**תרגיל בית 3 – מבוא לביואינפורמטיקה**

**מגישים: אלון הרשקוביץ 203854443, אפרת לבקוביץ 204536270**

**DNA sequencing and variant calling**

1. --
2. הגן הוא Tmc1 בעכבר. מצאנו אותו ע"י פתיחת קובץ ה-fasta בעורך טקסט, והעתקת הרצף למנוע החיפוש של blastn, יחד עם הרמז ואחוזי ההתאמה הגבוהים אין ספק שמדובר בגן הזה.

* השוני נמצא בעמדה 127 (כפי שמוצג בתמונה בסוף הסעיפים).
* כן, חירשות, כפי שראינו בת"ב 1. עם זאת, עם המידע הנתון לנו בתרגיל זה, חיפוש בבלאסט לא מניב מידע לגבי המוטציה, והעמדה הנתונה לנו היא יחסית למקטע הקצר שקבענו כגנום, כך שקשה לקבוע האם משויך לשוני פנוטיפ כלשהו.
* בדוגמה הראשונה:

A – 98, C – 0, G – 1, T – 893

בדוגמה השניה:

A – 643, C – 1, G – 1, T – 349

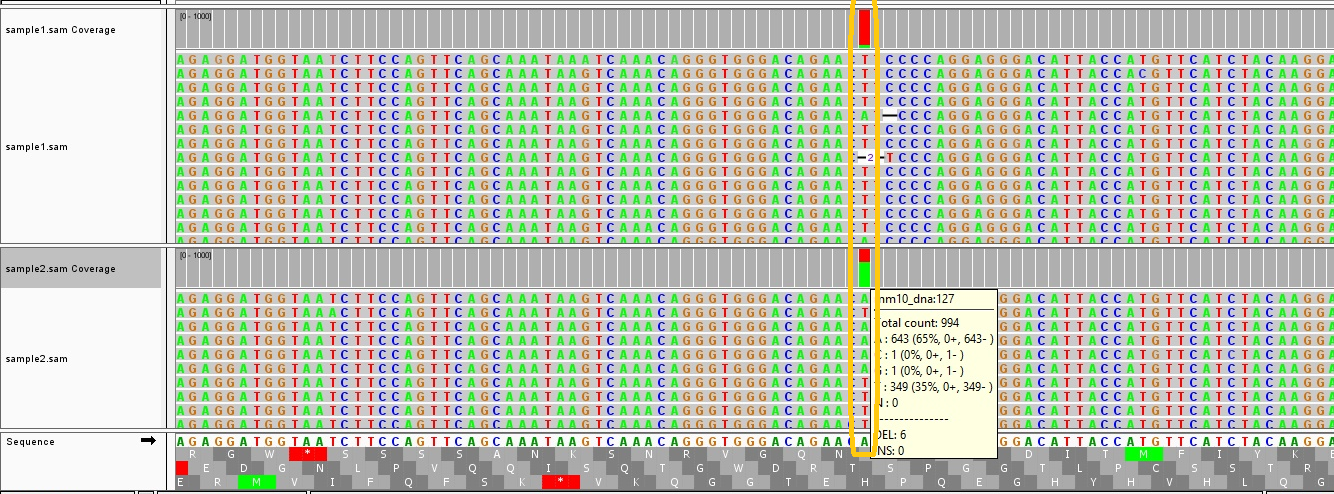
הריבוי השונה בין הדוגמאות – יתכן שנובע מכך שדוגמה 1 הינה של הומוזיגוט ודוגמה 2 הינה של הטרוזיגוט.

* (תמונה בסוף הסעיפים)

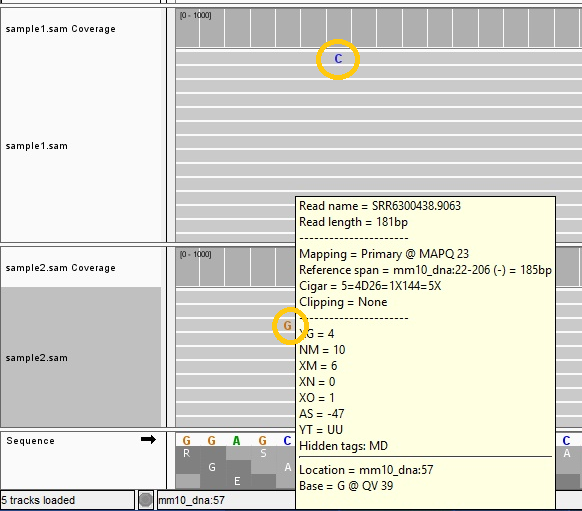
"מוטציה" כזו יכולה להגרם משגיאה בתהליך הריצוף בה מוכנס בטעות נוקלאוטיד לא נכון ונשאר לאורך הריצוף.

* (תמונה בסוף הסעיפים)
* לא ניכר שיש מיקומים ספציפיים, אבל מהסתכלות על כלל הדוגמאות הם לרוב מופיעים ליד השוני המשמעותי מתחילת הסעיף, או יותר באיזור הקצוות (או קרוב לקצוות).
* בקצוות התופעה יכולה פשוט לנבוע מהתהליך עצמו, שפחות יעיל ככל שמתקרבים לקצה/בתחילת הרצף. באיזור השוני ניזכר שזהו אזור מטרה של CRISPR שהינו מערכת עריכת גנים, כלומר ה-deletion נגרם ע"י מערכת עריכת הגנים.
* עקרונית תופעות מסוג deletion אינן צפויות מאחר ומשנות את מסגרת הקריאה ועלולות לגרום למוטציות נוספות, כמו גם להשתקת המוטציה.

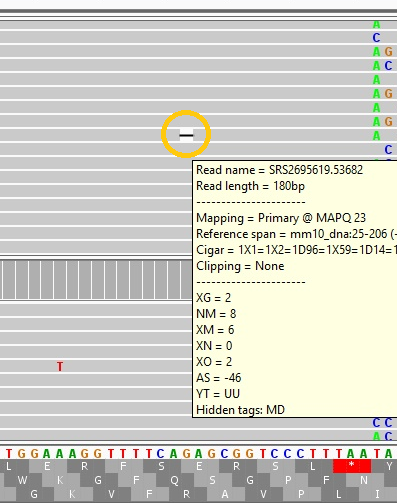
significant variation:



point variation (shows 2 different ones in both samples):



indel:



**Gene expression – theoretical questions**

1. הקושי יכול לנבוע ממגוון סיבות – שימוש במקטעים גדולים מדי שאינם מאפשרים זיהוי אקסונים נפרדים, או חלופין שכיחות נמוכה של איזופורמים כך שלא מגיעים לכמות הדרושה בשביל שתזוהה בשיטה. כמו כן, דמיון גבוה בין הרצפים מאחר ומדובר באותו מקטע דנ"א שמשועתק, כך שזיהוי באמצעות אינדיקטורים על הגן פחות יעיל.

רוב הבעיות הנ"ל מאפיינות בעיקר את שיטת ה-microarray, שדורש כמות גדולה יחסית של RNA וכן מסתמך על הצמדת גלאים.

בשיטת ה-RNA-seq יש הרבה פחות קושי בעניין זה נעשה גם פירוק למקטעים קטנים המאפשרים לזהות חיבורים בין אקסונים, וגם מקטעים גדולים יותר או ביצוע הריצוף בצורה של pair-end שנותנים מידע על חיבור של בין מס' אקסונים, וכן נדרשת כמות קטנה יותר של RNA ואין שימוש בגלאים כי אם בריצוף. אך הקושי העיקרי הוא עיבוד המידע שנוצר בתהליך שכן ב-RNA-seq מתקבלת כמות גדולה מאוד של מידע.

1. ניתן לפתור את הבעיות בשיטת ה-microarray באמצעות שיטה המכונה exon junction array המשתמשת בגלאים הנקשרים לאיזורים הצפויים לשמש כאיזורי חיתוך בקצוות האקסונים, כלומר מאפשרת לסמן את גבולות האקסונים וע"י כך לאפיין ביטוי איזומורפים של הגן. כמו כן בשיטה הנקראת exon array הקושרת גלאים לאקסונים עצמם ניתן לזהות איזופורמים שונים.

עבור שיטת ה-RNA-seq ניתן ליצור אלגוריתמים מהירים למציאה וספירה של איזופורמים, למשל כפי שמיושם ע"י [cufflinks](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3146043/) המשתמשת באלגוריתמים הסתברותיים למטרה זו.