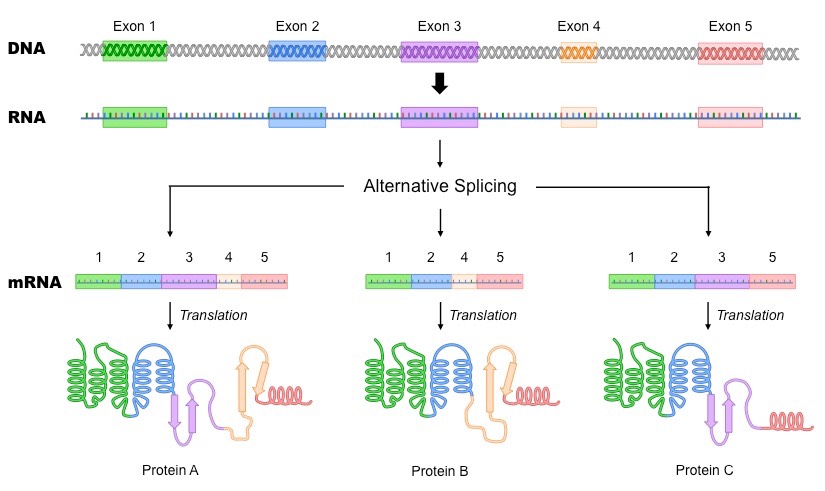
**מידת ההשפעה של תהליך השחבור החליפי**

**רקע:**

תהליך השחבור החליפי

גנים הינם יחידות מידע המקודדות באמצעות רצף של חומצות בסיס (נוקלאוטידים) בדנא המאוחסן בגרעין של כל תא. כל גן מכיל בתוכו את המידע הנחוץ ליצירת חלבון. בתהליך יצירת החלבון, מקטע הגן משועתק לרצף רנא המעביר את המידע מהדנא בגרעין התא החוצה לצורך יצירת החלבון. רצף הרנא הראשוני מורכב מחלקים שאינם מקודדים לחלבון (אינטרונים) ומחלקים המקודדים לחלבון (אקסונים). האינטרונים מוסרים מרצף הרנא בתהליך הקרוי שחבור. בעבר סברו שתהליך זה הינו תהליך מדויק ואחיד, אך היום ידוע שמתהליך זה יכולות להיווצר מולקולות רנא סופיות שונות עבור אותו גן המכונות טרנסקריפטים ועשויות להכיל אקסונים שונים עבור אותו גן , תהליך זה קרוי **"שחבור חליפי".**



**מטרות הפרוייקט:**

* חיפוש טרנסקריפטים דומים/שונים-

למרות שטרנסקריפטים שונים בעלי הרכב אקסונים שונה יתכן שבפועל אין השפעה על החלבון או ההיפך, שקיים שוני קטן אך הוא משמעותי בהשפעתו על החלבון. ננסה לכמת באמצעות שלושה מדדים את השוני וכמה הוא משמעותי.

* אילו גנים מושפעים יותר מהשחבור החליפי-

ישנם גנים שעוברים שינויים יותר משמעותיים בעקבות תהליך השחבור החליפי. אותם גנים הם אלה בהם קיים שוני רב בין טרנסקריפטים השייכים לגן. השוני יכול להיות בכמות האקסונים, המיקומים שלהם ובדומיינים.

* באילו מינים הגן מושפע יותר מהשחבור החליפי-

ישנם גנים שמקודדים למגוון רחב יותר של חלבונים במינים מסויימים. עבור כל גן אפשר לראות אם יש שמירה על מבנה דומה או שנוצרים טרנסקריפטים שונים זה מזה.

* משמעויות נוספות (אבולוציוניות, אופי הגן)-

גנים שמקבלים אחוזי השפעה דומים במספר מינים יעידו שמדובר באופי מסויים של גן. באופן דומה, במקרים מסויימים יתכן שעבור גן יהיה מין מסויים שיקבל תוצאה חריגה משאר המינים מה שיעיד על תפקיד שונה במין זה בניגוד לשאר המינים.

חשוב להבהיר שהנתונים המוצגים הם על פי נתונים שנחקרו בקהילה המדעית והם נוטים לשקף בצורה לא מדוייקת את המציאות כיוון שלא כל הגנים והמינים נחקרים ברמה אחידה.

**כיצד הושגה המטרה:**

בניית אתר אינטרנטי שמאפשר 3 פונקציונליות עיקריות :

1. השוואת בין שני טרנסקריפטים באותו הגן ובאותו המין על פי דמיון בין האקסונים
2. השוואת כל הטרנסקריפטים של אותו הגן באותו המין
3. השוואת כל הטרנסקריפטים של אותו הגן במינים שונים

**נתונים:**

נתונים מתבססים על מערכת DoChaP שבנינו במסגרת פרוייקט גמר (<https://dochap.bgu.ac.il/>)

המערכת מבוססת על ממספר בסיסי נתונים שמאוגדים לידי בסיס נתונים יחיד. מקורות הנתונים הם:

* מסד הנתונים RefSeq של [NCBI](https://www.ncbi.nlm.nih.gov/), מרבית הנתונים הינם בפורמט של flatfiles
* טבלאות רלוונטיות ממסד הנתונים [UCSC](https://genome.ucsc.edu/) המכיל מידע אינטגרטיבי ממקורות שונים

נשמרים נתונים אודות גנים, טרנסקריפטים, חלבונים, אקסונים, דומיינים ואיזורי חיתוך בתוך דומיינים.

**שלושה מדדי דימיון שבנינו:**

1. מיקום האקסון

אקסונים שהמיקום שלהם חופף יכילו קטעים חופפים מה שעשוי להשפיע על רמת הדימיון – גם אם אינם זהים. הערה: אם אקסון מתחיל ומסתיים כמו אקסון אחר אז הם זהים לחלוטין ואין צורך במדדים נוספים לבחינת דימיון.

**נוסחא לחישוב:**

1. אורך בCDS

בתחילת או בסיום הטרנסקריפט, יש UTR שלא ישפיע על קידוד החלבון ולכן יש לשקלל רק את החלק שכן ישפיע.

**נוסחא לחישוב:**

1. דומיינים משותפים

אקסונים המקודדים לאותם דומיינים הם בעלי משמעות דומה אז ייחשבו דומים.

**נוסחא לחישוב:**

\*המערכת מאפשרת למשתמש לעדכן משקולות ולתת משקל גבוה למדד שחשוב בעייניו!

**תוצאות:**

תצוגת התוצאות מוצגת בצורת מטריצות שמראה את אחוזי הדימיון. לצד האחוז מוצגת תצוגה גרפית באמצעות עיגולים שמראים את היחסים בגודל.

**תוצאת חיפוש מספר 1 - השוואת בין שני טרנסקריפטים באותו הגן ובאותו המין על פי דמיון בין האקסונים**

בחיפוש זה נציג מטריצה שמציגה את הדימיון בין כל זוג אקסונים של טרנסקריפט אחד מול הטרנסקריפט השני. לרוב באיזור האלכסון אמורה להיות את ההתאמה הרבה כיוון שהאקסונים מסודרים לפי הסדר הגנומי. רמת הדימיון מצויינת באחוזים על פי משקול המדדים שנבחר.

כדי להסיק את רמת הדימיון בין שני הטרנסקריפטים ניקח את הערכים הכי גבוהים בכל עמודה ובכל שורה (כלומר לכל אקסון נמצא את האקסון עם ההתאמה הגבוהה ביותר) ונעשה ממוצע.

**תוצאת חיפוש מספר 2 - השוואת כל הטרנסקיפטים של אותו הגן באותו המין**

בחיפוש זה נציג מטריצה שמציגה את הדימיון בין כל זוג טרנסקריפטים של הגן. באלכסון ניתן לראות שכל טרנסקריפט מקבל דימיון מלא עם עצמו. רמת הדימיון המצויינת היא באחוזים על פי חישוב רמת דימיון הטרנסקריפט שהצגנו בתוצאות חיפוש מספר 1.

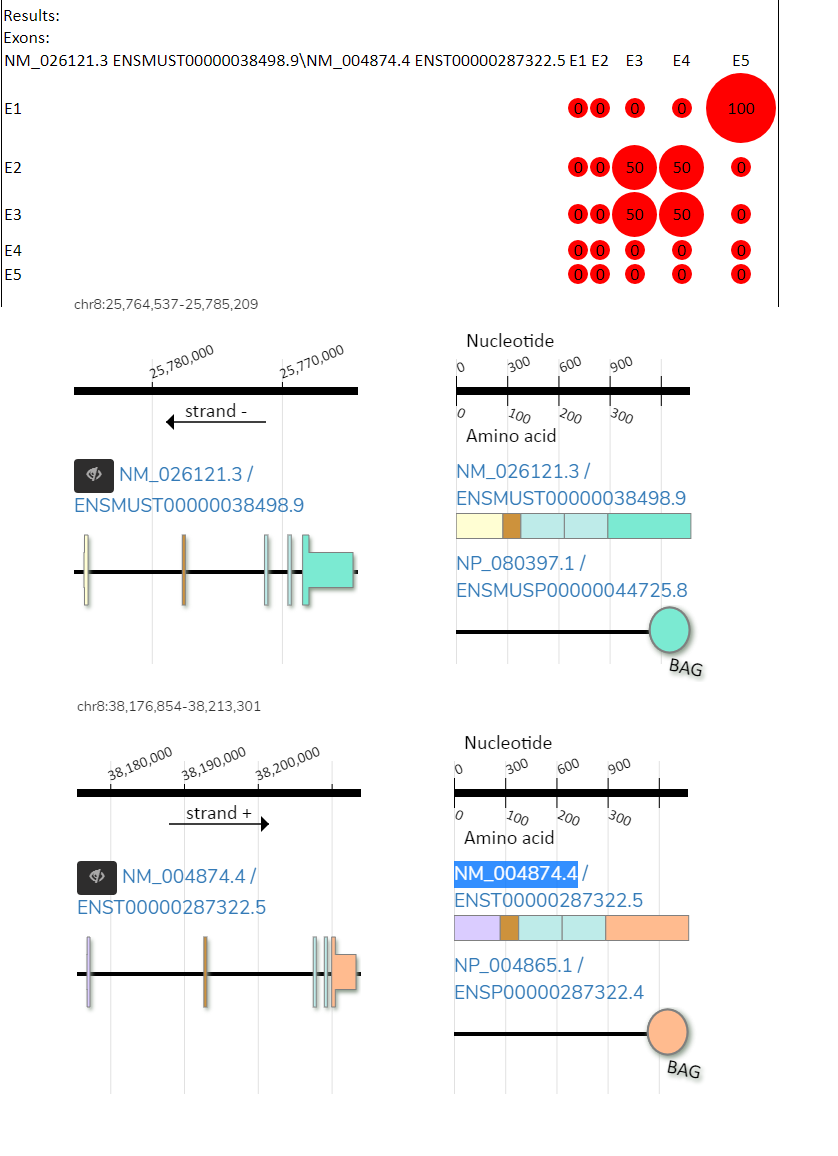
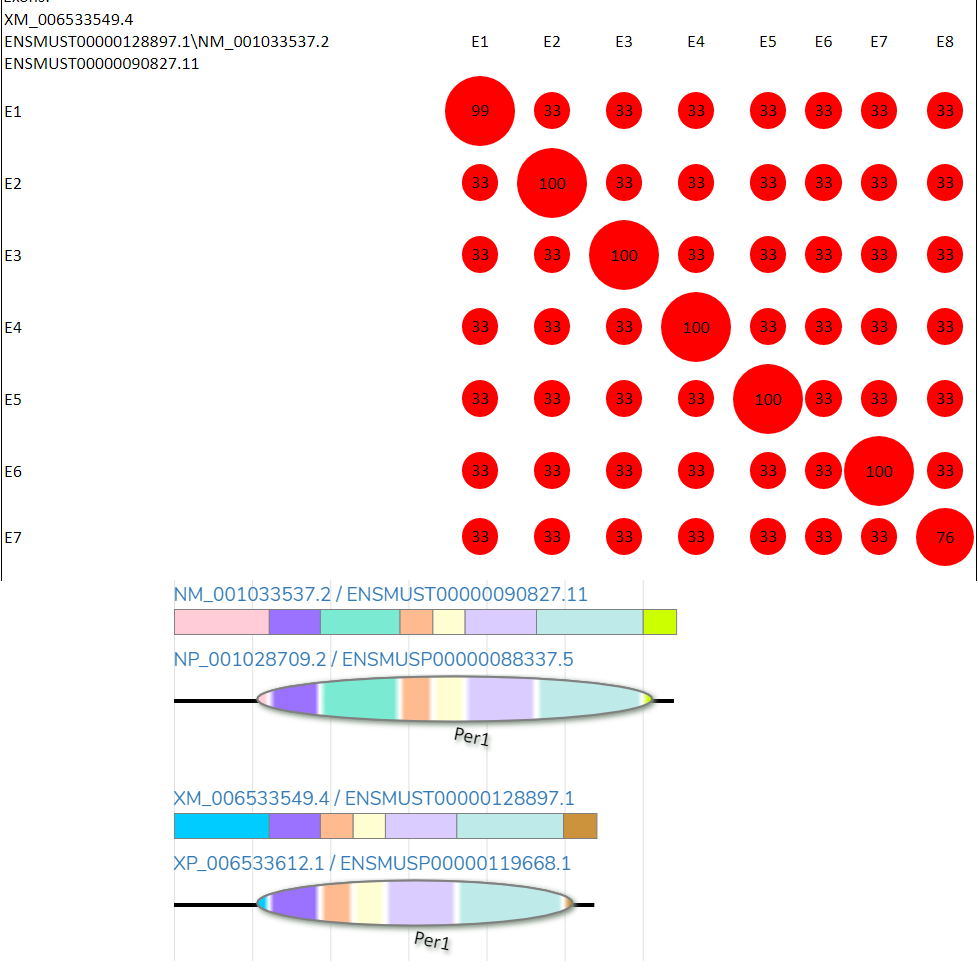
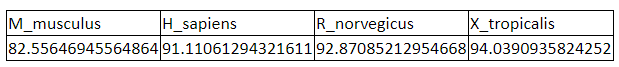
כדי להסיק את רמת ההשפעה של השחבור החליפי בגן המסויים נעשה ממוצע בין ערכי הדימיון המוצגים (נתעלם מזוגות שבהם הטרנסקריפט מושווה מול עצמו).

**תוצאת חיפוש מספר 3 - השוואת כל הטרנסקריפטים של אותו הגן במינים שונים**

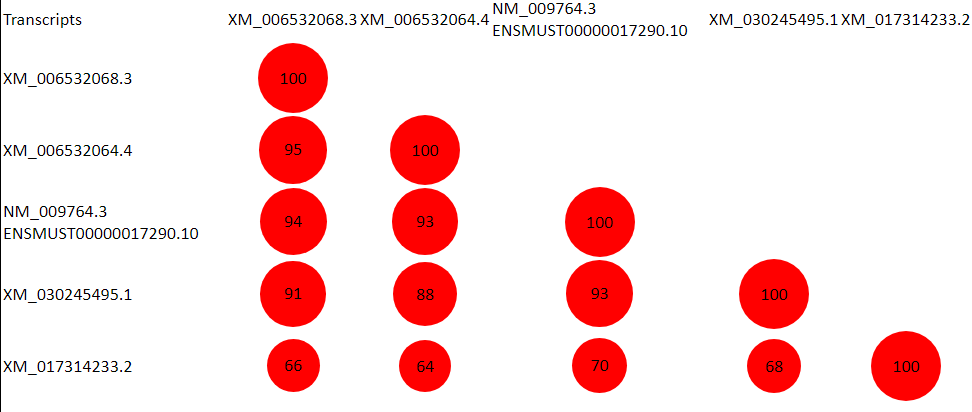
בחיפוש זה נציג רמת ההשפעה של השחבור החלופי בגן מסויים כפי שנמצאה בתוצאות חיפוש מספר 2 עבור כל המינים שבהם נמצא גן באותו השם.

**מסקנות:**

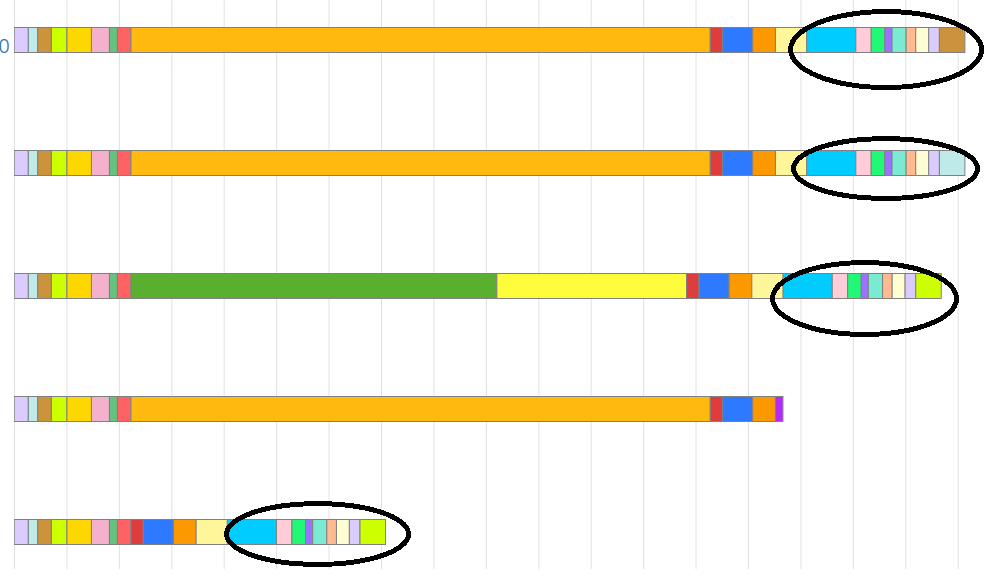
מצאנו כמה דברים מעניינים במהלך הבנייה ובחינת התוצאות:

* מדד המיקום הגנומי הוא המשמעותי ביותר. למרות שזה גם מה שחשבנו מההתחלה מסתבר שגם אם מאפסים את שאר המדדים הוא עדיין מצליח למצוא את האקסונים המתאימים. כמובן שהחסרונות העיקריים שלו מטופלים ע"י המדדים האחרים
* השוואה בין זוג טרנסקריפטים הנלקחים מגנים אורתולוגים – אם נאפס את מדד המיקום הגנומי, נוכל למצוא התאמה יחסית גבוהה למרות שמדובר במינים שונים. מוצגת דוגמא לגן BAG4 בעכבר ובבן אדם. הערה- האלכסון בכיוון ההפוך כי הגנים לא על אותו סטרנד בשני המינים. 
* מדד הדומיינים לא מספיק משוכלל – לעיתים לחלבון יש תפקיד פונקציונלי אחד שיסומן בדומיין אחד שמתפרס כמעט לאורך כל החלבון ומקודד ע"י כל האקסון. במקרים כאלו, על פי המדד שלנו, תהיה התאמה מלאה לפי מדד הדומיינים מה שייתן ניקוד 33% לפחות לכל זוג אקסונים – כפי שניתן לראות בתמונה המוצגת של טרנסקריפטים מתוך PGAP3 בבן אדם. 
* השוואה בין מינים- שלושת החיפושים משלימים זה את זה וביחד הם נורא שימושיים. הסיבה לכך היא שהם כל חיפוש מסביר את תוצאות קודמו. אם נסתכל על brca1 במינים שונים יתקבלו התוצאות הבאות. 

בקלות ניתן לבדוק מה גרם לעכבר לקבל ציון נמוך משמעותית מהשאר ע"י חיפוש הגן brca1 במין עכבר.



מהתוצאות עולה שטרנסקריפט XM\_017314233.2 אינו מספיק דומה כמו שאר הטרנסקריפטים. לאחר הסתכלות אפשר לראות שמדובר בטרנסקריפט שחסרים לו 9 אקסונים בסיום. אלו גם אקסונים שמקודדים לשני דומיינים שחסרים לו.



**הדרכת הרצה:**

זהו הקישור לגיט:

<https://github.com/haiatn/QInBioHW>

לאחר הורדת הקבצים יש ללחוץ על קובץ הhtml מתוך התיקייה באופן לוקלי שיפתח את האתר בדפדפן. כעת המשתמש יכול להזין קלטים ולצפות בתוצאות. זמן הריצה הצפוי הוא פחות מדקה.

הערה: מאחר שהאתר ניגש לשרת של DoChaP בשביל מידע על המחשב להיות מחובר לאינטרנט.