

## תרגיל 3 גנומיקה חישובית

### חלק א':

#### שאלה 1.

- א. ניתן לראות כי איכות הריצוף בכל הרידים יורדת משמעותית לאחר כ-60 בסיסים. נרצה לבצע גיזום (trimming) על כל הרידים ולחתוך אותם לאורך של 60 הבסיסים הראשונים, ואז נבצע בדיקת איכות נוספת לראות האם כעת כל הרידים באיכות טובה מספיק. ניתן לראות כי קיימים רידים שלכל אורך הריצוף נמצאים בחלק התחתון של הגרף הירוק, כך שנרצה לוודא שלאחר ביצוע גיזום הם יהיו באיכות טובה מספיק. אם לא, ייתכן שנרצה להשמיט אותם במלואם.
- א. ברצף דנ"א אמורה להיות התפלגות פחות או יותר אחידה בין סוגי הבסיסים השונים, כך שנצפה לראות כ-25% בסיסים מכל סוג. בתמונה זו המצב לא כזה, וישנם שינויים משמעותיים לאורך הריד. ניתן להסיק מכך שאיכות הריצוף בריד זה לא היתה טובה, וכי לא נוכל לסמוך עליו ולהשתמש במידע שקיבלנו עליו.
- ב. ניתן לראות כי אף על פי שהשכיחות נותרת קבועה יחסית עבור כל בסיס, היא אינה אחידה בין כולם. ניתן לראות שיש חוסר אחידות בבסיס 10, אך באזור זה קיימים פריימרים בדרך כלל ולכן אנחנו מצפים לראות חוסר אחידות. ייתכן שמדובר פה במספר רידים בעייתיים שמטים את התוצאות, ולכן ננסה לאתר אותם, לזרוק אותם ולעשות בדיקת איכות נוספת. יש לציין כי ייתכנו מקרים בהם אכן לא נצפה לקבל שכיחות אחידה בין הבסיסים השונים, משום שהספרייה שלנו לא היתה רנדומלית. במקרים כאלה נדע מראש שהגרף הזה פחות רלוונטי עבורנו.
- א. ניתן לראות כי קיים מספר גבוה מאוד של דופליקטים. זה עלול לנבוע מהטייה של ה-PCR של מקטעים מסוימים שהוגברו ברמה גבוהה מיתר המקטעים, כך שבריצוף נראה בעיקר אותם. במקרה כזה ניתן לנקות דופליקטים.
- ב. ניתן לראות שיש נוכחות גבוהה של אדפטורים ברצפים. ניתן לגזום אותם באמצעות תוכנה מתאימה כמו Trimmomatic, אך כיוון שיש להם נוכחות כל כך גבוהה משלב מאוד מוקדם, הגיזום יותיר לנו רידים קצרים מאוד שייתכן שנתקשה לבצע להם alignment לגנום ולהמשיך לעבוד איתם. ננסה לגזום אותם בכל זאת וננסה להשתמש במידע שמתקבל.

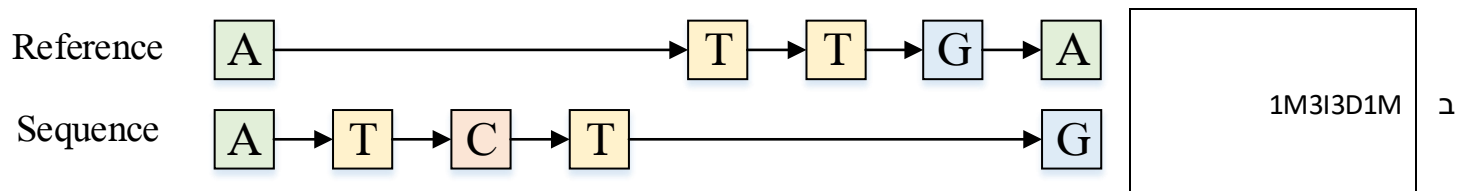
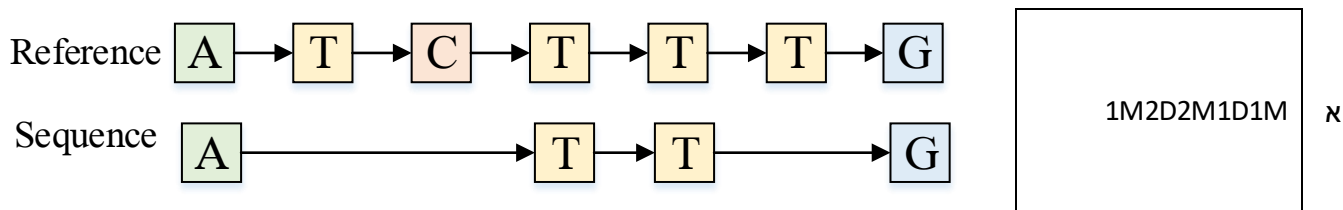
#### שאלה 2:

א.

רצף הCIGAR	אורך הרצף בגנום	אורך הריד המותאם	משמעות הCIGAR	
3S97M	97	100	97 התאמות/אי-התאמות, ורצף של 3 בסיסים שקיימים בשאילתא ולא קיימים בגנום.	א
3S30I40M10D30M	80	103	רצף של 3 שקיים בשאילתא ולא בגנום, 30 inserts, 40 התאמות/אי התאמות, 10 דלישנס, 30 התאמות/אי התאמות	ב
5H20M30I15M29D100N30M	194	100	רצף של 5 שמופיע בשאילתא ולא בגנום, 20 התאמות.אי התאמות, 30 הכנסות, 15 התאמות.אי התאמות, 29 מחיקות, גאפ באורך 100 (אינטרון?), 30 התאמות.אי התאמות	ג

בס"ד

ב.



שאלה 3:

ממה מורכב	בנוסף: התוצאה תקינה?	אם כן: מי הותאם? איפה (אם ידוע)? מי מה pair?	היה alignment (התאמה) (לגנום)?	קידוד	
0+1+6	כן, כי קיבלנו התאמה מלאה לשני הקצוות	שניהם עברו התאמה, והוא הראשון מהזוג	כן	67	א
0+1+5+6	כן, כי קיבלנו התאמה מלאה לשני הקצוות	שניהם עברו התאמה, והוא הראשון מהזוג. בן הזוג שלו עבר התאמה ברוורס.	כן	99	ב
2	לא 😞	לא	לא	4	ג
0+1+4+7	כן, כי קיבלנו התאמה מלאה לשני הקצוות	שניהם עברו התאמה, והוא השני מהזוג, ועבר התאמה ברוורס	כן	147	ד
1+4+5	קיבלנו תוצאות לא הגיוניות: בוצעה התאמה בזוג אע"פ שלא נעשה paired end, וכן יצא לנו שגם הריד וגם בן זוגו (הדמיוני) עברו התאמה ברוורס.	בעיה ההה	לא	50	ה

קידוד	הרשאות משתמש	הרשאות קבוצה	הרשאות אחרים	שימוש? (בנוס)
א	744	Rwx	r	אם נרצה לשמור על כימוס הקוד שלנו, אבל נרצה לתת לאנשים אפשרות להשתמש בו.
ב	444	R	R	כשנרצה שיהיה לנו קובץ קבוע ובטוח לא יערכו בו שינויים
ג	421	R	W	כשמישהו אחר כותב לקובץ ואנחנו בודקים אותו, אבל הוא לשימוש חיצוני כלשהו.
ד	404	R	---	Page not found
ה	763	rwX	rw	כשנרצה שליטה מלאה בקובץ, עם יכולת צפיה ועריכה של הקבוצה. הציבור מקבל יכולת כתיבה והרצה, אבל לא קריאה-לצורך אימונים בכתיבה עיוורת?

חלק ב':

מוגש בקבצים נפרדים.  
תם ונשלם הקונדס.