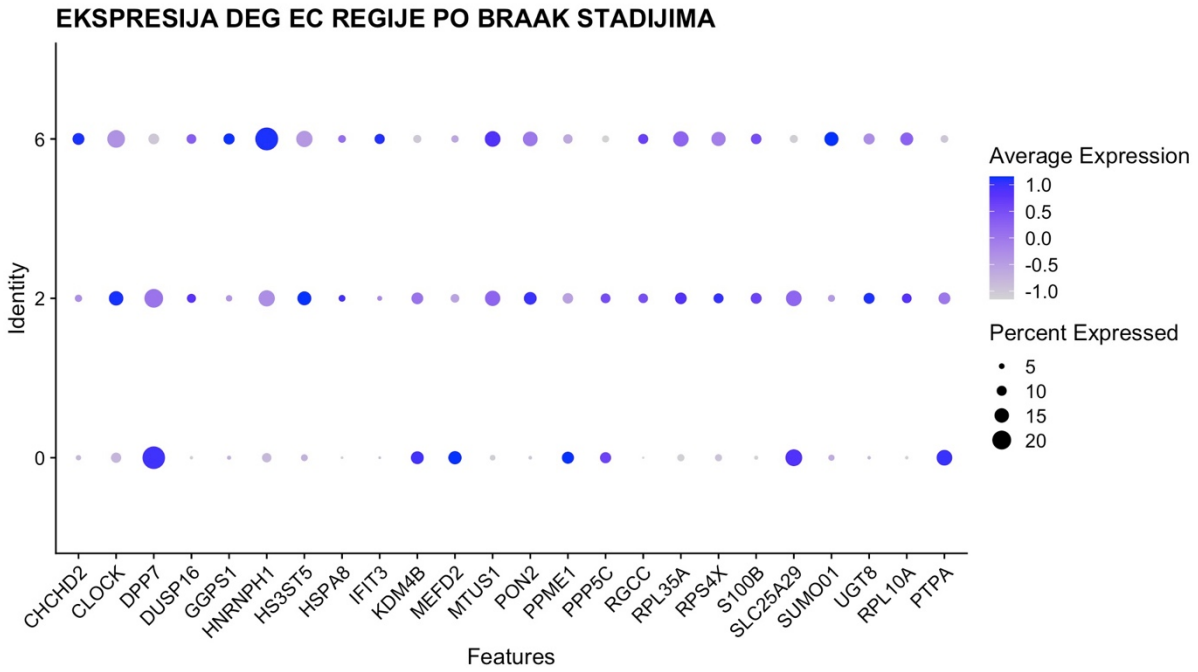
Komparativnom analizom prvo smo odabrali za gene koji je poznato da su povezani sa Alzheimerovom bolesti. Takvi geni su pokazivali najveću ekspresije. Od dobivenih 87 gena, 24 ih je odprije povezano s Alzheimerovom bolesti (Tablica 2).

Tablica 2. Tablični prikaz osnovnih podataka o 24 analiziranih DEG za poznatom asocijacijom sa AB

HGNC SIMBOL	ENSEMBEL GENE ID	PUNI NAZIV GENA	VRSTA GENA
CHCHD2	ENSG00000106153	NAMOTANA-ZAVOJNICA-HELIX-NAMOTANA ZAVOJNICA DOMENA 2	KODIRA ZA MITOHONDRIJSKI PROTEIN
CLOCK	ENSG00000134852	REGULATOR CIRKADIJALNOG RITMA	KODIRA ZA PROTEIN CIRKADIJALNOG RITMA
DPP7	ENSG00000176978	DIPEPTIDIL PEPTIDAZA 7	KODIRA ZA ENZIM SERINSKU PROTEAZU
DUSP16	ENSG00000111266	DVOSTRUKO SPECIF. FOSFATAZA 16	KODIRA ZA MAPK FOSFATAZU
GGPS1	ENSG00000152904	GERANILGERANIL DIFOSFAT SINTETAZA 1	KODIRA ZA PROTEIN GGPS1
HNRNPH1	ENSG00000169045	HETEROGENI NUKLEARNI RIBONUKLEOPROTEIN H1	KODIRA ZA PROTEINE hnRNP
HS3ST5	ENSG00000249853	HEPARAN-SULFAT-GLUKOZAMIN-3 SULFOTRANSFERAZA-5	KODIRA ZA PROTEIN HS3OST5
HSPA8	ENSG00000109971	PROTEIN TOPLINSKOG ŠOKA OBITELJI A (Hsp70) ČLAN 8	KODIRA ZA PROTEIN OBITELJI Hsp70
IFIT3	ENSG00000119917	INTERFERON INDUCIRANI PROTEIN S TETRATRICOPEPTIDNIM PONAVLJANJEM 3	KODIRA ZA IFIT3 PROTEIN
KDM4B	ENSG00000127663	LIZIN DEMETILAZA 4B	KODIRA ZA HISTONE H3 DEMETILAZE SA OBRNUTIM FUNKCIJAMA
MEF2D	ENSG00000116604	MIOCITNI POJAČIVAČ FAKTOR 2D	KODIRA ZA TRANSKRIPCIJSKI FAKTOR
MTUS1	ENSG00000129422	MIKROTUBULIMA POVEZAN SCAFFOLD PROTEIN 1	KODIRA ZA MIKROTUBLIMA POVEZAN PROTEIN 1S1
PON2	ENSG00000105854	PARAOKSONAZA 2	KODIRA ZA ČLANA PARAOKSONAZA 2 OBITELJI PROTEINA
PPME1	ENSG00000214517	PROTEIN FOSFATAZA METILESTERAZA 1	KODIRA ZA PME-1 U JEZGRI
PPP5C	ENSG00000011485	PROTEIN FOSFATAZA 5 KATALITIČKA PODJEDINICA	KODIRA ZA PP5C
RGCC	ENSG00000102760	REGULATOR STANIČNOG CIKLUSA	KODIRA ZA RGCC PROTEIN
RPL10A	ENSG00000198755	RIBOSOMALNI PROTEIN L10 A	KODIRA ZA 60S RIBOSOM
RPL35A	ENSG00000182899	RIBOSOMALNI PROTEIN L35A	KODIRA ZA 60S RIBOSOM
RPS4X	ENSG00000198034	RIBOSOMSKI PROTEIN S4 X-VEZAN	KODIRA ZA 40S RIBOSOM
S100B	ENSG00000160307	S100 KALCIJ VEZAJUĆI PROTEIN B	KODIRA ZA S100B PROTEIN
SLC25A29	ENSG00000197119	TRANSPORTER PROTEINA OBITELJI 25 ČLAN 29	KODIRA ZA MITOHONDRIJSKI TRANSPORTER PROTEIN
SUMO1	ENSG00000116030	MALI NALIK UBIKVITINU MODIFIKATOR 1	KODIRA ZA SUMO PROTEIN
UGT8	ENSG00000174607	UDP GLIKOZILTRANSFERAZA 8	KODIRA ZA UGT8 PROTEIN
PTPA	ENSG00000119383	PROTEIN FOSFATAZA 2 AKTIVATOR FOSFATAZE	KODIRA ZA PP2A PROTEIN

Kako bi vizualizirali ekspresijske vrijednosti 24 gena povezanih s Alzheimerovom bolesti prema Braak stadijima, generirali smo DotPlot u Seurat paketu (Slika 19, 20).

1. Geni povezani sa AB: Entorinalni korteks



Slika 19 Ekspresija 24 gena povezanih s AB u entorinalnom korteksu. X-os predstavljaju geni, na Y-osi navedeni su Braak stadiji predstavljajući progresiju bolesti od 0 do 6. Dijagram je napravljen u programskom jeziku R pomoću Seurat paketa.

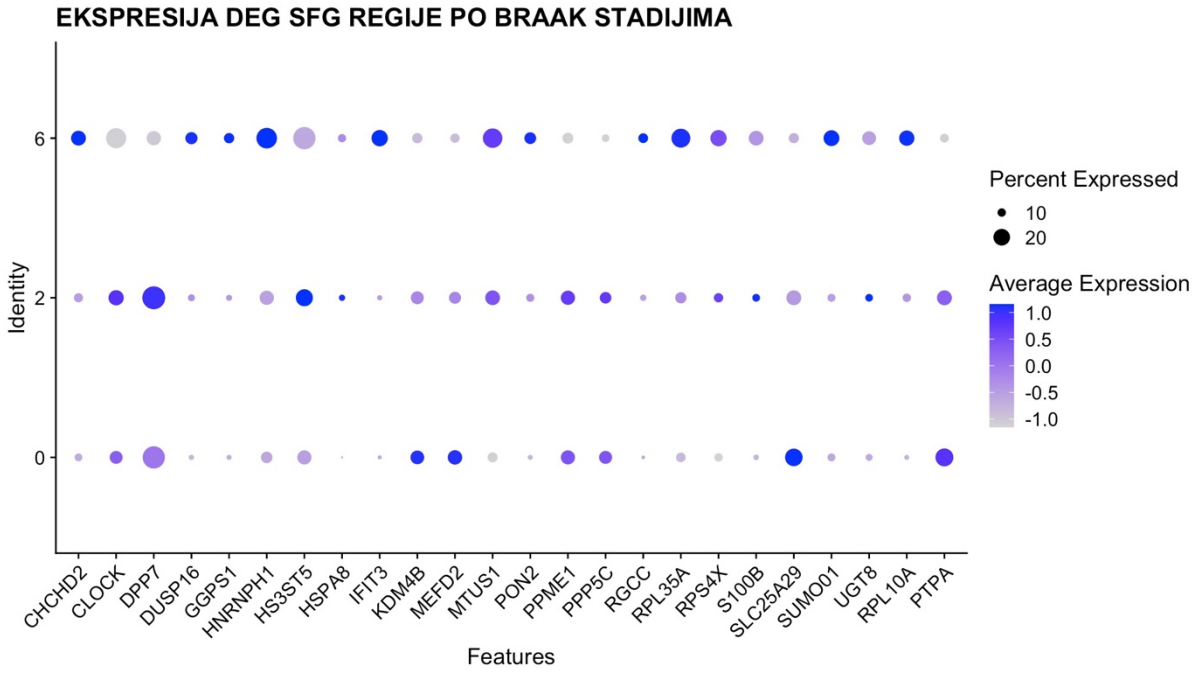
Braak stadij 0: većina gena pokazuje niske razine ekspresije ekspresije, što je označeno na dijagramu svijetlim bojama (razina ekspresije) i manjim točkicama (postotak stanica koje eksprimiraju određeni gen).
Opaženo je kako je gen DPP7 (Dipeptidil peptidaza 7) visoko eksprimiran, a pretpostavljamo da je to radi njegove uloge u održavanju proteolitičke ravnoteže i mogulacije imunosnog odgovora.
Pojačanu ekspresiju u zdravom tkivu pokazuju i KDM4B, MEF2D, PPME1, PPP5C, SLC25A29 i PTPA.

Braak stadij 2: kako bolest napreduje, vidljivo je kako više gena pokazuje povećanje ekspresije što očitujemo povećanjem točkica, te njihovom

tamnijom bojom. Najveću promjenu u razini ekspresije pokazuju geni CLOCK, CHCHD2, PON2, RPS4X i UGT8.

Braak stadij 6: Razine ekspresije određenih gena značajno su povišene, označenih tamnoplavim, većim točkama što ukazuje na snažnu uključenost ovih gena u progresiji bolesti u entorinalnom korteksu mozga.
Gen HNRNPH1 pokazuje najvišu ekspresiju u Braak stadiju 6, a sa nešto manjim postotkom stanica značajno su eksprimirani i geni SUMO1, MTUS1, CHCHD2 i GGPS1.
S druge strane, geni KDM4B, MEF2D, PPME1, PPP5C, SLC25A29, PTPA i DPP7 pokazuju značajnu redukciju u ekspresiji, što sugerira kako progresijom bolesti ovi geni gube svoju ulogu ili su suprimirani kao odgovor na bolest.

2. Geni povezani sa AB: superiorni frontalni girus



Slika 20 Ekspresija 24 gena povezanih s AB u superiornom frontalnom girusu.. X-os predstavljaju geni, na Y-osi navedeni su Braak stadiji predstavljajući progresiju bolesti od 0 do 6. Dijagram je napravljen u programskom jeziku R uz uporabom Seurat paketa.

Braak stadij 0: Većina gena pokazuje relativno niske prosječne razine ekspresije (svijetlije nijanse), a male točkice ukazuju na to da mali postotak stanica SFG eksprimira te gene.
Slično kao i u EC, prosječnu ekspresiju u zdravom tkivu pokazuju geni KDM4B, MEF2D, PPME1, PPP5C, SLC25A29 i PTPA i DPP7, s time da DPP7 pokazuje nešto manju ekspresiju u odnosu na EC.

Braak stadij 2: U usporedbi s EC uočene su veće promjene ekspresije i postotka stanica koje eksprimiraju te gene.
Također je uočeno kako geni CLOCK, DPP7, CHCHD2, HS3ST5, HSPA8, S100B i UGT8 pokazuju višu ekspresiju u odnosu na Braak stadije 0 i 6 što potencijalno može ukazivati na kompenzacijske mehanizme gena u odgovoru na neurodegenerativne promjene ili može odražavati promjene u staničnom sastavu jedinstvenom za ovu fazu bolesti.

Braak stadij 6: Vidljive tamnije i veće točke ukazuju na visoku razinu ekspresije i veći postotak stanica koje te gene eksprimiraju u uznapredovaloj fazi Alzheimerove bolesti. A posebno se ističu geni HNRNPH1, PPME1, MTUS1, RPL35A, RPS4X, SUMO1, RPL10A i CHCHD2 koji pokazuju veću ekspresiju i veći postotak stanica u kojima su ti geni eksprimirani u odnosu na ranije Braak stadije, slično kao i u EC, međutim s određenim varijacijama u razini ekspresije.
Najvišu razinu ekspresiju pokazuju geni HNRNPH1 i RPL35A što ukazuje da su ti geni u izravnoj korelaciji sa patološkim procesima napredovanja bolesti.
Gen CLOCK pokazuje nižu razinu ekspresije, ali je broj stanica koje taj gen eksprimiraju povećan. To se može pripisati njegovoj ulozi regulatora cirkadijalnog ritma koji obično funkcionira na bazalnoj razini u većini tipova stanica. Negativna ekspresija može biti posljedica statističkih obrada podataka poput log transformacije koja nakon normalizacije vrlo niskih vrijednosti ekspresije (blizu nule) može rezultirati negativnim, odnosno nižim vrijednostima.

Isto kao i u EC regiji, geni KDM4B, MEF2D, PPME1, PPP5C, SLC25A29, PTPA i DPP7 pokazuju značajnu redukciju u ekspresiji ili gotovo nisu eksprimirani, što sugerira kako progresijom bolesti geni gube funkciju ili su suprimirani kao odgovor na bolest.

Nadalje, istražili smo razinu ekspresije ovih gena u mozgu u normalnom stanju, ulogu u organizmu, te kako Alzheimerova bolest utječe na ekspresiju ova 24 gena s ciljem bolje povezanosti sa patologijom AB (Tablica 3).

Tablica 3. Tablični prikaz daljnjeg istraživanja 24 DEG povezanih sa AB uključujući ekspresiju u mozgu, funkciju u organizmu, ponašanje u AB i literaturni izvor.

HGNC SIMBOL	EKSPRESIJA U MOZGU	ULOGA U ORGANIZMU	POVEZANOST SA AB	IZVOR O POVEZANOSTI S AB
CHCHD2	UMJERENA	KONTROLA MITOHONDRIJA, NEUROZAŠTITA	DA – GENETSKA MUTACIJA	Zhou, Wei et al. "PD-linked CHCHD2 mutations impair CHCHD10 and MICOS complex leading to mitochondria dysfunction." *Human molecular genetics* vol. 28,7 (2019) [66]
CLOCK	UMJERENA, UBIKVITARNA	REGULACIJA CIRKADIJALNOG RITMA	DA - GENETSKA MUTACIJA, PROMJENA EKSPRESIJE	Thompson, L I et al. "Digital Clock Drawing as an Alzheimer's Disease Susceptibility Biomarker: Associations with Genetic Risk Score and APOE in Older Adults." *The journal of prevention of Alzheimer's disease* vol. 11,1 (2024) [67]
DPP7	UMJERENA, UBIKVITARNA	METABOLIZAM PEPTIDA, STANIČNA SIGNALIZACIJA	DA – POVIŠENA EKSPRESIJA	Mantle, D et al. "Comparison of cathepsin protease activities in brain tissue from normal cases and cases with Alzheimer's disease, Lewy body dementia, Parkinson's disease and Huntington's disease." *Journal of the neurological sciences* vol. 131,1 (1995) [68]
DUSP16	UMJERENA, UBIKVITARNA	REGULACIJA SIGNALNIH PUTEVA, NEUROZAŠTITA, UPALA	DA – POVIŠENA EKSPRESIJA	Matsuzaki Tada, Asuka et al. "Pharmaceutical Potential of Casein-Derived Tripeptide Met-Lys-Pro: Improvement in Cognitive Impairments and Suppression of Inflammation in APP/PS1 Mice." *Journal of Alzheimer's disease : JAD* vol. 89,3 (2022) [69]
GGPS1	UMJERENA, UBIKVITARNA	BIOSINTEZA PRENILCISTEIN MODIFIKACIJA NA PROTEINIMA	DA – POVIŠENA EKSPRESIJA	Hooff, Gero P et al. "Modulation of cholesterol, farnesylpyrophosphate, and geranylgeranylpyrophosphate in neuroblastoma SH-SY5Y-APP695 cells: impact on amyloid beta-protein production." *Molecular neurobiology* vol. 41,2-3 (2010)[70]
HNRNPH1	UMJERENA, UBIKVITARNA	METABOLIZAM RNA, REGULACIJA EKSPRESIJE GENA	DA – PROMJENA EKSPRESIJE	Fisette, Jean-François et al. "A G-rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing." *Journal of neurochemistry* vol. 121,5 (2012) [71]
HS3ST5	VISOKA, UBIKVITARNA	BIOSINTEZA HEPARAN SULFATA, RAZVOJ NEURONA	DA – SMANJENA EKSPRESIJA	Wainberg, Michael et al. "Shared genetic risk loci between Alzheimer's disease and related dementias, Parkinson's disease, and amyotrophic lateral sclerosis." *Alzheimer's research & therapy* Jun. 2023 [72]

HSPA8	VISOKA, UBIKVITARNA (RPKM 443.0)	ŠAPERONSKA ULOGA, REGULACIJA OKSIDATIVNOG STREA	DA – SMANJENA EKSPRESIJA	Silva, Patricia Natalia et al. "Analysis of HSPA8 and HSPA9 mRNA expression and promoter methylation in the brain and blood of Alzheimer's disease patients." *Journal of Alzheimer's disease : JAD* vol. 38,1 (2014) [73]
IFIT3	UMJERENA, UBIKVITARNA	ANTIVIRUSNI ODGOVOR, UPALA	DA – SMANJENA EKSPRESIJA	Garces, Armando et al. "Differential expression of interferon-induced protein with tetratricopeptide repeats 3 (IFIT3) in Alzheimer's disease and HIV-1 associated neurocognitive disorders." *Scientific reports* Feb. 2023 [74]
KDM4B	UMJERENA, UBIKVITARNA	REGULACIJA GENSKE EKSPRESIJE	DA – GENETSKA MUTACIJA	Park, Sung Yeon et al. "Targeted Downregulation of *kdm4a* Ameliorates Tau-engendered Defects in *Drosophila melanogaster*." *Journal of Korean medical science 2019. [75]*
MEF2D	VISOKA, UBIKVITARNA (RPKM 15.5)	KONTROLA MIŠIĆA, RAZVOJ I DIFERENCIJACIJA NEURONA	DA – SMANJENA EKSPRESIJA	Chu, Yaping et al. "α-synuclein aggregation reduces nigral myocyte enhancer factor-2D in idiopathic and experimental Parkinson's disease." *Neurobiology of disease* vol. 41,1 (2011) [76]
MTUS1	UMJERENA, UBIKVITARNA	REGULACIJA MIKROTUBULA, NEURALNI RAZVOJ	DA – PROMJENA EKSPRESIJE , UZROČNIK BOLESTI	Sun, Yanfa et al. "Identification of candidate DNA methylation biomarkers related to Alzheimer's disease risk by integrating genome and blood methylome data." *Translational psychiatry* vol. 13,1 387. 13 Dec. 2023. [77]
PON2	VISOKA, UBIKVITARNA (RPKM 81.0)	ANTIOKSIDATIVNA ULOGA, METABOLIZAM ORGANOFOSFATA	DA – GENETSKA MUTACIJA	Parween, Fauzia et al. "Association between human paraoxonase 2 protein and efficacy of acetylcholinesterase inhibiting drugs used against Alzheimer's disease." *PloS one 2021. [78]*
PPME1	VISOKA, UBIKVITARNA (RPKM 17.5)	REGULACIJA FOSFORILACIJE, NEURALNI RAZVOJ,	DA – PROMJENA EKSPRESIJE	Staniszewski, Agnieszka et al. "Reduced Expression of the PP2A Methylesterase, PME-1, or the PP2A Methyltransferase, LCMT-1, Alters Sensitivity to Beta-Amyloid-Induced Cognitive and Electrophysiological Impairments in Mice." *The Journal of neuroscience : the official journal of the Society for Neuroscience* vol. 40,23 (2020), [79]
PPP5C	VISOKA, UBIKVITARNA (RPKM 17.4)	REGULACIJA FOSFORILACIJE PROTEINA	DA – ABNORMALNA ESPRESIJA	Zhang, Hengheng et al. "Dual function of protein phosphatase 5 (PPP5C): An emerging therapeutic target for drug discovery." *European journal of medicinal chemistry* vol. 254 (2023), [80]
RGCC	NISKA	REGULACIJA STANIČNOG CIKLUSA	DA – POVIŠENA EKSPRESIJA	Counts, Scott E, and Elliott J Mufson. "Regulator of Cell Cycle (RGCC) Expression During the Progression of Alzheimer's Disease." *Cell transplantation* vol. 26,4 (2017), [81]
RPL10A	NISKA, UBIKVITARNA	SINTEZA PROTEINA	DA – POVIŠENA EKSPRESIJA	Suzuki, Masayoshi et al. "Upregulation of ribosome complexes at the blood-brain barrier in Alzheimer's disease patients." *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism* vol. 42,11 (2022), [82]
RPL35A	NISKA, UBIKVITARNA	SINTEZA PROTEINA	DA – POVIŠENA EKSPRESIJA	Suzuki, Masayoshi et al. "Upregulation of ribosome complexes at the blood-brain barrier in Alzheimer's disease patients." *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism* vol. 42,11 (2022), [82]
RPS4X	NISKA UBIKVITARNA	SINTEZA PROTEINA	DA – POVIŠENA EKSPRESIJA	Suzuki, Masayoshi et al. "Upregulation of ribosome complexes at the blood-brain barrier in Alzheimer's disease patients." *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism* vol. 42,11 (2022), [82]

S100B	PRISTRANA EKSPRESIJA (RPKM 150.1)	SIGNALIZACIJA KALCIJA, NEUROTROPNE PROMJENE	DA – PREKOMJERNA EKSPRESIJA	Zaręba-Kozioł, Monika et al. "Intracellular Protein S-Nitrosylation-A Cells Response to Extracellular S100B and RAGE Receptor." *Biomolecules* Apr.2022., [83]
SLC25A29	NISKA	METABOLIZAM MITOHONDRIJA, HOMEOSTAZA ENERGIJE NEURONA	DA – PROMJENA EKSPRESIJE	Huang, August Yue et al. "Somatic cancer driver mutations are enriched and associated with inflammatory states in Alzheimer's disease microglia." *bioRxiv Jan.2024., [84]*
SUMO1	VISOKA, UBIKVITIRANA	TRANSPORT JEZGRI, REGULACIJA TRANSKRIPCIJE	DA - UZROČNIK	Cho, Sun-Jung et al. "SUMO1 promotes Aβ production via the modulation of autophagy." *Autophagy* vol. 11,1 (2015), [85]
UGT8	PRISTRANA (RPKM 15.7)	FORMACIJA MIJELINSKE OVOJNICE	DA - GENETSKA MUTACIJA	Moll, Tobias et al. "Disrupted glycosylation of lipids and proteins is a cause of neurodegeneration." *Brain : a journal of neurology* vol. 143,5 (2020), [86]
PTPA	VISOKA UBIKVITIRANA (RPKM 26.2)	REGULACIJA FOSFORILACIJE, NEURALNI RAZVOJ	DA – SMANJENA EKSPRESIJA	Ando, Sana et al. "Age-related alterations in protein phosphatase 2A methylation levels in brains of cynomolgus monkeys: a pilot study." *Journal of biochemistry* vol. 173,6 (2023), [87]

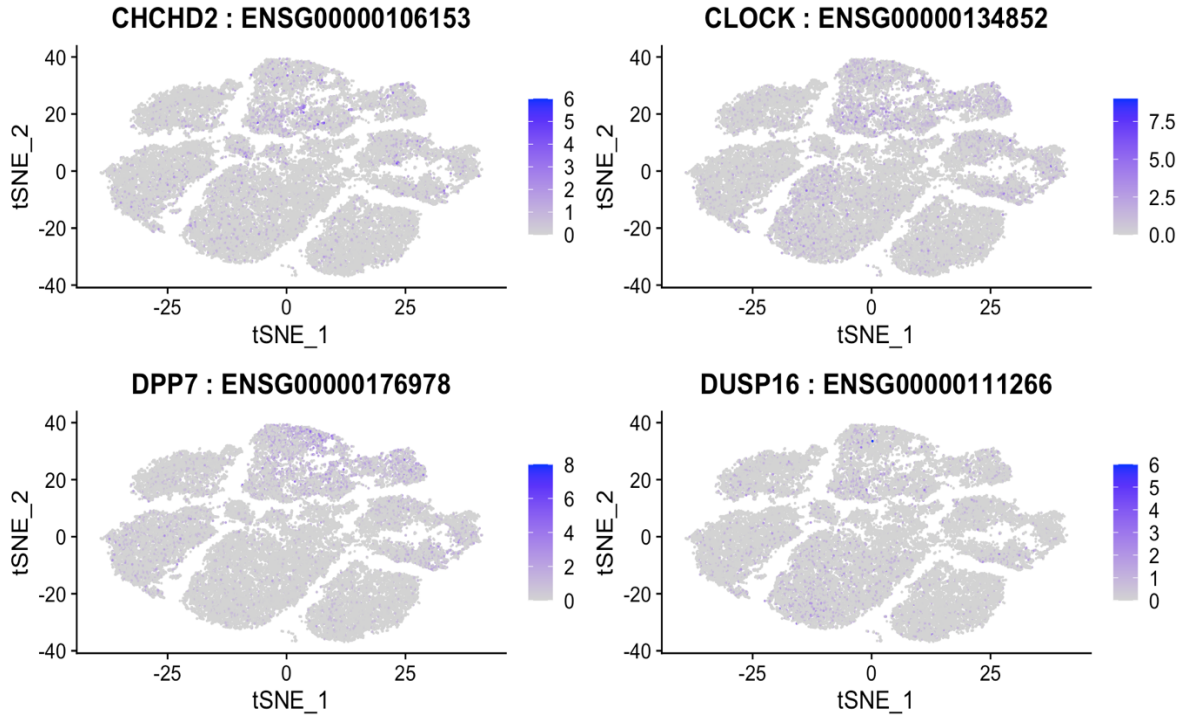
Kako bismo dobili uvid koji tipovi stanica eksprimiraju navedene gene i u kojem obujmu, za daljnju analizu, napravili smo Feature Plot analize pomoću *Seurat* paketa za vizualizaciju ekspresije 24 DEG u različitim tipovima stanica EC i SFG.

Uz Feature Plot dijagrame, napravili smo i Dot Plot dijagrame ekspresije gena u različitim tipovima klastera stanica mozga na stanici CELLXGENE (slika 21) na temelju znanstvenog istraživanja Leng i sur. [65] iz obje baze podataka (EC i SFG), kako bi dodatno potvrdili koji su geni posebno izraženi u određenim tipovima stanica u Alzheimerovoj bolesti.

Slika 21 U Dot Plot-u intenzitet boje označava razinu ekspresije gena, a raspodjela točkica odražava prisutnost ekspresije gena u različitim klasterima.Napravljeno na stranici: https://cellxgene.cziscience.com/gene-expression

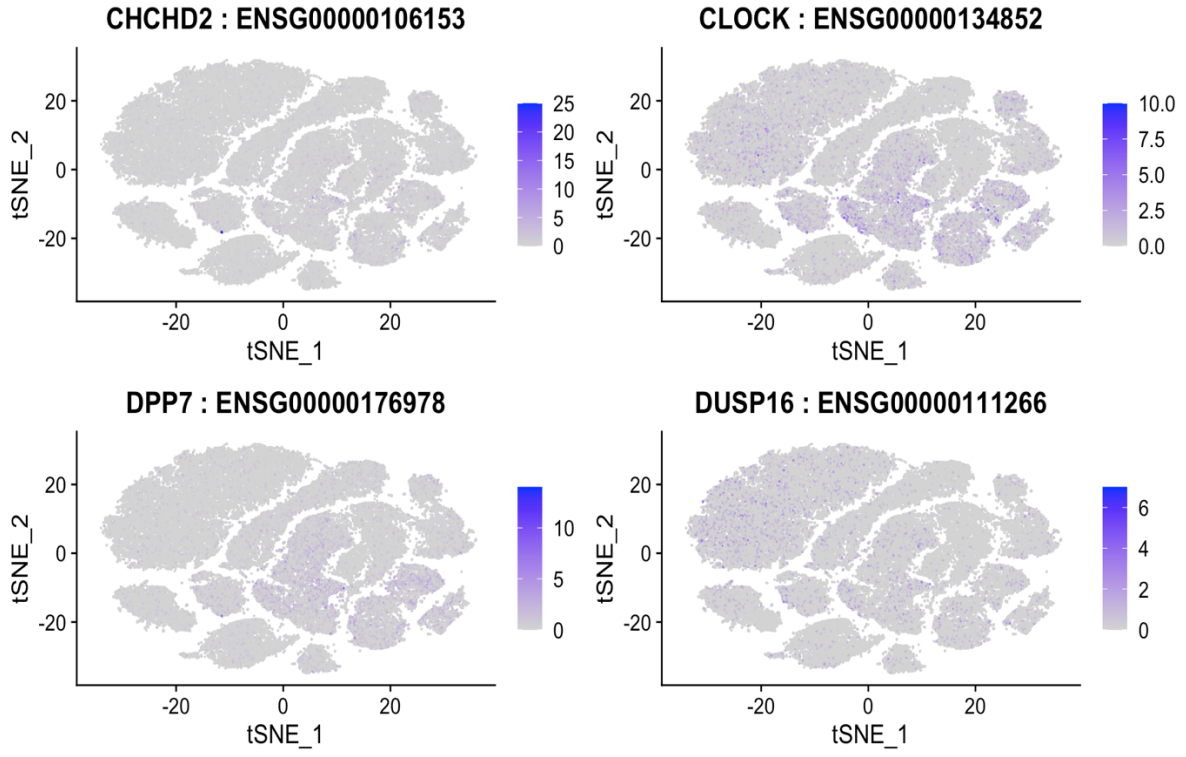
U FeaturePlot prikazima intenzitet boje označava razinu ekspresije gena, a raspodjela točkica odražava prisutnost ekspresije gena u različitim stanicama. Komparativnom analizom usporedili smo ekspresiju i raspodjelu gena u klasterima u entorinalnom korteksu i superiornom frontalnom girusu kako bi dodatno potvrdili u kojim stanicama su značajno eksprimirani dobiveni geni, te mogu li potencijalno služiti kao biomarkeri za Alzheimerovu bolest (Slike 22- 33).

- Entorinalni korteks 1



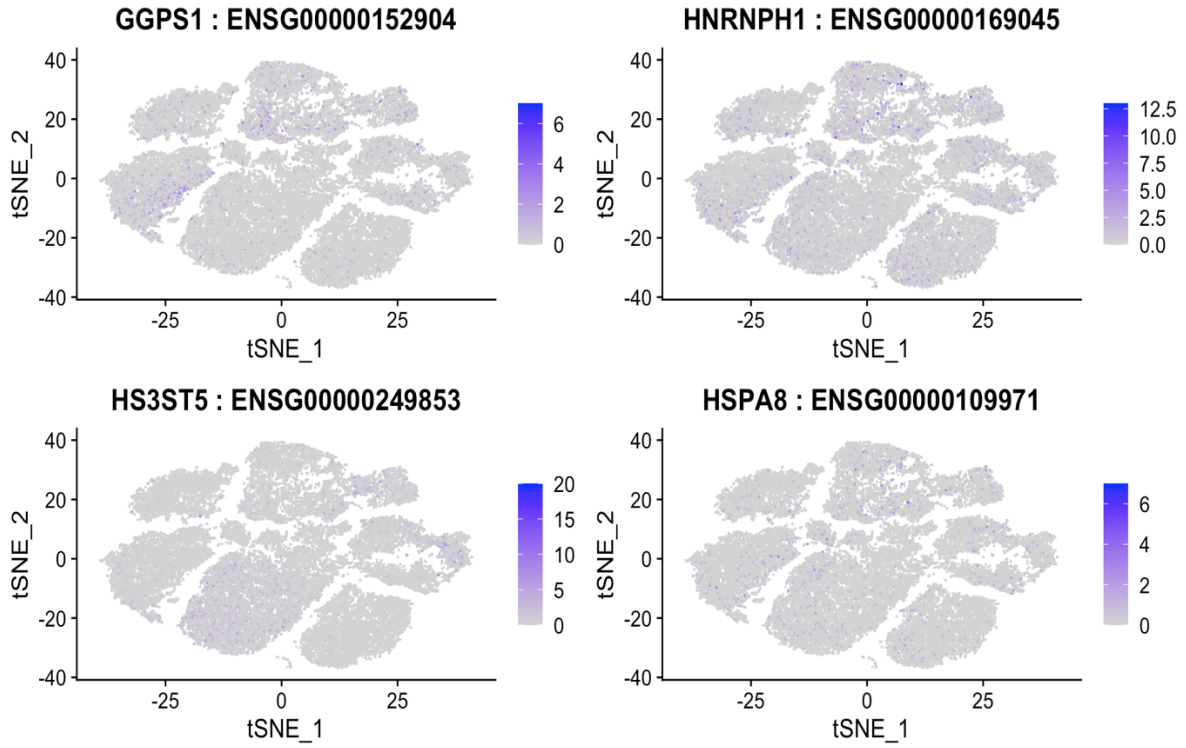
Slika 22 Prikaz tSNE Feature Plot obrazaca ekspresije CHCHD2, CLOCK, DPP7, DUSP16 gena u EC.Dijagrami su napravljeni u programskom jeziku R pomoću Seurat paketa.

- Superioni frontlni girus 1



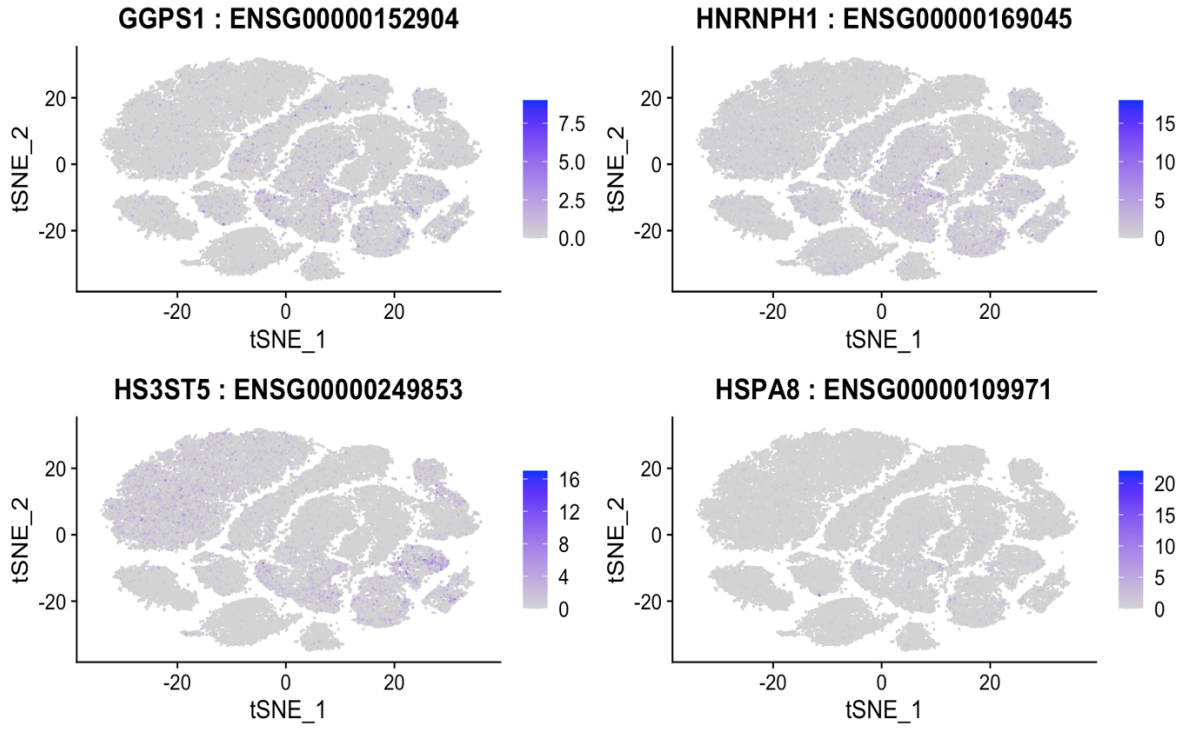
Slika 23 Prikaz tSNE Feature Plot obrazaca ekspresije CHCHD2, CLOCK, DPP7, DUSP16 gena u SFG.Dijagrami su napravljeni u R-u pomoću Seurat paketa.

- Entorinalni korteks 2



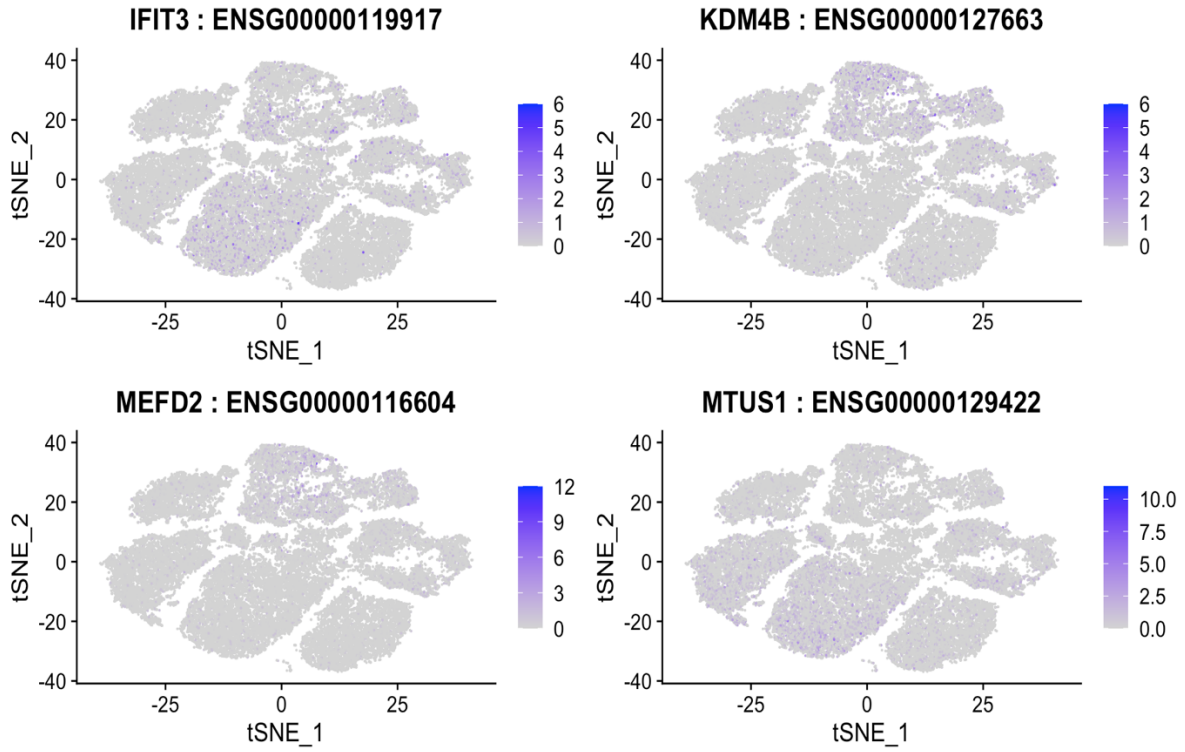
Slika 24 Prikaz tSNE Feature Plot obrazaca ekspresije GGPS1, HNRNPH1, HS3ST5 i HSPA8 gena u EC. Dijagrami su napravljeni u R-u pomoću Seurat paketa.

- Superiorni frontalni girus 2



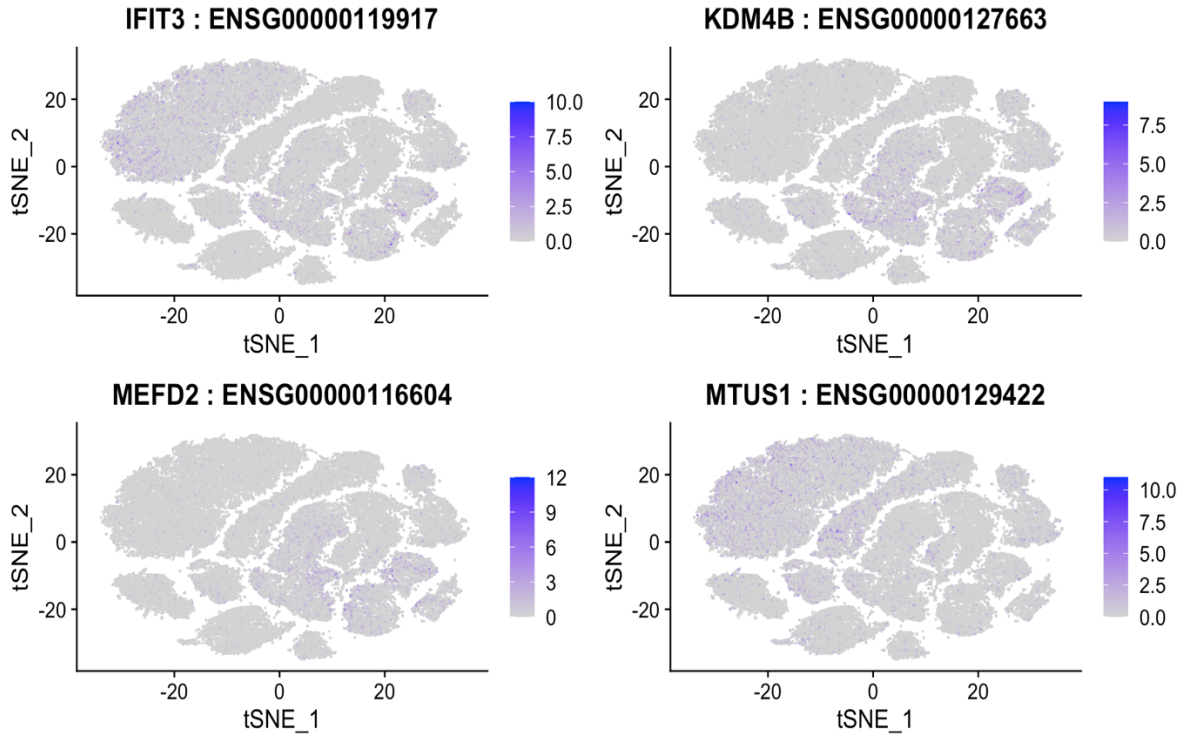
Slika 25 Prikaz tSNE Feature Plot obrazaca ekspresije GGPS1, HNRNPH1, HS3ST5 i HSPA8 gena u SFG Dijagrami su napravljeni u R-u pomoću Seurat paketa.

- Entorinalni korteks 3



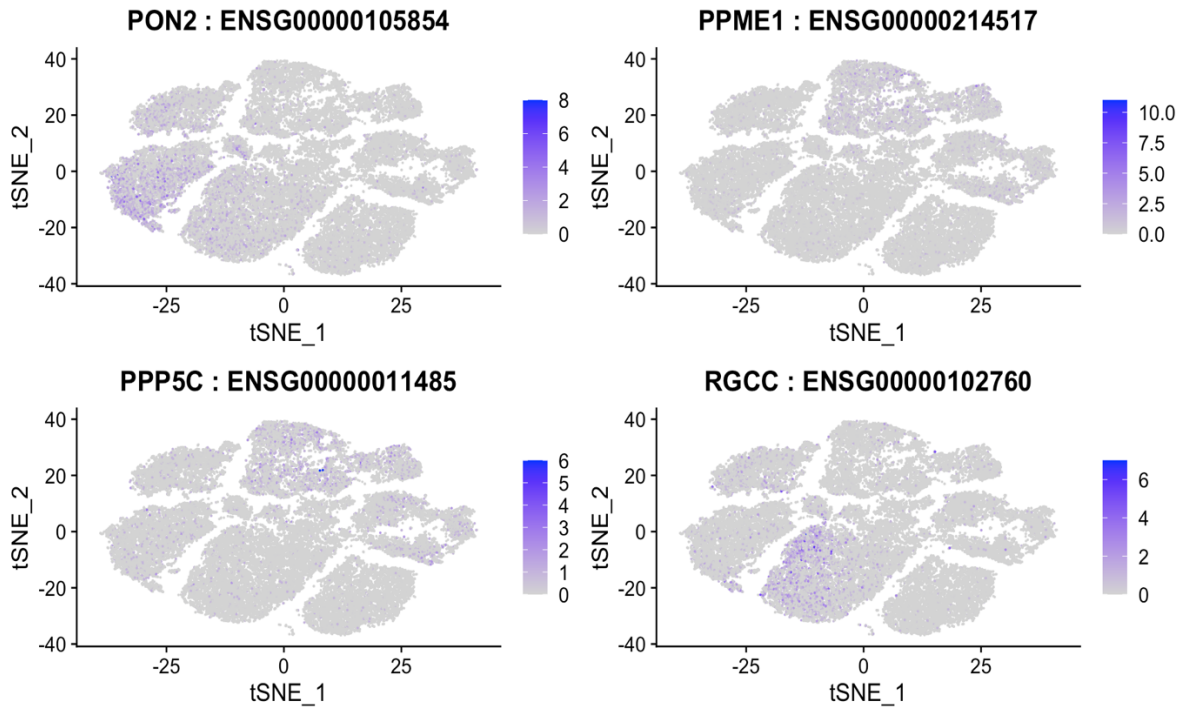
Slika 26 Prikaz tSNE Feature Plot obrazaca ekspresije IFIT3, KDM4B, MEFD2, MTUS1 gena u EC. Dijagrami su napravljeni u R-u pomoću Seurat paketa.

- Superiorni frontalni girus 3



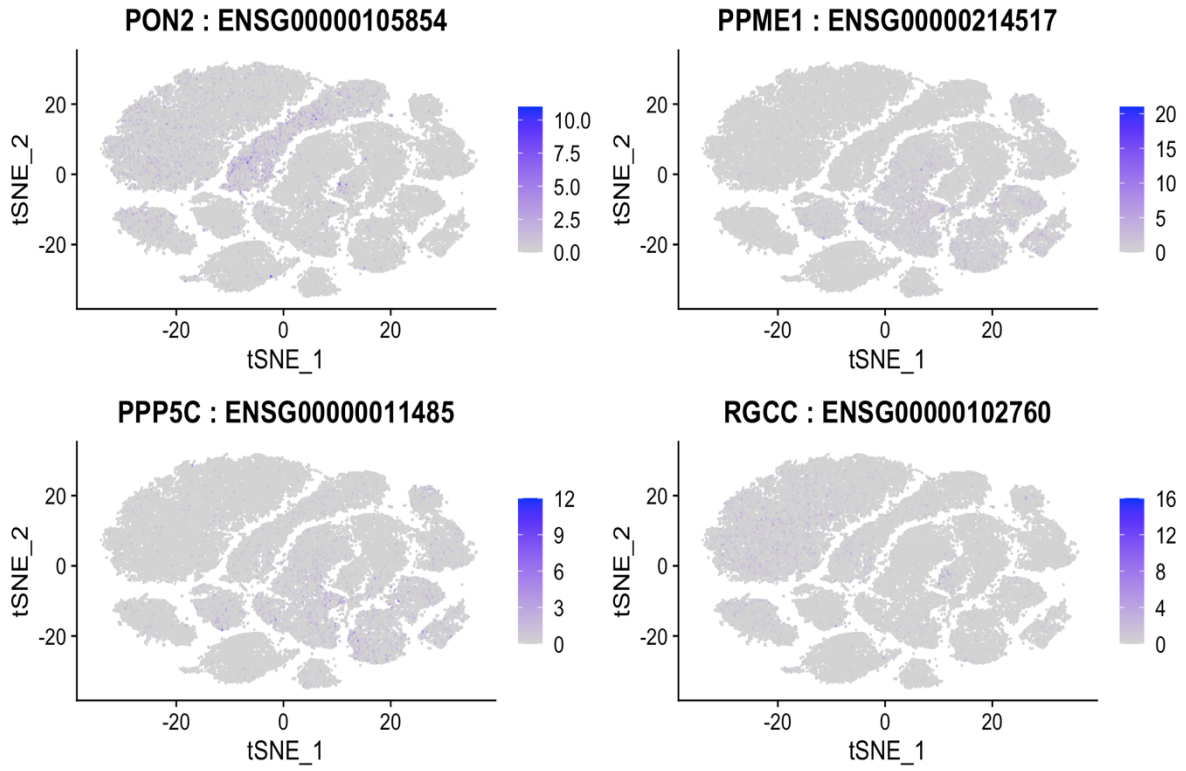
Slika 27 Prikaz tSNE Feature Plot obrazaca ekspresije IFIT3, KDM4B, MEFD2, MTUS1 gena u SFG. Dijagrami su napravljeni u R-u pomoću Seurat paketa.

- Entorinalni korteks 4



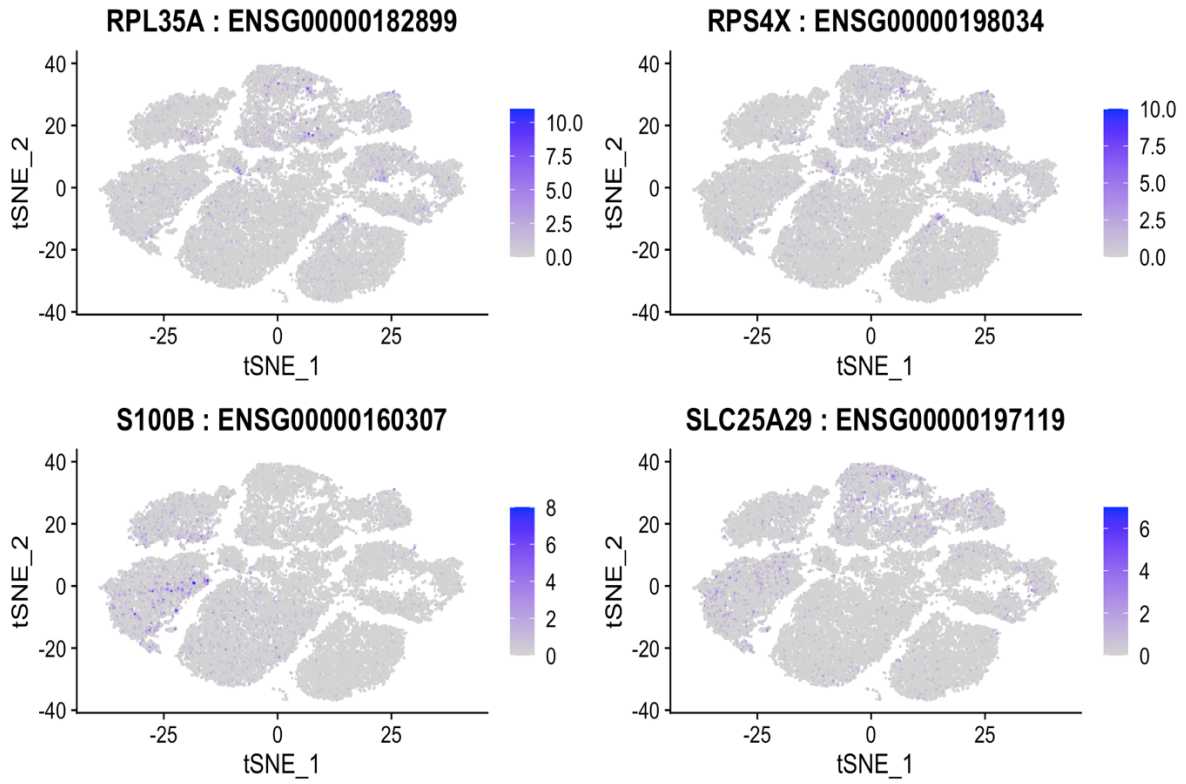
Slika 28 Prikaz tSNE Feature Plot obrazaca ekspresije PON2, PPME1, PPP5C i RGCC gena u EC.

- Superiorni frontalni girus 4



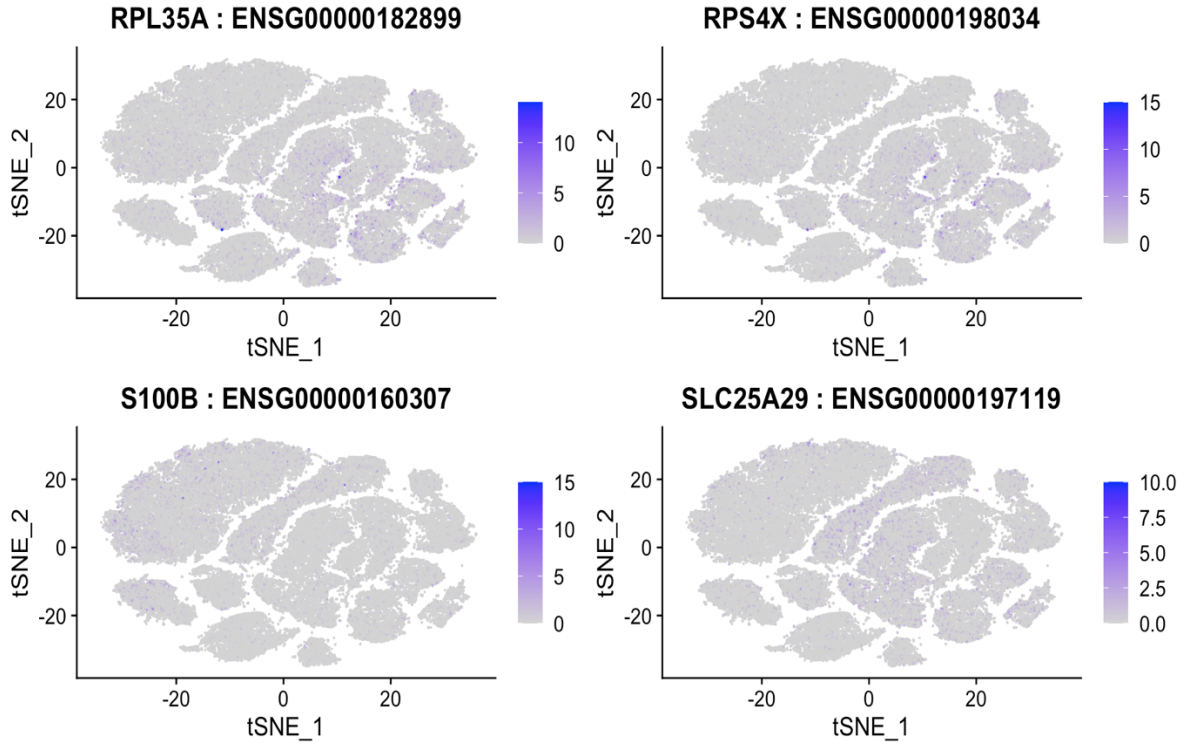
Slika 29 Prikaz tSNE Feature Plot obrazaca ekspresije PON2, PPME1, PPP5C i RGCC gena u SFG. Dijagrami su napravljeni u R-u pomoću Seurat paketa.

- Entorinalni korteks 5



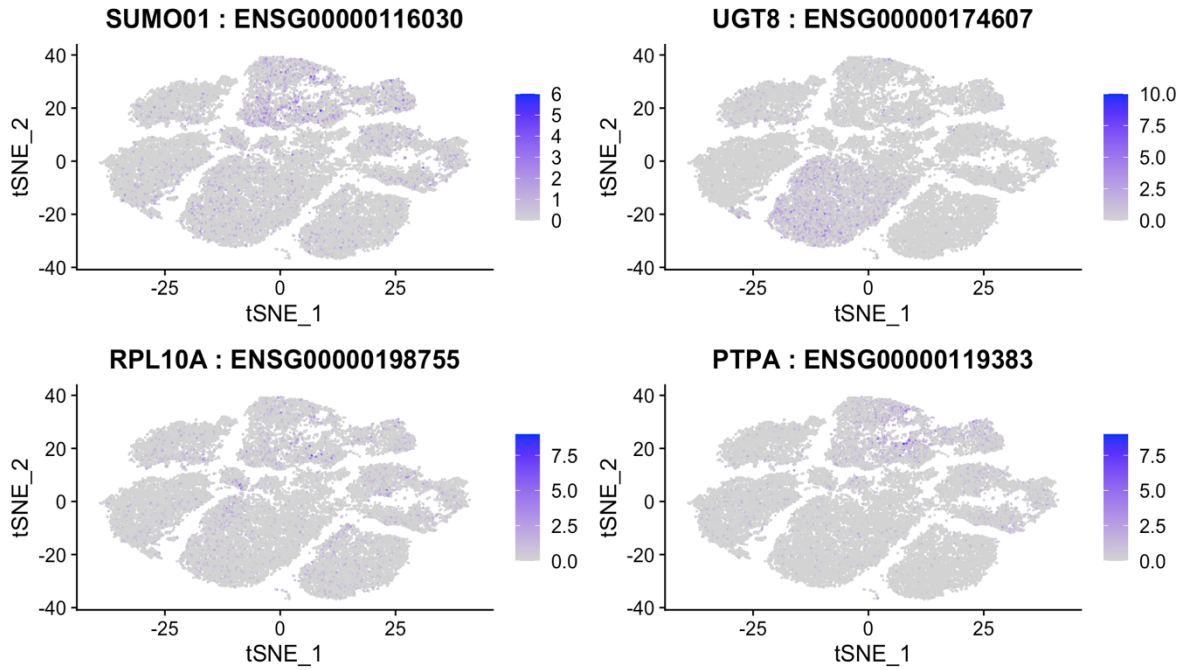
Slika 30 Prikaz tSNE Feature Plot obrazaca ekspresije RPL35A, RPS4X, S100B i SLC25A29 gena u EC. Dijagrami su napravljeni u R-u pomoću Seurat paketa.

- Superiorni frontalni girus 5



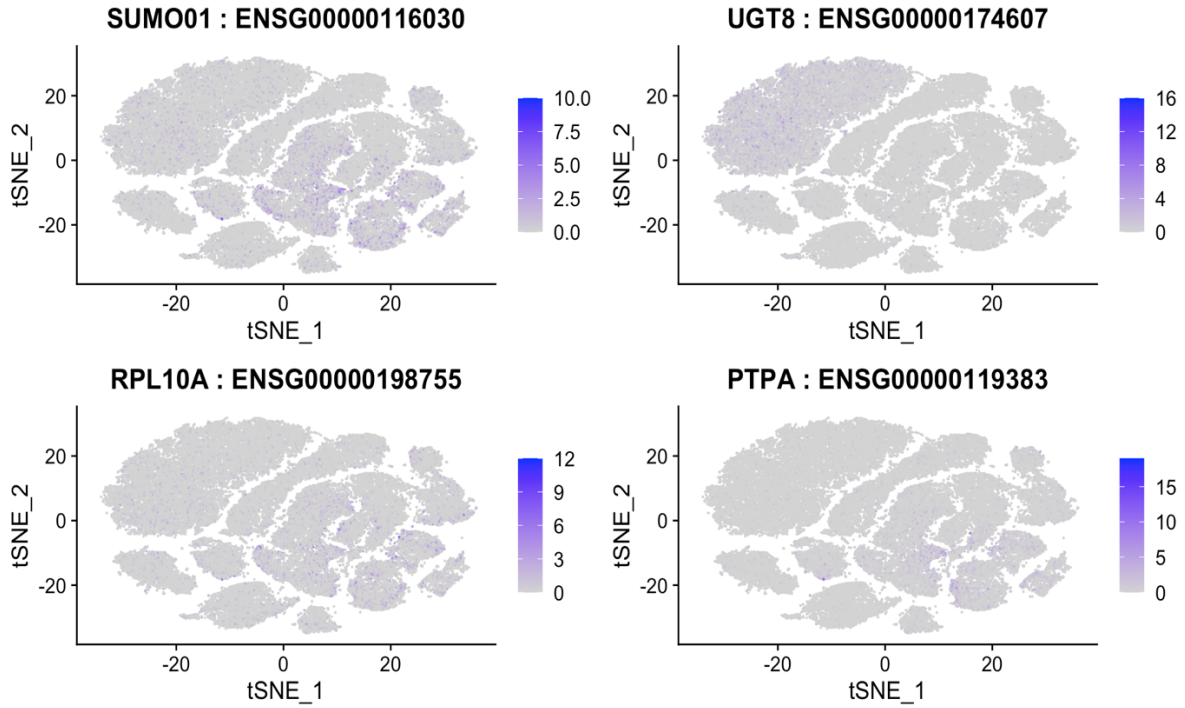
Slika 31 Prikaz tSNE Feature Plot obrazaca ekspresije RPL35A, RPS4X, S100B i SLC25A29 gena u SFG. Dijagrami su napravljeni u R-u pomoću Seurat paketa.

- Entorinalni korteks 6



Slika 32 Prikaz tSNE Feature Plot obrazaca ekspresije SUMO1, UGT8, RPL10A i PTPA gena u EC. Dijagrami su napravljeni u R-u pomoću Seurat paketa.

- Superiorni frontalni girus 6



Slika 33 Prikaz tSNE Feature Plot obrazaca ekspresije SUMO1, UGT8, RPL10A i PTPA gena u SFG. Dijagrami su napravljeni u R-u pomoću Seurat paketa.

Vizualizacije dijagrama Feature Plot i Dot Plot potvrdile su kako je većina gena u obje istraživane regije široko rasprostranjena po klasterima, što upućuje na kasnije faze Alzheimerove bolesti.

Gen CHCHD2 pokazuje intenzivnu ekspresiju u EC i smanjenu u SFG što potvrđuje rezultate Dot Plot-a s obzirom da je EC regija zahvaćena u ranijim stadijima bolesti.
Većina DEG je široko rasprostranjena po svim stanicama, uz najveću ekspresiju u glutamateričkim (ekscitatornim) neuronima.

Gen CLOCK u EC regiji najviše je eksprimiran u glutamateričkim neuronima, oligodendrocitima, te GABAergičkim neuronima. U SFG regiji također je najizraženiji u stanicama oligodendrocita, te ekscitatornim neuronima.
Uloga različitih vrsta stanica mozga opisana je u tablici 4.
Obzirom na zahvaćenost obje regije, vrste stanica i ulogu gena u organizmu, ovaj gen je izražen u uznapredovalom stadiju bolesti (stadij 6).

Gen HS3ST5 također je najviše eksprimiran u oligodendrocitima, te ekscitatornim neuronima, sa intenzivnijom ekspresijom u SFG što također povezujemo sa kasnijim stadijem Alzheimerove bolesti.

Geni MTUS1, PON2 i S100B eksprimirani su u oligodendrocitima i astrocitima u EC i SFG regijama, sa slabijim intenzitetom ekspresije u SFG što upućuje na srednji stadij progresije bolesti.

RGCC i UGT8 geni pretežito eksprimiraju oligodendrociti i prekursori oligodendrocita sa jačom ekspresijom u EC.

Gen IFIT3 nalazi se u stanicama oligodendrocita i ekscitatornim stanicama u obje regije.

Geni KDM4B, MEFD2, PTPA, PPME1 i PP5C eksprimiraju ekscitatorni neuroni u obje regije.

Geni HNRNPH1, RPL35A, RPL10A, RPS4X i SUMO1 eksprimirani su u svim stanicama, najviše u ekscitatornim neuronima, s jačom ekspresijom u EC, a slabijom u SFG, što ukazuje na kasniji stadij bolesti.

Gen SLC25A29 eksprimiraju astrociti i ekscitatorne stanice u obje regije.

Tablica 4. Tablični prikaz osnovnih informacija o stanicama pogođenih u Alzheimerovoj bolesti.

NAZIV	ULOGA U MOZGU	PATOLOGIJA AB
GLUTAMATERGIČNI NEURONI /EKSCITATORNI NEURONI	OTPUŠTANJE GLUTAMATA-GLAVNOG EKSCITATORNOG NEUROTRANSMITERA ODGOVORNOG ZA PROCESE UČENJA, KOGNICIJE I MEMORIJE	EKSCITOTOKSIČNOST GLUTAMATA – GUBITAK PAMĆENJA I KOGNITIVNO OŠTEĆENJE
GABAergički NEURONI / INHIBITORNI NEURONI	OTPUŠTANJE GAMA-AMINOMASLAČNE KISELINE - GLAVNOG INHIBITORNOG NEUROTRANSMITERA. ODRŽAVANJE RAVNOTEŽE IZMEĐU EKSCITACIJE I INHIBICIJE U MOZGU	NERAVNOTEŽA EKSCITACIJE/INHIBICIJE – KOGNITIVNE POTEŠKOĆE, ABNORMALNE AKTIVNOSTI MOZGA
OLIGODENDROCITI	STVARANJE MIJELINA ZA PROVOĐENJE ŽIVČANOG IMPULSA	OŠTEĆENJE MIJELINA – POREMEĆAJ POVEZANOSTI NEURONA, KOGNITIVNA OŠTEĆENJA
ASTROCITI	FUNKCIONALNA POTPORA NEURONA, ODRŽAVANJE KRVNO-MOŽDANE BARIJERE, MODULACIJA SINAPTIČKE AKTIVNOSTI	REAKTIVNI ASTROCITI-ASTROGLIOZA, OŠTEĆENJE KMB – KOGNITIVNI GUBITAK, POVEĆAN RIZIK OD NAPADAJA, MOTORIČKA OŠTEĆENJA, POREMEĆAJ SPAVANJA
PREKURSOR OLIGODENDROCITA	STVARANJE OLIGODENDROCITA	DEMIJALINIZACIJA I DEGENERACIJA BIJELE TVARI – KOGNITIVNI GUBITAK
MIKROGLIJA	IMUNOLOŠKE STANICE – UKLANJANJE ŠTETNIH TVARI (PLAKOVA), MODULATORI UPALE	KRONIČNA AKTIVACIJA, NEUROINFLAMACIJA KOJA DOVODI DO GUBITKA NEURONA, POJAČAVA SE POVEĆANJEM PLAKOVA
ENDOTELNE STANICE U MOZGU	ODRŽAVANJE KRVNO-MOŽDANE BARIJERE	DISFUNKCIJA STANICA, DOVODI DO OŠTEĆENJA KMB I NAKUPLJANJA AMILOIDNIH PLAKOVA

Na temelju ekspresije u stanicama možemo potvrditi kako bi određeni geni mogli biti potencijalni markeri za određene stadije bolesti:

- Biomarkeri ranog stadija AB: CHCHD2 (NAMOTANA-ZAVOJNICA-HELIX-NAMOTANA ZAVOJNICA DOMENA 2), RGCC (REGULATOR STANIČNOG CIKLUSA) i UGT8 (UDP GLIKOZILTRANSFERAZA 8)
- Biomarkeri srednjeg stadija AB: MTUS1 (MIKROTUBULIMA POVEZAN SKELOM PROTEIN 1), PON2 (PARAOKSONAZA 2) i S100B(S100 KALCIJ VEZAJUĆI PROTEIN B)
- Biomarkeri kasnog stadija AB: CLOCK (REGULATOR CIRKADIJALNOG RITMA), HS3ST5 (HEPARAN-SULFAT-GLUKOZAMIN-3 SULFOTRANSFERAZA-5), HNRNPH1 (HETEROGENI NUKLEARNI RIBONUKLEOPROTEIN H1), RPL35A (RIBOSOMALNI PROTEIN L35A), RPL10A (RIBOSOMALNI PROTEIN L10A), RPS4X (RIBOSOMSKI PROTEIN S4 X-VEZAN), SUMO1 (MALI NALIK UBIKVITINU MODIFIKATOR 1), KDM4 (LIZIN DEMETILAZA 4B) i MEFD2 (MIOCITNI POJAČIVAČ FAKTOR 2D)

5. DISKUSIJA

Analize napravljene u ovom istraživanju pružile su uvid u klasifikaciju statusa bolesti putem strojnog učenja na podacima sekvenciranja na razini pojedinačne stanice. Identificirali smo nekoliko značajnih gena pomoću kojih bi mogli klasificirati stadij bolesti.

SFG i EC regije pokazuju trend povećanja ekspresije gena od Braak faze 0 do Braak faze 6. Takav je obrazac u skladu s idejom da kako Alzheimerova bolest napreduje, više gena postaje aktivno uključeno u patološki proces.

Geni CHCHD2, RGCC i UGT8 potencijalni su biomarkeri klasifikacije ranog stadija bolesti s obzirom da su više izraženi u entorinalnom korteksu za kojeg je potvrđeno kako je to regija mozga koja je pogođena u početnim fazama bolesti i pokazuju staničnu homogenu lokalizaciju. Ovi geni su izraženi ponajviše u oligodendrociti čija je uloga, stvaranje i održavanje mijelinske ovojnice za provođenje živčanog impulsa duž neurona. Oštećenje mijelinske ovojnice događa se u Alzheimerovoj bolesti i posljedično dovodi do prvih simptoma i smanjenja kognitivnih sposobnosti.

Geni MTUS1, PON2 i S100B geni mogli bi se smatrati biomarkerima srednje faze obzirom da astrociti i oligodendrociti eksprimiraju te gene, te njihovim intenzitetima ekspresija u EC i SFG regijama mozga. Također je važno spomenuti značaj gena MTUS1 (mikrotubulima povezan skelom protein 1) koji se smatra jednim od doprinositeljima uzroka bolesti što dodatno potvrđuje njegov značaj kao potencijalnog biomarkera za srednju fazu bolesti.

Geni CLOCK, HS3ST5, HNRNPH1, RPL35A, RPL10A, RPS4X, SUMO1, KDM4 i MEFD2 geni mogli bi se smarati biomarkerima kasnog stadija bolesti, obzirom na proširenu ekspresiju u gotovo svim stanicama, iako sa slabijim intenzitetom u obje regije mozga.

Podaci o ekspresiji diferencijalno izraženih gena otkrivaju nekoliko značajnih uvida u molekularne temelje Alzheimerove bolesti. Varijacije u razinama ekspresije istaknutih gena sugeriraju da bolest može uključivati

više različitih bioloških procesa, uključujući metabolizam, regulaciju staničnog ciklusa i kontrolu transkripcije.

Otkrića ovog rada u skladu su s prethodnim istraživanjima koja su identificirala slične promjene ekspresije gena kod Alzheimerove bolesti.

Iako su modeli strojnog učenja bili robusni, moraju se naglasiti ograničenja istraživanja. Mali broj uzoraka (10 osoba) i velike baze podataka (42,528 za entorinalni korteks i 63,608 za superiorni frontalni girus) i kompleksnost bolesti uvelike utječu na ishod i vjernost rezultata. Osim toga, obrasci ekspresije sami po sebi ne utvrđuju uzročnost i bitno je razmotriti ove rezultate u kontekstu drugih bioloških podataka.

Daljnja istraživanja trebala bi se usredotočiti na potvrđivanje već poznatih gena povezanih s Alzheimerovom bolesti na većim kohortama, te na ntegraciju različitih molekularnih pristupa, kako bi se dobila cjelovitija slika molekularne osnove i identificirali pouzdani biomarkeri za lakšu klasifikaciju progresije bolesti. Također, razvoj novih modela strojnog učenja, mogli bi bi pružiti dublji uvid u biološki značaj identificiranih obrazaca genske ekspresije.

6. ZAKLJUČAK

Alzheimerova bolest progresivni je neurodegenerativni poremećaj s kompleksom molekularnom osnovom, karakteriziran gubitkom pamćenja, kognitivnih sposobnosti i ozbiljnim oštećenjem neurona. Glavnu ulogu u patologiji bolesti imaju izvanstanični plakovi nastali nakupljanjem amiloida β i unutarstanični neurofibrilarni čvorovi Tau proteina.
Za utvrđivanje progresije neurofibrilarnih čvorova u mozgu koristimo klasifikaciju zvanu Braakovi stadiji. Ova klasifikacija dijeli bolest dijeli u 6 stadija, prema zahvaćenosti mozga, pri čemu 0 označava zdravo stanje, a stadij 6 predstavlja uznapredovalu fazu Alzheimerove bolesti.

U ovom radu je istražena klasifikacija statusa Alzheimerove bolesti, pri čemu su korišteni podaci RNA sekvenciranja pojedinačnih stanica dobiveni pomoću platforme 10X Genomics Chromium. Ovaj nam je pristup omogućio analizu transkriptomskih profila pojedinačnih stanica u zdravim i Alzheimerom zahvaćenom moždanom tkivu, pružajući detaljan prikaz stanične heterogenosti. Klasifikaciju smo izveli strojnim učenjem modelima za klasifikaciju; logističkom regresijom i algoritmom nasumičnih šuma (eng. *Random Forest*) kako bi se pronašli statistički najznačajniji geni.

Dobiveni rezultati u skladu su sa opće prihvaćenim saznanjima o Alzheimerovoj bolesti. Identificirana je nekolicina gena koji su bili značajno diferencijalno eksprimirani između zdravih i patoloških stanica, uključujući gene CLOCK, HS3ST5, HNRNPH1, koji su se pokazali kao jedni od najizraženijih.

Rezultati ovog istraživanja pokazali su molekularnu kompleksnost bolesti i proširuju naše razumjevanje mehanizma razvoja bolesti, te naglašavaju potencijal upotrebe RNA sekvenciranja pojedinačnih stanica, zajedno s naprednim bioinformatičkim tehnikama strojnog učenja, za identifikaciju novih potencijalnih biomarkera.

7. LITERATURA

1. demencija. Hrvatska enciklopedija, mrežno izdanje. Leksikografski zavod Miroslav Krleža, 2013. – 2024. Pristupljeno 6.8.2024. <https://enciklopedija.hr/clanak/demencija>.
2. Qiu, C., Kivipelto, M., & von Strauss, E. (2009). Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention. Dialogues in Clinical Neuroscience, 11(2), 111–128. https://doi.org/10.31887/DCNS.2009.11.2/cqiu
3. Ramirez-Bermudez J. Alzheimer's disease: critical notes on the history of a medical concept. Arch Med Res. 2012 Nov;43(8):595-9. doi: 10.1016/j.arcmed.2012.11.008. Epub 2012 Nov 21. PMID: 23178566.
4. Cipriani G, Dolciotti C, Picchi L, Bonuccelli U. Alzheimer and his disease: a brief history. Neurol Sci. 2011 Apr;32(2):275-9. doi: 10.1007/s10072-010-0454-7. Epub 2010 Dec 11. PMID: 21153601.
5. World Health Organization, 2023. Dementia. [online] Available at: https://www.who.int/news-room/fact-sheets/detail/dementia [Accessed 6 August 2024].
6. Breijyeh Z, Karaman R. Comprehensive Review on Alzheimer's Disease: Causes and Treatment. Molecules. 2020 Dec 8;25(24):5789. doi: 10.3390/molecules25245789. PMID: 33302541; PMCID: PMC7764106.
7. Ferrari C, Sorbi S. The complexity of Alzheimer's disease: an evolving puzzle. Physiol Rev. 2021 Jul 1;101(3):1047-1081. doi: 10.1152/physrev.00015.2020. Epub 2021 Jan 21. PMID: 33475022.
8. Monteiro AR, Barbosa DJ, Remião F, Silva R. Alzheimer's disease: Insights and new prospects in disease pathophysiology, biomarkers and disease-modifying drugs. Biochem Pharmacol. 2023 May;211:115522. doi: 10.1016/j.bcp.2023.115522. Epub 2023 Mar 28. PMID: 36996971.

9. O'Brien RJ, Wong PC. Amyloid precursor protein processing and Alzheimer's disease. Annu Rev Neurosci. 2011;34:185-204. doi: 10.1146/annurev-neuro-061010-113613. PMID: 21456963; PMCID: PMC3174086.
10. Trejo-Lopez JA, Yachnis AT, Prokop S. Neuropathology of Alzheimer's Disease. Neurotherapeutics. 2022 Jan;19(1):173-185. doi: 10.1007/s13311-021-01146-y. Epub 2021 Nov 2. PMID: 34729690; PMCID: PMC9130398.
11. Chen GF, Xu TH, Yan Y, Zhou YR, Jiang Y, Melcher K, Xu HE. Amyloid beta: structure, biology and structure-based therapeutic development. Acta Pharmacol Sin. 2017 Sep;38(9):1205-1235. doi: 10.1038/aps.2017.28. Epub 2017 Jul 17. PMID: 28713158; PMCID: PMC5589967.
12. Thal DR, Rüb U, Orantes M, Braak H. Phases of A beta-deposition in the human brain and its relevance for the development of AD. Neurology. 2002 Jun 25;58(12):1791-800. doi: 10.1212/wnl.58.12.1791. PMID: 12084879.
13. Bear MF, Connors BW, Paradiso MA. Neuroscience: Exploring the Brain. 4th ed. Burlington, MA: Jones & Bartlett Learning; 2017.
14. Wegmann S, Biernat J, Mandelkow E. A current view on Tau protein phosphorylation in Alzheimer's disease. Curr Opin Neurobiol. 2021 Aug;69:131-138. doi: 10.1016/j.conb.2021.03.003. Epub 2021 Apr 21. PMID: 33892381.
15. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol. 1991;82(4):239-59. doi: 10.1007/BF00308809. PMID: 1759558.
16. Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., … Cummings, J. L. (2014). *Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. The Lancet Neurology, 13(6), 614–629.* doi:10.1016/s1474-4422(14)70090-0

17. Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, Holtzman DM, Jagust W, Jessen F, Karlawish J, Liu E, Molinuevo JL, Montine T, Phelps C, Rankin KP, Rowe CC, Scheltens P, Siemers E, Snyder HM, Sperling R; Contributors. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement. 2018 Apr;14(4):535-562. doi: 10.1016/j.jalz.2018.02.018. PMID: 29653606; PMCID: PMC5958625.
18. Van Cauwenberghe C, Van Broeckhoven C, Sleegers K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. Genet Med. 2016 May;18(5):421-30. doi: 10.1038/gim.2015.117. Epub 2015 Aug 27. PMID: 26312828; PMCID: PMC4857183.
19. Kim J, Basak JM, Holtzman DM. The role of apolipoprotein E in Alzheimer's disease. Neuron. 2009 Aug 13;63(3):287-303. doi: 10.1016/j.neuron.2009.06.026. PMID: 19679070; PMCID: PMC3044446.
20. Giau VV, Bagyinszky E, An SS, Kim SY. Role of apolipoprotein E in neurodegenerative diseases. Neuropsychiatr Dis Treat. 2015 Jul 16;11:1723-37. doi: 10.2147/NDT.S84266. PMID: 26213471; PMCID: PMC4509527.
21. Madnani RS. Alzheimer's disease: a mini-review for the clinician. Front Neurol. 2023 Jun 22;14:1178588. doi: 10.3389/fneur.2023.1178588. PMID: 37426432; PMCID: PMC10325860.
22. Singh R, Sadiq NM. Cholinesterase Inhibitors. 2023 Jul 17. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan–. PMID: 31335056.
23. Sharma K. Cholinesterase inhibitors as Alzheimer's therapeutics (Review). Mol Med Rep. 2019 Aug;20(2):1479-1487. doi: 10.3892/mmr.2019.10374. Epub 2019 Jun 11. PMID: 31257471; PMCID: PMC6625431.

24. Galimberti D, Ghezzi L, Scarpini E. Immunotherapy against amyloid pathology in Alzheimer's disease. J Neurol Sci. 2013 Oct 15;333(1-2):50-4. doi: 10.1016/j.jns.2012.12.013. Epub 2013 Jan 5. PMID: 23299047.

25. Jack CR Jr, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, Shaw LM, Vemuri P, Wiste HJ, Weigand SD, Lesnick TG, Pankratz VS, Donohue MC, Trojanowski JQ. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol. 2013 Feb;12(2):207-16. doi: 10.1016/S1474-4422(12)70291-0. PMID: 23332364; PMCID: PMC3622225.

26. Xia C, Dickerson BC. Tau PET: the next frontier in molecular imaging of dementia. Int Psychogeriatr. 2016 Sep;28(9):1403-6. doi: 10.1017/S1041610216000880. Epub 2016 Jun 23. PMID: 27334648; PMCID: PMC5690860.

27. Khozyainova AA, Valyaeva AA, Arbatsky MS, Isaev SV, Iamshchikov PS, Volchkov EV, Sabirov MS, Zainullina VR, Chechekhin VI, Vorobev RS, Menyailo ME, Tyurin-Kuzmin PA, Denisov EV. Complex Analysis of Single-Cell RNA Sequencing Data. Biochemistry (Mosc). 2023 Feb;88(2):231-252. doi: 10.1134/S0006297923020074. PMID: 37072324; PMCID: PMC10000364.

28. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., … Surani, M. A. (2009). *mRNA-Seq whole-transcriptome analysis of a single cell. Nature Methods, 6(5), 377–382.* doi:10.1038/nmeth.1315

29. Jovic D, Liang X, Zeng H, Lin L, Xu F, Luo Y. Single-cell RNA sequencing technologies and applications: A brief overview. Clin Transl Med. 2022 Mar;12(3):e694. doi: 10.1002/ctm2.694. PMID: 35352511; PMCID: PMC8964935.

30. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, Leonhardt H, Heyn H, Hellmann I, Enard W. Comparative

Analysis of Single-Cell RNA Sequencing Methods. Mol Cell. 2017 Feb 16;65(4):631-643.e4. doi: 10.1016/j.molcel.2017.01.023. PMID: 28212749.

31. Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, Anavy L, Gennert D, Li S, Livak KJ, Rozenblatt-Rosen O, Dor Y, Regev A, Yanai I. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. 2016 Apr 28;17:77. doi: 10.1186/s13059-016-0938-8. PMID: 27121950; PMCID: PMC4848782.

32. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015 May 21;161(5):1202-1214. doi: 10.1016/j.cell.2015.05.002. PMID: 26000488; PMCID: PMC4481139.

33. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015 May 21;161(5):1187-1201. doi: 10.1016/j.cell.2015.04.044. PMID: 26000487; PMCID: PMC4441768.

34. Gracz AD, Williamson IA, Roche KC, Johnston MJ, Wang F, Wang Y, Attayek PJ, Balowski J, Liu XF, Laurenza RJ, Gaynor LT, Sims CE, Galanko JA, Li L, Allbritton NL, Magness ST. A high-throughput platform for stem cell niche co-cultures and downstream gene expression analysis. Nat Cell Biol. 2015 Mar;17(3):340-9. doi: 10.1038/ncb3104. Epub 2015 Feb 9. PMID: 25664616; PMCID: PMC4405128.

35. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, Gregory MT, Shuga J, Montesclaros L, Underwood JG, Masquelier DA, Nishimura SY, Schnall-Levin M, Wyatt PW, Hindson CM, Bharadwaj R, Wong A, Ness

KD, Beppu LW, Deeg HJ, McFarland C, Loeb KR, Valente WJ, Ericson NG, Stevens EA, Radich JP, Mikkelsen TS, Hindson BJ, Bielas JH. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017 Jan 16;8:14049. doi: 10.1038/ncomms14049. PMID: 28091601; PMCID: PMC5241818.

36. Svensson V, Natarajan KN, Ly LH, Miragaia RJ, Labalette C, Macaulay IC, Cvejic A, Teichmann SA. Power analysis of single-cell RNA-sequencing experiments. Nat Methods. 2017 Apr;14(4):381-387. doi: 10.1038/nmeth.4220. Epub 2017 Mar 6. PMID: 28263961; PMCID: PMC5376499.

37. **10x Genomics.** The next generation of single-cell RNA-seq: An introduction to GEM-X technology [Internet]. 10x Genomics Blog; 2024 Aug 7 [cited 2024 Aug 9]. Available from: https://www.10xgenomics.com/blog/the-next-generation-of-single-cell-rna-seq-an-introduction-to-gem-x-technology

38. Bageritz J, Raddi G. Single-Cell RNA Sequencing with Drop-Seq. Methods Mol Biol. 2019;1979:73-85. doi: 10.1007/978-1-4939-9240-9_6. PMID: 31028633.

39. 10x Genomics. Information Guide: An Overview of Our Services [Internet]. Utrecht: 10x Genomics; 2023 [cited 2024 Aug 12]. Available from: https://app.hubspot.com/documents/4935197/view/728088454?accessId=d75f75

40. 10x Genomics. Chromium X Brochure: The most flexible way to single cell [Internet]. Utrecht: 10x Genomics; 2022 [cited 2024 Aug 12]. Available from: https://pages.10xgenomics.com/rs/446-PBO-704/images/10x_LIT000139_Chromium-X_Brochure_Digital.pdf

41. 10x Genomics, Inc. *Inside Chromium single cell technology*. Rev M. 2024. [brochure on the internet]. Available from: https://pages.10xgenomics.com/rs/446-PBO-

704/images/10x_LIT000025_Brochure_Chromium-products_Letter_Digital.pdf

42. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015 May;33(5):495-502. doi: 10.1038/nbt.3192. Epub 2015 Apr 13. PMID: 25867923; PMCID: PMC4430369.

43. Gisbrecht, A., Schulz, A., & Hammer, B. (2015). *Parametric nonlinear dimensionality reduction using kernel t-SNE. Neurocomputing, 147, 71–82.* doi:10.1016/j.neucom.2013.11.045

44. Zhou B, Jin W. Visualization of Single Cell RNA-Seq Data Using t-SNE in R. Methods Mol Biol. 2020;2117:159-167. doi: 10.1007/978-1-0716-0301-7_8. PMID: 31960377.

45. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. Cell. 2019 Jun 13;177(7):1888-1902.e21. doi: 10.1016/j.cell.2019.05.031. Epub 2019 Jun 6. PMID: 31178118; PMCID: PMC6687398.

46. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, Srivastava A, Molla G, Madad S, Fernandez-Granda C, Satija R. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024 Feb;42(2):293-304. doi: 10.1038/s41587-023-01767-y. Epub 2023 May 25. PMID: 37231261; PMCID: PMC10928517.

47. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Comput Sci. 2021;2(3):160. doi: 10.1007/s42979-021-00592-x. Epub 2021 Mar 22. PMID: 33778771; PMCID: PMC7983091.

48. Shastry K. A., Sanjay H. A. (2020). "Machine learning for bioinformatics," in *Statistical Modelling and Machine Learning Principles for Bioinformatics Techniques, Tools, and Applications Algorithms for Intelligent Systems* eds. Srinivasa K. G., Siddesh G.

M., Manisekhar S. R. (Singapore: Springer Singapore;), 25–39. 10.1007/978-981-15-2445-5_3

49. Forman, G. (2002, July). Incremental Machine Learning to Reduce Biochemistry Lab Costs in the Search for Drug Discovery. In *BIOKDD* (pp. 33-36).

50. Chase, R. J., Harrison, D. R., Burke, A., Lackmann, G. M., & McGovern, A. (2022). A machine learning tutorial for operational meteorology. Part I: Traditional machine learning. *Weather and Forecasting*, *37*(8), 1509-1529.

51. Sidey-Gibbons, J. A., & Sidey-Gibbons, C. J. (2019). Machine learning in medicine: a practical introduction. *BMC medical research methodology*, *19*, 1-18.

52. Paruchuri, H. (2021). Conceptualization of machine learning in economic forecasting. *Asian Business Review*, *11*(1), 51-58

53. Mitchell JB. Machine learning methods in chemoinformatics. Wiley Interdiscip Rev Comput Mol Sci. 2014 Sep 1;4(5):468-481. doi: 10.1002/wcms.1183. PMID: 25285160; PMCID: PMC4180928.

54. Roy, N., Posner, I., Barfoot, T., Beaudoin, P., Bengio, Y., Bohg, J., ... & Van de Panne, M. (2021). From machine learning to robotics: Challenges and opportunities for embodied intelligence. *arXiv preprint arXiv:2110.15245*

55. Chakraborty, D., Başağaoğlu, H., & Winterle, J. (2021). Interpretable vs. noninterpretable machine learning models for data-driven hydro-climatological process modeling. *Expert Systems with Applications*, *170*, 114498

56. Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M., & Klein, M. (2002). *Logistic regression* (p. 536). New York: Springer-Verlag.

57. LeCessie S, Van Houwelingen JC. Ridge estimators in logistic regression. J R Stat Soc Ser C (Appl Stat). 1992;41(1):191–201.

58. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.

59. Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*, 197-227.

60. Satija Lab. Guided tutorial: Analyzing PBMC scRNA-seq data [Internet]. Seurat; [cited year Month day]. Available from: https://satijalab.org/seurat/articles/pbmc3k_tutorial.html
61. Van Hoesen GW, Hyman BT, Damasio AR. Entorhinal cortex pathology in Alzheimer's disease. Hippocampus. 1991 Jan;1(1):1-8. doi: 10.1002/hipo.450010102. PMID: 1669339.
62. Li W, Qin W, Liu H, Fan L, Wang J, Jiang T, Yu C. Subregions of the human superior frontal gyrus and their connections. Neuroimage. 2013 Sep;78:46-58. doi: 10.1016/j.neuroimage.2013.04.011. Epub 2013 Apr 13. PMID: 23587692.
63. El-Baba RM, Schury MP. Neuroanatomy, Frontal Cortex. 2023 May 29. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan–. PMID: 32119370.
64. Chan Zuckerberg Initiative. Human Alzheimer's Disease Brain Collection [Internet]. cellxgene; [cited 2024 Aug 19]. Available from: https://cellxgene.cziscience.com/collections/180bff9c-c8a5-4539-b13b-ddbc00d643e6
65. . Leng K, Li E, Eser R, Piergies A, Sit R, Tan M, Neff N, Li SH, Rodriguez RD, Suemoto CK, Leite REP, Ehrenberg AJ, Pasqualucci CA, Seeley WW, Spina S, Heinsen H, Grinberg LT, Kampmann M. Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. Nat Neurosci. 2021 Feb;24(2):276-287. doi: 10.1038/s41593-020-00764-7. Epub 2021 Jan 11. PMID: 33432193; PMCID: PMC7854528.
66. Zhou W, Ma D, Sun AX, et al. PD-linked CHCHD2 mutations impair CHCHD10 and MICOS complex leading to mitochondria dysfunction. *Hum Mol Genet*. 2019;28(7):1100-1116. doi:10.1093/hmg/ddy413
67. Thompson LI, Cummings M, Emrani S, et al. Digital Clock Drawing as an Alzheimer's Disease Susceptibility Biomarker: Associations with Genetic Risk Score and APOE in Older Adults. *J Prev Alzheimers Dis*. 2024;11(1):79-87. doi:10.14283/jpad.2023.48

68. Mantle D, Falkous G, Ishiura S, Perry RH, Perry EK. Comparison of cathepsin protease activities in brain tissue from normal cases and cases with Alzheimer's disease, Lewy body dementia, Parkinson's disease and Huntington's disease. *J Neurol Sci*. 1995;131(1):65-70. doi:10.1016/0022-510x(95)00035-z

69. Matsuzaki Tada A, Hamezah HS, Pahrudin Arrozi A, Abu Bakar ZH, Yanagisawa D, Tooyama I. Pharmaceutical Potential of Casein-Derived Tripeptide Met-Lys-Pro: Improvement in Cognitive Impairments and Suppression of Inflammation in APP/PS1 Mice. *J Alzheimers Dis*. 2022;89(3):835-848. doi:10.3233/JAD-220192

70. Hooff GP, Peters I, Wood WG, Müller WE, Eckert GP. Modulation of cholesterol, farnesylpyrophosphate, and geranylgeranylpyrophosphate in neuroblastoma SH-SY5Y-APP695 cells: impact on amyloid beta-protein production. *Mol Neurobiol*. 2010;41(2-3):341-350. doi:10.1007/s12035-010-8117-5

71. Fisette JF, Montagna DR, Mihailescu MR, Wolfe MS. A G-rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing. *J Neurochem*. 2012;121(5):763-773. doi:10.1111/j.1471-4159.2012.07680.x

72. Wainberg M, Andrews SJ, Tripathy SJ. Shared genetic risk loci between Alzheimer's disease and related dementias, Parkinson's disease, and amyotrophic lateral sclerosis. *Alzheimers Res Ther*. 2023;15(1):113. Published 2023 Jun 16. doi:10.1186/s13195-023-01244-3

73. Silva PN, Furuya TK, Braga IL, et al. Analysis of HSPA8 and HSPA9 mRNA expression and promoter methylation in the brain and blood of Alzheimer's disease patients. *J Alzheimers Dis*. 2014;38(1):165-170. doi:10.3233/JAD-130428

74. Garces A, Martinez B, De La Garza R, et al. Differential expression of interferon-induced protein with tetratricopeptide repeats 3 (IFIT3) in Alzheimer's disease and HIV-1 associated

neurocognitive disorders. *Sci Rep*. 2023;13(1):3276. Published 2023 Feb 25. doi:10.1038/s41598-022-27276-7

75. Park SY, Seo J, Chun YS. Targeted Downregulation of *kdm4a* Ameliorates Tau-engendered Defects in *Drosophila melanogaster*. *J Korean Med Sci*. 2019;34(33):e225. Published 2019 Aug 26. doi:10.3346/jkms.2019.34.e225

76. Chu Y, Mickiewicz AL, Kordower JH. α-synuclein aggregation reduces nigral myocyte enhancer factor-2D in idiopathic and experimental Parkinson's disease. *Neurobiol Dis*. 2011;41(1):71-82. doi:10.1016/j.nbd.2010.08.022

77. Sun Y, Zhu J, Yang Y, et al. Identification of candidate DNA methylation biomarkers related to Alzheimer's disease risk by integrating genome and blood methylome data. *Transl Psychiatry*. 2023;13(1):387. Published 2023 Dec 13. doi:10.1038/s41398-023-02695-w

78. Parween F, Hossain MS, Singh KP, Gupta RD. Association between human paraoxonase 2 protein and efficacy of acetylcholinesterase inhibiting drugs used against Alzheimer's disease. *PLoS One*. 2021;16(10):e0258879. Published 2021 Oct 29. doi:10.1371/journal.pone.0258879

79. Cunningham CL, Martínez-Cerdeño V, Noctor SC. Microglia regulate the number of neural precursor cells in the developing cerebral cortex. *J Neurosci*. 2013;33(10):4216-4233. doi:10.1523/JNEUROSCI.3441-12.2013

80. Zhang H, Zhang Q, Tu J, You Q, Wang L. Dual function of protein phosphatase 5 (PPP5C): An emerging therapeutic target for drug discovery. *Eur J Med Chem*. 2023;254:115350. doi:10.1016/j.ejmech.2023.115350

81. Counts SE, Mufson EJ. Regulator of Cell Cycle (RGCC) Expression During the Progression of Alzheimer's Disease. *Cell Transplant*. 2017;26(4):693-702. doi:10.3727/096368916X694184

82. Suzuki M, Tezuka K, Handa T, et al. Upregulation of ribosome complexes at the blood-brain barrier in Alzheimer's disease patients. *J Cereb Blood Flow Metab*. 2022;42(11):2134-2150. doi:10.1177/0271678X221111602
83. Zaręba-Kozioł M, Burdukiewicz M, Wysłouch-Cieszyńska A. Intracellular Protein S-Nitrosylation-A Cells Response to Extracellular S100B and RAGE Receptor. *Biomolecules*. 2022;12(5):613. Published 2022 Apr 20. doi:10.3390/biom12050613
84. Huang AY, Zhou Z, Talukdar M, et al. Somatic cancer driver mutations are enriched and associated with inflammatory states in Alzheimer's disease microglia. Preprint. *bioRxiv*. 2024;2024.01.03.574078. Published 2024 Jan 4. doi:10.1101/2024.01.03.574078
85. Cho SJ, Yun SM, Jo C, et al. SUMO1 promotes Aβ production via the modulation of autophagy. *Autophagy*. 2015;11(1):100-112. doi:10.4161/15548627.2014.984283
86. Moll T, Shaw PJ, Cooper-Knock J. Disrupted glycosylation of lipids and proteins is a cause of neurodegeneration. *Brain*. 2020;143(5):1332-1340. doi:10.1093/brain/awz358
87. Ando S, Sakurai M, Shibutani S, et al. Age-related alterations in protein phosphatase 2A methylation levels in brains of cynomolgus monkeys: a pilot study. *J Biochem*. 2023;173(6):435-445. doi:10.1093/jb/mvad006
88. cellxgene.cziscience.com. Cellxgene Data Explorer [Internet]. Chan Zuckerberg Initiative; 2024 [cited 2024 Sep 5]. Available from: https://cellxgene.cziscience.com/collections/180bff9c-c8a5-4539-b13b-ddbc00d643e6
89. Zekić P. ROC analiza [završni rad]. Rijeka: Sveučilište u Rijeci, Tehnički fakultet; 2023.

8. ŽIVOTOPIS

OPĆI PODACI:

Ime i prezime: Klara Huzjak
Datum i mjesto rođenja: 05.12.1998., Varaždin, Republika Hrvatska
Državljanstvo: Hrvatsko
Adresa stanovanja: Ive Režeka 8., 42000 Varaždin
Kontakt: klara.huzjak1@gmail.com

OBRAZOVANJE:

2005. – 2013. III. Osnovna škola Varaždin
2013. – 2017. Prva privatna gimnazija s pravom javnosti Varaždin
2019. – 2022. Sveučilište u Splitu, Odjel zdravstvenih studija, preddiplomski sveučilišni studij medicinsko laboratorijske dijagnostike, Split
2022. – 2024. Sveučilište u Rijeci, Fakultet biotehnologije i razvoja lijekova, diplomski studij biotehnologija u medicini, Rijeka

RADOVI:
2022. Završni rad: Molekularna dijagnostika virusa SARS-CoV-2

STRANI JEZICI I OSTALO:
Engleski jezik: B2 razina
Francuski jezik: A2 razina
Talijanski jezik: A1 razina
Aktivno služenje MS Office-om
Vozačka dozvola kategorije B2