

# Analiza staničnog ekosustava tkiva štitnjače sekvenciranjem jednostanične RNK

---

Ilić, Klara

Master's thesis / Diplomski rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Medicine Osijek / Sveučilište Josipa Jurja Strossmayera u Osijeku, Medicinski fakultet Osijek**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:152:570903>

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

Download date / Datum preuzimanja: **2025-01-31**



Repository / Repozitorij:

[Repository of the Faculty of Medicine Osijek](#)



**SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU**

**MEDICINSKI FAKULTET OSIJEK**

**DIPLOMSKI SVEUČILIŠNI STUDIJ MEDICINSKO**

**LABORATORIJSKA DIJAGNOSTIKA**

**Klara Ilić**

**ANALIZA STANIČNOG EKOSUSTAVA  
TKIVA ŠTITNJAČE SEKVENCIRANJEM  
JEDNOSTANIČNE RNK**

**Diplomski rad**

**Osijek, 2022.**

**SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU**

**MEDICINSKI FAKULTET OSIJEK**

**DIPLOMSKI SVEUČILIŠNI STUDIJ MEDICINSKO**

**LABORATORIJSKA DIJAGNOSTIKA**

**Klara Ilić**

**ANALIZA STANIČNOG EKOSUSTAVA  
TKIVA ŠTITNJAČE SEKVENCIRANJEM  
JEDNOSTANIČNE RNK**

**Diplomski rad**

**Osijek, 2022.**

Rad je ostvaren u: Medicinski fakultet Osijek, Katedra za nuklearnu medicinu i onkologiju

Mentor rada: izv. prof. dr. sc. Mario Štefanić

Rad ima 43 lista, 1 tablicu i 23 slike.

*Hvala mojoj obitelji na stalnoj podršci tijekom cijelog obrazovanja.*

*Zahvaljujem mentoru, izv. prof. dr. sc. Mariju Štefaniću, na svim savjetima, stručnosti, strpljenju i velikoj pomoći tijekom izrade diplomskog rada.*

*Hvala mojim prijateljima i kolegama koji su učinili ovaj period ljepšim.*

# SADRŽAJ

|   |    |
|---|----|
| 1. UVOD.....  | 1  |
| 1.1. Štitnjača .....                                  | 1  |
| 1.1.1. Anatomija štitnjače .....                      | 1  |
| 1.1.2. Histologija štitnjače .....                    | 1  |
| 1.1.3. Funkcija štitnjače.....                        | 2  |
| 1.1.4. Stanični atlas .....                           | 3  |
| 1.2. Sekvenciranje jednostanične RNK (scRNA-Seq)..... | 4  |
| 1.2.1. Tehnike izolacije stanica .....                | 4  |
| 1.2.2. Preprocesiranje podataka .....                 | 5  |
| 1.2.3. Redukcija dimenzionalnosti .....               | 7  |
| 1.2.4. Grupiranje .....                               | 8  |
| 1.2.5. Vizualizacija.....                             | 8  |
| 1.2.6. Anotacija stanica.....                         | 9  |
| 1.2.7. Primjena scRNA-Seq.....                        | 9  |
| 2. HIPOTEZA .....                                     | 11 |
| 3. CILJ ISTRAŽIVANJA .....                            | 12 |
| 4. MATERIJALI I METODE .....                          | 13 |
| 5. REZULTATI.....                                     | 15 |
| 6. RASPRAVA .....                                     | 31 |
| 7. ZAKLJUČCI.....                                     | 33 |
| 8. SAŽETAK .....                                      | 34 |
| 9. ABSTRACT .....                                     | 35 |
| 10. LITERATURA .....                                  | 36 |
| 11. ŽIVOTOPIS .....                                   | 43 |

## POPIS KRATICA

|        |   |
|--------|---|
| Bmem   | memorijske B stanice  |
| CaSTLe | klasifikacija stanica prijenosnim učenjem (engl. <i>Classification of Single cells by Transfer Learning</i> )                         |
| cDC    | konvencionalne dendritičke stanice  |
| CIDR   | klasteriranje kroz imputaciju i redukciju dimenzionalnosti (engl. <i>Clustering through Imputation and Dimensionality Reduction</i> ) |
| CIPR   | prediktor identiteta klastera (engl. <i>Cluster Identity Predictor</i> )  |
| DCA    | engl. <i>Deep Count Autoencoder</i>   |
| DIT    | dijodotirozin (engl. <i>diiodotyrosine</i> )  |
| Dmaps  | mape difuzije (engl. <i>Diffusion maps</i> )  |
| EC     | endotelne stanice   |
| EGF    | epidermalni čimbenik rasta (engl. <i>epidermal growth factor</i> )  |
| FACS   | sortiranje stanica protočnom citometrijom (engl. <i>Flow-Activated Cell Sorting</i> )   |
| fcLVM  | engl. <i>Factorial Single-cell Latent Variable Model</i>  |
| GEO    | engl. <i>Gene Expression Omnibus</i>  |
| HLA    | Humani Leukocitni Antigen   |
| KEGG   | Kyoto Enciklopedija Gena i Genoma   |
| kNN    | k-najbliži susjedi (engl. <i>k-nearest neighbour</i> )  |
| LAmbDA | engl. <i>Label Ambiguous Domain Adaptation</i>  |
| LCM    | laserska mikrodisekcija (engl. <i>Laser Capture Microdissection</i> )   |
| LLE    | lokalno linearno ugnježđenje (engl. <i>Locally Linear Embedding</i> )   |
| MAGIC  | identifikacija gena markera (engl. <i>Marker Genes Identification</i> )   |
| MIT    | monojodotirozin (engl. <i>moniodotyrosine</i> )   |
| MNN    | uzajamno najbliži susjedi (engl. <i>Mutual Nearest Neighbour</i> )  |
| Moana  | multirezolucijska analiza transkriptoma (engl. <i>Multi-Resolution Transcriptome Analysis</i> )                                       |
| NTRK2  | tirozin-kinaza neurotrofnog receptora 2 (engl. <i>Neurotrophic Receptor Tyrosine Kinase 2</i> )                                       |
| PAGODA | analiza prekomjerne disperzije seta gena i putova (engl. <i>Pathway and Gene</i> )  |

|            |   |
|------------|---|
|            | <i>Set Overdispersion Analysis</i> )  |
| pDC        | plazmocitoidne dendritičke stanice  |
| PCA        | analiza glavnih komponenti (engl. <i>Principal Component Analysis</i> )   |
| SCANPY     | jednostanična analiza u Pythonu (engl. <i>Single-cell analysis in Python</i> )  |
| scATAC-Seq | jednostanična analiza kromatina dostupnog transpozazi (engl. <i>Single-cell Sequencing Assay for Transposase-Accessible Chromatin</i> )                               |
| scCATCH    | jednostanični anotacijski alat za staničnu heterogenost baziran na klasterima (engl. <i>Single-cell Cluster-based Annotation Toolkit for Cellular Heterogeneity</i> ) |
| SCINA      | polu-nadzirana kategorijska identifikacija i raspodjela (engl. <i>Semi-supervised Category Identification and Assignment</i> )  |
| scRNA-Seq  | sekvenciranje jednostanične RNK (engl. <i>Single-cell RNA Sequencing</i> )  |
| SCSA       | engl. <i>Single-cell Score Annotation</i>   |
| SMC        | glatke mišićne stanice (engl. <i>smooth muscle cells</i> )  |
| scVI       | engl. <i>Single-cell Variational Inference</i>  |
| T3         | trijodtironin (engl. <i>triiodothyronine</i> )  |
| T4         | tiroksin (engl. <i>thyroxine</i> )  |
| Tg         | tireoglobulin (engl. <i>thyroglobulin</i> )   |
| TCM        | centralno-memorijske T stanice  |
| TEM        | efektorsko-memorijske T stanice   |
| TFC        | folikularne stanice štitnjače (engl. <i>thyroid follicular cells</i> )  |
| TNF        | čimbenik nekroze tumora (engl. <i>tumor necrosis factor</i> )   |
| TSH        | tireostimulin (engl. <i>Thyroid Stimulating Hormone</i> )   |
| TR         | receptor za tiroidne hormone (engl. <i>thyroid hormone receptor</i> )   |
| t-SNE      | t-distribuirano stohastičko ugrađivanje susjeda (engl. <i>t-Stochastic Neighbour Embedding</i> )  |
| UMAP       | engl. <i>Uniform Manifold Approximation and Projection</i>  |
| UMI        | jedinstveni molekularni identifikatori (engl. <i>Unique Molecular Identifiers</i> )   |
| ZINB-Wave  | engl. <i>Zero-Inflated Negative Binomial-based Wanted Variation Extraction</i>  |



## POPIS TABLICA

Tablica 1. Broj (N) i udio (%) pojedinih staničnih populacija u integriranom skupu.....18

## POPIS SLIKA

|  |    |
|--|----|
| Slika 1. Stanične subpopulacije, UMAP ( <i>Uniform Manifold Approximation and Projection</i> ) reprezentacija .....  | 15 |
| Slika 2. Razdioba populacija u glavnim staničnim razredima. UMAP = <i>Uniform Manifold Approximation and Projection</i> .....                                    | 16 |
| Slika 3. Izražaj kanoničkih staničnih biljega (normalizirane vrijednosti) u populacijama.....  | 16 |
| Slika 4. Prikaz stanica u S i G2M fazi staničnog ciklusa .....   | 17 |
| Slika 5. Izražaj linijski specifičnih biljega (A) i 10 vodećih razlikovnih biljega (B) u glavnim staničnim razredima.....  | 17 |
| Slika 6. Učešće pojedinih uzoraka u staničnim populacijama .....   | 18 |
| Slika 7. Točkasti dijagram ekspresije razlikovnih biljega u pojedinim subpopulacijama mijeloidnih stanica .....  | 19 |
| Slika 8. Točkasti dijagram ekspresije razlikovnih biljega u pojedinim subpopulacijama dendritičkih stanica (DC).....   | 20 |
| Slika 9. Modul procesiranja i prezentacije antigena.....   | 21 |
| Slika 10. Razlikovni izražaj gena u M1 i M2 populaciji makrofaga (Bland-Altman reprezentacija) .....   | 22 |
| Slika 11. UMAP ( <i>Uniform Manifold Approximation and Projection</i> ) T staničnih populacija.. .....   | 22 |
| Slika 12. UMAP ( <i>Uniform Manifold Approximation and Projection</i> ) reprezentacija ekspresije CD3D, CD4, CD8A i TRGC2 gena u staničnim subpopulacijama ..... | 23 |
| Slika 13. Točkasti dijagram ekspresije razlikovnih biljega u pojedinim subpopulacijama CD4+ centralno-memorijskih T (TCM) stanica.....                           | 23 |
| Slika 14. UMAP ( <i>Uniform Manifold Approximation and Projection</i> ) reprezentacija ekspresije biljega disfunkcionalnih, iscrpljenih T stanica .....          | 24 |

|   |    |
|---|----|
| Slika 15. UMAP ( <i>Uniform Manifold Approximation and Projection</i> ) reprezentacija B staničnih populacija .....   | 24 |
| Slika 16. Točkasti dijagram ekspresije razlikovnih biljega u pojedinim subpopulacijama B stanica .....  | 25 |
| Slika 17. Dvodimenzionalna (UMAP, <i>Uniform Manifold Approximation and Projection</i> ) reprezentacija epitelnih subpopulacija.....  | 25 |
| Slika 18. Točkasti dijagram ekspresije razlikovnih biljega u pojedinim subpopulacijama folikularnih epitelnih stanica (TFC) .....   | 26 |
| Slika 19. UMAP ( <i>Uniform Manifold Approximation and Projection</i> ) reprezentacija ekspresije biljega mezenhimske tranzicije .....  | 27 |
| Slika 20. Violinske mape izražaja NTRK2 (tirozin-kinaza neurotrofnog receptora 2) u folikularnim epitelnim stanicama štitnjače (TFC) i njihovim subpopulacijama.....  | 27 |
| Slika 21. Dvodimenzionalna (UMAP, <i>Uniform Manifold Approximation and Projection</i> ) reprezentacija ekspresije razlikovnih biljega epitelnih subpopulacija .....  | 28 |
| Slika 22. Dvodimenzionalna (UMAP, <i>Uniform Manifold Approximation and Projection</i> ) reprezentacija izražaja razlikovnih biljega mezenhimskih stanica .....   | 29 |
| Slika 23. Dvodimenzionalna (UMAP, <i>Uniform Manifold Approximation and Projection</i> ) reprezentacija (A) izražaja razlikovnih biljega endotelnih stanica (EC). Biljezi arterijskih, kapilarnih i venskih endotelnih stanica (B)..... | 30 |

## 1. UVOD

### 1.1. ŠTITNJAČA

#### 1.1.1. ANATOMIJA ŠTITNJAČE

Štitnjača je neparna endokrina žlijezda smještena u prednjem dijelu vrata. U normalnim okolnostima štitnjača na čovjeku nije vidljiva niti opipljiva. Žlijezda ima oblik slova H – sastoji se od dva postranična režnja, *lobus dexter* i *lobus sinister*, povezanih suženjem, *isthmus*. *Lobus pyramidalis*, uski izdanak koji izlazi iz istmusa ili jednog od režnjeva, pričvršćen je vezivnim tkivom na podjezičnu kost. Vezivna ovojnica koja obavija štitnjaču, *capsula fibrosa*, stvara vezivne pregrade kojima dijeli tkivo štitnjače na režnjeve (*lobuli glandulae thyroideae*) te je pričvršćuje za grkljan i dušnik, a dodatno je učvršćuju vezivni snopovi koji prolaze od istmusa do prstenaste hrskavice, perivaskularno vezivno tkivo žila koje ulaze u štitnjaču te pomično, rahlo vezivno tkivo koje je povezuje s okolnim strukturama. Dvije parne arterije koje opskrbljuju štitnjaču, *arteria thyroidea inferior* i *arteria thyroidea superior*, koje se granaju u mrežu između vanjske i unutarnje ovojnice te njihovi ogranci ulaze u tkivo žlijezde. U 10 % slučajeva do istmusa dolazi neparna *arteria thyroidea ima*. Vensku mrežu čine *venae thyroideae superiores*, *venae thyroideae mediae* i *venae thyroideae inferiores* (1, 2). Štitnjača je jedan od najbolje vaskulariziranih organa, a endotelne stanice kapilara su fenestrirane, što olakšava prelaženje hormona u cirkulaciju (3). Tanka vezivna stroma sadrži i opsežnu mrežu živaca i limfnih žila koje omogućuju migraciju i recirkulaciju imunskih stanica između štitnjače i regionalnih limfnih čvorova. Inervacija štitnjače dolazi preko srčanih i laringealnih ogranaka vagusnog živca, a postganglijska vlakna djeluju na krvne žile i preko njih posredno na štitnjaču. Zajedno s folikularnim epitelom štitnjače, navedene komponente tvore složeni stanični ekosustav čiji proizvod, hormoni štitnjače, djeluju gotovo na sve stanice u tijelu, upravljajući brojnim aspektima rasta, razvoja, stanične energetike i dugovječnosti (1, 4, 5).

#### 1.1.2. HISTOLOGIJA ŠTITNJAČE

Štitnjača je izgrađena od folikula, mjehurića različite veličine (50-300 µm) obloženih specijaliziranim jednoslojnim epitelom. Lumen folikula ispunjen je koloidom, homogenom želatinoznom tvari koja se sastoji od glikoproteina velike molekularne mase, tireoglobulina (Tg), koji sadržava hormone štitnjače (3, 6).

## 1. UVOD

Folikularne stanice mogu biti pločaste, kubične ili cilindrične. Oblik stanica folikularnog epitela određuje njihova funkcionalna aktivnost. Pločasti i kubični folikularni epitel predstavljaju fazu relativnog mirovanja, dok cilindrični epitel predstavlja visoku sintetsku aktivnost. Glavni regulator aktivnosti štitnjače je *tireostimulin* ili tireotropin (TSH) koji se izlučuje u hipofizi i potiče sintezu hormona, a time i porast visine folikularnog epitela, smanjenje folikula i količine koloida u njima (2, 3). U razvojnog smislu, folikularni epitel je produkt regionalne specijalizacije i migracije endoderma primitivne probavne cijevi (6). Malobrojne parafolikularne stanice ili C stanice nešto su veće od folikularnih stanica. Nalaze se unutar folikularnog epitela ili su u odvojenim nakupinama izvan njega. Brojna zrnca unutar stanica sadrže hormon kalcitonin koji snižava kalcij u krvi zbog suprimiranja razgradnje kosti, a izlučuje se porastom koncentracije kalcija u krvi (3).

### 1.1.3. FUNKCIJA ŠTITNJAČE

Glavna je funkcija štitnjače izlučivanje hormona tiroksina (T4), trijodtironina (T3) i kalcitonina. Iako tiroksin (T4) čini 90 % količine hormona štitnjače, on se gotovo cijeli u tkivima pretvara u trijodtironin (T3) koji djeluje brže i snažnije od T4. Hormoni štitnjače povećavaju staničnu metaboličku aktivnost, imaju utjecaj na krvožilni sustav, na rast organizma i razvoj živčanog sustava tijekom fetalnog razvoja (3).

Proces stvaranja hormona štitnjače započinje sintezom tireoglobulina u endoplazmatskom retikulumu i Golgijevom aparatu te njegovim izlučivanjem u lumen folikula. Folikularne stanice pomoću natrij-jodnog sunosača unose jodid iz krvi koji se zatim aktivira oksidacijom. Oksidacija jodida se događa pomoću enzima peroksidaze te se prenosi u lumen folikula ionskim nosačem pendrinom. U lumenu slijedi proces koji katalizira enzim peroksidaza – organifikacija tireoglobulina, odnosno vezanje joda s molekulom tireoglobulina, pri čemu nastali T3 i T4 postaju dio molekule tireoglobulina. Na poticaj TSH izlučenog iz hipofize, u lizosomima se djelovanjem proteaza kidaju peptidne veze između molekula tireoglobulina i jodiranih ostataka, nakon čega se T3, T4, dijodtirozin (DIT) i monojodtirozin (MIT) oslobađaju u citoplazmu (3, 7). T3 i T4 se otpuštaju u kapilare čiji endotel pokriva 25-50 % površine folikula, a MIT i DIT uklanjaju jod djelovanjem jodotirozin dehalogenaze te se tako nastali jod i tirozin ponovno iskorištavaju u folikularnim stanicama za sintezu hormona. U perifernim tkivima, T4 i T3 ulaze u stanice putem specifičnih transportera, nakon čega ostvaruju glavnu ulogu vezanjem na nuklearne TR $\alpha$  i TR $\beta$  receptore. U najopćenitijem smislu, sinteza, pohrana i otpuštanje T4 i T3 u cirkulaciju koordinirani je pothvat koji

## 1. UVOD

zahtijeva suradnju brojnih specijaliziranih vrsta stanica s različitim ulogama u održavanju tkivne funkcije i integriteta (3, 8, 9).

Posljedično, cjelovit i detaljan opis staničnog krajobraza štitnjače i dalje je nepotpun, a štitnjača ostaje jedno od rijetkih tkiva za koje još uvijek nedostaje sveobuhvatan atlas jednostanične rezolucije u zdravlju i bolesti. U drugim tkivima, ovakvi katalozi temeljito su izmijenili razumijevanje biologije i embriologije tkiva, ali za štitnjaču, precizno mapiranje kojim bi se nadoknadile manjkavosti postojećih staničnih atlasa tek predstoji (10, 11).

### 1.1.4. STANIČNI ATLASI

Referentni stanični atlas (Human Cell Atlas, Tabula Sapiens, Tabula Muris) danas predstavljaju poredbeni standard u seciranju ekosustava oboljelih, upalno promijenjenih i maligno preobraženih tkiva. Najčešće, identitet stanica je definiran genskim izražajem karakterističnih biljega, u čemu prednjači sekvenciranje jednostanične RNK (engl. *Single-cell RNA Sequencing* – scRNA-Seq), a slijedi ju jednostanična analiza otvorenog kromatina (engl. *Single-cell sequencing Assay for Transposase-Accessible Chromatin* – scATAC-Seq) (12, 13, 14). Naime, unatoč činjenici da većina stanica ima identičan genom, tkiva i tipovi stanica ipak se međusobno razlikuju po izražaju specifičnih gena što omogućava prisutnost velikog opsega staničnih fenotipova u organizmu. Razvoj scRNA-Seq i scATAC-Seq omogućio je određivanje profila genskog izražaja i genske regulacije svake stanice što otkriva heterogenost staničnih populacija te daje značajniji uvid u ponašanje stanice (15, 16). Obje tehnike omogućuju simultanu analizu tisuća pojedinačnih stanica na cjelogenomskoj razini omogućujući sistematsko otkrivanje i karakterizaciju nepoznatih stanica u složenim tkivima (14).

### 1.2. SEKVENCIRANJE JEDNOSTANIČNE RNK (scRNA-Seq)

Prvotno objavljena 2009. godine, analiza jednostaničnog transkriptoma (bazirana na metodi sekvenciranja sljedeće generacije) opisala je karakter stanica od ranih stupnjeva razvoja, nakon čega je počeo rasti interes za heterogenost stanica. Metodom scRNA-Seq mogu se otkriti kompleksne i rijetke populacije stanica, regulatorni međustanični odnosi, može se pratiti razvoj različitih staničnih linija. Unatoč ograničenjima, tehnologija za analizu se stalno poboljšava te omogućuje izradu detaljnog atlasa genskog izražaja svake stanice u ljudskom organizmu. Ukratko, svaki protokol metode scRNA-Seq uključuje tri glavna koraka: izolaciju pojedinačnih stanica (disocijacija uzorka i razvrstavanje u mikrojažice), pripremu knjižnice (reverzna transkripcija i amplifikacija cijele glasničke RNK (engl. *messenger*, mRNK) ili jednog kraja mRNK i sekvenciranje (25 000 – 5 000 000 sekvenci po stanici) (16 – 18).

#### 1.2.1. TEHNIKE IZOLACIJE STANICA

U upotrebi su mnoge metode izolacije koje omogućavaju analizu velikog broja stanica. Limitirana dilucija je tehnika koja izolira pojedinačne stanice pipetom, no nije efikasna. Mikromanipulacija je tehnika izolacije kapilarnom pipetom uz pomoć mikroskopa za uzorke s malim brojem stanica (embrionalne stanice, nekultivirani mikroorganizmi), ali je niskoprotočna i dugo traje. Sortiranje protočnom citometrijom, FACS (engl. *Flow-Activated Cell Sorting*), se primjenjuje za izolaciju stanica koje izražavaju male količine markera označavajući ih fluorescentnim monoklonalnim protutijelom koje prepoznaje specifične markere na površini, ali zahtijeva veliki početni volumen uzorka. Laserska mikrodisekcija, LCM (engl. *Laser Capture Microdissection*), je tehnika koja koristi laser vođen kompjutorskim sustavom kako bi se izolirale stanice iz solidnih uzoraka. Mikrofluidika je tehnika koja koristi malu količinu uzorka, jeftinija je te koristi nanolitarske volumene što smanjuje rizik od vanjske kontaminacije. Trenutno najkorištenija metoda je tehnika mikrokapljica (engl. *microdroplet-based microfluids*) koja se temelji na monodisperziji vodenih kapljica u kontinuiranoj uljnoj fazi (17, 19). Ova metoda je visokoprotočna te koristi manju količinu uzorka. Na ovoj metodi se temelji rad komercijalnih platformi kao što su inDrop, Drop-seq, Chromium (10xGenomics), Nadia. Uspoređujući, primjerice, Chromium platformu s Drop-seq i inDrop platformama, prednosti Chromium platforme su nešto bolja kvaliteta jer može detektirati više gena po stanici, veća osjetljivost, manja tehnička buka, veća efikasnost, lakše korištenje, ali nedostatak ove platforme je veća cijena u odnosu na druge.

## 1. UVOD

Nakon liziranja stanica kako bi se izdvojila RNK, slijedi reverzna transkripcija RNK u komplementranu, cDNK koja se zatim amplificira kako bi se pripremila knjižnica sekvenci (16, 20, 21).

### 1.2.2. PREPROCESIRANJE PODATAKA

Proizvod scRNA-Seq sadrži sve sekvencirane komplementarne DNK. Takvi sirovi, odnosno izvorni podatci moraju proći određene korake predanalize koji pretvaraju sirove podatke u lakše upotrebljivi oblik (15).

#### Izrada tablice genskog izražaja

Izrada tablice genskog izražaja (engl. *gene expression matrix*) prvi je korak scRNA-Seq analize i polazište za ostale korake u analizi. Tablica genskog izražaja se sastoji od stupaca (stanice) i redova (geni), a uneseni brojevi prikazuju broj molekula svakog detektiranog gena u svakoj stanici (15).

#### Kontrola kvalitete i isključenje dubleta

S obzirom na to da je tijekom izdvajanja stanica i sekvenciranja moguća pojava stanične smrti i stvaranja grupa (klastera) mrtvih stanica, potrebno je napraviti kontrolu kvalitete čime se isključuju sekvence koje ne pripadaju živim stanicama ili nemaju značaj. Jedan način je izračunati prag minimalnog broja jedinstvenih molekularnih identifikatora, UMI-a (engl. *Unique Molecular Identifiers*). Drugi je način razlikovanje živih stanica od oštećenih ili mrtvih stanica. Stanice sa prisutnošću mitohondrijskih sekvenci (indikator stanične smrti) ili niskim cDNK očitanjem ili brojem UMI-a se filtriraju. Uz to se može postaviti prag za minimalni broj stanica u kojima je detektiran određeni gen, ali se pritom ne postavlja previsoki prag kako se ne bi isključile rijetke vrste stanica.

Pojava dubleta može rezultirati stvaranjem umjetnih klastera. Dubleti se stvaraju kada su dvije ili više stanica uhvaćene u jednu kapljicu ili komoru te se moraju pažljivo identificirati kako se ne bi zamijenile za nove stanične tipove. S obzirom na veliki broj tipova stanica, njihovo ručno uklanjanje predstavlja izazov, stoga se koriste simulirani dubleti (kombinirane stanice iz različitih grupa) te se pronalaze stanice koje imaju njima slične profile genskog izražaja. U tu svrhu se često koriste programi scrublet i DoubletFinder (15, 18, 22).



## 1. UVOD

### Normalizacija podataka

Količina RNK po stanici može značajno varirati ovisno o stadiju staničnog ciklusa i drugim biološkim faktorima, čak i kod stanica istog tipa. Ako nisu normalizirane, razlike u ukupnoj količini UMI-a koje mogu nastati zbog tehničkih faktora (efikasnost reverzne transkripcije, efikasnost vezanja početnice, greške povezane s raspadanjem UMI-a) mogu utjecati na ostale korake analize. Primjerice, moguće je povezivanje stanica sa sličnom količinom UMI-a u klastere umjesto povezivanje stanica sa sličnim genskim izražajem. Ovaj korak je bitan i za mogućnost uspoređivanja genskog izražaja unutar uzorka i između uzoraka. Korištenje sustava kao što su SCnorm ili sctransform omogućava normalizaciju podataka bez prevelike korekcije manje izraženih gena, čime se ne utječe na vidljivost heterogenosti stanica (15, 21, 22).

### Stabilizacija varijance

Stabilizacija varijance je isključivanje utjecaja jačine genskog izražaja na gensku varijancu. Ovaj korak omogućava fokusiranje na određene gene na osnovi njihove relevantnosti, a ne samo na osnovi jačine izražaja. Jedan način stabilizacije je logaritamska transformacija normaliziranih očitavanja kojom se smanjuje razlika između gena s visokim i niskim izražajem. Takvim pristupom se koriste sustavi Seurat, Pagoda i SCANPY (engl. *Single-Cell Analysis in Python*). Drugi pristup je negativna binominalna distribucija koju koriste sustavi kao što su ZINB-Wave (engl. *Zero-Inflated Negative Binomial-based Wanted Variation Extraction*), scVI (engl. *Single-cell Variational Inference*), DCA (engl. *Deep Count Autoencoder*) (15).

### Skriveni slijedovi (*batch* efekti) i integracija podataka

Analiza više skupova podataka dobivenih drugačijom tehnologijom, od više pacijenata ili uzoraka, iz različitih eksperimenata i slično omogućava analizu većeg broja stanica. Međutim, to može rezultirati razlikama u genskom izražaju unutar skupa (zbog razlike između tipova stanica i biološki relevantnih faktora) ili između skupova (zbog tehničkih razlika). Razlike se mogu pojaviti zbog varijabilnosti pacijenata, uzoraka, protokola. Također, veliki su izazov prilikom spajanja podataka iz različitih eksperimenata. Alternativna metoda je njihovo uspoređivanje, što je korisno posebno kada je jedan od skupova podataka jako velik, primjerice stanični atlas. Jedan od načina uklanjanja skrivenih slijedova je pretpostavka da svi

## 1. UVOD

skupovi imaju jednak prosječan izražaj gena svih stanica te uklanjanje svih razlika u skupovima koristeći model regresije. Regresija može procijeniti povezanost genskog izražaja i tehničkih ili bioloških faktora. Nedostatak ove metode je moguća prevelika korekcija kada skupovi nisu identični, primjerice kada su proporcije tipova stanica među skupovima različite. Metode specifične za integraciju scRNA-Seq podataka omogućavaju očuvanje razlike u proporciji tipova stanica među skupovima tijekom uklanjanja skrivenih slijedova. Koncept metoda je pronalaženje parova stanica između skupova koje odgovaraju istom tipu ili stanju stanice. Parovi MNN (engl. *Mutual Nearest Neighbours* – najbliži susjedi) su stanice iz različitih skupova koje imaju najbliži genski izražaj. Za njihove međusobne razlike se smatra da su posljedica skrivenih slijedova te se one ispravljaju (15, 22).

### 1.2.3. REDUKCIJA DIMENZIONALNOSTI

Metoda scRNA-Seq pruža informacije o svim genima, što je jako bitno za nova saznanja u biologiji. Međutim, simultana analiza tisuća gena predstavlja izazove u statističkoj obradi. Kada se uspoređuju stanice u visoko dimenzionalnom prostoru genskog izražaja, udaljenosti između stanica postaju homogenije, što otežava odvajanje razlika između populacija od varijabilnosti unutar populacije (18). Dva su glavna pristupa rješavanju ovog problema. Redukcija dimenzionalnosti je sumiranje velikog broja varijabli, ali zadržavanje što više podataka. Dimenzionalnost skupa podataka je broj gena mjerenih za svaku stanicu. Broj zadržanih komponenti za daljnju analizu će ovisiti o kompleksnosti skupa podataka. Reducirane dimenzije se obično nazivaju ugnježđenja (engl. *embedding*). Ovaj korak nije nužan za tijek analize, ali je od pomoći zbog ubrzanja analize. Neki od modela redukcije dimenzionalnosti su linearna redukcija dimenzionalnosti (engl. *Principal Component Analysis* – PCA), nelinearna redukcija dimenzionalnosti (engl. *Locally Linear Embedding* – LLE, *Diffusion maps* – Dmaps, *Factorial Single-cell Latent Variable Model* – fcLVM, *Single-cell Variational Inference* – scVI) i inflacija nule (engl. *zero inflation*; takvi modeli su MAGIC (engl. *Marker Genes Identification*), kNN (engl. *k-nearest neighbour*), CIDR (engl. *Clustering through Imputation and Dimensionality Reduction*), scImpute) (15, 22). Izbor obilježja (engl. *feature selection*) identificira gene s najjačim biološkim signalom u odnosu na tehničku buku. Analiza se ograničava na gene koji nose najviše relevantnih informacija čime se smanjuje efekt dimenzionalnosti, smanjuje se buka i pojednostavljuje se metoda. Najčešće korištena metoda je uzimanje visoko varijabilnih gena. Ograničenje mnogih metoda za selekciju je to da promatraju opću varijabilnost u cijelom skupu podataka, što može dovesti do toga da se ne detektiraju rijetki stanični tipovi jer takve stanice imaju samo mali doprinos

## 1. UVOD

ukupnoj varijabilnosti. Metode za postupanje s visoko dimenzioniranim podacima se međusobno ne isključuju te se često provodi više metoda (22).

### 1.2.4. GRUPIRANJE

Klasteriranje stanica (engl. *clustering*) podrazumijeva identifikaciju i grupiranje stanica iz heterogenog skupa podataka prema sličnostima u obrascima genskog izražaja. Klasteri obično odgovaraju različitim tipovima stanica u skupu podataka. Pojedine od ovih grupa mogu predstavljati tipove stanica, a ostale se mogu gledati kao prijelazna stanja stanice, primjerice faze staničnog ciklusa (15, 22). Svaki klaster obično izražava set gena koji nisu izraženi u stanicama drugih klastera. K-means je jedna od korištenijih metoda klasteriranja stanica koja dodjeljuje stanice najbližem klasteru. S obzirom na to da je potrebno predodrediti broj klastera, moguće je određivanje većeg broja klastera od pretpostavljenog, a potom dijeljenje većih klastera ili spajanje susjednih (15, 19). Hijerarhijski način grupiranja (CIDR, BackSPIN) je potreban za veće skupove podataka ( $> 100\,000$  stanica) s puno različitih tipova stanica. Ovaj način obuhvaća inicijalnu identifikaciju definiranih tipova stanica koji se potom dijele u podgrupe. k-means i hijerarhijska metoda klasteriranja su spore za veće količine podataka i ograničeni su tipovi klastera koji se mogu detektirati. Seurat, PAGODA (engl. *Pathway And Gene set Overdispersion Analysis*), SCANPY i CellRanger klasteriraju algoritmima koji se temelje na grafovima, što je brža metoda i stvara biološki relevantne klastere. Graf se izrađuje povezivanjem svake stanice s njezinim najbližim susjedom. Takve metode kao što su Leiden ili Louvain funkcioniraju dobro i poprilično brzo. Drugi načini klasteriranja uključuju SC3Consensus Clustering koji koristi konsenzus više metoda klasteriranja, što povećava točnost, a koristi se kod vrlo heterogenih staničnih populacija koje je teško interpretirati, primjerice kod stanica karcinoma (15).

### 1.2.5. VIZUALIZACIJA

Vizualizacija je izrada 2D ili 3D skice raspršivanja stanica (engl. *scatterplot*), a najpoznatiji modeli vizualizacije su t-SNE (engl. *t-Stochastic Neighbour Embedding*) te UMAP (engl. *Uniform Manifold Approximation and Projection*). UMAP prikazuje topologiju podataka koristeći veze između najbližih susjeda (stanica iz različitih skupina koje imaju najbliži genski izražaj) te odabire podatke umanjene dimenzionalnosti koji održavaju globalnu strukturu. t-SNE stvara 2D vizualizacije koje održavaju separaciju staničnih tipova, ali ne održava globalnu strukturu (15, 22).

### 1.2.6. ANOTACIJA STANICA

Standardni protokol identifikacije staničnih tipova je pronalaženje gena koji su svojstveni za svaki klaster i uočavanje podudarnosti s listom kanoničkih markera za tipove stanica. U Seurat i Pagoda protokolima najčešće se koristi Wilcoxonov test. S obzirom na to da ručno proučavanje liste gena markera oduzima puno vremena te je podložno subjektivnosti, razvijeni su automatizirani načini anotacije stanica. Jedan od načina identifikacije je korištenje javno dostupnih lista gena markera specifičnih za određeni tip stanice koje mogu biti izvedene iz podataka prijašnjih scRNA-Seq analiza ili izrađene ručno (engl. *Single-cell Cluster-based Annotation Toolkit for Cellular Heterogeneity* – scCATCH, *Single-Cell Score Annotation* – SCSA, *Semi-supervised Category Identification and Assignment* – SCINA, Cell Assign). Bolji je način uspoređivanje genskog izražaja s poznatim referentnim anotiranim podacima scRNA-Seq analiza. Uspoređuju se ili pojedinačne stanice (scmap, SingleR, scMatch) ili klasteri (*Cluster Identity Predictor* – CIPR, ClustifyR). Anotacija nadziranom klasifikacijom koristi metode strojnog učenja za uvježbavanje identifikatora na referentnim scRNA-Seq podacima te se zatim koristi na neidentificiranim tipovima stanica (engl. *Classification of Single Cells by Transfer Learning* – CaSTLe, *Label Ambiguous Domain Adaptation* – LAmbDA, *Multi-resolution Transcriptomic Analysis* – Moana, superCT) – takve automatizirane metode su pogodne, ali zahtijevaju postojeće referentne scRNA-Seq skupove podataka tako da je ručna anotacija potrebna za nove tipove stanica. Također, uz referentne skupove podataka, ručna provjera gena markera je nužna za validaciju tipova stanica (15, 23).

### 1.2.7. PRIMJENA scRNA-Seq

Primjena metode scRNA-Seq je revolucionalizirala biomedicinska istraživanja analizom transkriptoma pojedinačnih stanica iz različitih uzoraka tkiva i sve češće se koristi u raznim područjima medicine:

- proučavanje heterogenosti stanica tumora (tumorske, vaskularne, imunosne stanice, fibroblasti), pronalazak boljih dijagnostičkih i prognostičkih biomarkera za tumore, učinkovitije antitumorske terapije, smanjenje otpornosti na lijekove)
- istraživanje genskih izražaja koji su pozadina razvoja organa i različitih putova diferencijacije stanica
- proučavanje patogenih procesa

## 1. UVOD

- stvaranje atlasa humanih stanica (engl. *Human Cell Atlas*) koji bi zabilježio 35 trilijuna stanica ljudskog organizma, postojeće tipove stanica te identificirao nove stanične tipove, međustanične reakcije, obrasce genskog izražaja i posljedične razlike u razvoju i diferencijaciji stanica, molekularne mehanizme (16, 21)

Na tragu posljednjeg, u ovom istraživanju napraviti će se cenzus staničnih (sub)populacija u ljudskom tkivu štitnjače na jednostaničnoj rezoluciji koristeći javno dostupne sirove scRNA-Seq podatke tkivnih uzoraka benigne histologije.

## 2. HIPOTEZA

Tkivo štitnjače složeni je stanični ekosustav s epitelnom, mezenhimskom, vaskularnom i imunosnom komponentom.

### 3. CILJ ISTRAŽIVANJA

Cilj istraživanja je anotacija i cenzus staničnih subpopulacija ljudskog tkiva štitnjače na jednostaničnoj rezoluciji uporabom scRNA-Seq.

## 4. MATERIJALI I METODE

Za ovo istraživanje korišteni su cjelogenomski scRNA-Seq podatci (ekspresijske matrice, brojevi očitavanja) pet neovisnih uzoraka ljudskog tkiva štitnjače benigne histologije [GSE134355: GSM4008659+GSM4008660 (Drop-Seq, Illumina HiSeqXTen); GSE158291: GSM4796530 (3 donora, BD Rhapsody Single-Cell Analysis System, Illumina NovaSeq 6000)] preuzeti iz *Gene Expression Omnibus* (GEO) repozitorija. Anonimizirani podatci su javno dostupni, a procesni detalji su ranije publicirani (24, 25). Svi korišteni podatci pripadaju javnoj domeni. Za istraživanje je pribavljena suglasnost Etičkog povjerenstva Medicinskog fakulteta u Osijeku (br. 2158-61-46-22-91, 30. 04. 2022.).

### Obrada podataka

Nakon isključenja dubleta [DoubletFinder, (26)] i analize stanične kontaminacije [celda/deconX, (27)], za daljnju obradu korišten je Seurat v4.0.2 (28). Za filtriranje stanica korišten je slijedeći kriterij:  $200 < \text{nFeature\_RNA} < 3000$  (GSE134355),  $800 < \text{nFeature\_RNA} < 5000$ ,  $2000 < \text{nCount\_RNA} < 20\ 000$  (GSE158291). Uklonjenje su sve stanice s više od 20 % mitohondrijskih gena. Nakon normalizacije ekspresije i stabilizacije varijance regulariziranom negativnom binomnom regresijom [SCTransform funkcija, (28)], izdvojeno je vodećih 2000 varijabilnih gena (FindVariableFeatures funkcija) i provedena je integracija pojedinačnih scRNA-Seq setova uporabom FindIntegrationAnchors (dims = 1:50) → IntegrateData funkcije (29). Nakon analize glavnih komponenti, grupiranje stanica provedeno je uporabom FindAllClusters funkcije [graf k-najbližih susjeda, Louvain algoritam, (30)], a vizualizacija primjenom nelinearne redukcije dimenzionalnosti [*Uniform Manifold Approximation and Projection*, rezolucija 0,8-1,2, prvih 20 glavnih komponenti, (31)]. Analiza razlikovno izraženih gena u pojedinim staničnim subpopulacijama provedena je Wilcoxonovim testom (FindConservedMarkers funkcija, > 25 % stanica, omjer izražaja  $|\log_2\text{FC}/\text{fold change}| > 0,25$ , Bonferroni  $P < 0,05$ ). Produkt ove analize, lista staničnih biljega za svaku subpopulaciju, upotrijebljena je za ručnu i asistiranu anotaciju uz pomoć *Cluster Identity Predictor* algoritma (32) i Azimuth 0.4.6 aplikacije [azimuth.hubmapconsortium.org, 23-08-2022, 10x Genomics, v3 PBMC10k, I.3 anotacija, *The Human BioMolecular Atlas Program* (13), National Institutes of Health, Bethesda, Maryland, SAD]. Za analizu faza staničnog ciklusa (CellCycleScoring funkcija) korištena je lista biljega iz rada Tirosh i sur. (33). Za analizu biljega disocijacijskog stresa (AddModuleScore funkcija) korišteni su genski potpisi iz rada van der Brink i sur. (34). Biljezi procesiranja i prezentacije antigena preuzeti su



#### 4. MATERIJALI I METODE

iz Chen i sur. (35)., a lista biljega epitelno-mezenhimske tranzicije preuzeta je iz Cook i sur. (36). Biljezi endotelinih stanica preuzeti su iz (37-39). Analiza bioloških putova koji su zastupljeni u listi staničnih biljega napravljena je uporabom H (Hallmark) i C2 seta [*Kyoto Encyclopedia of Genes and Genomes*, KEGG, (40)] MSigDB zbirke v7.4 [*The Molecular Signatures Database* (41), University of California San Diego / Broad Institute, Cambridge, Massachusetts, SAD]. Prag značajnosti definiran je kao FDR (*false discovery rate*)  $< 0,05$ , i to za genske skupove s minimalno 10 gena (*Gene Set Enrichment* algoritam).

Za genska imena korišteni su HGNC (*The HUGO Gene Nomenclature Committee*) simboli.

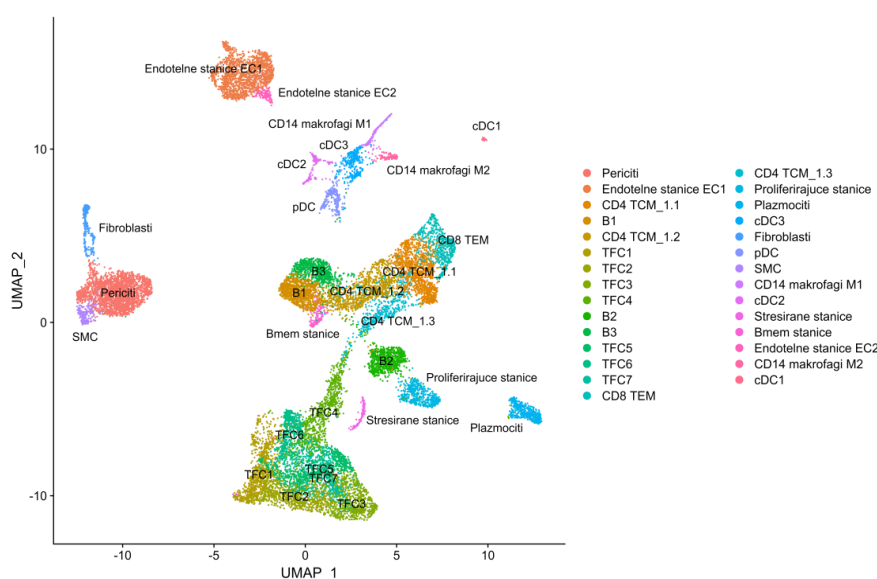
#### Statističke metode

Deskriptivna analiza slijedi uobičajena pravila prikaza podataka. Prosječni izražaj odgovara aritmetičkoj sredini. Analitički pristup scRNA-Seq podacima slijedi preporučena pravila (17, 27). Ako nije drugačije naznačeno, preprocesiranje, anotacija i vizualizacija rezultata (ComplexHeatmap v2.6.2, dittoSeq v1.6.0) provedeni su u R4.0.3 programu (The R Project for Statistical Computing, [www.r-project.org](http://www.r-project.org)).

## 5. REZULTATI

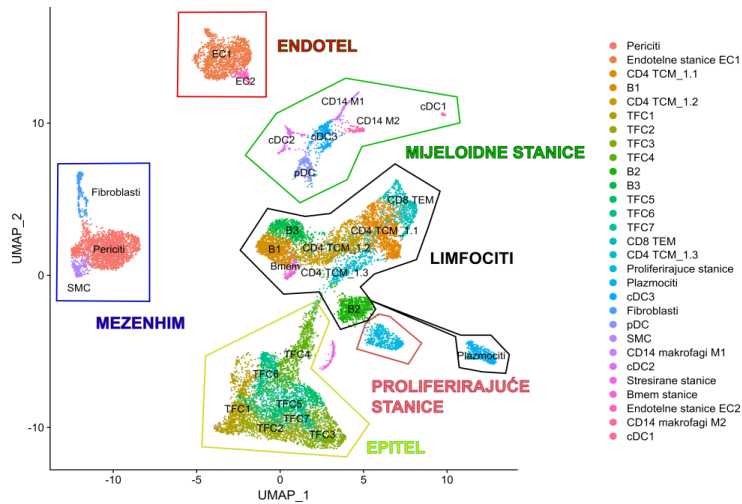
Deskriptivna analiza

Reanalizirani su scRNA-Seq podatci iz eksplantiranih štitnjača. Svi su donori ( $n = 5$ ) ženskog spola, u dobi 29-61 g. Nakon kontrole kvalitete, za konačnu analizu zadržano je 25 740 gena i 18 713 stanica. Na temelju ekspresije linijskih biljega, stanice su razvrstane u 29 subpopulacija (Slika 1) iz šest glavnih razreda (stanice u diobi, mezenhimske, epitelne, limfoidne, endotelne i mijeloidne stanice; Slika 2, Slika 3). Kanonički biljezi glavnih staničnih vrsta su prikazani na Slici 3. Proliferirajuće stanice (Slika 4) dijelom zrcale ekspresiju limfoidnih biljega i vjerojatno većim dijelom sadrže limfatičke stanice u diobi (Slika 5). Jedna skupina stanica (vjerojatno epitelnog podrijetla; Slika 1, Slika 5) isključena je iz daljnjeg razmatranja zbog masivnog izražaja ribosomalnih gena i biljega staničnog stresa. Na Slici 6 prikazano je učešće pojedinih uzoraka u staničnim linijama, kao i broj zadržanih stanica iz svakog uzorka: endotelna i epitelna grupa napučene su stanicama iz sva tri eksperimenta; mezenhimske, limfoidne i mijeloidne stanice uglavnom pripadaju GSE158251 i AT2.GSE134355 uzorku.



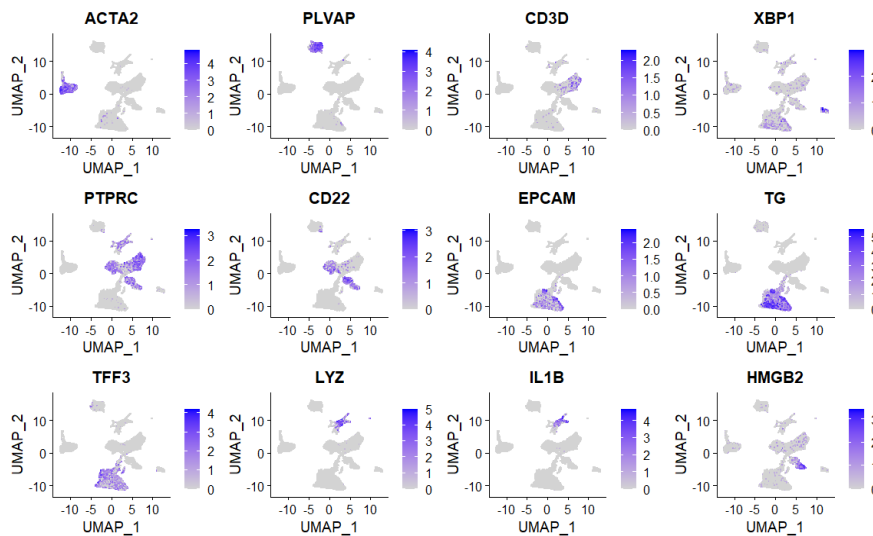
Slika 1. Stanične subpopulacije, UMAP (*Uniform Manifold Approximation and Projection*) reprezentacija. C(E)M centralni (efektorski) memorijski, TFC folikularne stanice štitnjače (*thyroid follicular cells*), cDC konvencionalne dendritičke stanice, pDC plazmocitoidne dendritičke stanice, Bmem memorijske B stanice, SMC glatke mišićne stanice (*smooth muscle cells*).

## 5. REZULTATI



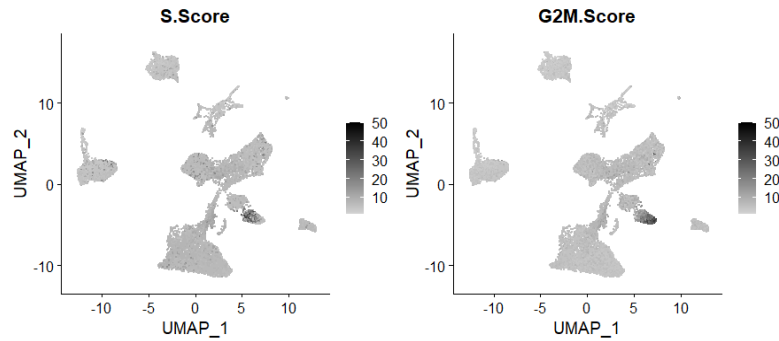
Slika 2. Razdioba populacija u glavnim staničnim razredima. UMAP = *Uniform Manifold Approximation and Projection*.

C(E)M centralni (efektorski) memorijski, TFC folikularne stanice štitnjače (*thyroid follicular cells*), cDC konvencionalne dendritičke stanice, pDC plazmocitoidne dendritičke stanice, Bmem memorijske B stanice, SMC glatke mišićne stanice (*smooth muscle cells*).

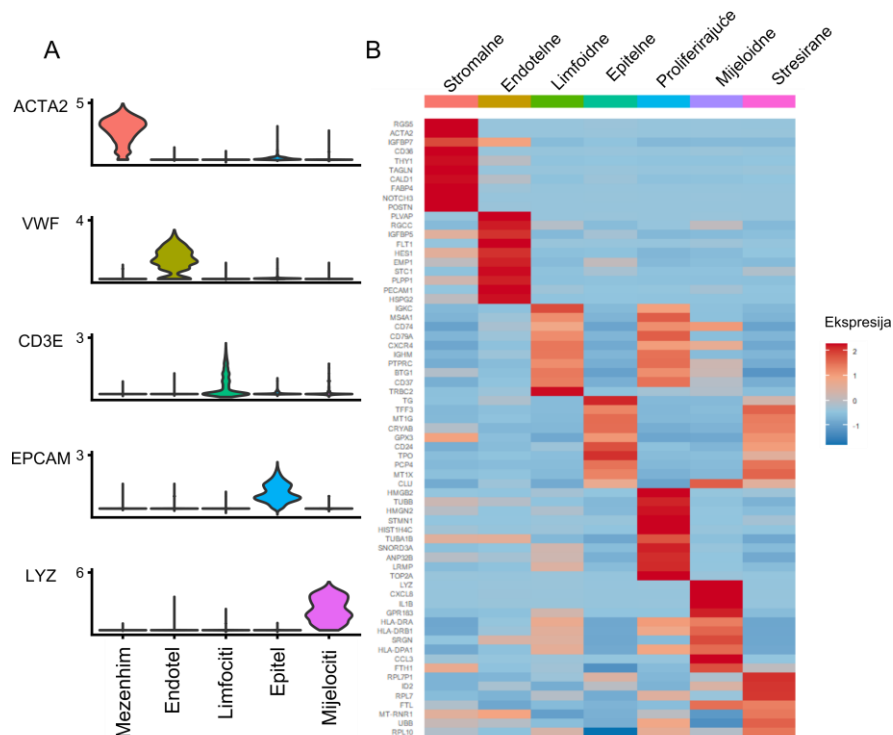


Slika 3. Izražaj kanoničkih staničnih biljega (normalizirane vrijednosti) u populacijama. ACTA2 (mezenhimske stanice), PLVAP (endotelne stanice), CD3D (T stanice), XBP1 (plazma stanice), PTPRC (imunosne stanice), CD22 (B stanice), EPCAM, TG, TFF3 (epitelne stanice), LYZ, IL1B (monociti-makrofagi), HMGB2 (proliferirajuće stanice). Za sve slike, rječnik genskih simbola dostupan je na [www.genecards.org](http://www.genecards.org) (GeneCards: The Human Gene Database, v 5.11, 21-06-2022, Weizmann Institute of Sciences, Rehovot, Izrael). UMAP = *Uniform Manifold Approximation and Projection*.

## 5. REZULTATI

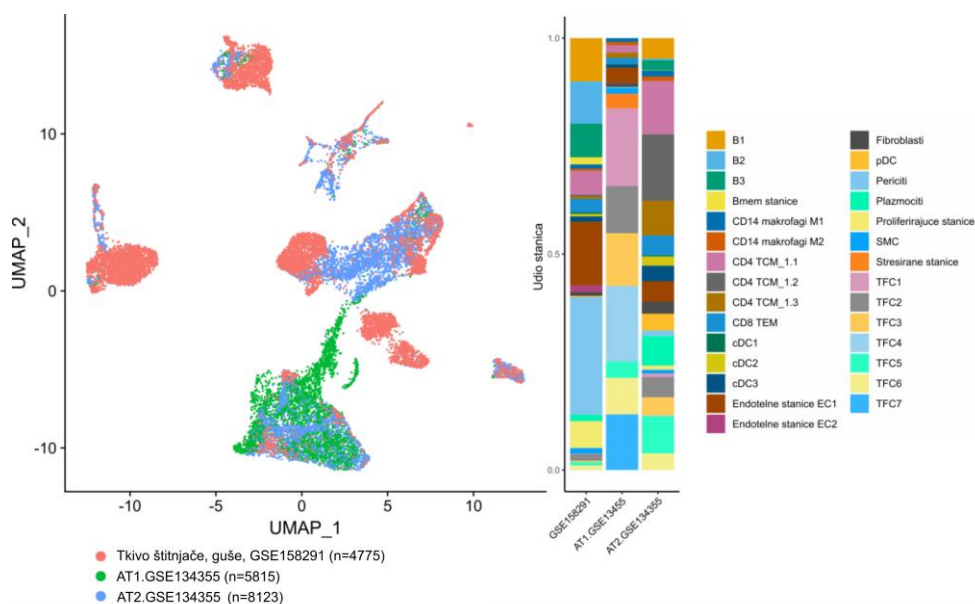


Slika 4. Prikaz stanica u S i G2M fazi staničnog ciklusa. Za analizu faza staničnog ciklusa korištena je lista biljega iz rada Tirosh i sur. (33). UMAP = *Uniform Manifold Approximation and Projection*.



Slika 5. Izražaj linijski specifičnih biljega (A) i 10 vodećih razlikovnih biljega (B) u glavnim staničnim razredima.

## 5. REZULTATI



Slika 6. Učešće pojedinih uzoraka u staničnim populacijama. Legenda odgovara Slici 1 i 2. UMAP = *Uniform Manifold Approximation and Projection*.

Točan broj stanica u svakoj subpopulaciji prikazan je u Tablici 1. Daljnjom raščlambom prvobitnih 29 staničnih subpopulacija (Slika 1) identificirali smo 8200 razlikovno izraženih gena (6294 jedinstvena gena, Tablica 1, cjelovit popis dostupan je na razuman zahtjev). Najveća raznolikost zabilježena je među epitelnim, mijeloidnim i limfoidnim stanicama.

Tablica 1. Broj (N) i udio (%) pojedinih staničnih populacija u integriranom skupu

| Stanična linija | Vrsta stanica         | N    | %     | DEG (N) |
|-----------------|-----------------------|------|-------|---------|
| Mezenhim        | Periciti              | 2294 | 12.26 | 935     |
|                 | Fibroblasti           | 277  | 1.48  | 113     |
|                 | SMC                   | 219  | 1.17  | 248     |
| Endotel         | Endotelne stanice EC1 | 1635 | 8.74  | 563     |
|                 | Endotelne stanice EC2 | 134  | 0.71  | 671     |
| Limfoidna       | CD4 TCM_1.1           | 1255 | 6.71  | 76      |
|                 | CD4 TCM_1.2           | 928  | 4.96  | 13      |
|                 | CD4 TCM_1.3           | 569  | 3.04  | 17      |
|                 | CD8 TEM               | 587  | 3.14  | 108     |
|                 | B1                    | 1092 | 5.83  | 419     |
|                 | B2                    | 816  | 4.36  | 1122    |
|                 | B3                    | 767  | 4.1   | 386     |
|                 | Bmem stanice          | 139  | 0.74  | 864     |
|                 | Plazmociti            | 520  | 2.78  | 50      |
| Mijeloidna      | cDC1                  | 49   | 0.26  | 412     |
|                 | cDC2                  | 174  | 0.93  | 157     |
|                 | cDC3                  | 343  | 1.83  | 151     |

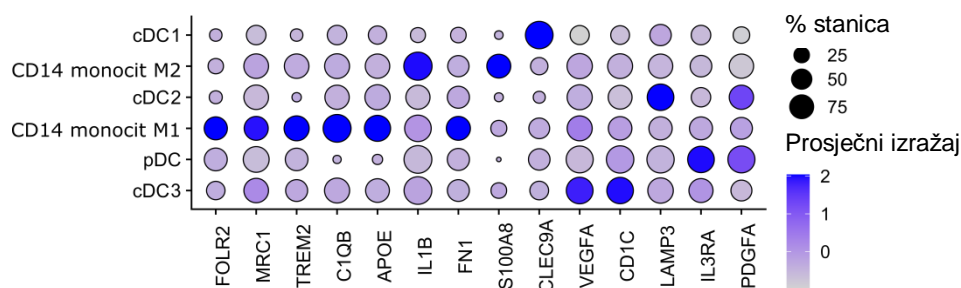
## 5. REZULTATI

| Stanična linija | Vrsta stanica           | N   | %    | DEG (N) |
|-----------------|-------------------------|-----|------|---------|
| Mijeloidna      | pDC                     | 249 | 1.33 | 45      |
|                 | CD14 makrofagi M1       | 203 | 1.08 | 230     |
|                 | CD14 makrofagi M2       | 133 | 0.71 | 157     |
| Epitel          | TFC1                    | 918 | 4.9  | 36      |
|                 | TFC2                    | 913 | 4.88 | 74      |
|                 | TFC3                    | 854 | 4.56 | 53      |
|                 | TFC4                    | 849 | 4.53 | 12      |
|                 | TFC5                    | 741 | 3.96 | 43      |
|                 | TFC6                    | 711 | 3.8  | 63      |
|                 | TFC7                    | 617 | 3.3  | 24      |
| Ostalo          | Proliferirajuće stanice | 565 | 3.02 | 1166    |
|                 | Stresirane stanice      | 162 | 0.86 | 77      |

Razlikovno izraženi geni (Wilcoxon test, DEG/*differentially expressed genes*), SMC – vaskularne glatke mišićne stanice (*smooth muscle cells*), TCM centralno-memorijske T stanice, TEM efektorsko-memorijske stanice T stanice, Bmem memorijske B stanice, cDC konvencionalne dendritičke stanice, pDC plazmocitoidne dendritičke stanice, TFC folikularne stanice štitnjače (*thyroid follicular cells*)

U nastavku, pobliže su opisane pojedine subpopulacije i njihova glavna transkripcijska obilježja.

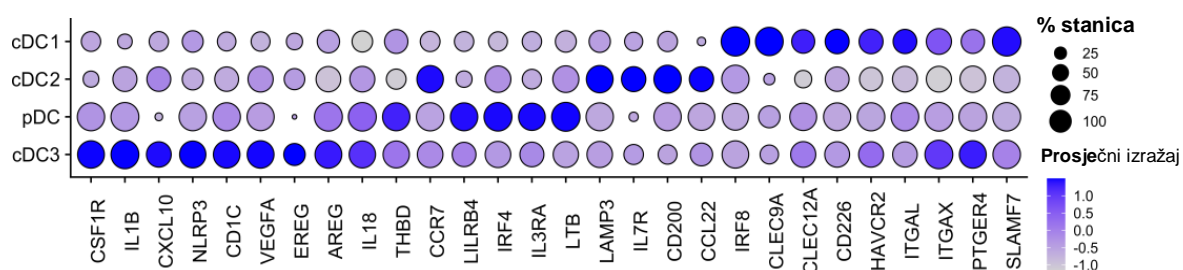
### Mijeloidna loza



Slika 7. Točkasti dijagram ekspresije razlikovnih biljega u pojedinim subpopulacijama mijeloidnih stanica. Intenzitet boje i veličina kruga proporcionalni su razini ekspresije (Prosječni izražaj) i postotnom udjelu stanica (% stanica). Nazivi stanica dostupni su u legendi Slike 2.

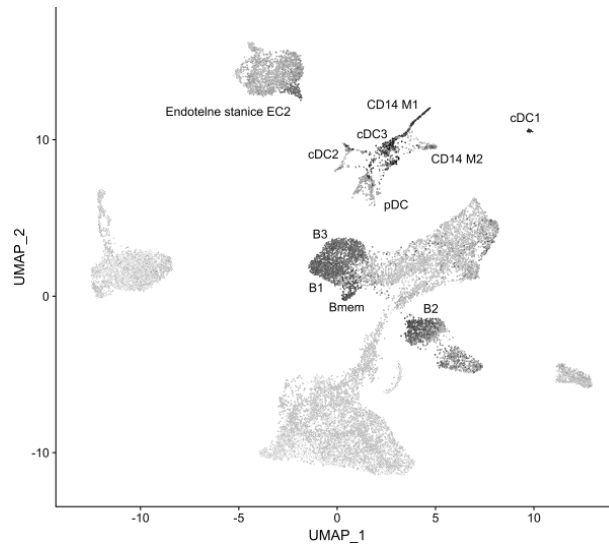
## 5. REZULTATI

Stanice mijeloidne loze grupirane su u 6 populacija (Slika 7): dvije skupine CD14<sup>+</sup> makrofaga, jednu grupu plazmocitoidnih dendritičkih stanica (pDC, IL3RA<sup>+</sup>, IRF4<sup>+</sup>, LILRB4<sup>+</sup>) i tri subpopulacije tipičnih konvencionalnih dendritičkih stanica (Slika 8): cDC1 (XCR1<sup>+</sup>, CLEC9A<sup>+</sup>), cDC2 (LAMP3<sup>+</sup>) i cDC3 (CD1C<sup>+</sup>, VEGFA<sup>+</sup>). cDC3 populacija nosi biljege migrirajućih (42) cDC stanica (CLEC10A / CD301b) i obilježena je snažnim izražajem fibrogenih citokina (CXCL2, CCL3, CCL4, IL1B) i molekula uključenih u pregradnju ekstracelularnog matriksa (TIMP1, MMP9). cDC2 populacija posjeduje biljege aktiviranih stanica (CCR7, FSCN1). cDC1 populacija odlikuje se vršnim kapacitetom procesiranja i prezentacije antigena, a slijedi ju cDC3 populacija s CD14<sup>+</sup> M1 makrofagima i B stanicama (Slika 9).



Slika 8. Točkasti dijagram ekspresije razlikovnih biljega u pojedinim subpopulacijama dendritičkih stanica (DC). Intenzitet boje i veličina kruga proporcionalni su razini ekspresije (Prosječni izražaj) i postotnom udjelu stanica (% stanica). Nazivi stanica dostupni su u legendi Slike 2.

## 5. REZULTATI

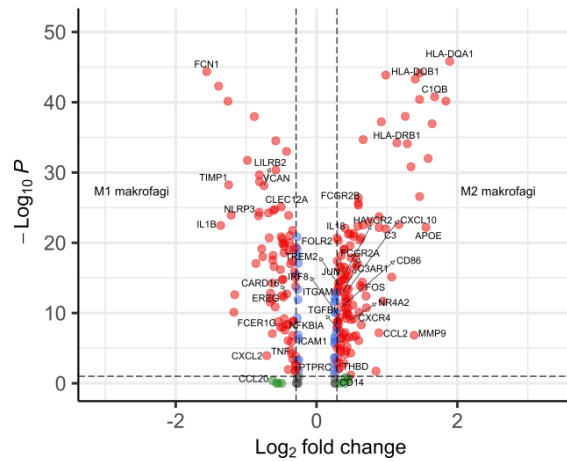


Slika 9. Modul procesiranja i prezentacije antigena. Izdvojene su dendritičke stanice (DC), makrofagi (M), B stanice (B) i endotelne stanice (EC). Intenzitet sive skale razmjernan je intenzitetu izražaja. Nazivi stanica dostupni su u legendi Slike 2. Biljezi procesiranja i prezentacije antigena preuzeti su iz Chen i sur. (35). UMAP = *Uniform Manifold Approximation and Projection*

CD14 makrofagi su podijeljeni u 2 kanoničke subpopulacije: brojniju M2 (TREM2<sup>+</sup>, C1QB<sup>+</sup>, APOE<sup>+</sup>) i manju M1 (IL1B<sup>+</sup>, FCN1<sup>+</sup>, S100A8<sup>+</sup>) grupu. M2 populacija obilježena je ekspresijom triju transkripcijskih čimbenika s poznatom ulogom u alternativnoj polarizaciji makrofaga (KLF4, KLF 10 i NR4A2). U citokinskom panelu M2 stanica dominira ekspresija IL18, proupalnog kemoatraktanta s ulogom u citokinskoj oluji i angiogenezi. Unutar M2 skupine prevladava ekspresija gena povezanih s fagocitozom (43) apoptotskih stanica (FCGR2B, FGL2, MRC1, CD163, TREM2, FTL, FTH1, ITGAM, NCKAP1L), prometom lipida (APOE, APOC1, NPC, GM2A, GLA), pregradnjom izvanstaničnog matriksa (SPP1, MMP9, PLAU, FPR3, CTSL) i fibrogenozom (SERPINA1, HMOX1, IL1B, CCL3, CCL2, SKIL CXCL2, CXCL8, PLAU, MMP9). Nasuprot M2 stanicama, ekspresija amfiregulina (AREG) i epiregulina (EREG) je razlikovno obilježje M1 stanica. AREG i EREG su članovi EGF (engl. *epidermal growth factor*) obitelji molekula s važnom ulogom u tkivnoj homeostazi, regeneraciji epitelnih stanica i obnovi vaskularne barijere nakon ozljede tkiva (44). U istoj skupini, izražaj monocitnih biljega (VCAN, CD300E, AQP9) ukazuje na prisustvo monocita u različitim diferencijacijskim fazama, u prilog kontinuiranom sazrijevanju makrofaga u tkivu.



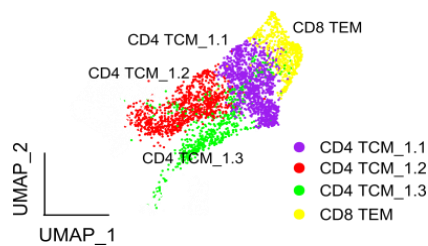
## 5. REZULTATI



Slika 10. Razlikovni izražaj gena u M1 i M2 populaciji makrofaga (Bland-Altman reprezentacija). Linije omeđuju abs (FC, omjer izražaja (*fold change*)) > 1,5 i korigirani P < 0,05 (*Benjamini-Hochberg*, *Wilcoxon test*).

### T stanice

U populaciji T stanica (Slika 1, Slika 11), dominira razdioba u centralno-memorijski (T<sub>CM</sub>) i efektorsko-memorijski fenotip (T<sub>EM</sub>).

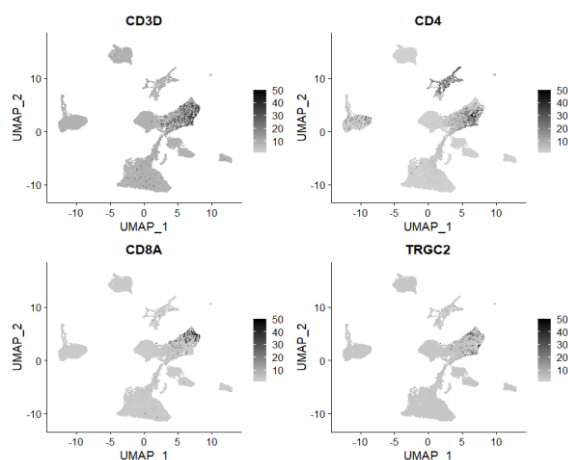


Slika 11. UMAP (*Uniform Manifold Approximation and Projection*) T staničnih populacija. TCM centralno-memorijske T stanice, TEM efektorsko-memorijske stanice T stanice.

T<sub>EM</sub> stanice pretežito su CD8A<sup>+</sup> i CD8B<sup>+</sup>, RUNX3<sup>+</sup>, s dominantnom ekspresijom citotoksičnih molekula poput GZMK, GZMA, NKG7, PRF1 i GNLY. Uz efektorske citokine (IL32, CCL5, IFNG, XCL1 i XCL2), izražena je ekspresija inhibicijskih biljega (45) i biljega disfunkcionalnih, iscrpljenih (T<sub>EX</sub>) stanica u anergiji (TOX2, ICOS, TIGIT, CTLA4). U cjelini, prevladava potpis GZMK<sup>+</sup>T<sub>EM</sub> i terminalnih GZMK<sup>+</sup>T<sub>EX</sub> stanica (46). U istoj T<sub>EM</sub>

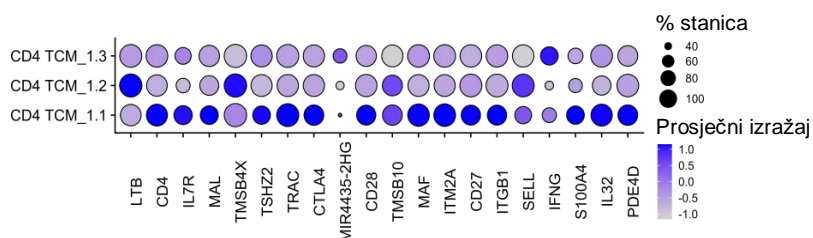
## 5. REZULTATI

skupini nalazi se i dio prirodno-bilačkih [KLRD1<sup>+</sup>, P = 3,7 x 10<sup>-162</sup>) i  $\gamma\delta$  T stanica (TRGC2<sup>+</sup>, P = 7,7 x 10<sup>-131</sup>, Wilcoxon test, Slika 12, (47)].



Slika 12. UMAP (*Uniform Manifold Approximation and Projection*) reprezentacija ekspresije CD3D, CD4, CD8A i TRGC2 gena u staničnim subpopulacijama. Uz ekspresiju CD4 u T stanicama, vidljiva je i konstitutivna ekspresija CD4 u dijelu makrofaga.

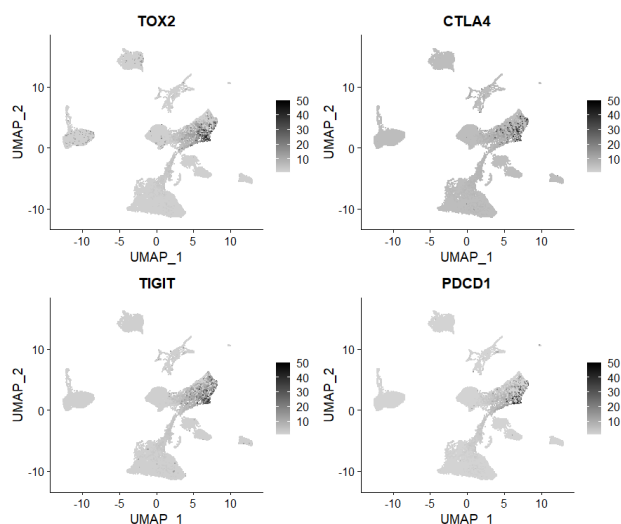
T<sub>CM</sub> stanice pretežito su CD4<sup>+</sup> i tvore tri subpopulacije, T<sub>CM</sub>1-3 (Slika 13).



Slika 13. Točkasti dijagram ekspresije razlikovnih biljega u pojedinim subpopulacijama CD4<sup>+</sup> centralno-memorijskih T (TCM) stanica. Intenzitet boje i veličina kruga proporcionalni su razini ekspresije (Prosječni izražaj) i postotnom udjelu stanica (% stanica).

T<sub>CM</sub>1 i T<sub>CM</sub>2 stanice vodeći su izvor su limfotoksina (LTB), a T<sub>CM</sub>3 stanice proizvode interferon- $\gamma$  (IFNG). U prilog kroničnoj stimulaciji, T<sub>CM</sub>1 subpopulacija (CD3E<sup>+</sup>, IL7R<sup>+</sup>, TCF7<sup>+</sup>, CD69<sup>+</sup>, CD27<sup>+</sup>, CD28<sup>+</sup>) obilježena je robustnom ekspresijom inhibicijskih biljega (TOX2, NR4A2, TIGIT, i CTLA4, Slika 14) i biljega disfunkcionalnih stanica (CD200).

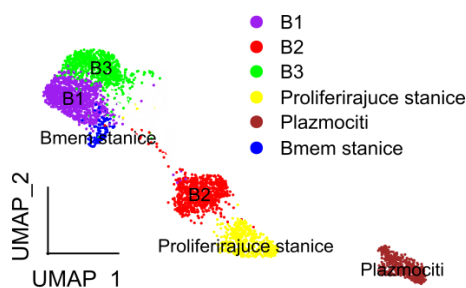
## 5. REZULTATI



Slika 14. UMAP (*Uniform Manifold Approximation and Projection*) reprezentacija ekspresije biljega disfunkcionalnih, iscrpljenih T stanica.

T<sub>CM2</sub> stanice su heterogena populacija, s obilježjima efekorskog (CXCR4<sup>+</sup>, LTB<sup>+</sup>, PFN1<sup>+</sup>) i naivnog (T<sub>N</sub>, IL7R<sup>-</sup>T<sub>N</sub>) / memorijskog (SELL<sup>+</sup>) programa (48). U manjem broju stanica (< 20 %), prisutni su medijatori funkcijskog iscrpljenja (TOX2, TIGIT, PDCD1).

### B stanice

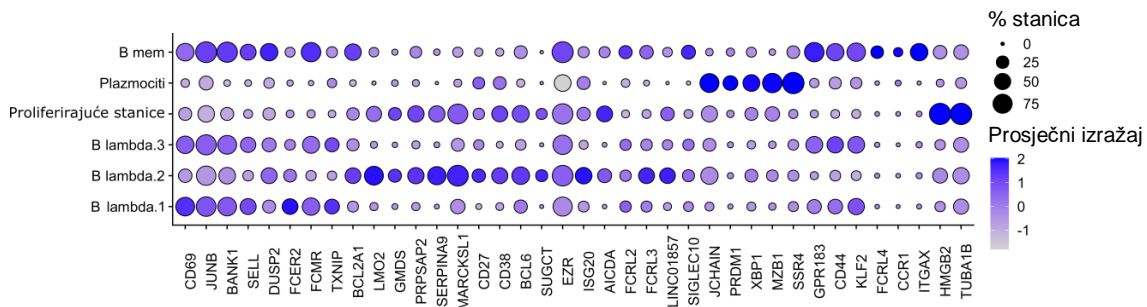


Slika 15. UMAP (*Uniform Manifold Approximation and Projection*) reprezentacija B staničnih populacija.

U ovoj populaciji zastupljene su sve stanice iz glavnih faza sazrijevanja B limfocita (Slika 15, Slika 16): aktivirane B stanice, stanice svijetle (*light zone*) i tamne (*dark*) zone germinativnih centara (GC), memorijske B stanice, plazmociti i proliferirajuće stanice. Profil stanica

## 5. REZULTATI

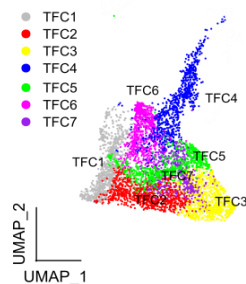
upućuje na imunološki aktivno limfatičko tkivo, odgovarajući ranije opisanom fenomenu limfoidne neoorganogeneze u štitnjači.



Slika 16. Točkasti dijagram ekspresije razlikovnih biljega u pojedinim subpopulacijama B stanica. Intenzitet boje i veličina kruga proporcionalni su razini ekspresije (Prosječni izražaj) i postotnom udjelu stanica (% stanica).

B stanice razdijeljene su u 3 populacije intermedijarnih lambda B stanica ( $MS4A1^+$ ,  $CD79A^+$ , B1-3), jednu memorijsku populaciju i jednu plazmocitnu populaciju ( $PRDM1^+$ ,  $XBP1^+$ ,  $SSR4^+$ ,  $MZB1^+$ ,  $JCHAIN^+$ , Slika 16). U B1 i B2 populaciji prevladavaju aktivirane B stanice i B stanice žarišta klijanja. B3 populaciju čine uglavnom aktivirane memorijske  $AICDA^+$  B stanice. U skupini proliferirajućih stanica, ( $HMGB2$ ,  $TUBA1B$ ), ekspresija  $AICDA$  i  $BCL6$  znak je prisustva mitotski aktivnih B stanica podrijetla GC.

### Epitelne stanice

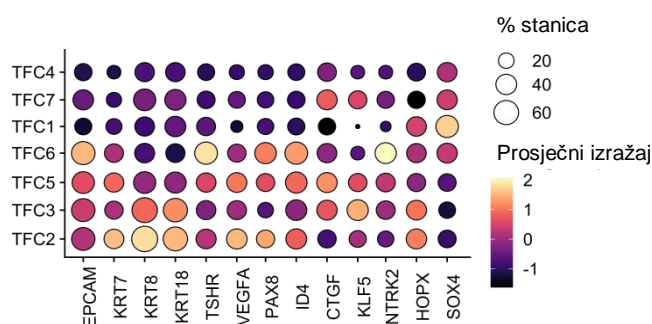


Slika 17. Dvodimenzionalna (UMAP, *Uniform Manifold Approximation and Projection*) reprezentacija epitelnih subpopulacija.

Epitelne (folikularne) stanice štitnjače razdijeljene su u 7 skupina s brojnim hibridnim i tranzicijskim obilježjima koja više upućuju na kontinuitet staničnih stanja nego na zasebne populacije (Slika 17, Slika 18).

## 5. REZULTATI

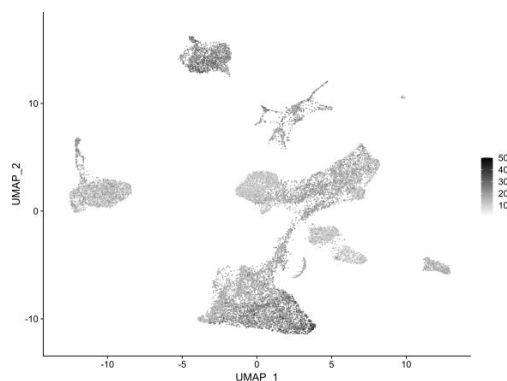
TFC2 i TFC3 skupina (Slika 18) uključuju terminalno diferencirane, EPCAM<sup>+</sup>, KRT7<sup>+</sup>, KRT8<sup>+</sup>, KRT18<sup>+</sup>, CDH1<sup>+</sup> stanice. TFC2 / 3 stanice su obilježene prominentnom ekspresijom apikalnih anionskih transportera uključenih u luminalni efluks joda (SLC26A7, SLC26A4 / pendrin) i gena uključenih u proizvodnju i metabolizam hormona štitnjače (IYD, TG, DIO2, PAX8, CTSB, TPO). Metabolizam glutationa (GSTP1, MGST1, GPX3) i detoksifikacija (PRDX1, MT1G, MT1E, MT1X, PRDX5, GPX3) očekivano prate oksidativni stres tijekom sinteze tiroksina, biokemijskog procesa koji u osnovi počiva na seriji redoks-reakcija.



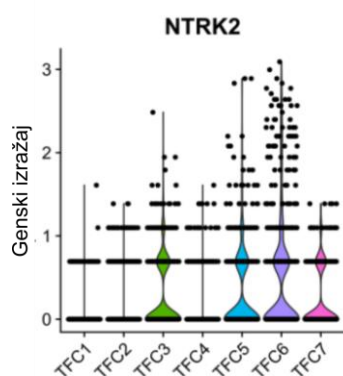
Slika 18. Točkasti dijagram ekspresije razlikovnih biljega u pojedinim subpopulacijama folikularnih epitelnih stanica (TFC). Intenzitet boje i veličina kruga proporcionalni su razini ekspresije (Prosječni izražaj) i postotnom udjelu stanica (% stanica). Redoslijed populacija na ordinati prilagođen je broju biljega u stanicama.

Za razliku od TSHR<sup>+</sup>, VEGFA<sup>+</sup>, PAX8<sup>+</sup>, ID4<sup>+</sup> TFC2 stanica, u TFC3 skupini prevladava TSHR<sup>lo</sup>, CTGF<sup>+</sup>, KLF5<sup>+</sup> fenotip s ekspresijom biljega mezenhimske tranzicije (Slika 19). PAX8 je jedan od četiri transkripcijska čimbenika s poznatom ulogom u morfogenezi i homeostazi TFC. Zajedno s NKX2-1, FOXE1 i HOXA3, PAX8 je biljeg progenitorskih stanica štitnjače u kojima regulira folikulogenezu, poglavito koordinacijom izražaja adhezijskih i spojničkih molekula (49). KLF5 je transkripcijski čimbenik i biljeg endodermalnih matičnih stanica podrijetla probavne cijevi (50) s ulogom u angioinvasivnom potencijalu papilarnog karcinoma štitnjače (Slika 20).

## 5. REZULTATI



Slika 19. UMAP (*Uniform Manifold Approximation and Projection*) reprezentacija ekspresije biljega mezenhimske tranzicije. Lista biljega epitelno-mezenhimske tranzicije preuzeta je iz Cook i sur. (36).



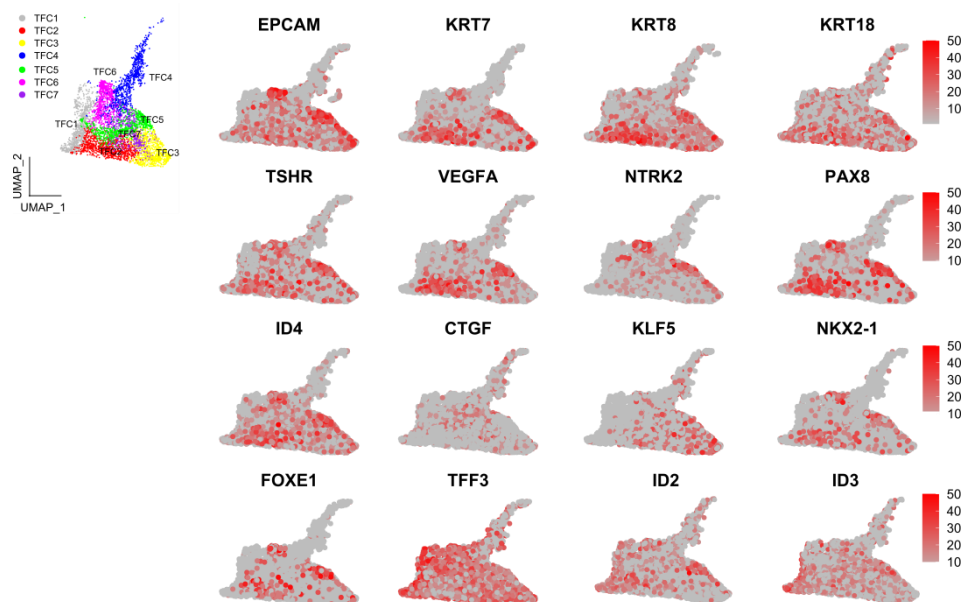
Slika 20. Violinske mape izražaja NTRK2 (tirozin-kinaza neurotrofnog receptora 2) u folikularnim epitelnim stanicama štitnjače (TFC) i njihovim subpopulacijama.

TFC5 stanice dijele obilježja TFC2 i TFC3 skupine, vjerojatno kao tranzicijski fenotip. Posjeduju tipične epitelne biljege (EPCAM<sup>+</sup>, KRT7<sup>+</sup>, KRT8<sup>+</sup>, KRT18<sup>+</sup>, CDH1<sup>+</sup>) i hibridni TSHR<sup>+</sup>, VEGFA<sup>+</sup>, PAX8<sup>+</sup>, TFF3<sup>+</sup>, ID4<sup>+</sup>, CTGF<sup>+</sup> profil. Kao i TFC5 stanice, TFC6 i TFC7 stanice odlikuju se nizom prijelaznih obilježja. Ekspresija NTRK2 (*Neurotrophic Receptor Tyrosine Kinase 2*) je vodeće razlikovno obilježje TFC5 i TFC6 stanica (Slika 20, Slika 21), NTRK2 (51) je kolesterol-senzitivni membranski receptor za BDNF (*brain-derived neurotrophic factor*) i NTF3/4 (neurotrofin-3/4). NTRK2 protein doista je prisutan u tkivu štitnjače (The *Human Protein Atlas*, v21.1, 31-05-2022, [www.proteinatlas.org](http://www.proteinatlas.org)), ali njegova uloga u folikularnom epitelu nije do sada istražena.

Osobitost TFC1 stanica je ekspresija različitih meta FOXE1, transkripcijskog čimbenika s ulogom u folikulogenezi (52). Među njima, dominira metabolizam glutationa (GPX3, MGST1, GSTP1) i ekspresija različitih podjedinica mitohondrijskih ATP sintaza s ulogom u

## 5. REZULTATI

oksidativnoj fosforilaciji ( $P = 1,5 \times 10^{-11}$ , Hallmark). TFC1 stanice lišene su svih molekula iz puta sinteze tiroksina. Od transkripcijskih čimbenika zastupljeni su ID2, ID3 i SOX4 (Slika 21). SOX4 je udružen s invazijom, G1/S s progresijom i inhibicijom apoptoze (53) u epitelnim stanicama papilarnog karcinoma štitnjače (Slika 18).



Slika 21. Dvodimenzionalna (UMAP, *Uniform Manifold Approximation and Projection*) reprezentacija ekspresije razlikovnih biljega epitelnih subpopulacija. Kolorirane vrpce kodiraju intenzitet izražaja.

### Mezenhimske stanice

Mezenhimske stanice grupirane su u dvije populacije: grupu stromalnih stanica i jedan oblak vaskularnih glatkih mišićnih stanica (Slika 22).

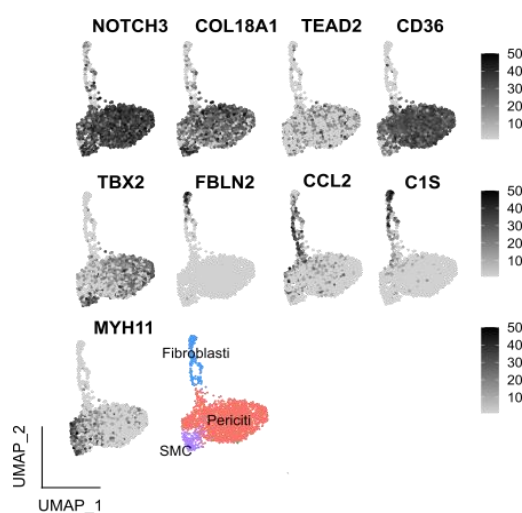
Stromalne stanice ( $TAGLN^+$ ,  $BGN^+$ ,  $FBN1^+$ ,  $VIM^+$ ,  $COL^+$ ) su dalje podijeljene u 2 podgrupe stanica, pericite i fibroblaste.

Periciti [ $RGS5$ ,  $NOTCH3$ ,  $HIGD1B$ ,  $COX4I2$ , (54)] se odlikuju vrlo bogatim popisom razlikovnih gena (876 prijepisa) koji predvode biljezi miofibroblasta ( $ACTA2$ ,  $POSTN$ ). Osobito je prominentna ekspresija  $NOTCH3$  (Slika 23), receptora koji kodira pozicijski identitet perivaskularnih fibroblasta i regulira njihov upalni odgovor. U kolagenom profilu, dominiraju  $COL4A1$  i  $COL18A1$ , dva kolagena bazalne membrane epitelnih i endotelih stanica, uz pratnju brojnih laminina ( $LAMA4/B1/B2/C1$ ). U trećini pericita (38 %) nalazimo izražaj  $LOXL2$ , lizil-oksidaze s ulogom u kapilarnoj tubulogenezi, koja u međustaničnom

## 5. REZULTATI

prostoru katalizira umrežavanje kolagena i elastina (55). Komponente Hippo signalnog puta (56) su također snažno zastupljene (YAP1, TEAD1/3, BMP2, BMPR2). Među drugim biološkim procesima, bogato su zastupljene komponente autofagije (57), a u metaboličkom smislu, dominira oksidativna fosforilacija i metabolizam lipida (LPL, ANGPTL4, CD36, FABP4, PPARG, KLF9, PLSCR1).

U skupini fibroblasta (DCN<sup>+</sup>, LUM<sup>+</sup>, FBLN<sup>+</sup>, PDGFA<sup>+</sup>, PI16<sup>+</sup> stanice), prisutna je robustna ekspresija proupalnih i fibrogenih citokina (IL6, CCL2, CXCL2, CXCL12, CXCL14), komponenti komplekta (C1S/Q, C3, C7, Slika 22) i mreže elastičnih vlakana (FBN1, FBLN5, LTBP4).



Slika 22. Dvodimenzionalna (UMAP, *Uniform Manifold Approximation and Projection*) reprezentacija izražaja razlikovnih biljega mezenhimskih stanica. SMC glatke mišićne stanice (*smooth muscle cells*).

### Glatke mišićne stanice

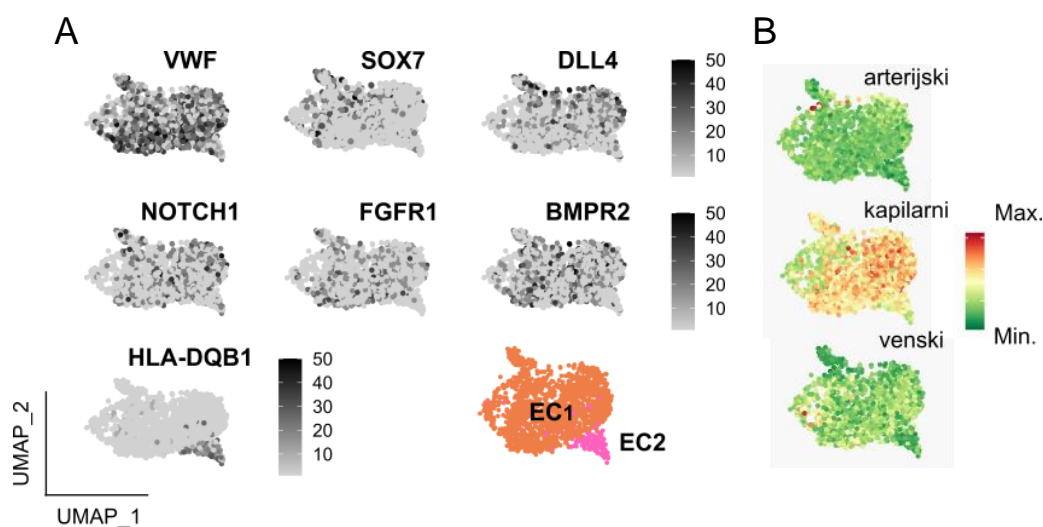
Vaskularne glatke mišićne stanice (Slika 23) obilježene su izrazitom ekspresijom kontraktilnih elemenata (MYH11<sup>+</sup>, ACTA2<sup>+</sup>, TAGLN<sup>+</sup>, CNN1<sup>+</sup>, VCL<sup>+</sup>). U ostalim detaljima, prevladavaju strukturne i komunikacijske molekule mezenhimske linije, komponente ekstracelularnog matriksa (58), ekspresija mehanosenzitivnih transduktora (YAP1) i ekspresija ZEB2, transkripcijskog čimbenika s ulogom u epigenetskom remodeliranju kontraktilnog programa.



Endotelne stanice

Endotelne stanice čine dva skupa (Slika 24) koji međusobno dijele 351 biljeg. U panendotelnim obilježjima, dominiraju prototipski endotelni biljezi (PLVAP, KDR/VEGFR2, VWF, PECAM1, CDH5, TM4SF1, Slika 23A), komponente fokalnih adhezija ( $P = 1,2 \times 10^{-13}$ , KEGG), regulatori transendotelne migracije ( $P = 5,6 \times 10^{-14}$ , KEGG) i HLA I molekule (humani leukocitni antigen). Glavina endotelnih stanica nosi kapilarne biljege (Slika 23B). Sve endotelne populacije prepisuju transkripcijske čimbenike ERG i SOX18 (37-39).

U razlikovnom dijelu, EC1 stanice obilježene su ekspresijom čimbenika koji su odgovorni za održavanje vaskularnog identiteta (SOX7, DLL4-NOTCH1 kompleks, Slika 23A).



Slika 23. Dvodimenzionalna (UMAP, *Uniform Manifold Approximation and Projection*) reprezentacija (A) izražaja razlikovnih biljega endotelnih stanica (EC). Biljezi arterijskih, kapilarnih i venskih endotelnih stanica (B). Genski biljezi preuzeti su iz ref. (37-39).

Nasuprot EC1 stanicama, EC2 stanice obilježene su ekspresijom HLA II klase molekula (Slika 23), odnosno biljezima izlaganja interferonu- $\gamma$  i čimbeniku nekroze tumora, TNF- $\alpha$ , kao znak sudjelovanja u aktivnom imunološkom nadzoru.

## 6. RASPRAVA

U tradicionalnom narativu, štitnjača je jednostavno strukturirani organ monotone građe i ne osobito raznolike stanične populacije. U odrasloj dobi, mitotska aktivnost u štitnjači je niska, a izmjena stanica spora. U direktnoj suprotnosti s takvim staničnim poimanjem, tkivo štitnjače je dom brojnim vrstama upalnih, regulacijskih i stromalnih stanica čija raznolikost upućuje na složen i umrežen žljezdani krajobraz. Ukupno, ovim istraživanjem je identificirano najmanje 28 različitih staničnih tipova koristeći jednostaničnu RNK > 18 000 stanica. Unutar ovog mikrookoliša, folikularne epitelne stanice, kao najbrojnija komponenta, čine heterogeni skup stanica s nizom prijelaznih stanja koja potencijalno odgovaraju podjeli posla unutar i između žljezdanih folikula (59). U tom smislu, rezultati ovog istraživanja čine mali korak prema boljem razumijevanju stanične preraspodjele poslova tijekom sinteze, pohrane i oslobađanja koloida iz folikula. U skladu s modernim rezultatima, potporna stromalna komponenta obilježena je diverzifikacijom i specijalizacijom fibroblasta (60-62) i endotela (37-39). Osobita stanična raznolikost je zabilježena među mijeloidnim i limfoidnim upalnim stanicama, među kojima se izdvajaju najmanje četiri subpopulacije kanoničkih dendritičkih stanica i značajna monocitno-makrofagna populacija omeđena TREM2<sup>+</sup> (M2) i IL1B<sup>+</sup> (M1) makrofagima. Veći dio ove makrofagne skupine posjeduje obilježja nedavno opisanih regulacijskih makrofaga (63). Među T stanicama, podjednako su brojna CD4<sup>+</sup> efektorско-memorijska i CD8<sup>+</sup> centralno-memorijska subpopulacija s obilježjima citotoksičnih, prirođenih  $\gamma\delta$  i iscrpljenih, disfunkcionalnih T<sub>EX</sub> limfocita. B limfociti čine osobito fenotipski šaroliku populaciju, obuhvaćajući cijeli niz prijelaznih stanja karakterističnih za žarišno klijanje, odnosno *in situ* proliferaciju, selekciju i sazrijevanje B stanica. Rezultat je značajna lokalna populacija zrelih plazma stanica (64-65). U skladu s tim, infiltracija T i B stanicama već je opisana u tkivu štitnjače sličnog podrijetla – tkivu guša i osobito Hashimotovom tireoiditisu (66). Dopunjavajući cenzus stanica, atlas koji proizlazi iz ovog istraživanja dodatno nudi detaljan uvid u transkripcijska, signalna i metabolička obilježja pojedinih staničnih tipova. Naposljetku, atlas nudi i polazišnu osnovu za analizu stanične komunikacije receptor-ligand parova, analizu pseudotemporalnih putanja i analizu staničnih regulona, što ostavljamo za budući prikaz. Time je definiran i translacijski potencijal istraživanja, posebice imajući u vidu visoku populacijsku prevalenciju autoimunih bolesti, gušavosti i papilarnog karcinoma štitnjače (67-69).

Unatoč bogatijem uvidu u ekosustav štitnjače u usporedbi sa ranijim fragmentiranim studijama, ovo istraživanje ima i nekih nedostataka. Ponajprije, to je nedostatak ortogonalne

## 6. RASPRAVA

validacije staničnih subpopulacija imunohistokemijom, potom i ograničen potencijal detekcije rijetkih stanica zbog skromnog broja uzoraka (npr. nisu detektirane C stanice). Također, nedostaje i rekonstrukcija prostornih odnosa staničnih vrsta, npr. uporabom prostorne transkriptomike. U ostalim detaljima, izdvaja se srednja rezolucija koja je korištena za hijerarhijsko grupiranje stanica u ovoj prigodi, čime svakako nije iscrpljen potencijal detekcije staničnih tipova. U tu svrhu, prethodno obogaćivanje populacija selekcijom i sortiranjem ima veći potencijal, a moguće je koristiti i dodatne cikluse grupiranja. Nadalje, uzorci pripadaju uglavnom kasnoj odrasloj dobi i ženama, što utemeljuje potrebu za budućim profiliranjem tkiva iz cijelog razvojnog raspona, od fetalne dobi do starosti, u oba kromosomska spola. Na tom putu, svakako treba primijetiti i potrebu za ciljanom optimizacijom uzorkovanja, prezervacije i disocijacije tkiva štitnjače. U međuvremenu, izvrsno suglasje više detektiranih staničnih podvrsta i njihovih obilježja sa postojećom literaturom ohrabruje opreznu upotrebu ovog atlasa.

Sveukupno, rezultat je ovog istraživanja cenzus staničnog susjedstva štitnjače na jednostaničnoj rezoluciji kao potencijalno vrijedan biološki resurs i izvor uvida u fiziologiju štitnjače u zdravlju i bolesti.

### 7. ZAKLJUČCI

Temeljem provedenog istraživanja i dobivenih rezultata mogu se izvesti sljedeći zaključci:

- Štitnjača je organ složene strukture, čiji se raznoliki stanični krajobraz može prikazati sekvenciranjem jednostanične RNK.
- Štitnjača je građena od stanica koje pripadaju grupi endotelnih, epitelnih, mezenhimskih, limfoidnih, mijeloidnih stanica te stanica u diobi, a unutar tih grupa stanice su razvrstane u različite populacije i subpopulacije koje doprinose strukturnoj i funkcionalnoj heterogenosti tkiva štitnjače.

## 8. SAŽETAK

**CILJ ISTRAŽIVANJA.** Anotacija i cenzus staničnih subpopulacija ljudskog tkiva štitnjače na jednostaničnoj rezoluciji uporabom scRNA-Seq (bioinformatička analiza).

**METODE.** U tu svrhu, integrirani su i analizirani javno dostupni podatci sekvenciranja jednostanične RNK pet štitnih žlijezda odraslih osoba (*Gene Expression Omnibus Identifier* GSE134355, GSE158291, Seurat v4 alat), što je omogućilo raščlambu staničnog susjedstva štitnjače na srednjoj rezoluciji. Uključene su samo stanice benignih tkiva štitnjače. Nakon diferencijalne analize genskog izražaja (Wilcoxonov test) uslijedila je analiza obogaćenih putova (Hallmark, *Kyoto Encyclopedia of Genes and Genomes, Broad Molecular Signatures Database* v7.4).

**REZULTATI.** Ukupno je analizirano > 18 000 stanica visoke kvalitete te je identificirano 28 staničnih stanja (7 epitelnih, 9 limfoidnih, 6 mijeloidnih, 3 stromalna, 2 endotelna i jedan proliferativni klaster). Okarakterizirani su genski izražaji specifični za stanične tipove i biološki programi za sve veće stanične linije štitnjače koji pružaju informacije o njihovim transkripcijskim, signalizacijskim i metaboličkim svojstvima.

**ZAKLJUČAK.** Sveobuhvatni jednostanični atlas humane štitne žlijezde je vrijedan resurs koji povećava naše razumijevanje o staničnoj heterogenosti i mikrookolišima specifičnih tkiva.

**Ključne riječi:** analiza sekvenci; jednostanična analiza; profiliranje genskog izražaja; računalna biologija; RNK; štitnjača

## 9. ABSTRACT

„The thyroid cell ecosystem analyzed by single-cell RNA sequencing”

**AIM.** To obtain a high-resolution single-cell type map of human thyroid tissue using RNA sequencing (bioinformatics analysis).

**METHODS.** To help fill this gap, we integrated and analyzed publicly available single-cell RNA sequencing data from 5 adult thyroids (Gene Expression Omnibus Identifier GSE134355, GSE158291, Seurat v4 pipeline), which enabled us to capture the underlying cellular neighborhood at medium resolution. Only cells from benign thyroid tissues were included. Differential gene expression analysis was performed (Wilcoxon’s test), followed by pathway enrichment profiling (Hallmark, Kyoto Encyclopedia of Genes and Genomes, Broad Molecular Signatures Database v7.4).

**RESULTS.** In total, > 18 000 high-quality cells were recovered, from which we identified 28 cell states (7 epithelial, 9 lymphoid, 6 myeloid, 3 stromal, 2 endothelial and one proliferating cluster). We characterized cell-type-specific gene expression and biological programs for all major thyroid cell lineages, providing information on their transcriptional, signaling and metabolic properties.

**CONCLUSIONS.** Our comprehensive single-cell atlas of the human thyroid gland provides a valuable resource, enhancing our understanding of cellular heterogeneity and specific tissue microenvironments.

**Keywords:** Computational Biology; Gene Expression Profiling; RNA; Sequence Analysis; Single-Cell Analysis; Thyroid Gland

## 10. LITERATURA

1. Krmpotić-Nemanić J. Anatomija čovjeka. Zagreb: Medicinska naklada; 2007.
2. Vinter I, ur. Waldeyerova anatomija čovjeka. 17. izd. Zagreb: Golden marketing-tehnička knjiga; 2009, str. 337-40.
3. Bradamante Ž, Kostović-Knežević L, ur. Osnove histologije. Zagreb: Školska knjiga; 2005, str. 423-28.
4. Garasto S, Montesanto A, Corsonello A, Lattanzio F, Fusco S, Passarino G, i sur. Thyroid hormones in extreme longevity. *Mech Ageing Dev.* 2017 Jul;165(Pt B):98-106.
5. Bano A, Dhana K, Chaker L, Kavousi M, Ikram MA, Mattace-Raso FUS, i sur. Association of Thyroid Function With Life Expectancy With and Without Cardiovascular Disease: The Rotterdam Study. *JAMA Intern Med.* 2017 Nov 1;177(11):1650-1657.
6. Fagman H, Nilsson M. Morphogenetics of early thyroid development. *J Mol Endocrinol.* 2011 Feb;46(1):R33-42.
7. Guyton AC, Hall JE. Medicinska fiziologija. Zagreb: Medicinska naklada; 2012, str. 907-10.
8. Groeneweg S, van Geest FS, Peeters RP, Heuer H, Visser WE. Thyroid Hormone Transporters. *Endocr Rev.* 2020 Apr 1;41(2):146-201.
9. Ortiga-Carvalho TM, Sidhaye AR, Wondisford FE. Thyroid hormone receptors and resistance to thyroid hormone disorders. *Nat Rev Endocrinol.* 2014 Oct;10(10):582-91.
10. Pu W, Shi X, Yu P, Zhang M, Liu Z, Tan L, i sur. Single-cell transcriptomic analysis of the tumor ecosystems underlying initiation and progression of papillary thyroid carcinoma. *Nat Commun.* 2021 Oct 18;12(1):6058.

## 10. LITERATURA

11. Zhang QY, Ye XP, Zhou Z, Zhu CF, Li R, Fang Y, i sur. Lymphocyte infiltration and thyrocyte destruction are driven by stromal and immune cell components in Hashimoto's thyroiditis. *Nat Commun.* 2022 Feb 9;13(1):775.
12. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, i sur. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science.* 2017 Apr 21;356(6335):eaah4573.
13. HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature.* 2019 Oct;574(7777):187-192.
14. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, i sur. Integrated analysis of multimodal single-cell data. *Cell.* 2021 Jun 24;184(13):3573-3587.e29.
15. Wu Y, Zhang K. Tools for the analysis of high-dimensional single-cell RNA sequencing data. *Nat Rev Nephrol.* 2020 Jul;16(7):408-421.
16. Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat Rev Nephrol.* 2018 Aug;14(8):479-492.
17. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med.* 2018 Aug 7;50(8):1-14.
18. Andrews TS, Hemberg M. Identifying cell populations with scRNASeq. *Mol Aspects Med.* 2018 Feb;59:114-122.
19. Chen G, Ning B, Shi T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Front Genet.* 2019 Apr 5;10:317.
20. Chen H, Ye F, Guo G. Revolutionizing immunology with single-cell RNA sequencing. *Cell Mol Immunol.* 2019 Mar;16(3):242-249.
21. Li L, Xiong F, Wang Y, Zhang S, Gong Z, Li X, i sur. What are the applications of single-cell RNA sequencing in cancer research: a systematic review. *J Exp Clin Cancer Res.* 2021 May 11;40(1):163.



## 10. LITERATURA

22. Andrews TS, Kiselev VY, McCarthy D, Hemberg M. Tutorial: guidelines for the computational analysis of single-cell RNA sequencing data. *Nat Protoc.* 2021 Jan;16(1):1-9.
23. Pasquini G, Rojo Arias JE, Schäfer P, Busskamp V. Automated methods for cell type annotation on scRNA-seq data. *Comput Struct Biotechnol J.* 2021 Jan 19;19:961-969.
24. Han X, Zhou Z, Fei L, Sun H, Wang R, Chen Y, i sur. Construction of a human cell landscape at single-cell level. *Nature.* 2020 May;581(7808):303-309.
25. Peng M, Wei G, Zhang Y, Li H, Lai Y, Guo Y, i sur. Single-cell transcriptomic landscape reveals the differences in cell differentiation and immune microenvironment of papillary thyroid carcinoma between genders. *Cell Biosci.* 2021 Feb 15;11(1):39.
26. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* 2019 Apr 24;8(4):329-337.e4.
27. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol.* 2019 Jun 19;15(6):e8746.
28. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 2019 Dec 23;20(1):296.
29. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, i sur. Comprehensive Integration of Single-Cell Data. *Cell.* 2019 Jun 13;177(7):1888-1902.e21.
30. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep.* 2019 Mar 26;9(1):5233.
31. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, i sur. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol.* 2018 Dec 3. doi: 10.1038/nbt.4314.
32. Mabbott NA, Baillie JK, Brown H, Freeman TC, Hume DA. An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genomics.* 2013 Sep 20;14:632.

## 10. LITERATURA

33. Tirosh I, Izar B, Prakadan SM, Wadsworth MH 2nd, Treacy D, Trombetta JJ, i sur. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. 2016 Apr 8;352(6282):189-96.
34. van den Brink SC, Sage F, Vártesy Á, Spanjaard B, Peterson-Maduro J, Baron CS, i sur. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat Methods*. 2017 Sep 29;14(10):935-936.
35. Chen B, Scurrah CR, McKinley ET, Simmons AJ, Ramirez-Solano MA, Zhu X, i sur. Differential pre-malignant programs and microenvironment chart distinct paths to malignancy in human colorectal polyps. *Cell*. 2021 Dec 22;184(26):6262-6280.e26.
36. Cook DP, Vanderhyden BC. Transcriptional census of epithelial-mesenchymal plasticity in cancer. *Sci Adv*. 2022 Jan 7;8(1):eabi7640.
37. Goveia J, Rohlenova K, Taverna F, Treps L, Conradi LC, Pircher A, i sur. An Integrated Gene Expression Landscape Profiling Approach to Identify Lung Tumor Endothelial Cell Heterogeneity and Angiogenic Candidates. *Cancer Cell*. 2020 Jan 13;37(1):21-36.e13. Erratum in: *Cancer Cell*. 2020 Mar 16;37(3):421.
38. Schupp JC, Adams TS, Cosme C Jr, Raredon MSB, Yuan Y, Omote N, i sur. Integrated Single-Cell Atlas of Endothelial Cells of the Human Lung. *Circulation*. 2021 Jul 27;144(4):286-302.
39. Winkler EA, Kim CN, Ross JM, Garcia JH, Gil E, Oh I, i sur. A single-cell atlas of the normal and malformed human brain vasculature. *Science*. 2022 Mar 4;375(6584):eabi7377.
40. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017 Jan 4;45(D1):D353-D361.
41. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015 Dec 23;1(6):417-425.

## 10. LITERATURA

42. Tatsumi N, Codrington AL, El-Fenej J, Phondge V, Kumamoto Y. Effective CD4 T cell priming requires repertoire scanning by CD301b<sup>+</sup> migratory cDC2 cells upon lymph node entry. *Sci Immunol*. 2021 Dec 10;6(66):eabg0336.
43. Kobori T, Hamasaki S, Kitaura A, Yamazaki Y, Nishinaka T, Niwa A, i sur. Interleukin-18 Amplifies Macrophage Polarization and Morphological Alteration, Leading to Excessive Angiogenesis. *Front Immunol*. 2018 Mar 6;9:334.
44. Kumagai S, Koyama S, Nishikawa H. Antitumour immunity regulated by aberrant ERBB family signalling. *Nat Rev Cancer*. 2021 Mar;21(3):181-197.
45. Li H, van der Leun AM, Yofe I, Lubling Y, Gelbard-Solodkin D, van Akkooi ACJ, i sur. Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. *Cell*. 2019 Feb 7;176(4):775-789.e18.
46. Zheng L, Qin S, Si W, Wang A, Xing B, Gao R, i sur. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science*. 2021 Dec 17;374(6574):abe6474.
47. Pizzolato G, Kaminski H, Tosolini M, Franchini DM, Pont F, Martins F, i sur. Single-cell RNA sequencing unveils the shared and the distinct cytotoxic hallmarks of human TCRV $\delta$ 1 and TCRV $\delta$ 2  $\gamma\delta$  T lymphocytes. *Proc Natl Acad Sci U S A*. 2019 Jun 11;116(24):11906-11915.
48. Kurtulus S, Madi A, Escobar G, Klapholz M, Nyman J, Christian E, i sur. Checkpoint Blockade Immunotherapy Induces Dynamic Changes in PD-1<sup>+</sup>CD8<sup>+</sup> Tumor-Infiltrating T Cells. *Immunity*. 2019 Jan 15;50(1):181-194.e6.
49. Abu-Bonsrah KD, Newgreen DF, Dottori M. Development of Functional Thyroid C Cell-like Cells from Human Pluripotent Cells in 2D and in 3D Scaffolds. *Cells*. 2021 Oct 26;10(11):2897.
50. Pratheeshkumar P, Siraj AK, Divya SP, Parvathareddy SK, Siraj S, Diaz R, i sur. Prognostic Value and Function of KLF5 in Papillary Thyroid Cancer. *Cancers (Basel)*. 2021 Jan 7;13(2):185.
51. Castrén E, Monteggia LM. Brain-Derived Neurotrophic Factor Signaling in Depression and Antidepressant Action. *Biol Psychiatry*. 2021 Jul 15;90(2):128-136.

## 10. LITERATURA


52. Ozolek JA. Selected Topics in the Pathology of the Thyroid and Parathyroid Glands in Children and Adolescents. *Head Neck Pathol.* 2021 Mar;15(1):85-106.
53. Luo Q, Guo F, Fu Q, Sui G. hsa\_circ\_0001018 promotes papillary thyroid cancer by facilitating cell survival, invasion, G<sub>1</sub>/S cell cycle progression, and repressing cell apoptosis via crosstalk with miR-338-3p and SOX4. *Mol Ther Nucleic Acids.* 2021 Feb 24;24:591-609.
54. Tsukui T, Sun KH, Wetter JB, Wilson-Kanamori JR, Hazelwood LA, Henderson NC, i sur. Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nat Commun.* 2020 Apr 21;11(1):1920.
55. Bignon M, Pichol-Thievend C, Hardouin J, Malbouyres M, Bréchet N, Nasciutti L, i sur. Lysyl oxidase-like protein-2 regulates sprouting angiogenesis and type IV collagen assembly in the endothelial basement membrane. *Blood.* 2011 Oct 6;118(14):3979-89.
56. Kato K, Diéguez-Hurtado R, Park DY, Hong SP, Kato-Azuma S, Adams S, i sur. Pulmonary pericytes regulate lung morphogenesis. *Nat Commun.* 2018 Jun 22;9(1):2448.
57. Rudnick JA, Monkkonen T, Mar FA, Barnes JM, Starobinets H, Goldsmith J, i sur. Autophagy in stromal fibroblasts promotes tumor desmoplasia and mammary tumorigenesis. *Genes Dev.* 2021 Jul 1;35(13-14):963-975.
58. Wirka RC, Wagh D, Paik DT, Pjanic M, Nguyen T, Miller CL, i sur. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat Med.* 2019 Aug;25(8):1280-1289.
59. Nitsch L, Wollman SH. Thyrotropin preparations are mitogenic for thyroid epithelial cells in follicles in suspension culture. *Proc Natl Acad Sci U S A.* 1980 May;77(5):2743-7.
60. Muhl L, Genové G, Leptidis S, Liu J, He L, Mocci G, i sur. Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat Commun.* 2020 Aug 7;11(1):3953.

## 10. LITERATURA

61. Sharma V, Letson J, Furuta S. Fibrous stroma: Driver and passenger in cancer development. *Sci Signal*. 2022 Mar 8;15(724):eabg3449.
62. Buechler MB, Pradhan RN, Krishnamurty AT, Cox C, Calviello AK, Wang AW, i sur. Cross-tissue organization of the fibroblast lineage. *Nature*. 2021 May;593(7860):575-579.
63. Sanin DE, Ge Y, Marinkovic E, Kabat AM, Castoldi A, Caputa G, i sur. A common framework of monocyte-derived macrophage activation. *Sci Immunol*. 2022 Apr 15;7(70):eabl7482.
64. Marinkovic T, Garin A, Yokota Y, Fu YX, Ruddle NH, Furtado GC, i sur. Interaction of mature CD3+CD4+ T cells with dendritic cells triggers the development of tertiary lymphoid structures in the thyroid. *J Clin Invest*. 2006 Oct;116(10):2622-32.
65. Tang H, Zhu M, Qiao J, Fu YX. Lymphotoxin signalling in tertiary lymphoid structures and immunotherapy. *Cell Mol Immunol*. 2017 Oct;14(10):809-818.
66. Biczysko M, Stajgis P, Zeromski J. Lymphocytes in the structure of non-neoplastic goitre. *Pol J Pathol*. 2002;53(2):73-81.
67. McGrogan A, Seaman HE, Wright JW, de Vries CS. The incidence of autoimmune thyroid disease: a systematic review of the literature. *Clin Endocrinol (Oxf)*. 2008 Nov;69(5):687-96.
68. Zimmermann MB, Andersson M. Assessment of iodine nutrition in populations: past, present, and future. *Nutr Rev*. 2012 Oct;70(10):553-70.
69. Pizzato M, Li M, Vignat J, Laversanne M, Singh D, La Vecchia C, i sur. The epidemiological landscape of thyroid cancer worldwide: GLOBOCAN estimates for incidence and mortality rates in 2020. *Lancet Diabetes Endocrinol*. 2022 Apr;10(4):264-272.

## 11. ŽIVOTOPIS

**OSOBNE INFORMACIJE**

 **Klara Ilić**  
 Zagorska 13, 21220 Trogir  
 (+385) 92 221 93 74  
 klara.ilic15@gmail.com

**OBRAZOVANJE I OSPOSOBLJAVANJE**

**2020.-** Medicinski fakultet Osijek, Diplomski sveučilišni studij  
 Medicinsko laboratorijska dijagnostika  
**2017.-2020.** Sveučilišni odjel zdravstvenih studija Split, Preddiplomski  
 sveučilišni studij Medicinsko laboratorijska dijagnostika  
**2013.-2017.** Srednja škola Ivana Lucića, Trogir; opća gimnazija  
**2005.-2013.** Osnovna škola Majstora Radovana, Trogir

**RADNO ISKUSTVO**

**2020.** KBC Split – administrativni poslovi, prijem uzoraka,  
 upisivanje pacijenata u BIS i LIS (studentski ugovor)

**OSOBNE VJEŠTINE**

| <b>Strani jezici</b> | <b>RAZUMIJEVANJE</b> | <b>GOVOR</b> | <b>PISANJE</b> |
|----------------------|----------------------|--------------|----------------|
| engleski             | C1                   | B2           | B2             |
| njemački             | A2                   | A1           | A1             |
| talijanski           | A2                   | A1           | A1             |
| španjolski           | A2                   | A1           | A1             |