

UNIVERZITET U BEOGRADU
MATEMATIČKI FAKULTET



Milica D. Simić

KLASTEROVANJE PODATAKA DOBIJENIH
STEREO-SEQ METODOM

master rad

Beograd, 2023.

Mentor:

dr Jovana KOVAČEVIĆ, docent
Univerzitet u Beogradu, Matematički fakultet

Članovi komisije:

dr Vladimir KOVAČEVIĆ, BGI

dr Mladen NIKOLIĆ, vanredni profesor
Univerzitet u Beogradu, Matematički fakultet

Datum odbrane: _____

Naslov master rada: Klasterovanje podataka dobijenih Stereo-seq metodom

Rezime:

Ključne reči:

Sadržaj

1	Uvod	1
2	Kako radi Stereo-seq metoda	3
2.1	Ekspresija gena	3
2.2	Stereo-seq metoda	4
3	Zaključak	8
	Bibliografija	9

Glava 1

Uvod

Stereo-seq (eng. *SpaTial Enhanced REsolution Omics - SEQuencing*) je inovativna metoda pomoću koje na osnovu uzorka tkiva dobijamo informacije o pojedinačnim ćelijama koje se nalaze u datom tkivu, o njihovom položaju u ravni i genima koji su ekspresovani (aktivni) u tim ćelijama. Razvijena je od strane BGI (eng. *Beijing Genomics Institute*), jednog od vodećih svetskih centara za genomska istraživanja i razvoj novih tehnologija za sekvenciranje genoma i analizu bioloških podataka.

geneID	x	y	MIDCounts	cell
Arpc1b	19881	9379	3	170675
B230118H07Rik	19881	9379	2	170675
Bcl2l12	19881	9379	1	170675
Ccnd3	19881	9379	1	170675
Cdk8	19881	9379	3	170675
Commd4	19881	9379	1	170675
Cops9	19881	9379	4	170675
Creg1	19881	9379	3	170675
Dhx36	19881	9379	4	170675
Emc7	19881	9379	2	170675
Emg1	19881	9379	1	170675
Fam111a	19881	9379	1	170675
Fbrs	19881	9379	1	170675

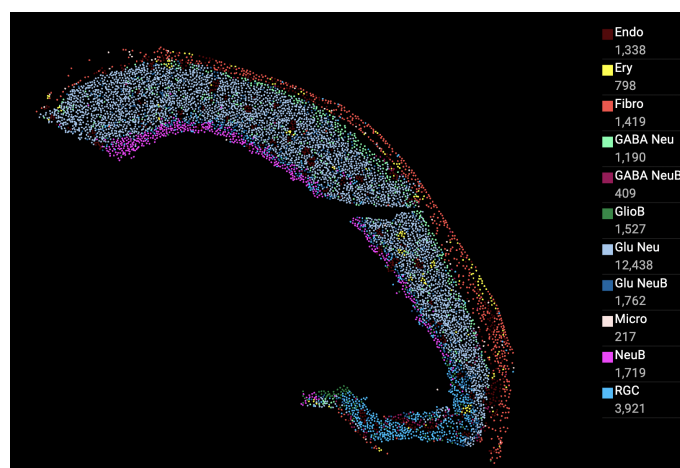
Slika 1.1: Podaci dobijeni Stereo-seq metodom

Podaci, dobijeni ovom metodom, u svakom redu sadrže (slika 1.1):

- oznaku gena (geneID),
- njegove koordinate (x i y),
- nivo ekspresije tog gena (MIDCounts) i
- oznaku ćelije u kojoj je taj gen ekspresovan (cell)

Oni su veoma korisni za dublje razumevanje odnosa između morfologije ćelija, njihove genske ekspresije i okoline u kojoj se ćelije nalaze.

U jednom tkivu se mogu nalaziti ćelije različitih tipova. Postoji očekivanje da se na osnovu položaja ćelije i drugih ćelija u njenoj okolini i njene genske ekspresije može zaključiti kom tipu ona pripada. Nekada su ćelije istog tipa blizu, a nekada su udaljene, nekada imaju sličnu gensku ekspresiju, a nekada različitu. Podaci dobijeni StereoSeq metodom ne daju direktno podatke o tipu ćelije, međutim postoje javno dostupni podaci o tipovima ćelija u različitim uzorcima.



Slika 1.2: Mozak embriona miša starosti 14 dana

Na slici 1.2 su ćelije raspoređene u skladu sa njihovim tipom. Primećujemo da su neke ćelije istog tipa (npr. Gliob) prostorno blizu, dok su neke prostorno udaljene (npr. GluNeu).

Cilj ovog master istraživanja je da se ispita zavisnost između koordinata ćelija i njihove genske ekspresije, kao i da se pomoću metoda klasterovanja nad podacima dobijenim Stereo-seq metodom ispita da li dobijeni klasteri odgovaraju tipovima ćelija.

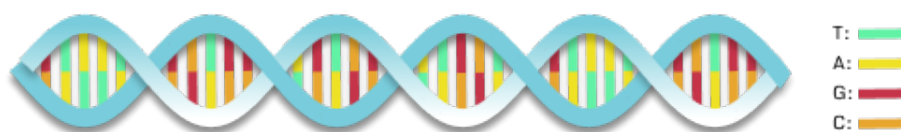
Glava 2

Kako radi Stereo-seq metoda

2.1 Ekspresija gena

Dezoksiribonukleinska kiselina (DNK) je molekul koji se nalazi u jezgru svake ćelije i prenosi genetske informacije sa jedne generacije na drugu putem reprodukcije. Sastoji od dva komplementarna lanca nukleotida. Nukleotide se razlikuju po azotnoj bazi koja se nalazi u njihovoj strukturi. Postoje 4 azotne baze u DNK nukleotidima i one su adenin (A), timin (T), guanin (G) i citozin (C).

Baze se uparuju prema određenom obrascu, gde adenin formira vezu sa timinom (A-T), a guanin formira vezu sa citozinom (G-C). Ove baze spajaju dva lanca DNK zajedno, a redosled baza na jednom lancu DNK predstavlja genetski kod koji određuje strukturu i funkciju proteina i drugih molekula u organizmu.



Slika 2.1: DNK molekul

Proces nastanka proteina (sinteza proteina) se odvija u dve glavne faze: transkripcija i translacija. Ovaj proces uključuje prenos genetske informacije iz DNK u RNK¹ molekule, koji zatim služe kao šablon za sintezu proteina.

- Transkripcija: Prepisivanje informacija iz DNK molekula u gRNK molekul (glasnička RNK). Odvija se duž molekula DNK, tako što enzim RNK polime-

¹Za razliku od DNK molekula, RNK je jednolančani molekul, i njegove nukleotide umesto timina (T) imaju uracil (U), koji se uparuje sa adeninom (A-U).

raza čita nukleotide i sintetiše komplementarni lanac gRNK. Kada polimeraza dođe do sekvence za zaustavljanje, proces transkripcije se završava, a gRNK molekul se odvaja od DNK.

- Procesiranje gRNK: Pre nego što napusti jezgro ćelije, gRNK se mora obraditi. U ovoj fazi se dešava dodavanje poli-A repa (sekvenca adenina) na gRNK.
- Translacija: U citoplazmi ćelije, gRNK dolazi do ribozoma, gde se informacija iz gRNK prevode u aminokiselinske sekvence proteina. Transfer RNK molekuli (tRNK) prepoznaju kodone (triplete nukleotida) na gRNK i donose odgovarajuće aminokiseline. Ribozom pomaže u povezivanju aminokiselina, stvarajući polipeptidni lanac koji će se presaviti u funkcionalni protein. Kada se dođe do zaustavnog kodona na gRNK, proces translacije se završava, a novosintetisani protein se oslobađa.

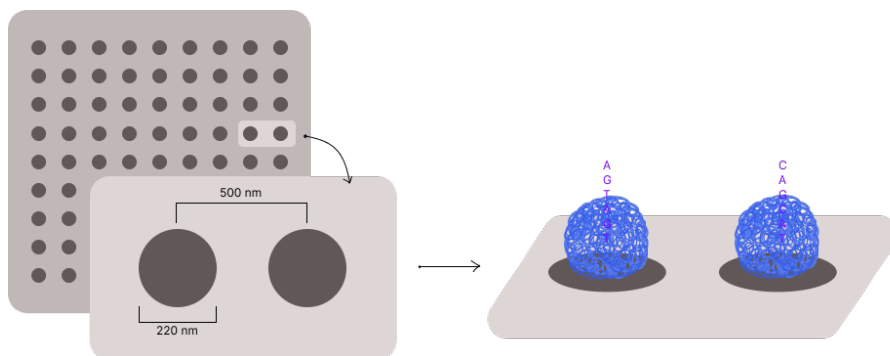
Svaka ćelija istog organizma sadrži isti skup gena, delova DNK molekula, koji se mogu ispoljiti (ekspresovati), ali i ne moraju. Ekspresija gena predstavlja kvantifikaciju nivoa transkripta (gRNK) ili proteina koji se proizvode iz tog gena.

2.2 Stereo-seq metoda

Stereo-seq je metoda koja omogućava istraživačima da analiziraju ekspresiju gena na nivou tkiva, uz istovremeno pružanje informacija o njihovom prostornom rasporedu. Ova metoda kombinuje uređene nizove DNK nano-kuglica² i in situ RNK sekvenciranje (sekvenciranje RNK na licu mesta) i može se predstaviti u nekoliko koraka:

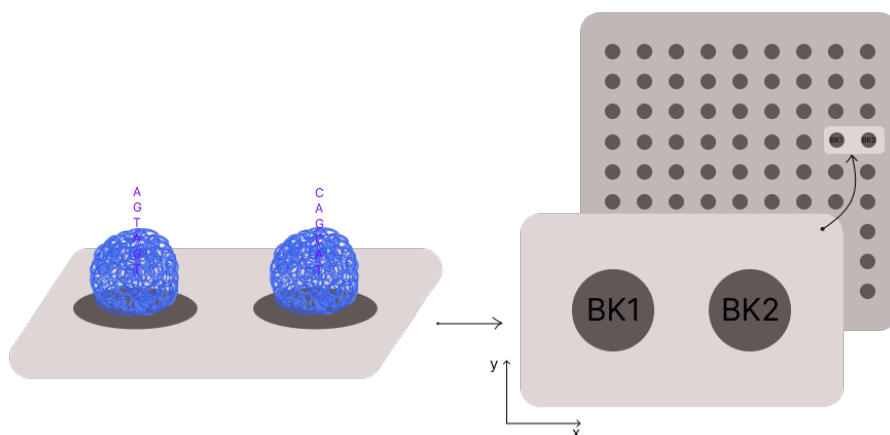
1. Priprema čipa uređenih nizova DNK nano-kuglica i formiranje matrice koordinatnih identiteta
 - Koriste se standardni čipovi, sa prostorom za DNK nano-kuglice prečnika od približno 220 nm i rastojanje od centra do centra od 500nm. Na svako mesto na čipu se postavljaju DNK nano-kuglice koje sadrže nasumične barkodove (sekvenca nukleotida). (slika 2.2)

²DNK nano-kuglice (engl. *DNA nanoballs*) su sferne strukture, izgrađene od kopija jednog DNK molekula, koje su složene u kompaktan i stabilan oblik.



Slika 2.2: Postavljanje DNK nano-kuglica na čip

- Izvršava se sekvenciranje kako bi se pročitali nasumični barkodovi prisutni na svakom mestu i onda se na osnovu tih informacija formira matrica koordinatnih identiteta. Svaki element u matrici predstavlja jedinstvenu lokaciju na čipu i sadrži odgovarajući barkod, koji ćemo zvati i koordinatni identitet - KID (eng. *Coordinate Identity*) (slika 2.3)

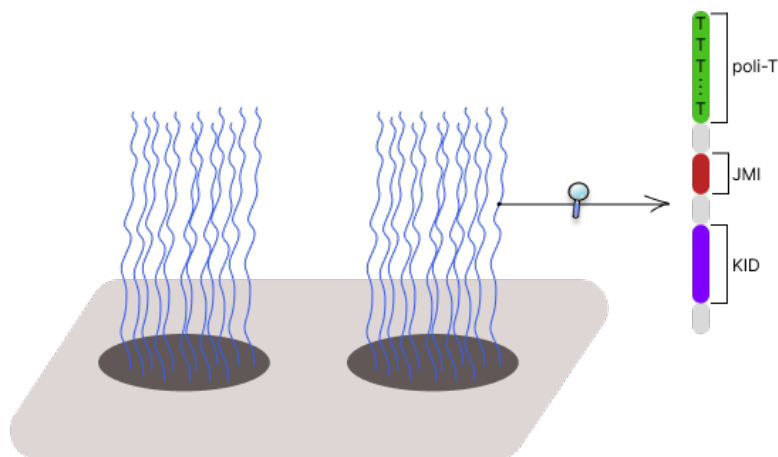


Slika 2.3: Sekvenciranje barkodova koji jedinstveno određuju svako mesto na čipu

2. In citu RNK sekvenciranje

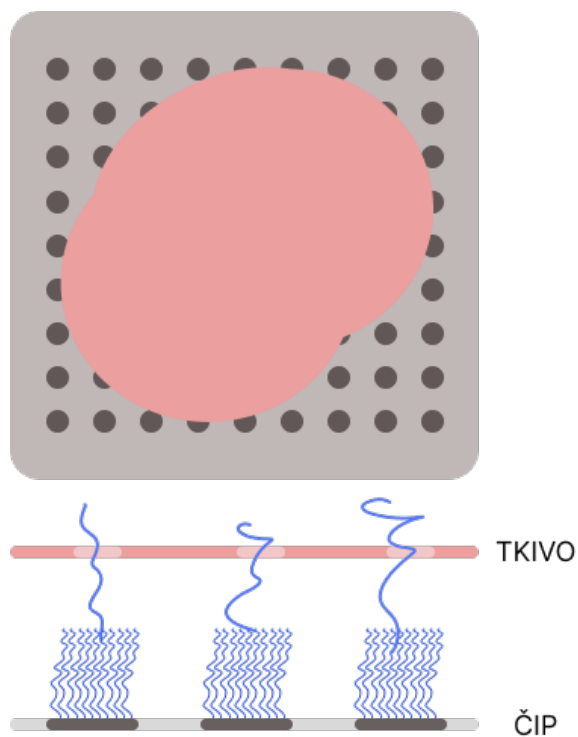
- Na svako mesto na čipu za KID (barkod) se vežu:
 - jedinstveni molekularni identifikatori (JMI) koji služe za obeležavanje individualnih molekula, što omogućava istraživačima da tačno kvantifikuju i razlikuju izvorne molekule od onih koji su rezultat amplifikacije i time odrede njegovu ekspresiju i

- oligonukleotidi sa poli-T sekvencama (sekvenca timina), kako bi se poli-A repovi RNK vezali za njih (slika 2.4)



Slika 2.4: Na čip se dodaju JMI i poli-T sekvenca

- Zamrznuto tkivo se postavlja na čip, fiksira i permeabilizira, omogućavajući difuziju poli-A ukrašenih RNK molekula iz njega. (slika 2.5)



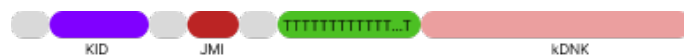
Slika 2.5: Difuzija RNK molekula iz tkiva

- Vezuje se poli-A ukrašeni RNK molekul za poli-T sekvencu i pomoću reverzne transkripcije na osnovu tog RNK molekula se sintetiše kDNK (komplementarna DNK). (slika 2.6) Zatim se vrši amplifikacija, čime se povećava količina materijala potrebnog za sekvenciranje.



Slika 2.6: Sintetisanje kDNK

- Nakon amplifikacije, vrši se sekvenciranje, tj. sakuplja se kDNK (koja sadrži informacije o poli-A ukrašenom RNK molekulu), jedinstveni molekularni identifikator (JMI) i koordinatni identitet (KID) koji odgovara položaju na čipu. (slika 2.7)



Slika 2.7: Sekvenciranje KID, JMI i kDNK

3. Analiza podataka dobijenih in citu RNK sekvenciranjem

- Na osnovu kDNK molekula i JMI sekvence koja nam pomaže prilikom kvantifikovanja tog kDNK molekula dobijamo nivo ekspresije gena koji odgovara tom kDNK molekulu
- Na osnovu koordinatnog identiteta i matrice koordinatnih identiteta određujemo koordinate u ravni ćelije u kojoj je taj gen ekspresovan

Glava 3

Zaključak

Bibliografija

Biografija autora