

## גנומיקה חישובית- תרגיל 4

מגישות: תמר סעד  
אור ארבל

### חלק א':

#### שאלה 1:

הורדנו את הקבצים שלנו מהאתר  
<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE158952>  
מעיים, ולקחנו את שני הקבצים:  
[SRR12765537](#) , [SRR12765536](#)

הורדנו את הקבצים באמצעות הפקודה:

```
fastq-dump --gzip -X 1000000 -O $WANTED_DIR --split-files SRR12765536
```

הורדנו את כרומוזום 7 ההומני באמצעות הפקודה:

```
wget  
'https://hgdownload.soe.ucsc.edu/goldenPath/hg38/chromosomes/chr7.fa.gz'
```

בחרנו בכרומוזום 7 משום ש:

א. כל השביעין חביבין ב. שבע הוא המספר הקסום ביותר ג. כך עלה במחשבה לפנינו

את הפקודה והכרומוזום מצאנו באתר UCSC:

הפקודה להורדת קובץ הכתובה לעיל -> Sequence data by chromosome-> Human-> Downloads

לחילוץ הקובץ מהקובץ המכוון השתמשנו בפקודה: `zcat chr7.fa.gz > chr7.fa`

#### שאלה 2:

בדקנו את מספר השורות שלנו בקובץ ה-multiqc שמוסבר למטה כיצד יצרנו אותו:

SRR12765536_1	51.0%	52%	1.0
SRR12765536_2	50.5%	52%	1.0
SRR12765537_1	43.6%	51%	1.0
SRR12765537_2	41.1%	51%	1.0

כך שניתן לראות שאכן יש לנו מיליון רידים בכל קובץ.

לא ראינו צורך בחילוץ הקבצים.

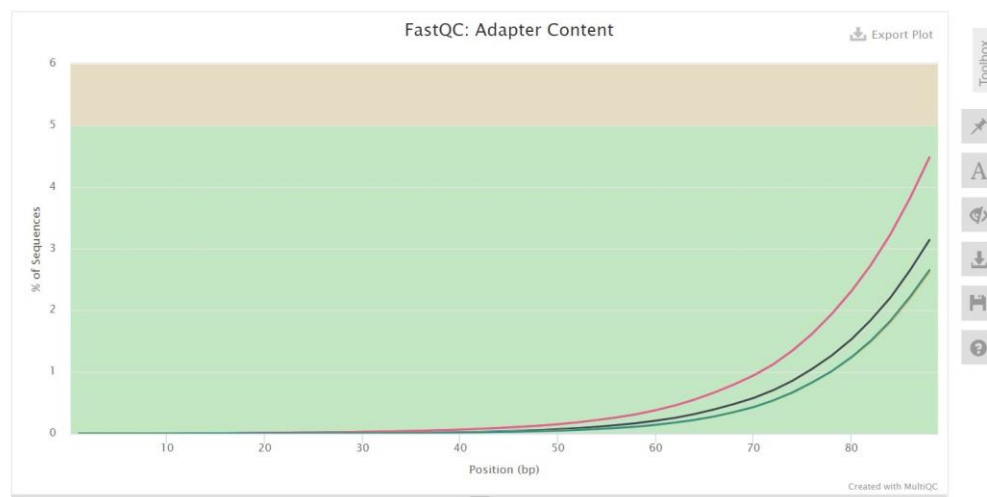
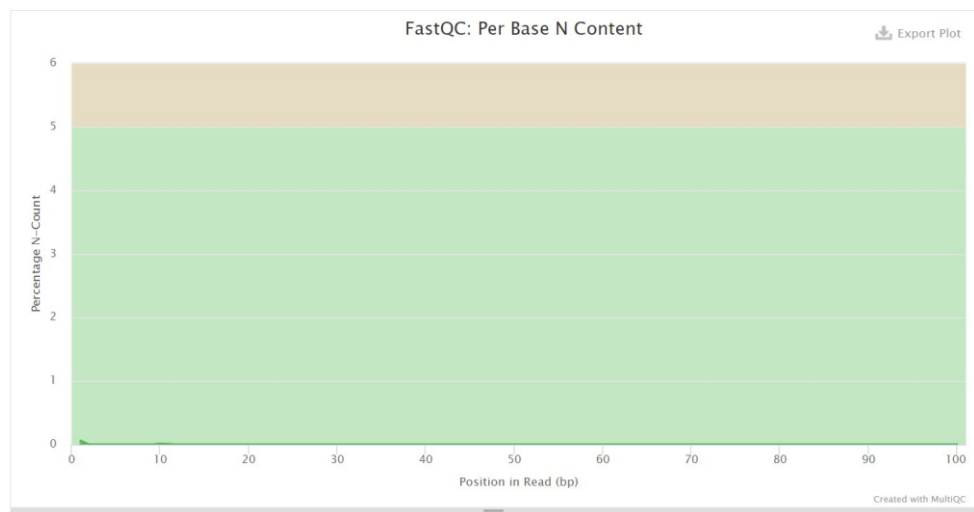
יצרנו קבצי fastqc שיראה לנו את איכות הריצוף, באמצעות הפקודה:

```
fastqc -o ${WANTED_DIR} ${WANTED_DIR}"SRR12765536_1.fastq.gz"  
${WANTED_DIR}"SRR12765536_2.fastq.gz" ${WANTED_DIR}"SRR12765537_1.fastq.gz"  
"${WANTED_DIR}"SRR12765537_2.fastq.gz
```

הרצנו על קבצי ה-fastqc את התוכנה multiqc:

```
home/alu/kobish/MULTIQC/multiqc/multiqc --in_dir $WANTED_DIR --out_dir /  
$WANTED_DIR
```

ניתן לראות כי הקבצים יצאו באיכות טובה:



כפי שניתן לראות, רמת הריצוף טובה וכן אין צורך בביצוע גזום (Trimming) לקבצים.

### שאלה 3:

1. יצרנו אינדקס עם הפקודה: `bwa-0.7.4 index chr7.fa`
2. ביצענו התאמה לגנום עם הפקודה:  
`bwa-0.7.4 mem "${WANTED_DIR}"/chr7.fa "${WANTED_DIR}"/SRR12765536_1.fastq.gz "${WANTED_DIR}"/SRR12765536_2.fastq.gz > SRR12765536.sam`
3. כדי לגלות כמה רידים התמפו לגנום באופן ייחודי כתבנו:

samtools flagstat SRR12765536.sam

וקיבלנו את המספרים הבאים:

```
[aluguest@alu12 Ex4Genomics]$ samtools-1.9 flagstat "/home/alu/aluguest/Project2022_Renana_Or/Ex4Genom
cs/SRR12765536.sam"
2010643 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 secondary
0 + 0 supplementary
0 + 0 duplicates
453535 + 0 mapped (22.56% : N/A)
2010643 + 0 paired in sequencing
1005127 + 0 read1
1005516 + 0 read2
398527 + 0 properly paired (19.82% : N/A)
427069 + 0 with itself and mate mapped
26466 + 0 singletons (1.32% : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)

2013839 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 secondary
0 + 0 supplementary
0 + 0 duplicates
430470 + 0 mapped (21.38% : N/A)
2013839 + 0 paired in sequencing
1006655 + 0 read1
1007184 + 0 read2
367231 + 0 properly paired (18.24% : N/A)
402941 + 0 with itself and mate mapped
27529 + 0 singletons (1.37% : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
[aluguest@alu12 Ex4Genomics]$
```

כלומר, בקובץ הראשון:

453535

בקובץ השני: 430470

4. בקובץ הראשון: 22.56%

בקובץ השני: 21.38%

#### שאלה 4:

1. על מנת להמיר את קבצי ה-SAM לקבצי BAM השתמשנו בפקודה:  
"samtools-1.9 view -Sb "SRR12765536.sam" > "SRR12765536.bam"
2. כיוון שקבצי BAM בינאריים, למחשב יותר קל לעבוד איתם ולכן הפעולות מתרחשות בזמן ריצה קצר יותר.
3. כדי למצוא את מספר הקריאות בקובץ ה-BAM כתבנו:  
samtools-1.9 view "SRR12765537.bam" | wc -l  
ויצא לנו: 2013839 עבור קובץ אחד, ו- 2010643 עבור הקובץ השני.  
נשים לב שבשאלה 2 בדקנו את מספר הקריאות עבור 4 קבצים, אך כאן יש לנו שני קבצים שמורכבים משניים מהקבצים האלה, כיוון שהיו לנו קבצים paired end. לכן החיבור שלהם אמור לצאת כ-2 מיליון.
4. על מנת לסנן את הקריאות שהתמפו בלבד, השתמשנו בפקודה:  
samtools-1.9 view -b -F 4 "SRR12765536.bam" | samtools-1.9 sort > "SRR12765536\_noUnmappad.bam"

5. בוצע כבר בסעיף 4 (מוליטאסקריות או מה?)

#### שאלה 5:

1. `bcftools-1.8 mpileup -Ov -f ${WANTED_DIR}/chr7.fa "SRR12765536_noUnmappad.bam" | bcftools-1.8 call -mv -o "" "${WANTED_DIR}/SRR12765536.vcf"`  
על מנת למצוא הבדלים בין הקבצים הרצונו:
2. `bcftools-1.8 stats ${WANTED_DIR}/SRR12765537.vcf > stats37.txt`  
וקיבלנו את ההבדלים הבאים:  
בקובץ SRR12765536:  
מספר סניפים:

```
26 SN → 0→number·of·SNPs:→71658
```

מספר שורות:

```
24 SN → 0→number·of·records: → 2986253
```

בקובץ SRR12765537:

מספר סניפים:

```
26 SN → 0→number·of·SNPs:→70442
```

מספר שורות:

```
24 SN → 0→number·of·records: → 2800613
```

## חלק ב':

שלב מקדים:

יצרנו קובץ vcf חדש שמכיל רק את המקומות בהם היו מוטציות A-G או C-T:

```
cat SRR12765536.vcf | awk '!/^ *#/ {print $0}' | awk '($4 == "A" && $5 == "G")||($4 == "C" && $5 == "T") {print $0}' > SRR12765536edit.vcf
```

## שאלה 6:

1. את הקבצים שקיבלנו מהשלב המקדים שלחנו ל-ANNOVAR. השתמשנו באתר האינטרנטי, ועל כן אין לנו פקודת הרצה.
2. דוגמאות לתוצאות שקיבלנו:

chr7	2360847	2360847	A	G	exonic	EIF3B
chr7	193450	193450	A	G	exonic	FAM20C
chr7	195676	195676	A	G	exonic	FAM20C

## שאלה 7:

1. חיפשנו את מחלת המעיים קרוהן, ובחרנו 5 גנים: IL6, NOD2, TNF, IL23R, IL10.
2. ביצענו חיפוש ב-Gene and Gene Prediction של הגנים הכתובים לעיל. את החיפוש הכנסנו בשדה ה-Name שבפילטר. הורדנו את תוצאת השאילתא בתור קובץ BED.
3. על מנת להפוך את קובץ ה-VCF לקובץ BED השתמשנו בפקודה:  
`"vcf2bed < "SRR12765536edit.vcf" > "SRR12765536edit_bed.txt"`  
לחיתוך השתמשנו בפקודה:  
`intersectBed -a SRR12765536edit_bed.txt -b chosengenes`  
התקבל קובץ ריק עבור שני הקבצים שלנו, ולכן לא קיים חיתוך.

4. הורדנו מה-table browser את טבלת הסניפים שקיימים בכרומוזום 7.

את החיתוך עשינו עם אותה הפקודה:

intersectBed -a SRR12765536edit\_bed.txt -b clinVar

פה קיבלנו תוצרי חיתוך לשני הקבצים:

:SRR12765536edit\_bed

```
chr7 44645448 44645449 . 126 A G . DP=10;VDB=0.265379;SGB=-0.556411;RPB=0.810265;
MQB=1.01283;MQSB=1;BQB=0.810265;MQOF=0;ICB=1;HOB=0.5;AC=1;AN=2;DP4=2,1,3,1;MQ=60 GT:PL 0/1:159,0,115
chr7 75985687 75985688 . 128 C T . DP=5;VDB=0.640281;SGB=-0.590765;MQSB=1;MQOF=0;
AC=2;AN=2;DP4=0,0,4,1;MQ=60 GT:PL 1/1:158,15,0
chr7 87509328 87509329 . 225 A G . DP=14;VDB=0.131006;SGB=-0.651104;MQSB=1;MQOF=0;
;AC=2;AN=2;DP4=0,0,7,1;MQ=60 GT:PL 1/1:255,24,0
chr7 87550284 87550285 . 137 A G . DP=7;VDB=0.00595258;SGB=-0.616816;MQSB=0.25;MQ
OF=0;AC=2;AN=2;DP4=0,0,2,4;MQ=35 GT:PL 1/1:167,18,0
chr7 92012613 92012614 . 11.7172 A G . DP=1;SGB=-0.379885;MQOF=0;AC=2;AN=2;DP4=0,0,0,
1;MQ=60 GT:PL 1/1:41,3,0
chr7 92085596 92085597 . 30.4183 C T . DP=2;SGB=-0.379885;MQOF=0;AC=2;AN=2;DP4=0,0,1,
0;MQ=60 GT:PL 1/1:60,3,0
chr7 94428303 94428304 . 19.4765 A G . DP=10;VDB=0.26;SGB=-0.453602;RPB=0.2;MQB=0.4;M
QSB=0.810265;BQB=1;MQOF=0;ICB=1;HOB=0.5;AC=1;AN=2;DP4=1,4,2,0;MQ=57 GT:PL 0/1:53,0,149
chr7 107773973 107773974 . 225 A G . DP=37;VDB=0.357893;SGB=-0.692562;MQOF=0;AC=2;A
N=2;DP4=0,0,22,0;MQ=60 GT:PL 1/1:255,66,0
chr7 107782808 107782809 . 93 C T . DP=18;VDB=0.0107983;SGB=-0.556411;RPB=0.579578
;MQB=0.579578;MQSB=0.105399;BQB=0.463395;MQOF=0;ICB=1;HOB=0.5;AC=1;AN=2;DP4=6,5,
3,1;MQ=56 GT:PL 0/1:126,0,255
chr7 114629920 114629921 . 12.6566 A G . DP=5;VDB=0.510154;SGB=-0.556411;MQSB=0;MQOF=0.
4;AC=2;AN=2;DP4=0,0,2,2;MQ=9 GT:PL 1/1:42,12,0
chr7 116797151 116797152 . 11.7172 A G . DP=1;SGB=-0.379885;MQOF=0;AC=2;AN=2;DP4=0,0,1,
0;MQ=60 GT:PL 1/1:41,3,0
chr7 116797449 116797450 . 11.7172 A G . DP=1;SGB=-0.379885;MQOF=0;AC=2;AN=2;DP4=0,0,0,
1;MQ=60 GT:PL 1/1:41,3,0
chr7 128837235 128837236 . 43.4147 C T . DP=2;VDB=0.02;SGB=-0.453602;MQOF=0;AC=2;AN=2;D
P4=0,0,2,0;MQ=40 GT:PL 1/1:73,6,0
chr7 128848639 128848640 . 45.4146 A G . DP=2;VDB=0.56;SGB=-0.453602;MQOF=0;AC=2;AN=2;D
P4=0,0,2,0;MQ=56 GT:PL 1/1:75,6,0
chr7 142750560 142750561 . 72 C T . DP=3;VDB=0.0221621;SGB=-0.511536;MQOF=0;AC=2;A
N=2;DP4=0,0,3,0;MQ=60 GT:PL 1/1:102,9,0
chr7 150977847 150977848 . 121 A G . DP=6;VDB=0.0221621;SGB=-0.511536;MQOF=0;AC=2;A
N=2;DP4=0,0,0,3;MQ=60 GT:PL 1/1:151,9,0
chr7 156681988 156681989 . 4.37899 C T . DP=2;SGB=-0.379885;RPB=1;MQB=1;MQSB=1;BQB=1;MQ
```

:SRR12765537edit\_bed

```
chr7 75985687 75985688 . 93 C T . DP=5;VDB=0.228191;SGB=-0.511536;MQSB=1;MQOF=0;
AC=2;AN=2;DP4=0,0,2,1;MQ=54 GT:PL 1/1:123,9,0
chr7 92083383 92083384 . 105 A G . DP=6;VDB=0.690245;SGB=-0.511536;RPB=1;MQB=1;MQ
SB=1;BQB=1;MQOF=0;ICB=1;HOB=0.5;AC=1;AN=2;DP4=1,0,2,1;MQ=60 GT:PL 0/1:139,0,29
chr7 92084657 92084658 . 4.38466 C T . DP=1;SGB=-0.379885;MQOF=0;AC=2;AN=2;DP4=0,0,1,
0;MQ=60 GT:PL 1/1:32,3,0
chr7 92085596 92085597 . 85 C T . DP=3;VDB=0.0389723;SGB=-0.511536;MQSB=1;MQOF=0
;AC=2;AN=2;DP4=0,0,2,1;MQ=60 GT:PL 1/1:115,9,0
chr7 92086338 92086339 . 11.7172 C T . DP=1;SGB=-0.379885;MQOF=0;AC=2;AN=2;DP4=0,0,0,
1;MQ=60 GT:PL 1/1:41,3,0
chr7 94427036 94427037 . 125 C T . DP=8;VDB=0.0513395;SGB=-0.590765;RPB=0.1;MQB=0
.2;MQSB=0.2;BQB=1;MQOF=0;ICB=1;HOB=0.5;AC=1;AN=2;DP4=1,1,1,4;MQ=48 GT:PL 0/1:159,0,61
chr7 94428303 94428304 . 99 A G . DP=24;VDB=0.00923447;SGB=-0.590765;RPB=0.84530
7;MQB=0.00673795;MQSB=0.526752;BQB=0.449329;MQOF=0;ICB=1;HOB=0.5;AC=1;AN=2;DP4=8,4,5,0;MQ=58 GT:PL 0/1:132,0,255
chr7 107773973 107773974 . 225 A G . DP=69;VDB=0.143003;SGB=-0.693147;MQSB=1;MQOF=0
;AC=2;AN=2;DP4=0,0,54,2;MQ=60 GT:PL 1/1:255,169,0
chr7 107782808 107782809 . 191 C T . DP=61;VDB=1.04003e-07;SGB=-0.686358;RPB=0.2297
14;MQB=1;MQSB=1;BQB=0.999939;MQOF=0;ICB=1;HOB=0.5;AC=1;AN=2;DP4=35,0,13,1;MQ=60 GT:PL 0/1:224,0,255
chr7 107929071 107929072 . 42.4147 A G . DP=2;VDB=0.02;SGB=-0.453602;MQSB=1;MQOF=0;AC=2
;AN=2;DP4=0,0,1,1;MQ=36 GT:PL 1/1:72,6,0
chr7 114629920 114629921 . 35.4663 A G . DP=18;VDB=4.42314e-08;SGB=-0.676189;RPB=1;MQB=
1;MQSB=1;BQB=1;MQOF=0.0555556;ICB=1;HOB=0.5;AC=1;AN=2;DP4=1,0,10,1;MQ=18 GT:PL 0/1:69,0,24
chr7 116797449 116797450 . 11.7172 A G . DP=1;SGB=-0.379885;MQOF=0;AC=2;AN=2;DP4=0,0,1,
0;MQ=60 GT:PL 1/1:41,3,0
chr7 128394574 128394575 . 11.7172 C T . DP=1;SGB=-0.379885;MQOF=0;AC=2;AN=2;DP4=0,0,0,
1;MQ=60 GT:PL 1/1:41,3,0
chr7 128830918 128830919 . 45.4146 A G . DP=2;VDB=0.38;SGB=-0.453602;MQOF=0;AC=2;AN=2;D
P4=0,0,2,0;MQ=60 GT:PL 1/1:75,6,0
chr7 128848639 128848640 . 71 A G . DP=3;VDB=0.1;SGB=-0.453602;MQSB=1;MQOF=0;AC=2;
AN=2;DP4=0,0,1,1;MQ=60 GT:PL 1/1:101,6,0
chr7 142750560 142750561 . 45.4146 C T . DP=2;VDB=0.02;SGB=-0.453602;MQOF=0;AC=2;AN=2;D
P4=0,0,2,0;MQ=60 GT:PL 1/1:75,6,0
```

## חלק ג' - תמר:

שלב מקדים:

זוהי המוטציה שלי:

chr7	4860183	4860183	A	G	exonic	PAPOLB	nonsynonymous SNV	PAPOLB:NM_020144:exon1:c.T1628C:p.M543T
------	---------	---------	---	---	--------	--------	-------------------	---

ב-ANNOVAR זה מופיע רחוק למדי ולכן זה לא מופיע בצילום המסך, אבל המוטציה לא נובעת מ-snp.

שאלה 9:



שם הגן הינו: PAPOLB

מיקום המוטציה בגן הינו: בנוקלאוטיד 4860183, באקסון.

מספר האקסונים: 1

מספר האינטרונים: 0

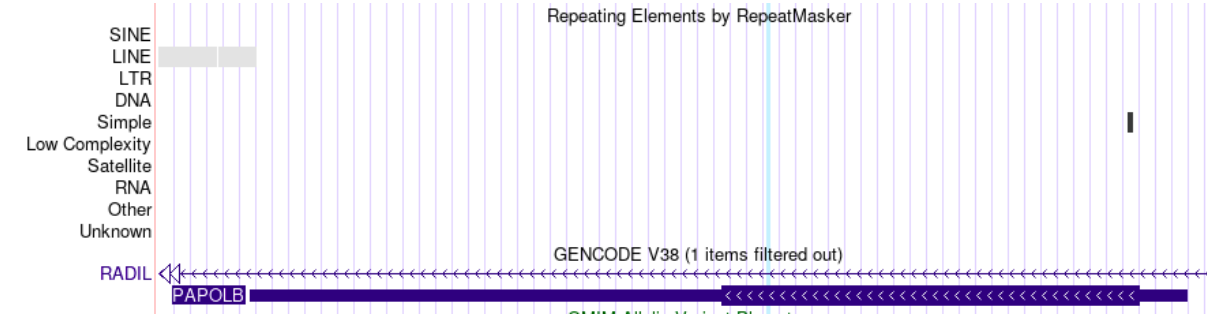
אורך הטרנסקריפט: 4,293, אורך האזורים המקודדים: 1,914

תפקידו הביולוגי של הגן: הוא polyA פולימראז אצל הומוסאפינס.

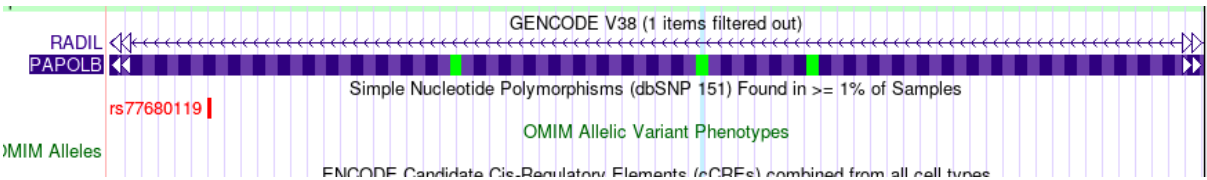


הנקודה אינה שמורה באבולוציה (המיקום מסומן בתכלת):

הנקודה קיימת באזור ללא אלמנטים חזרניים:

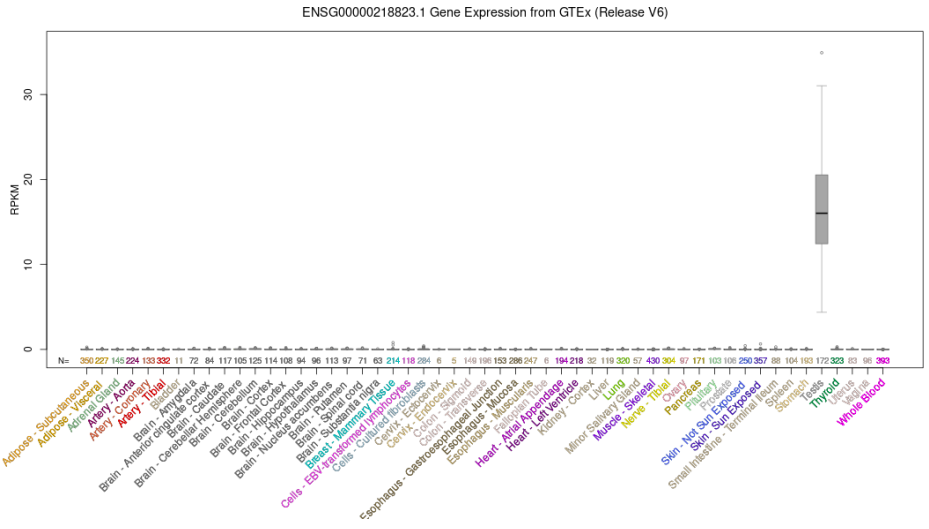


הנקודה היא לא ב-SNP:



תפקידו המולקולרי של הגן: הוא לוקח רצף רנ"א ומולקולת ATP, מפרק שני פוספטים ומחבר את האדנוזין לרצף הרנ"א.

הגן מתבטא ברקמת testis.



## שאלה 10:

- הנקודה אכן נמצאת באקסון. החומצה האמינית המקורית היא מתיונין, והיא הופכת לטריאנין, זה קורה במיקום 543 (כפי שניתן לראות במידע שניתן לנו ב-ANNOVAR. צילום מסך מופיע למעלה). לא קיימות מסגרות קריאה אלטרנטיביות, כיוון שקיים בגן אקסון אחד בלבד.
- סוג המוטציה הוא non-synonymous, כלומר גורם להחלפה של חומצת אמינו.
- מאחר שלגן יש אקסון אחד, אין לו איזופורמים כלל. (אחלה גן שבעולם!)

ניתן לראות כי ה-score שניתן למוטציה הוא נמוך (0.223), וכן הפרדיקציה שלה היא N שמשמעה פולימורפיזם בנוקלאוטיד, כלומר כנראה שהמוטציה תהיה לא מזיקה (אם כי יתכן שלאחר שיבצעו עליה מחקר מעמיק של מספר שנים, משהו יחשוב לקרוא לה "לא מזיקה ברובה").

MutationTaster converted rankscore	MutationTaster pred
0.588	D
0.090	N
0.223	N

## שאלה 11:

נרצה לקחת מדגם גדול של אנשים שחולים במחלות המעיים אותם חקרו בניסוי שבמידע שלו השתמשנו. ניקח מדגם גדול של אנשים חולים, יחד עם מדגם גדול של אנשים בריאים. נעשה ריצוף רנ"א אצל האנשים הללו של הגן, ונבדוק האם בנקודה הזו אנו רואים הבדל בהתפלגות של המוטציה באוכלוסיית החולים מול הבריאים. את הפקת הרנ"א נרצה לעשות מאזורים ספציפיים: נרצה לקחת את ה-rectum וה-ileum, שם נמצאה המוטציה מלכתחילה, וכן נרצה לקחת את ה-testis, ששם ראינו שהגן מתבטא משמעותית הכי הרבה.

נשים לב כי יש לעשות את האנליזה ברמת הרנ"א, כיוון שראינו שהמוטציה לא ממוקמת באזור של snps. משמע, זוהי לא מוטציה תורשתית ברמת הדנ"א, ועל כן יש לבדוק אותה ברמת הרנ"א ולא לעשות ריצוף דנ"א.

## חלק ג' - אור:

chr7	4860109	4860109	C	T	exonic	PAPOLB	nonsynonymous SNV
------	---------	---------	---	---	--------	--------	-------------------

## שאלה 9:

שם הגן הינו: PAPOLB

מיקום המוטציה בגן הינו: באקסון, במיקום 4860109

מספר האקסונים: 1

מספר האינטרונים: 0

אורך הגן: אורך הטרנסקריפט – 4293bp, אורך האזור המקודד – 1914bp.

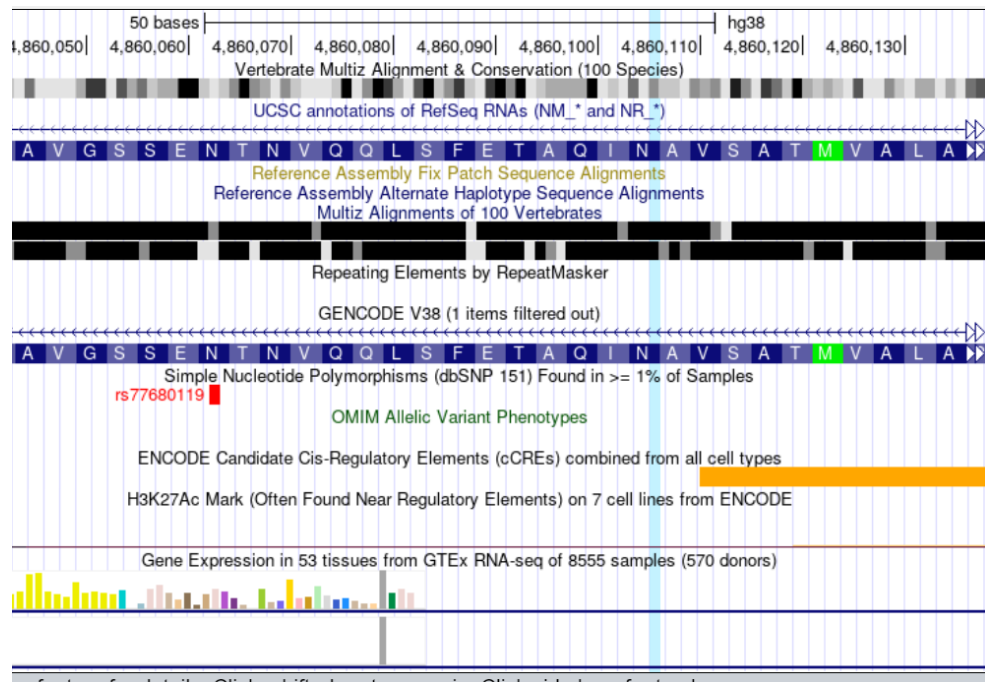
תפקידו הביולוגי של הגן: Poly A polymerase באדם.

הנקודה שמורה/אינה שמורה באבולוציה: הנקודה אינה שמורה, ציון השמירות שלה הוא 0.2.

הנקודה קיימת באלמנטים חזרניים, ב-ALU: הוספתי לgenome browser את RepeatMasker וניתן היה לראות שאין בנקודה אלמנטים חזרניים.

הנקודה היא לא ב-SNP: נכון.

הגן מתבטא ברקמת: testis.



## שאלה 10:

המוטציה מצויה באקסון ומשנה את החומצה אלאנין לטריאונין בעמדה 568. לגן אין מסגרות קריאה אלטרנטיביות מכיוון שיש לו אקסון אחד.

סוג המוטציה שקרתה הוא nonsynonymous, כלומר גורמת להחלפת חומצה אמינית.

לגן אין מסגרות איזופורמים שונים מכיוון שיש לו אקסון אחד.

ניתן לראות באתר annovar שציון mutationTaster הוא נמוך מאוד – 0.09, והוא N, כלומר – זוהי מוטציה שככל הנראה נובעת מפולימורפיזם. גם דרגת השמירות באיזור זה נמוכה מאוד, לכן ניתן להניח שמוטציה זו אינה חמורה.

MutationTaster converted rankscore	MutationTaster pred
0.588	D
0.090	N

## שאלה 11:

כדי להבין אם מוטציה זו אכן קשורה למחלה, ניקח דגימות ממספר גדול של אנשים החולים במחלות מעי דלקתיות (קוליטיס וקרוהן) וממספר גדול של אנשים בריאים. נרצף את הגן PAPOLB בכל אחת מהדגימות ונבדוק האם המוטציה שמצאנו קיימת בו. נבדוק אם תדירות המוטציה שונה באופן מובהק בין חולים



לבריאם. נרצה לוודא גם ששוני זה לא נגרם כתוצאה מגורם שלישי כלשהו, למשל – מוטציה קרובה בגנום שהיא בעלת השפעה על המחלה והמוטציה הנוכחית היא רק "טרמפיסטית" שלה. אם אכן יימצא שוני מובהק מבחינה סטטיסטית, נוכל להניח שיש קשר כלשהו בין המוטציה שבדקנו למחלה.

## חלק ד':

עבדנו על נתונים ממאגר ריצופים של אנשים עם מחלות מעיים. לקחנו דגימה אחת שנלקחה מ-rectum, והשנייה נלקחה מ-ileum.

ראשית הורדנו את קבצי ה-fastq וביצענו עליהם בדיקת איכות באמצעות fastqc. לאחר שוידאנו שהקבצים מרוצפים בצורה טובה, ביצענו alignment לגנום באמצעות התוכנה BWA. את קבצי ה-SAM שקיבלנו כפלט המרנו לקבצי BAM, שהכילו רק את הריידים שעברו מיפוי (). את קבצי ה-BAM המרנו לקבצי VCF, בהם הסתכלנו על סטטיסטיקות שונות- מספר הסניפים שקיימים בכל אחד, ומספר השורות.

בחלק השני סיננו את קבצי ה-VCF כך שיכילו רק מוטציות A-G ו-C-T. התבוננו בתוצאות הקובץ באתר ANNOVAR, וחפשנו מוטציות שלא נובעות מסניפים. לקחנו מ-genevar 5 גנים שקשורים למחלה אותה בדקנו, ורצינו לבדוק האם הם קיימים בקבצי ה-VCF שלנו. לצורך כך הורדנו מה-table browser קובץ bed של הגנים הללו, וכן הפכנו את קבצי ה-VCF שלנו לקבצי BED. בדקנו האם קיים חיתוך בין הטבלאות, וקיבלנו שלא. אחרי כן בדקנו אותו דבר עם טבלה שמכילה סניפים, ושם דווקא קיבלנו חיתוכים.

באנליזה ראינו שהגנים שבחרנו (שהיו גנים של מערכת החיסון שמעודדים דלקתיות) לא הופיעו בקובץ שלנו. תוצאה זאת לא מפתיעה, כיוון שייתכן שהם לא נמצאים על הכרומוזום שבחרנו, וייתכן גם שלא חלה בהם מוטציה מהסוג שבדקנו. עם זאת, לאחר ביצוע חיתוך עם טבלת הסניפים של clinVar ראינו כי מופיעים בקבצים שלנו snps. תוצאה זו חשובה לנו לאנליזת המשך, כיוון שנרצה לבדוק מקרים בהם מעבר מנוקלאוטיד אחד לאחר קרה כתוצר של עריכת רנ"א ולא של מוטציה ברמת הדנ"א.

## בנוסף:

### **תמר:**

בשאלה 6 חיפשתי את טבלת ה-VCF שמכילה את המוטציות A-G ו-C-T באתר ANNOVAR. באתר הפעלתי סינון כך שאראה מוטציות באזורים מקודדים ושהן non-synonymous. בבחירת המוטציה בדקתי בשדה dbSNP שהמוטציה לא נמצאת ב-snp. לאחר הסינונים הללו נותרו מספר גדול של מוטציות, ומהן בחרתי אחת באופן אקראי.

המוטציה היא מוטציה ממתיונין לטריאונין בגן PAPOLB, שהוא polyA פולימראז בבני אדם. המוטציה כאמור נמצאת באזור מקודד, ואינה נמצאת על snp. לפי MutationTaster המוטציה היא N, כלומר כנראה שלא תהיה מזיקה.

לאור מה שכתבתי לעיל, לא הייתי חושבת שהמוטציה הזו ספציפית מעניינת במיוחד. אמנם היה לה פתח מרשים (גם מקודדת וגם לא סניף), אבל נראה שהיא לא משפיעה ברמה משמעותית. עם זאת, כשבחרתי את המוטציה ב-ANNOVAR, לא יכולתי שלא לשים לב שהיו בגן הזה המון מוטציות מסוג דומה- non-synonymous, באזורים מקודדים ולא על סניף. לכן אולי היה מעניין לחקור את השפעות המוטציות על הגן הזה באופן כללי, בהקשר של המחלה אותה בדקנו.

### **אור:**

ב-annovar הפעלנו סינון כדי לקבל רק מוטציות המצויות באקסונים. בחרתי מוטציה שהיא nonsynonymous. וידאתי שהיא אינה נחשבת SNP מכיוון שסניפים הם יותר נפוצים ולכן פחות סביר שהם קשורים באופן ספציפי למחלה.

המוטציה היא מוטצית nonsynonymous באקסון של הגן PAPOLB שאחראי על הכנסת פוליא A במRNA. היא משנה את החומצה האמינית אלאנין לטריאונין. ציון הmutationTaster שלה נמוך מאוד, ולכן ככל הנראה היא אינה משפיעה על המחלה.

לסיכום, לדעתי אין קשר משמעותי בין מוטציה זו למחלות המעיים מכיוון שהיא לא מאוד משמעותית לתפקוד החלבון (היא קורית באיזור בעל שמירות נמוכה וקיבלה ציון נמוך). בנוסף, ראינו שחלבון זה לא מתבטא ברמות גבוהות במעי, ולכן ניתן להניח שמוטציה בו לא תשפיע על מחלות המעי.