

תשובות לתרגיל 4:

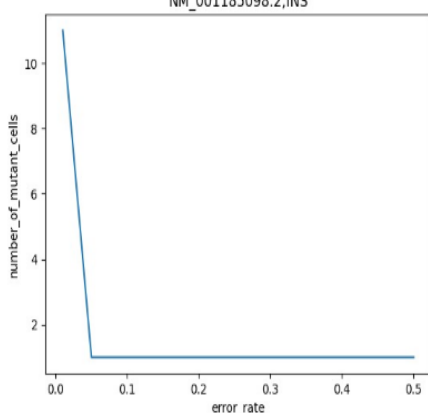
תמר סעד 207256991

רחל וינברגר 208812628

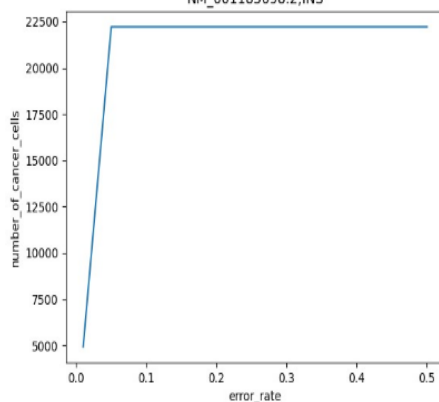
1. החשיבות של שלוש החזרות לכל קצב טעויות היא בגלל האקראיות של מיקום המוטציות, כך שבכל חזרה נוצר לנו חלבון שונה. אכן ניתן לראות שההבדל היחיד בין החזרות השונות קיים בממד של מספר החלבונים, וגם הוא נותר יחסית קבוע בין החזרות השונות (סטייה של בין 0 ל-1 בכל פעם).

גרפים שנוצרו ב- 5a:

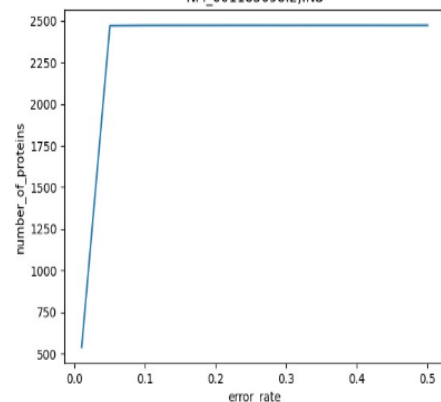
NM_001185098.2;INS



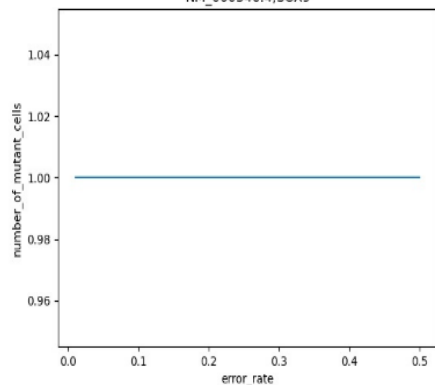
NM_001185098.2;INS



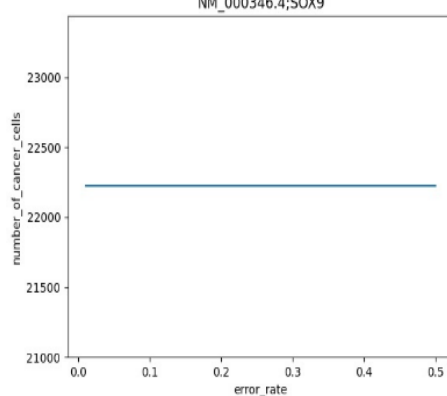
NM_001185098.2;INS



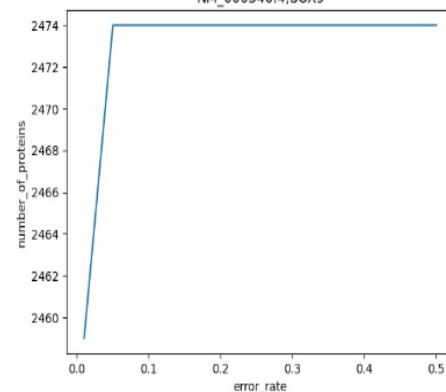
NM_000346.4;SOX9



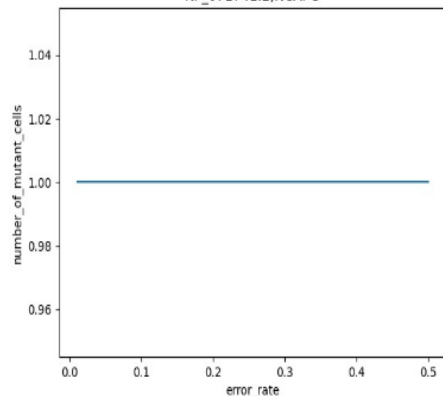
NM_000346.4;SOX9



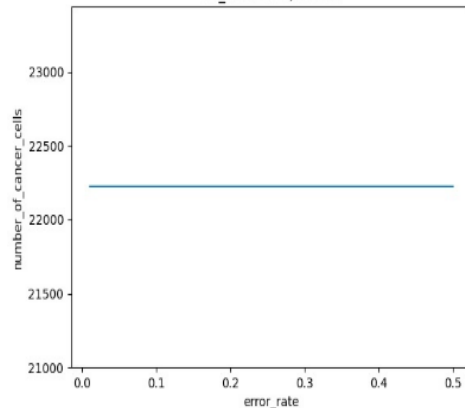
NM_000346.4;SOX9



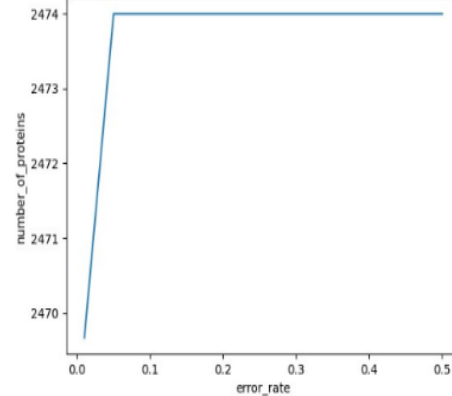
NP_071741.2;NCAPG



NP_071741.2;NCAPG



NP_071741.2;NCAPG



2. בחרנו לייצג את המידע בסעיף 5a באמצעות גרפים לינאריים. הסיבה לכך היא שבכל גרף קיימים לנו שני משתנים רציפים- משתנה בלתי תלוי (ציר x) ומשתנה תלוי (ציר y). גרף לינארי נותן לנו ייצוג חזותי מוצלח עבור נתונים מסוג זה.

3. הייצוג הוויזואלי בו בחרנו מתאים רק כאשר קיימים לנו שני משתנים כמתואר בסעיף הקודם (רציפים וש אחד תלוי בשני). במקרה בו נרצה להציג נתונים של יותר משני מדדים נשתמש בייצוגים גרפיים אחרים. לכן, אם נרצה לייצג בצורה גרפית את כל הנתונים וכל החזרות עבור כל הרצפים בגרף יחיד- גרף לינארי לא יהיה יעיל ונחפש סוג גרף מתאים יותר.

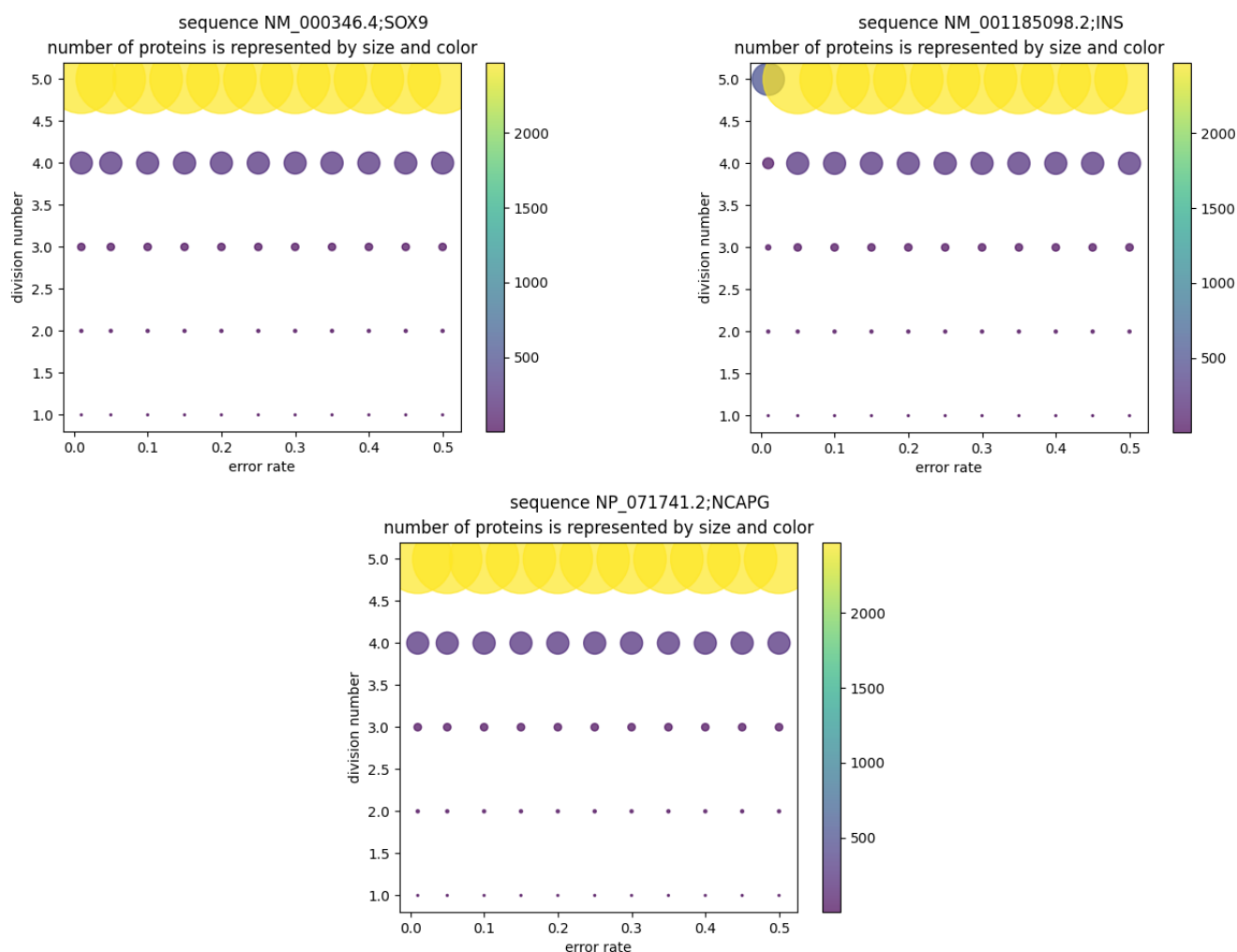
4. ניתן לראות הן בטבלת התוצאות והן בגרפים לעיל שקצב המוטציות לא משפיע על מספר החלבונים, מלבד קצב המוטציות הקטן ביותר. זה קורה משום שנשאלנו על מספר החלבונים השונים, ולא נשאלנו על עד כמה הם שונים זה מזה. גם בתא בעל קצב מוטציות נמוך וגם בתא בעל קצב מוטציות גבוה נוצר חלבון אחד בכל פעם, אבל מספר המוטציות בתוך אותו חלבון משתנה משמעותית. כלומר, מבחינה ביולוגית, קצב המוטציות אמנם פחות ישפיע על עצם היווצרות החלבון, אבל הוא ישפיע על האם החלבון יתפקד בסופו של דבר. (יש לציין גם שהמוטציות שאנו יצרנו בתרגיל מזיקות פחות ממוטציות רגילות: אם יש פגיעה בקודון עצירה או מוטציית שינוי במסגרת קריאה- בתא רגיל זה יגרום לאי היווצרות חלבון בעוד בתרגיל הנוכחי החלבון יתורגם עד סוף הרצף).

קצב המוטציות הקטן ביותר משפיע על מספר החלבונים בצורה מינורית בשני הרצפים האחרונים (הבדל של 15 חלבונים באחד ושל 4 חלבונים בשני), ובצורה משמעותית יותר ברצף INS (הבדל של קרוב ל-2000). ברצף הראשון ניתן לייחס זאת לכך שעדיין יש מספר נמוך משמעותית של תאים מאשר בשאר קצבי הטעות. בשני הרצפים האחרים זה כנראה משום שהמוטציות נפלו באזור הלא מקודד.

5. ניתן לראות כי רק עבור הרצף INS קצב המוטציות משפיע על מספר התאים הסרטניים והמוטנטיים (שאינם סרטניים). הסבר לכך הוא שקצב המוטציות הוא גורם משמעותי משפיע רק כאשר רצף הדנ"א קצר וקצב המוטציות קטן- 0.01. הרצף הראשון קצר יחסית, כך שבחלוקה הראשונה עם קצב מוטציות של 0.01 התאים הנוצרים ממנו הם עדיין מוטנטיים, אך לאחר מכן גם ממנו נוצרים תאים סרטניים בלבד. בשני החלבונים האחרים הרצף ארוך יותר כך שגם קצב מוטציות קטן הופך את התא לסרטני. כדי שתא מוטנט יהפוך לתא סרטני אנו צריכים מספר קבוע של מוטציות (בתרגיל שלנו- 10) ולכן החל ממספר חלוקה מסוים, מספר התאים המוטנטיים יישאר קבוע בעוד מספר התאים הסרטניים ילך ויגדל. מסיבה זו ייתכן שניתן לוותר על המידע של מספר התאים המוטנטיים ולהישאר עם המידע של מספר התאים הסרטניים בלבד.

6. נראה שבגדול מדד מספר החלבונים נראה דומה עבור כל הרצפים, מלבד הנתון של מספר החלבונים עבור קצב הטעות הראשון- 0.01. במדד זה אמנם מספר החלבונים של כל אחד מהרצפים נמוך יותר מבשאר קצבי החלוקה, אך בעוד ברצף הראשון הפער הוא משמעותי, בשני הרצפים הנותרים ההפרש מינורי. כפי שפירטנו בסעיף 4, כנראה שברצף הראשון מספר החלבונים הקטן נובע ממספר התאים הקטן בצורה משמעותית שקיים ברקמה (חלק זה מוסבר בסעיף 5). ברצפים השני והשלישי זה נובע כנראה ממוטציות שנפלו באזור הלא מקודד. מלבד חלק זה, כל הרצפים התנהגו בצורה דומה והגיעו למספר זהה של חלבונים בסופו של דבר. כפי שהסברנו בסעיף 4, מספר החלבונים הזהה נובע מכך שנשאלנו על מספר חלבונים שונים, כך ששני חלבונים ששונים ביניהם בחומצה אמינית אחת נספרים בצורה דומה לשני חלבונים ששונים ביניהם ב-30 חומצות אמינו, וכך לקצב המוטציות כמעט ואין השפעה על מספר החלבונים הסופי שנוצר.

גרפים שנוצרו ב- 5b:



7. בסעיף זה בחרנו לייצג את המידע באמצעות scatter plot (תרשים פיזור).

בחרנו בגרף זה משום שהיו לנו יותר משני משתנים אותם רצינו לייצג, וגרף זה התאים לצרכים שלנו. בחרנו את ציר X להיות קצב הטעות ואת ציר Y להיות מספר החלוקות. את מספר החלבונים ייצגנו הן באמצעות צבע והן באמצעות גודל, על מנת לתת המחשה ויזואלית נוחה ככל האפשר. תרשים פיזור מאפשר לנו להשתמש בשני הצירים כגורמים בלתי תלויים, ולתת משתנה נוסף בתור משתנה תלוי, כפי שרצינו להראות כאן.

בנוסף, בחרנו לייצג את הרצפים השונים בגרפים נפרדים משום שהתוצאות שלהם היו כל כך דומות. כשניסינו לייצג אותן על גרף בודד ניתן היה לחשוב שמצגות בו תוצאות של רצף אחד בלבד משום שכל רצף "דרס" את הרצף שלפניו. דווקא בייצוג נפרד ניתן לראות בבירור עד כמה כל רצף דומה בהתנהגותו ליתר הרצפים.

8. ניתן לראות שבכל הגרפים ציר X (קצב הטעות) נשאר קבוע, בעוד בציר Y (מספר החלוקות) קיימת עלייה שבאה לידי ביטוי הן בצבע והן בגודל הנקודות בגרף. יוצא הדופן היחיד הוא הרצף הראשון, בו קיים הבדל בין קצב הטעות הראשון לבין היתר. משמע, לקצב הטעות כמעט ואין השפעה על מספר החלבונים שהתקבל, לעומת מספר החלוקות לו יש השפעה משמעותית. את העובדה שלקצב הטעות כמעט ואין השפעה הסברנו באריכות בסעיפים הקודמים. תוצאות אלה שהתקבלו ביחד עם העובדה שלמספר החלוקות יש השפעה משמעותית על מספר החלבונים היא טריוויאלית: בכל חלוקה נוצרים לנו פי 10 תאים סרטניים חדשים, וכל תא שמתחלק יוצר לנו חלבון חדש. אמנם בתרגיל זה כל תא שהתחלק יצר לנו 10 תאים חדשים, אך הם זהים זה לזה ורק בדור הבא של החלוקה כל תא כזה ייצר תאים חדשים שאינם זהים אחד לשני. נתון זה משתקף במספר החלבונים: מספר החלבונים השונים שנקבל יהיה כעשירית ממספר התאים הסרטניים הקיימים.