

## **KOMISIJI ZA STUDIJE II STEPENA ELEKTROTEHNIČKOG FAKULTETA U BEOGRADU**

Komisija za studije II stepena, Elektrotehničkog fakulteta u Beogradu, na svojoj sednici održanoj 28.08.2018. godine imenovalo nas je u Komisiju za pregled i ocenu master rada dipl. inž. Nikole Jovanovića pod naslovom „Obrada podataka iz DNK sekvencera nove generacije: konverzija u FASTQ format“. Nakon pregleda materijala Komisija podnosi sledeći

### **IZVEŠTAJ**

#### **1. Biografski podaci kandidata**

Nikola Jovanović je rođen 30.03.1992. godine u Beogradu. Završio je osnovnu školu "Ujedinjene nacije" u Beogradu kao vukovac. Upisao je Trinaestu beogradsku gimnaziju u Beogradu koju je završio sa odličnim uspehom. Elektrotehnički fakultet upisao je 2011. godine. Diplomirao je na odseku za Računarsku tehniku i informatiku 2016. godine sa prosečnom ocenom 8,05. Diplomski rad odbranio je u julu 2016. godine sa ocenom 10. Diplomske akademske – master studije na Elektrotehničkom fakultetu u Beogradu, na Modulu za računarsku tehniku i informatiku upisao je u oktobru 2016. godine. Položio je sve ispite sa prosečnom ocenom 8,80.

#### **2. Opis master rada**

Master rad kandidata sadrži 32 strane teksta, zajedno sa slikama i dodacima. Rad sadrži 7 poglavlja i spisak literature. Spisak literature sadrži 23 reference.

Prvo poglavlje predstavlja uvod u kome su opisani predmet i cilj rada. Dati su najznačajniji pojmovi u bioinformatici i problemi sa kojima se ova oblast suočava. Jedan od problema jeste čuvanje i obrada podataka, kao i konverzija iz jednog formata u drugi kako bi se omogućila njihova obrada korišćenjem bioinformatičkih alata.

Drugo poglavlje detaljnije opisuje proces sekvenciranja DNK. Dat je kratak istorijat tehnologija sekvenciranja i opisane tehnologije sekvenciranja nove generacije, koje se danas najšire koriste.

U okviru trećeg poglavlja rad je fokusiran na sekvencere kompanije Illumina i dat je opis procesa sekvenciranja korišćenjem sekvencera ove kompanije i unapređenja tehnike sekvenciranja koje su uveli sekvenci ove kompanije.

Četvrto poglavlje opisuje format podataka koje generiše sekvencer Illumina NovaSeq. Opisana je direktorijum struktura, bitnije datoteke u okviru strukture i način na koji se podaci čuvaju u okviru datoteka. Dat je i opis formata FASTQ za čuvanje informacija o određenim

sekvencama. Ovaj standardni format koriste alati koji od kratkih sekvenci koje proizvode sekvenceri rekonstruišu genom ili delove genoma. Datoteke u ovom formatu se smatraju početnom tačkom složenih bioinformatičkih analiza.

U petom poglavlju je detaljno opisana implementacija softvera za konverziju datoteka unutar direktorijum strukture koje proizvodi sekvencer Illumina NovaSeq u FASTQ format. Dat je detaljan opis korišćenih klasa, objašnjenje grupisanja polja i metoda u klase i pregled optimizacija procesa konverzije i radi ubrzanja obrade i smanjenja korišćenih resursa. Opisane su datoteke neophodne za konverziju i podaci koji su od značaja u njima, kao i ponašanje konvertera u zavisnosti od dostupnih informacija.

Šesto poglavlje, objašnjava načine na koji je sama realizacija testirana i validirana sa setom podataka koji su korišćeni kao referentni. Opisan je još jedan drugi alat koji se može koristiti u svrhe konverzije ovakvih podataka i način na koji se on pokreće. Naznačene su i razlike u pristupu tog alata i implementacije i objašnjene razlike u proizvedenim izlazima.

### **3. Analiza rada sa ključnim rezultatima**

Master rad dipl. inž. Nikole Jovanovića se bavi problematikom čuvanja i obrade podataka proizvedenih od strane sekvencera i njihovom standardizacijom. Sekvenceri nove tehnologije su napravljeni kako bi vršili masovno paralelno određivanje sekvence nukleotidnih baza kratkih DNK lanaca. Prilikom ovog procesa generiše se velika količina podataka, a unapređenjem same tehnologije u istom postupku sekvenciranja je omogućena i obrada velikog broja različitih uzoraka, što dalje povećava količinu proizvedenih podataka. Kako se problem sa brzine sekvenciranja prebacuje na čuvanje velike količine podataka, kompanije koje proizvode sekvencere uvode nove formate za čuvanje podataka koje sekvenceri proizvode.

Jedna od ovih je i kompanija Illumina, Inc. koja je definisala novi format - BCL (Binary Base Call). Konverzija ovog formata u standardni format koji obrađuju bioinformatički alati koji se koriste za dalju analizu sekvenciranih podataka - FASTQ format je glavna tema i implementacija jednog konvertera je cilj ovog rada.

Osnovni doprinosi rada su:

- (a) implementacija konvertera podataka iz BCL direktorijuma podataka u standardni FASTQ format
- (b) primena konverzije i poređenje sa postojećim rešenjima
- (c) smernice za dalje unapređenje implementacije konverzije podataka.

### **4. Zaključak i predlog**

Kandidat Nikola Jovanović je u svom master radu uspešno izvršio implementaciju konvertera podataka iz BCL direktorijum strukture u standardni FASTQ format, čime je omogućena dalja obrada podataka koje generiše Illumina NovaSeq sekvencer. Prvi korak u daljoj obradi ovih

podataka predstavlja rekonstrukciju genoma jedinke sa koje je uzet uzorak za sekvenciranje i poređenjem sa referentnim genomom se može doći do zaključaka o varijantama u sekvenciranom genomu. Predložena poboljšanja mogu značajno da unaprede mogućnosti primene implementacije konvertera. Kandidat je iskazao samostalnost i sistematičnost u svome postupku kao i inovativne elemente u rešavanju problematike ovog rada. Na osnovu gore navedenog Komisija predlaže Nastavno-naučnom veću Elektrotehničkog fakulteta u Beogradu da prihvati rad „Obrada podataka iz DNK sekvencera nove generacije: konverzija u FASTQ format“ dipl. inž. Nikole Jovanovića kao master rad i odobri javnu usmenu odbranu.

Beograd, 03.09.2018.

Članovi komisije:



dr Zoran Jovanović



dr Žarko Stanisljević