

מטלה 2 – מעבדה בכלים בביואינפורמטיקה, ת.ז. - 209018209

1. Global alignment –

- ציון עימוד מיטבי = -1
- קיימים שלושה סידורים אפשריים בציון זה
- דוגמה לסידור אופטימלי AT -- G
ATGGG

	0	A	T	G	G	G
0	0	-2	-4	-6	-8	-10
A	-2	1	-1	-3	-5	-7
T	-4	-1	2	0	-2	-4
G	-6	-3	0	3	1	-1

Local alignment –

- ציון מקסימלי = 3
- קיים סידור אחד אופטימלי בציון ATGGG

	0	A	T	G	G	G
0	0	0	0	0	0	0
A	0	1	0	0	0	0
T	0	0	2	0	0	0
G	0	0	0	3	1	1

c. הגן FOXP1 יושב על כרומוזום 3 ומקודד לחלבון Forkhead box protein P1.

הגן שייך למשפחת חלבוני FOX – פקטורי שעתוק בעלי תפקיד מרכזי ברגולציה על גנים שקשורים בגדילה וחלוקה של תאים וחיוניים להתפתחות העובר¹. בנוסף הגן מעורב בוויסות תפקודי הריאות ועשוי להגן מפני הפרעות שונות בריאות (כמו ברונכיטיס כרוני)². חלבוני FOX מאופיינים ב-"forkhead motif" רצף של כ-80-100 חומצות אמינו. מוטציה בגן עלולה לגרום לתסמונת הנדירה FOXP1³. לרוב המוטציה תהיה ספונטנית בשלבי התפתחות מוקדמים ולא בתורשה מההורים.

e. ניתן לראות שברצפים של Human + Bos לעומת הרצף של Mus חסרות כ-30 חומצות אמינו מסוג Q – Glutamine. אין דרך לדעת האם מדובר במוטציה מסוג insertion/deletion – ניתן להניח שאם מדובר במחיקה המוטציה התרחשה בענף e2 ואם זו הייתה הוספה היא התרחשה בענף e1.

```
foxp1_human VS foxp1_mus - score = 1858, significance level = TRUE  
foxp1_human vs foxp1_bos- score = 1943, significance level = TRUE  
foxp1_mus vs foxp1_bos - score = 1808, significance level = TRUE
```

¹ <https://www.uniprot.org/uniprotkb/Q9H334/entry>, <https://en.wikipedia.org/wiki/FOXP1>

² [The FoxP1 gene regulates lung function, production of matrix metalloproteinases and inflammatory mediators, and viability of lung epithelia - PubMed \(nih.gov\)](#)

³ <https://www.foxp1.org/post/manage-your-blog-from-your-live-site>, [FOXP1 syndrome: a review of the literature and practice parameters for medical assessment and monitoring - PubMed \(nih.gov\)](#)

1. ניתן להסיק שהרצף הוא חלק מהגנום של SARS-COV-2 (100% התאמה).

2. נמצאה התאמה ב-98.31% לעטלף עם $E\text{-value} = 4e-18$ ו- $\text{Score} = 103$.

ניתן לראות ש- $E\text{-value} < 10^{-4}$ לכן ככה"נ מדובר בגנים הומולוגיים והסבירות למצוא התאמה טובה כמו זה באופן רנדומלי נמוכה מאוד.

3. עבור בן אדם , Sequence ID: OQ773592.1 , אורך הגנום = 29818 bp.

עבור עטלף , Sequence ID: MZ937003.2 , אורך הגנום = 29844 bp.

4. נניח שרצף 1 מתחיל "במקום הנכון" ונכיל את רצף 2 לפיו (כלומר נסובב רק את 2).

נגריל רנדומלית אינדקס התחלה (נבצע k סיבובי הגרלה). עבור כל אינדקס רנדומלי שהגרלנו נבצע עימוד רצפים גלובלי לרצף באורך W שמתחיל באינדקס זה למול W הנוקלאוטידים הראשונים ברצף 1.

נאתחל משתנה של max score ועבור כל עימוד רצפים שביצענו נבדוק האם ה-alignment score החדש גדול יותר מהמקסימום, אם כן נעדכן את המקסימום ואת אינדקס ההתחלה של הרצף עבורו ניקוד זה נקבע.

W = המינימום בין 2000 לבין 10% מאורך הגנום (גנום ממוצע של וירוס נע בין 2KB bp-1M⁴bp, לכן בחרתי חלון שיהיה גדול מספיק אך לא גדול מידי עבור רצפים ארוכים מאוד).

K = נגדיר את מס' ההגרלות להיות כרבע מאורך הרצף (ניסיתי לבחור מספר שייתן תוצאה מספיק טובה בזמן ריצה סביר, אלגוריתם היוריסטי ולא מדויק).

6. אפשר לראות שעבור שתי שיטות הניקוד השונות התוצאה עבור local alignment טובה ביחס לזו ב-global.

ניתן להסיק מהתוצאות שככל הנראה ישנם תתי-רצפים בעלי התאמה גבוהה בין שני הגנומים (אזורים שמורים).

	1,-1,-1,-1	1,-1,-3,-1
Global	25,061	24,750
local	26,626	26,485

⁴ <https://www.sciencedirect.com/topics/medicine-and-dentistry/genome-size>