# שחזור עץ פילוגנטי על תאי נאופלזמות מיאלופרוליפרטיביות

# שמות חברי הקבוצה:

* גל סזנה 318510633.
* נועה מרגוליס 208635334.
* עדי יפרואימסקי 326045515.
* יואל מרקו 325245280.
* איתן סמסון 214935165.

# רקע:

* המאמר: "Life histories of myeloproliferative neoplasms inferred from phylogenies".
* פרטים: המאמר פורסם בכתב העת Nature, כרך 602, בתאריך 3 בפברואר 2022. המחברים כוללים את ניקולס וויליאמס ואחרים ממוסדות כמו מכון וולקאם סנגר והמחלקה להמטולוגיה באוניברסיטת קיימברידג'.
* תקציר רקע: מחלות מיילופרוליפרטיביות (MPNs) הן סוג של סרטן דם כרוני הנגרם ממוטציות גנטיות סומטיות בתאי גזע המטופויאטיים (HSCs). . המאמר עוסק בזיהוי מוטציות סומטיות ותיעוד ההיסטוריה הגנומית של מושבות מתאי גזע של חולי MPN . זוהו 580,133 מוטציות סומטיות לצורך שחזור אילנות יוחסין המטופויאטיים ובחינת ההיסטוריות והתרחבותית של השיבוטים.
* המחקר מצביע על כך שרכישת מוטציות מובילות (driver mutation) מוקדמות והתפתחותן לכל אורך החיים הן הבסיס למחלות מיילופרוליפרטיביות במבוגרים, דבר שמעלה הזדמנויות להתערבות מוקדמת ומציע מודל חדש להתפתחות סרטן.

# שאלת מחקר ומטרות:

* **שאלת מחקר:**
  + כיצד מוטציות גנטיות מוקדמות ותהליכי התרחבות שיבוטיים משפיעים על התפתחותן של מחלות מיילופרוליפרטיביות לאורך החיים?
  + מתי בערך בחיים אדם עלול לקבל מוטציית דרייבר?
* **מטרות:**
  + נרצה לנצל את הידע הנלמד בקורס בנושא בניית עצים פילוגנטיים על מנת לזהות פיצולים (מוטציות) בהתפחות התאים שמובילות לסרטן להבין את התזמון של רכישת מוטציות מובילות (driver mutations) ואת התרחבותן.
  + שיחזור תוצאות המאמר עליו אנחנו מתבססים.
  + הכרת עבודה עם דאטה ביולוגי.

# דאטה:

* [קישור למקור הדאטה סט.](https://data.mendeley.com/datasets/hrmxybrd2n/1)
* המידע נלקח מתוך קובץ נתונים גנומיים בפורמט VCF (Variant Call Format), שמטרתו לתאר מוטציות גנטיות באזורים ספציפיים בגנום. במקרה זה, מדובר במוטציות סומטיות שנמצאו בתאי דם שנדגמו מ-12 פרטים. אלו מוטציות שנרכשו במהלך החיים ולא תורשתיות.
* ראו הרחבה על הדאטה ונספח ניתוח הפורמט.

# היפותזה:

מוטציות מניעות שנרכשות בשלבים מוקדמים בחיים מובילות לדינמיקות התרחבות משתנות לפני האבחנה הקלינית.

הבנה זו תוכל לשפר את הגילוי המוקדם ולתרום להתאמה אישית של טיפולים.

# שיטות חישוביות:

* שימוש באלגוריתמים של בניית עצים פילוגנטיים (MPBoot).
* בשיעור למדנו שתי שיטות מרכזיות למימוש האלגוריתם הנבדלות זו מזו בקריטריון לבחירת הזוג הכי קרוב של הקודקודים – האחת UPGMA השנייה Neighbor joining.
* מודלים בייזיאניים לחישוב קצבי התפשטות קלונית.
* ניתוח סטטיסטי להשוואת מוטציות מניעות בין קלונים.

# שלבי ביצוע:

1. ניתוח ועיבוד ראשוני של הנתונים ובדיקת איכות.
2. בניית עץ פילוגנטי עבור כל הדאטה המצורף במאמר:

* השוואת גנים של חולים ברצפים הייעודיים לגנים בריאים ותיעוד היסטוריית המוטציות.
* נבדוק איזו שיטה של בניית עץ פילו גנטי תניב תוצאות טובות יותר. NJ הוא ללא דרישה לאולטרמטריות.

1. ניתוח התוצאות מהעץ ומסקנות.

# בביבליוגרפיה:

Williams, N. et al. (2022). Life histories of myeloproliferative neoplasms inferred from phylogenies. \*Nature\*, 602, 162–168.

# גיוון הקבוצה:

* 3 סטודנטים לביולוגיה חישובית, סטודנט למדעי המחשב חד חוגי, סטודנטית למדעי המחשב עם חטיבה בביולוגיה.
* 3 עתודאים, 1 קצין במיל' (תותחנים), 1 קצינה בקבע (חיל הים) – שילוב הזרועות השונים.
* 2 דתיים, 3 חילוניים.
* ג'ינג'י אחד.
* גם בנים וגם בנות.
* 4 משקפופרים 1 ללא משקפיים.
* גיוון אתני – ישראל הראשונה עם ישראל השנייה.

הקבוצה שלנו מגוונת ברקע האקדמי, המקצועי והאישי של חבריה.

אנו כוללים סטודנטים ממסלולים שונים כמו ביולוגיה חישובית ומדעי המחשב, לצד עתודאים, קצין במילואים וקצינה בקבע מחילות שונים. בקבוצה יש ייצוג של דתיים וחילוניים, בנים ובנות, וכן גיוון אתני המשקף את החברה הישראלית. אלמנטים אישיים כמו ג'ינג'י אחד ומשקפופרים מוסיפים צבע וגיוון נוסף.

המגוון שלנו מאפשר שיתוף פעולה ייחודי וחשיבה רחבה ויצירתית.

**הרחבה על הדאטה**

**כל המוטציות שנתעסק איתם במאמר הם מוטציות סומטיות ולא מוטציות מתאי נבט, מסוג (סניפ) החלפה של בסיס בודד או מסוג (Indel) הכנסה ומחיקה של בסיס בודד או מספר בסיסים. בתהליך עושים single-colony whole-genome sequencing על תאים מהדם או ממוח עצם. כלומר, התאים שנלקחו הופרדו לתאים בודדים והם הופרדו לתרביות שונות, ושם נתנו לכל אחת להתחלק ולהתפתח לקולונה.**

**כל קולונה מייצגת לנו גנום של הבן אדם בשלב כלשהו בחיים שלו, עם מידע כלשהו של מוטציות שהוא כן סבר ומוטציות שהוא לא סבר. במחקר נבקש להבין איזה מסלול של צבירת מוטציות מוביל לסרטן ומתי בערך אמורה להיות המוטציית דרייבר.**

**יש בידנו מוטציות שמתאפיינות עם סרטן ולכן נוכל להגיד שקולונה עם מסלול מוטציות שמוביל למוטציה עם סרטן הוא מסלול מוטציות שמוביל לסרטן.**

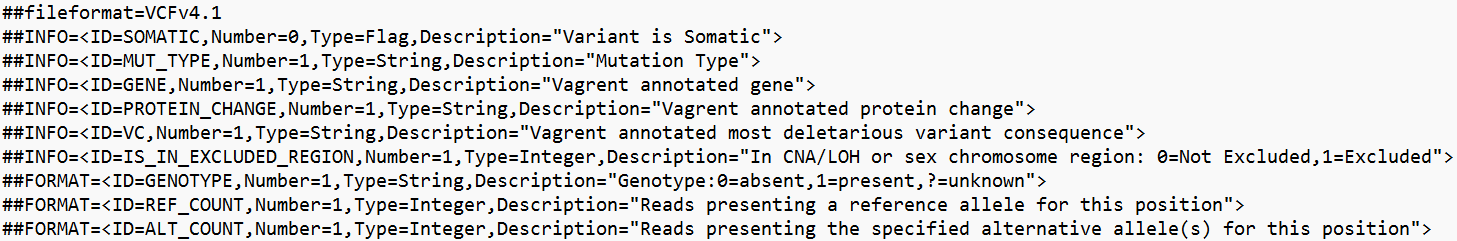
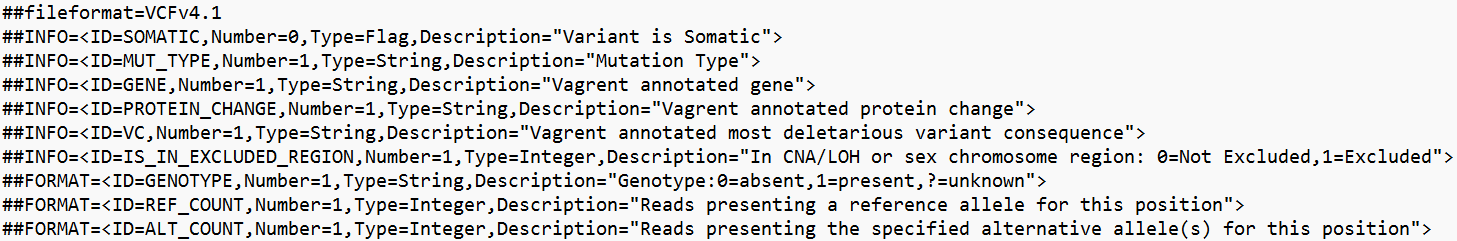
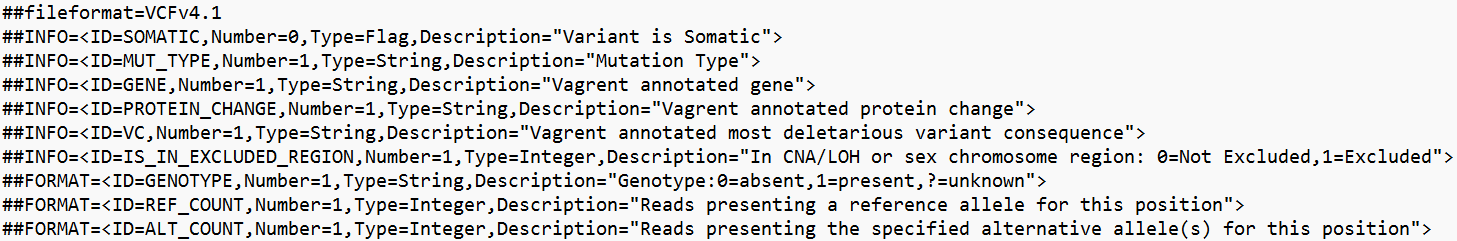
**בשביל שני הכיוונים נרצה ליצור עץ פילוגנטי שמכיל את המסלולים של מוטציות מסוימות בגוף.**

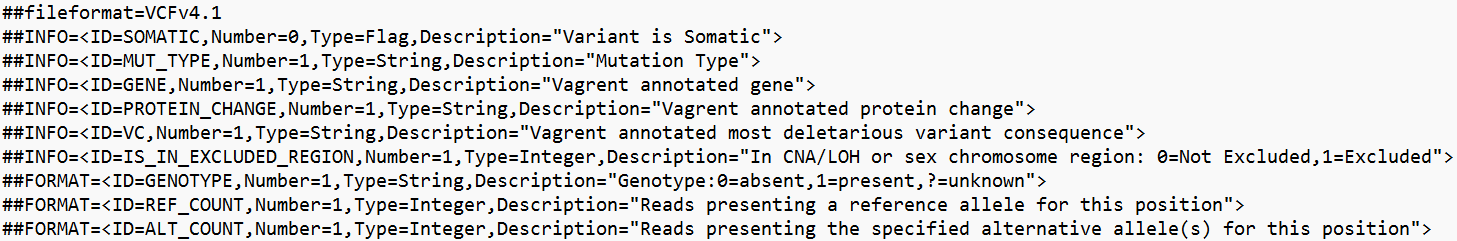
**אז מה צריך בשביל להפיק את העץ הפילוגנטי?**

**ראשית, צריך את המידע מהמחקר על איזה מוטציות כל קולונה צברה ואיזה היא לא צברה. יש לנו כבר מקומות שחשודים למוטציה שאנחנו יודעים עליהם. מכל קולונה לקחו כמה תאים ועשו להם WGS וקיבלו כמה גנומים מאותה קולונה ובדקו על הגנומים האלו במקומות החשודים במוטציה האם מופיעה אותה מוטציה או לא. ספרו את מספר הפעמים שהופיעה וכתבו "1" אם החליטו שיש, כתבו "0" אם החליטו שאין וכתבו "?" אם לא יכלו להסיק מסקנה חד משמעית. כל המידע הזה מאוגד לתוך קובץ מסוג VCF (הסבר על הבנת הקובץ בנספח א'), על הקובץ הזה עושים אנליזות בשביל להוציא את העץ הפילוגנטי.**

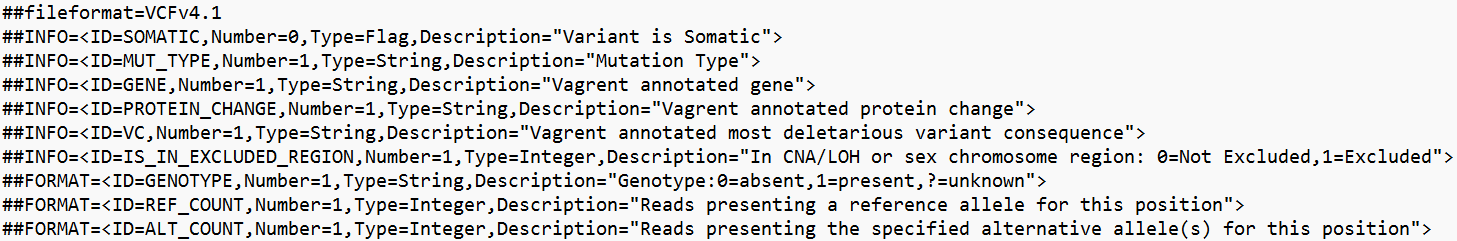
## נספח א' – ניתוח פורמט הדאטה

**הקובץ מתחיל בתיאור של המידע שמוכל בתוכו, הקטע של התיאור של הצגת הנתונים מתחיל ב##. אחריו מגיע טמפלייט של מבנה הצגת שורה. כל שורה היא בדיקה של וריאנט סומטי ספציפי. אחריהם יש את השורות של הבדיקות עצמם (הנתונים).**

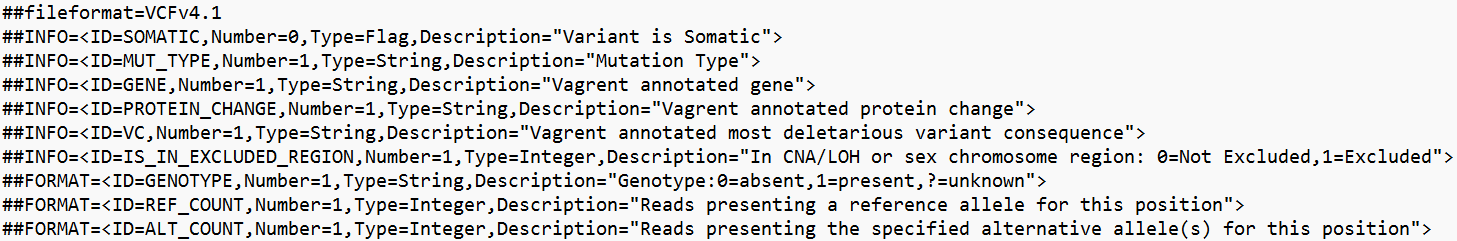
1. **הצהרת הגרסה של הקובץ  
   **
2. **פורמט המידע של הרצה של בדיקת קיום מוטציה על קולונות**
   1. **סוג הוריאנט: somatic או germline**
   2. **סוג המוטציה:   
      **
      1. **החלפת בסיס בודד – SNV/SNP**
      2. **הוספה ומחיקה – INDEL**
   3. **הגן שנבחן בהרצה זו**



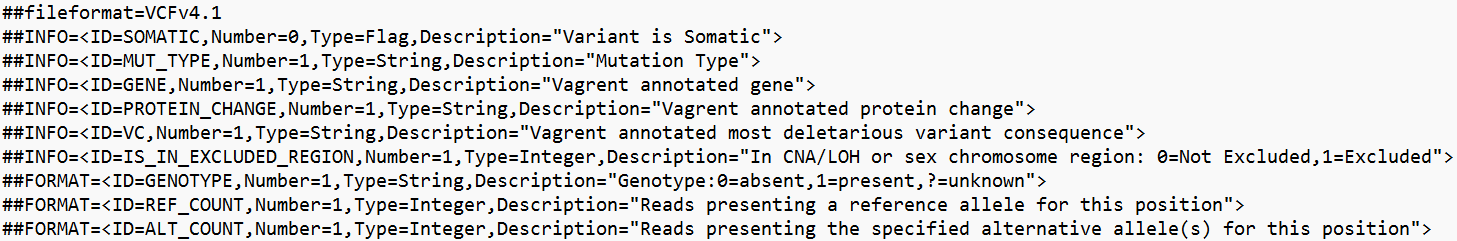
* 1. **לכל מוטציה, החלבון שבו נעשה המוטציה, השינוי שנעשה ואיפה בדיוק בחלבון זה התרחש**

****

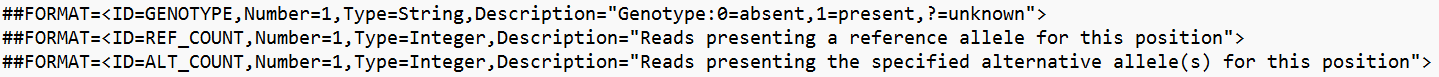
* 1. **לכל מוטציה** **ההשלכות שקרו בגללה, למשל missense, nonsense, 3 prime utr ועוד**



* 1. **גן CNA/LOH או גן בכרומוזום המין נכתוב 1 אחרת נכתוב 0 (ש-0 מסמן לכלול אותו באנליזות מסוימות ו-1 מסמן לא לכלול אותו)**

****

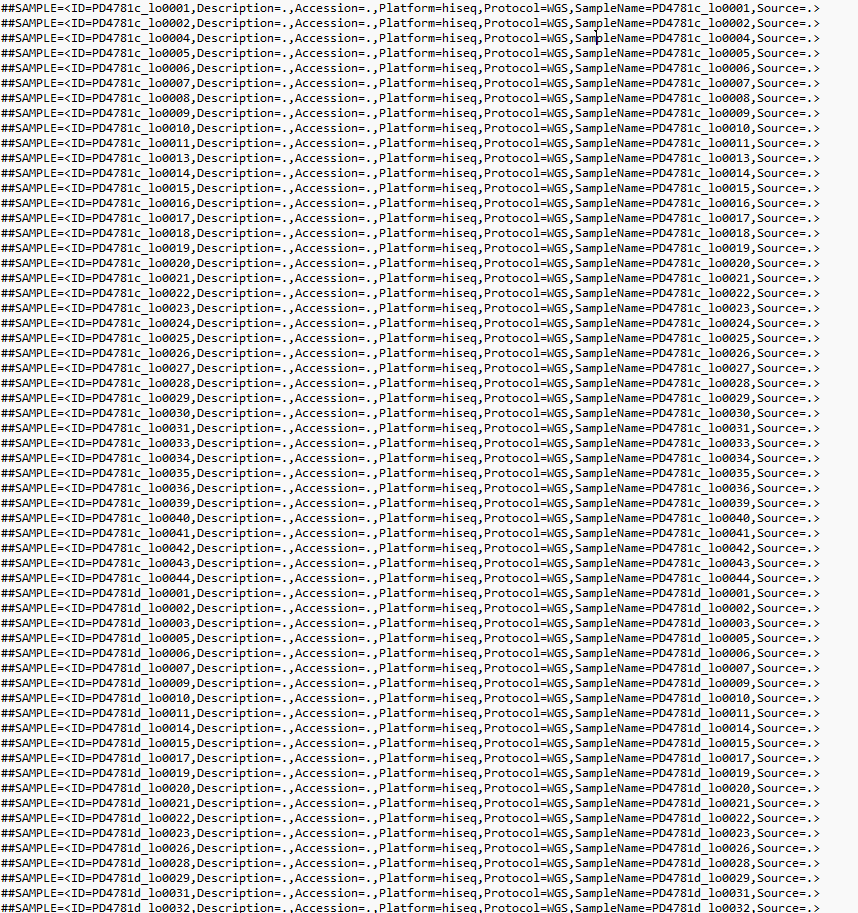
1. הפורמט (מיוצג כשלישיה) של הצגת הנתונים עבור בדיקה על קולונה. (כל שלישיה זו קולונה).
   1. האם המוטציה מופיעה (1 אם כן 0 אם לא : מספר הרידים שהייתה בהם מוטציה : מספר הרידים שלא היה בהם מוטציה

****

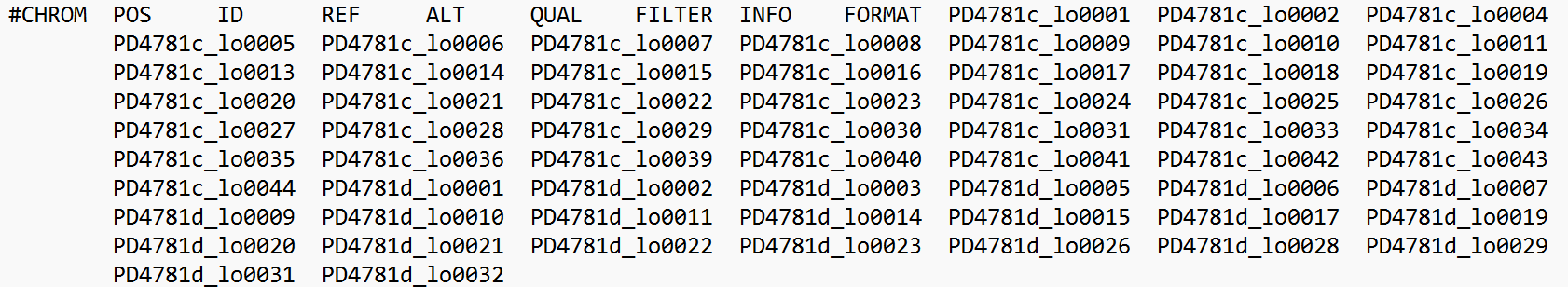
1. הכרומוזומים והאורכים שלהם:



1. המושבות:



1. הטמפלייט של הרצה על מוטציה:



* 1. מספר הכרומוזום שעליו נמצאת המוטציה
  2. המיקום של הבסיס שבו מתחילה המוטציה
  3. הזהות של הבן אדם שעליו אנחנו מריצים את זה
  4. מה שאמור להיות בלי מוטציה
  5. מה שאמור להיות אחרי מוטציה
  6. עד כמה אנחנו בטוחים בתוצאה
  7. לא בטוח מה זה הFILTER
  8. זה החלק של המידע של פורמט המידע שהסברתי בהרחבה ב2
  9. זה פורמט ההצגה של קולונה שהסברתי בהרחבה ב3
  10. אלו השלשות שמייצגות את הקולונות שכתובות ב5