

BIOINFORMATIKA

1. Gde se nalazi DNK u celiji kod eukariota?
 - U jedru celije
2. Objasniti razliku izmedju transkripcije i translacije.
 - Transkripcija - sinteza RNK na osnovu DNK.
 - Translacija - sinteza proteina na osnovu RNK zapisa.
3. Objasniti pojam alel
 - Aleli su razlicite varijante istog gena. Npr. gen za krvnu grupu.
4. Sta je prajmer i koju ulogu ima u procesu sekvenciranja?
 - Prajmer je delic DNK koji sluzi kao marker za pocetak lanca DNK za vreme sinteze.
5. PCR ciklus se sastoji iz sledeca 3 procesa: _____, _____ i _____.
 - Denaturation, Annealing, and Extension (or Elongation)
6. Ako sekvencirana baza ima Phred kvalitet 10, kolika je verovatnoca da je ta baza pogresno ocitana?
 - $P = 10^{(-Q/10)}$
 - $P = 10^{(-10/10)}$
 - $P = 10^{(-1)}$
 - $P = 1/10 = 0.1\%$
7. Kako je nastao referentni ljudski genom?
 - U projektu 'Human Genome Project'
8. Ljudski genom se sastoji od oko _____ baznih parova.
 - 3 milijarde
9. Koja je razlika izmedju FASTA i FASTQ fajlova?
 - FASTQ - fajlovi koji sadrze read-ove tokom sekvenciranja, a ti readovi nisu jos spojeni u duze sekvence ili ceo genom.
 - FASTA - fajlovi koji sadrze referentni genom ili njegove delove

10. Sta je asembli?
- To je sklapanje read-ova u neke duze sekvence (sa krajnjim ciljem da npr. sklopimo ceo genom)
11. Sta je alignment?
- Sklapanje read-ova u duze sekvence uz pomoc referentnog genoma.
12. Koja je razlika izmedju De Bruijn i string grafa?
- **String** grafovi čuvaju informacije o celom readu, koriste se za duže sekvence
 - **De Bruijn** su konceptualno jednostavniji i brži, dužina čvora im je jedinstvena i jedinstvenu dužinu informacije na ivici, koriste se za kratke readove. Kod De Bruijn-a ne mozemo da konstruisemo jedinstvenu sekvencu na osnovu samog grafa.
13. Nacrtati De Bruijn graf za sekvencu "AAACCCA"
- AAAC AACC ACCC CCCA
AAA → AAC → ACC → CCC → CCA
14. Koliko se poredjenja izvrsi u najgorem slucaju naivnog algoritma za pretragu niski gde je referenca duzine 'n', a rid duzine 'm'?
- $(n-m+1)*m$ je broj poredjenja
 - $n*m$ je slozenost
15. Pored naivnog, koji se jos algoritam za pretragu niski cesto koristi?
- Smith-Waterman, BWT, naivni
16. Navesti nekoliko (bar tri) tipa struktura podataka koje se cesto koriste kod modernog pristupa mapiranja gena:
- Nizovi, hes mape, stable, FM index
17. Navesti nekoliko (bar tri) tipa struktura podataka koje se cesto koriste kod modernog pristupa mapiranja gena:
- Delecija, insercija, inverzija, copy number variant, translokacija, duplikacija, duplikacija citavog genoma
18. Koji od navedenih tipova uticaja varijanti je uglavnom neskodljiv:
- a) Silent
 - b) Missense

- c) Nonsense
- d) Frameshift

- Silent su bezopasne

19. Izracunati VAF (variant allele frequency) ako je pokrivenost (depth of coverage) 30 i postoji 10 varijanti koje podrzavaju REF i 20 varijanti koje podrzavaju ALT. (naci formulu u prezentacijama - valjda je $\text{alt}/(\text{alt}+\text{ref})$)

- $\text{VAF} = \text{ALT} / (\text{REF} + \text{ALT})$
- $\text{VAF} = 20 / (10+20)$
- $\text{VAF} = 20/30 = 2/3$

20. Odrediti genotipe (0/0, 0/1, 1/1, 1/2) u sledecim primerima:

a) REF: A b) REF: A c) REF: A d) REF: A

-----	-----	-----	-----
A	C	T	A
A	C	T	A
A	C	T	A
A	G	T	T
A	G	T	T
A	G	T	T

- a) 0/0
- b) 1/2
- c) 1/1
- d) 0/1

21. Zaokruziti nebalansirane (unbalanced) strukturne varijante (nebalansirane strukturne varijante menjaju duzinu genoma, a balansirane ne menjaju):

- a) Insercije
- b) Alele
- c) Translokacija
- d) Pirimidini
- e) Delecije
- f) Izoformi
- g) Inverzije

h) Duplikacije

Insercija, delecija, duplikacije,

22. Dat je referentni genom (REF) i sekvencirani genom (OBS). Sekvenciranje je urađeno tehnikom uparenih read-ova (paired-end sequencing) gde je dužina svakog read-a 150 baznih parova, a prosečna veličina fragmenta je 600 baznih parova. Urađeno je mapiranje sekvenciranih read-ova na referentni genom. Grafički prikazati izgled mapiranja read-ova na referencu u oblasti gde se u sekvenciranom genomu nalazi delecija veličine 200 baznih parova. Prikazati read pair, read depth i split read signale. Da li je iz svakog od ovih signala ponaosob moguće odrediti koja je strukturna varijanta prisutna u sekvenciranom genomu?

/

23. Prekursorska informaciona RNK (eng. pre-mRNA) je prvi proizvod procesa transkripcije i sadrži _____ i _____. Od lanca ove RNK nastaje informaciona RNK (eng. mRNA) tako što se pojedini _____ zadržavaju, a svi _____ izbacuju.
- Egzoni, introni, egzoni, introni
24. RNK sekvenca koja nastaje kao rezultat transkripcije jednog gena naziva se _____ tog gena.
- Transkript
25. Šta je alternativno isecanje (eng. alternative splicing)?
- To je mogućnost da se od jednog segmenta gena dobije više različitih RNK u zavisnosti od toga koje egzome pokupimo
26. Navesti dve metode normalizacije koje se koriste u RNA-Seq eksperimentima (normalizuju u odnosu na dužinu transkripta i ukupnog broja sekvenciranih ridova):
- FPKM и TPM

27. Prilikom statističkog testiranja hipoteza, verovatnoća da se pri uslovu nulte hipoteze dobiju još ekstremniji rezultati naziva se _____.
- P-vrednost
28. Metod normalizacije koji prvo normalizuje u odnosu na broj sekvenciranih ridova a zatim u odnosu na dužinu transkripta naziva se _____.
- TPM
29. Šta je rak?
- Čelije koje se abnormalno ponašaju; ne obavljaju svoju funkciju, besmrtno su, metastaziraju, troše resurse organizma, nekontrolisano se razmnožavaju
30. Šta su tumor-supresorski geni?
- Geni koji paze o tome da ćelije umru kad treba
31. Zašto čistoća tumora (eng. tumor purity) predstavlja problem kod pronalaženja varijanti?
- U tom malignom tumoru postoje i zdrave i ćelije kancera, tako da je bitno koji je odnos između njih i koji udeo čega uzimamo jer rezultati variraju
32. Sta je epigenetika?
- Oblast koja se bavi pitanjem kako od istog genetskog materijala dobijamo različite RNK
33. Sta je euhromatin a sta heterohromatin?
- Euhromatin je kad je DNK dostupna za čitanje, a heterohromatin kad nije
34. Sta je centralna dogma molekularne biologije?
- Centralna dogma molekularne biologije se bavi detaljnim ostatak-po-ostatak transferom sekcenvijalne informacije. Ona navodi da se takve informacije ne mogu vratiti sa proteina na bilo drugi protein ili nukleinsku kiselinu.

DODATNA PITANJA

1. Koliko parova hromozoma ima čovek?
- 23
2. Koja se 2 tipa grafova najcesce koriste za asembli?
- De Bruijn graf, Overlap graf

3. Navesti bar 3 algoritma/pristupa poravnanja genomskih sekvenci?
 - Brute force, suffix-based, hash-based
4. Koji fajl format je najzastupljeniji za cuvanje genomskih varijanti?
 - VCF
5. Mutacije u genomu duzine vece od 50 baznih parova nazivamo?
 - Strukturnim varijantama (SV)
6. Dopuniti centralnu dogmu molekularne biologije: ____ -> ____ -> ____
 - DNK -> RNK -> Protein
7. Sta se osim samih sekvenci jos cuva u FASTQ formatu?
 - Kvalitet read-ova za svaku bazu (phred score)
8. U kojem fajl format se cuvaju mapirane (aligned) sekvence?
 - BAM
9. Navesti 3 tipa strukturnih varijanti:
 - Balansirane
 - Nebalansirane
10. Zasto je potrebna normalizacija podataka u RNA-Seq analizi?
 - Eliminacija tehnickih varijacija
 - Poredjenje ekspresije gena
 - Otkrivanje retkih ili slabo ekspresnih gena
 - ...
11. Kada celije raka imaju mogucnost da koriste krvotok za migraciju u udaljene delove tela tada kazemo da je ____ ?
 - metastazirao