

גנומיקה חישובית- תרגיל בית 2

מגישות: אור ארבל תמר סעד

בנוס אפשרי: חינויות יתר

שאלה 1:

חישוב	התו	איכות הבסיס	הסיכוי לטעות	טעות ב(%)	
$\frac{100}{100} = 1$ $Q_{phred} = -10 \log_{10} 1 = 0$ $0 + 33 = ' ! '$!	0	1	100%	1
$\frac{42}{100} = 0.42$, $Q_{phred} = -10 \log_{10} 0.42 = 3.76$, $4 + 33 = ' \% '$	%	$3.76 \Rightarrow 4$	0.42	42%	2
$\frac{7}{100} = 0.07$, $Q_{phred} = -10 \log_{10} 0.07 = 11.54$, $12 + 33 = ' - '$	-	$11.54 \Rightarrow 12$	0.07	7%	3
$\frac{0.001}{100} = 0.00001$, $Q_{phred} = -10 \log_{10} 0.00001 = 50$, $50 + 33 = ' S '$	S	50	0.00001	0.001%	4
$\frac{0.025}{100} = 0.00025$, $Q_{phred} = -10 \log_{10} 0.00025 = 36.02$, $36 + 33 = ' E '$	E	36	0.00025	0.025%	5
$62 - 33 = 29$, $10^{-\left(\frac{29}{10}\right)} = 1.26 * 10^{-3}$, $1.26 * 10^{-3} * 100 = 0.126$	>	29	$1.26 * 10^{-3}$	0.126%	6
$72 - 33 = 39$, $10^{-\left(\frac{39}{10}\right)} = 1.258 * 10^{-4}$, $1.26 * 10^{-4} * 100 = 0.0126$	H	39	$1.26 * 10^{-4}$	0.0126%	7
$40 - 33 = 7$, $10^{-\left(\frac{40}{10}\right)} = 0.2$, $0.2 * 100 = 20$	(7	0.2	20%	8
$47 - 33 = 14$, $10^{-\left(\frac{14}{10}\right)} = 0.04$, $0.04 * 100 = 4$	/	14	0.04	4%	9
$94 - 33 = 61$, $10^{-\left(\frac{61}{10}\right)} = 7.94 * 10^{-7}$, $7.94 * 10^{-7} * 100 = 7.94 * 10^{-5}$	^	61	$7.94 * 10^{-7}$	$7.94 * 10^{-5} \%$	10

הערה לבדוק האדיב:

טבלה זו הוכנה בשעת לילה מעניינת ותהליך ההכנה שלה היה מלאה, אם כי לא מסובך במיוחד. מסיבות אלו, אם קיימת טעות חישובית קטנה או שמא במשוואה כתוב ערך שגוי (שיתכן ויתכן שאינו הגיע מהמשוואה שמעליו, או ממוחה הקודח של הכותבות האדיבות גם הן)- עמך הסליחה.

שאלה 2:

חיפשנו את הגן NM_021920.4 ב-Genome browser בגנום ההומני המבוקש.

א. הגדרנו בפלט המבוקש פורמט BED, וקיבלנו אותו כפלט.

chr11 626308 627181 ENST00000176195.4 0 - 626430 627143 12,12,120 4 222,93,102,109, 0,388,579,764,

- ב. קיבלנו 4 אקסונים, כמוצג בעמודה 10 של פורמט ה-BED.
 ג. האזור המקודד הוא 626430-627143, כמוצג בעמודות 7-8 בפורמט ה-BED.
 ד. אורכי האקסונים הם 109 ו-222. ראינו זאת גם בעמודה ה-11 בפורמט BED וגם במידע על ריצוף האזורים המקודדים.

בנוסף: המשמעות היא כרומזום שאינו רגיל, שנוצר כתוצאה של טרנסלוקציה, לדוגמה. כרומזום כזה עשוי להיות בעל מידע גנטי חלקי בלבד, או שיהיה מורכב מחלקי כרומזומים שונים. מצב זה עלול לגרום לתסמונות שונות, וניתן לאבחן אותו באמצעות בדיקת קריוטיפ. לא לנסות בבית (את החלק של טרנסלוקציה, לא קריוטיפ).

שאלה 3:

השתמשנו בטבלת ה-'refGene table', הגדרנו את הtrack להיות 'NCBI RefSeq' ואת הקבוצה להיות ('Gene and Gene Prediction Tracks'). כתבנו בפילטר את הטקסט הבא:

(strand="-") AND ((chrom LIKE "chr??" AND chrom NOT LIKE "chr10" AND chrom NOT LIKE "chr11") OR (chrom LIKE "chrM")) AND (cdsStart != txStart) AND (txEnd-txStart<1001) AND (exonCount>1)

ובתוצאות קיבלנו:

refGene (refGene) Summary Statistics	
item count	66
item bases	35,311 (0.00%)
item total	50,807 (0.00%)
smallest item	330
average item	770
biggest item	996
block count	144
block bases	24,678 (0.00%)
block total	35,233 (0.00%)
smallest block	31
average block	245
biggest block	861

- א. מצאנו 66 גנים.
 ב. חיפשנו פורמט BED עבור כל הגנים שמצאנו. ראינו בעמודה 10 כי לגן בעל מספר האקסונים הגדול ביותר יש 4 אקסונים, ולקחנו את פורמט ה-BED שלו:

chr16 67482570 67483547 NM_001138 0 - 67482635 67483398 0 4 248,86,133,31, 0,454,698,946,

- ג. השינויים שביצענו על מנת להגיע לתוצאה:

Select dataset

clade: genome: assembly:

group: track:

table:

Note: Most dbSNP tables are huge. Trying to download them through the Table Browser usually leads to errors. Please see our [Data Access FAQ](#) on how to download dbSNP data.

Define region of interest

region: ☐ genome ☒ position

identifiers (names/accessions):

עבור ההגדרות האלה, קיבלנו:

item count	317
item bases	326 (33.37%)
item total	349 (35.72%)
smallest item	0
average item	1
biggest item	13

כלומר, 317 סניפים.

ד. הוספנו בפילטר שהגדיל צריך להיות שלילי, ושהמוטציה צריכה להיות מסוג missense. קיבלנו

item count	2
item bases	2 (0.20%)
item total	2 (0.20%)
smallest item	1
average item	1
biggest item	1

ניתן לראות שקיבלנו 2 מוטציות מיסנס.

שאלה 4:

- הגדרנו את הפרמטרים כנדרש, והוספנו לפילטר כרומוזום מבוקש 10. קיבלנו 591,247 תוצאות.
- יצרנו הצלבה, ובה חיפשנו את כל ההצלבות בין השאילתא הקודמת שלנו לבין מאגר המידע שמכיל גנים.

covidHgiGwasR4PvalA2 (covidHgiGwasR4PvalA2) Summary Statistics

item count	21,973
item bases	21,973 (0.00%)
item total	21,973 (0.00%)
smallest item	1
average item	1
biggest item	1

ג. שמרנו את התוצאות מהסעיף הקודם כטבלת custom track, וביצענו איתה הצלבה עם repeat masker.

item count	4,866
item bases	4,866 (0.00%)
item total	4,866 (0.00%)
smallest item	1
average item	1
biggest item	1

ד. יצרנו custom track חדש עבור אזורים חזרניים ממשפחת Alu. לאחר מכן, חיפשנו הצלבה של הטבלה מסעיף ב' עם הטבלה שאותה יצרנו, על מנת למצוא את החפיפה ביניהן.

item count	1,299
item bases	1,299 (0.00%)
item total	1,299 (0.00%)
smallest item	1
average item	1
biggest item	1

ה. שמרנו את הטבלה מסעיף ד' בתור custom track (בשם המקורי: seifD), וביצענו עבורו הצלבה עם clinvar, כנדרש.

item count	18
item bases	18 (0.00%)
item total	18 (0.00%)
smallest item	1
average item	1
biggest item	1

ו. א. יצרנו טבלה חדשה בה חיפשנו סניפים, עם סינון של פילטר לאדנזין בלבד (A). יצרנו טבלה באופן דומה גם עבור ציטוזין (C). לאחר מכן ביצענו הצלבה של כל אחת מהטבלאות עם הטבלה מסעיף ד', וקיבלנו:

C:

A:

item count	3
item bases	3 (0.00%)
item total	3 (0.00%)
smallest item	1
average item	1
biggest item	1

item count	5
item bases	5 (0.00%)
item total	5 (0.00%)
smallest item	1
average item	1
biggest item	1

ב. באפשרויות השונות לפלט בחרנו בהיפר-קישורים לדפדפן הגנומי. שם קיבלנו את שמות הגנים ואת שמות הסניפים.

rs568873828	DCLRE1C
rs7079776	CACNB2
rs12257322	MRPS16

rs55885853	IL2RA
rs75676336	DCLRE1C
rs56218110	CHST3
rs2018017	HOGA1
rs1140593	ACADSB

שאלה 5:

- א. ראשית נכנסנו לתיקייה עם הקבצים באמצעות פקודת cd.
 הרצנו את הפקודה: `wc -l covid.txt`
 וקיבלנו 152024 שורות.
 על מנת למצוא כמה מהרשומות נמצאות בקובץ המבוקש, הרצנו:
`intersectBed -u -a covid.txt -b whole_gene.txt | wc -l`
 וקיבלנו 89236.
- ב.
1. הרצנו: `intersectBed -wa -a exons.txt -b introns.txt | wc -l`
 קיבלנו: 62267
 2. הרצנו: `intersectBed -wa -a exons.txt -b introns.txt | sort | uniq | wc -l`
 קיבלנו: 8301
 3. `intersectBed -wa -a exons.txt -b introns.txt | sort | uniq > exons_introns_intersection.txt`
 קיבלנו קובץ טקסט בשם `exons_introns_intersection.txt` שמכיל את התוצאה של סעיף ב'.
- ג. הרצנו: `intersectBed -wa -a exons_introns_intersection.txt -b CDS.txt | sort | uniq > ei_CDS_intersection.txt`
 קיבלנו קובץ טקסט בשם `ei_CDS_intersection.txt` שמכיל את תוצאת החיתוך של הקובץ הקודם עם האזורים המקודדים.
 בדקנו את מספר השורות עם `wc -l ei_CDS_intersection.txt`, וקיבלנו 4509 תוצאות.
- ד.
1. ענינו בסעיף לעיל. (מציאת התשובה מושארת כתרגיל בית לקורא הנבון) (סתם- זה לא כזה מסובך, זה 4509, זה כתוב ממש פה!)

2. הרצנו:

```
bedtools sort -i ei_CDS_intersection.txt | bedtools merge | wc -l
```

קיבלנו: 900.

3. כדי לשמור את הפלט לקובץ השתמשנו ב:

```
bedtools sort -i ei_CDS_intersection.txt | bedtools merge > merge_ei_CDS.txt
```

ה.

1. הרצנו: `intersectBed -a covid.txt -b merge_ei_CDS.txt > covid_marge_ei_CDS.txt`ואז כדי לגלות את מספר השורות: `wc -l covid_marge_ei_CDS.txt`

קיבלנו: 2356

2. הרצנו: `head -n 20 covid_marge_ei_CDS.txt`

קיבלנו:

```
orarb1@LAPTOP-LDVN9SE8: /mnt/c/Users/orarb/OneDrive/BED_files$ head -n 20 covid_marge_ei_CDS.txt
chr21 14108835 14108836 21:14108836 0 . 14108835 14108836 10526975
chr21 14108846 14108847 rs62208635 0 . 14108846 14108847 10526975
chr21 14108912 14108913 rs7279700 0 . 14108912 14108913 10526975
chr21 14109043 14109044 rs7278737 0 . 14109043 14109044 10526975
chr21 14144626 14144627 rs2822432 0 . 14144626 14144627 10526975
chr21 14152599 14152600 rs74369337 0 . 14152599 14152600 16752800
chr21 14165267 14165268 rs61740029 0 . 14165267 14165268 10526975
chr21 14186024 14186025 rs397517 0 . 14186024 14186025 10526975
chr21 14189301 14189302 rs11909217 0 . 14189301 14189302 255
chr21 14221163 14221164 rs141953885 0 . 14221163 14221164 255
chr21 14221173 14221174 rs378280 0 . 14221173 14221174 10526975
chr21 14224450 14224451 21:14224451 0 . 14224450 14224451 10526975
chr21 14227032 14227033 rs139439630 0 . 14227032 14227033 255
chr21 14516982 14516983 rs34607574 0 . 14516982 14516983 16752800
chr21 14546255 14546256 rs7281104 0 . 14546255 14546256 16752800
chr21 14546279 14546280 rs118133999 0 . 14546279 14546280 10526975
chr21 14582206 14582207 21:14582207 0 . 14582206 14582207 10526975
chr21 14582271 14582272 rs2822786 0 . 14582271 14582272 16752800
chr21 14582273 14582274 21:14582274 0 . 14582273 14582274 255
chr21 14582279 14582280 rs139465416 0 . 14582279 14582280 10526975
```

בנוס:

חיפשנו בכרומזום 21 אקסונים באזורים מקודדים שחופפים לאינטרונים בגנים שקיימת קורלציה ביניהם לבין פיתוח מחלה חמורה בקורונה. כלומר, חיפשנו אתרי שחבור חלופי בגנים הללו, כנראה במטרה להסביר את הקשר בין הגנים לבין רמת ההדבקות במחלה כתוצאה של שחבור חלופי. אנחנו התחלנו פה את המחקר, ומכאן הדרך סלולה להמשך מחקר פורה ומשם הדרך קצרה לנובל. בהצלחה!