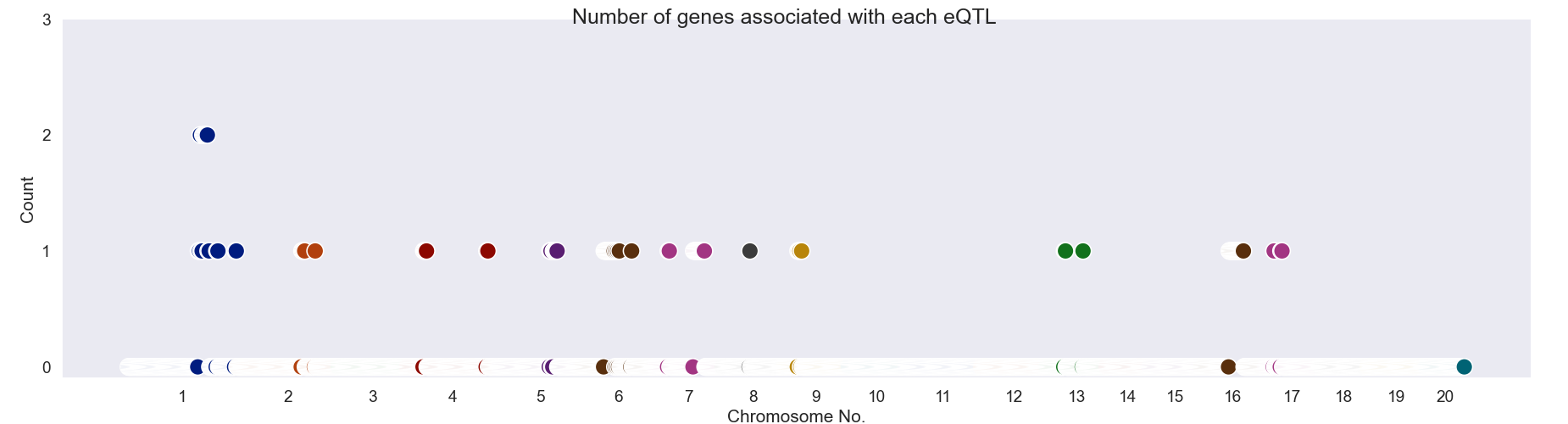
**מטלה 3 – גנטיקה בעידן הרפואה האישית**

יהלי בן דוד 315639898  
ניר בורגר 313580920

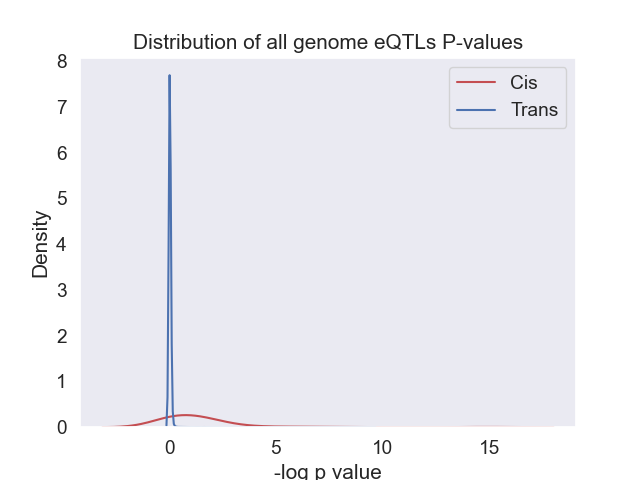
תהליך מקדים:  
מסד הנתונים שבחרנו הוא PAM.  
מיצענו את הנתונים על פרטים שונים ולאחר מכן סיננו את העמודות הרלוונטיות, קרי העמודות שמופיעות בשני מאגרי הנתונים. נשארנו עם 1525 לוקוסים רלוונטיים שיאפשרו לנו לחסוך בחישובים ומבחנים.

מבחן אסוציאציה:  
ביצענו מבחן רגרסיה (לשם חישובים מדויקים, נעזרנו בחבילת scipy). במהלך המבחן, חישבנו את ערך P-value של רגרסיה לינארית בין כל זוג של גן ו-SNP. חילקנו את הציונים כך שהטרוזיגוט (H) יקבל 1, הומוזיגוט (D) יקבל 0 והומוזיגוט (B) יקבל 2. לאחר מכן, ביצענו תיקון FDR להשערות מרובות, השמטנו גנים שהראו חוסר מובהקות, כלומר גנים שעבורם התקיים לכל הSNPs. בסוף התהליך נשארנו עם 21 גנים.

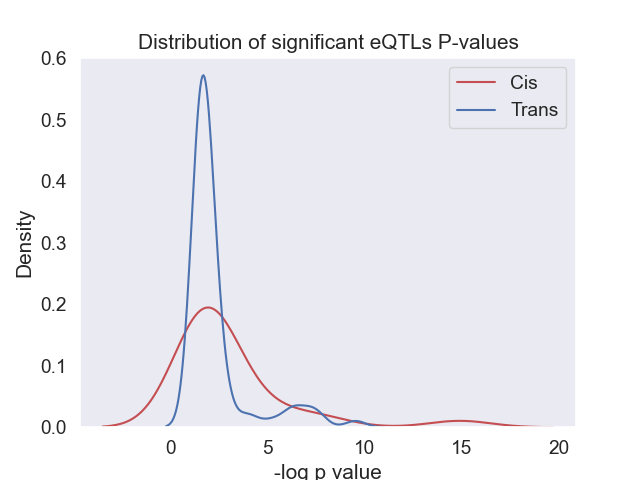
סעיף 1:  
כדי להחליט מהם ה-eQTL המובהקים, בחרנו את כל ה-eQTL שעבורם מתקיים . בנוסף, לפי ההגדרה של cis-trans (מרחקים של 2Mbps או יותר בין גן ל-SNP) מיפינו את הצימודים של גן ו-SNP להגדרה המתאימה.  
סה"כ קיבלנו 101 eQTL מובהקים, מתוכם 24 במצב cis ו-77 במצב trans.

סעיף 2:  
בציר x מוצגים מספרי הכרומוזומים. בציר y מופיע מספר הגנים שנמצאים באסוציאציה מובהקת לסניפים על אותו הכרומוזום.

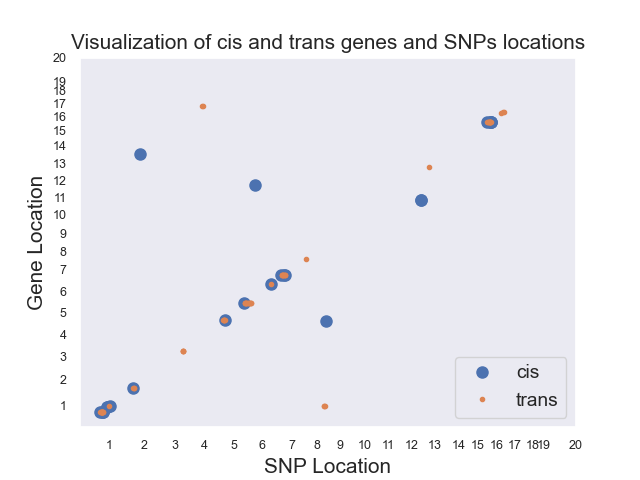
מהגרף ניתן להסיק שבכרומוזום 1 יש 7 סניפים סמוכים שמקושרים ל-2 גנים. עם זאת, ככל הנראה לא מדובר בנקודה חמה (Hot-spot) מכיוון שמספר הגנים שמקושרים לסניפים נמוך מדי בשביל ההגדרה הנ"ל. בנוסף ניתן להסיק שיש מספר גדול של סניפים הסמוכים זה לזה שמקושרים לגן אחד בכרומוזומים 1, 5, 6 ו-16.  
בניגוד לגרפים דומים שראינו, בהם יש מגוון נקודות חמות, בגרף זה אין, והאסוציאציה היא לכל היותר עבור 2 גנים. עם זאת, זו תוצאה סבירה משום שהסתכלנו על 21 גנים ולא על כל הגנום.

סעיף 3:

בגרף שמשמאל ניתן לראות את הצפיפות של ציס וטראנס על פני ערכי ה-P-value לאחר שהפעלנו עליהם טרנספורמציית -log, עבור כל הגנום. נבחין כי במקרה הזה יש משמעותית יותר טראנס מציס. זהו גרף הדומה לגרף שראינו בכיתה והוא עומד בתחזיות.



גם בגרף הזה ניתן לראות את צפיפות ערכי ה-P-value לאחר טרנספורמציית -log. במקרה הזה, לקחנו בחשבון אך ורק את הגנים שהראו מובהקות. ניתן לראות כי גם פה יש יותר טראנס מציס, אך כאן ניתן לראות גם עלייה בערכי ציס סביב . השוני נובע ככל הנראה מכך שיש משמעותית פחות דגימות (101 לעומת 1525).

סעיף 4:  
בציר ה-x מוצגים מיקומי הסניפים. בציר ה-y מיקומי הגנים.  
הפכנו את המיקומים לרציפים כדי ליצור בקלות את הגרף. כדי להפוך את המיקומים לרציפים, הגדרנו את המיקום של כל גן להיות הממוצע בין ההתחלה לסוף שלו, ולאחר מכן, לכל המיקומים של כל גן ו-SNP, הוספנו את המיקום המקסימלי מהכרומוזום הקודם.  
ניתן לראות שהעיגולים הכחולים הגדולים, המייצגים את הזוגות המובהקים בציס, מסודרים בקירוב על הישר האלכסוני, דבר הרומז על קורלציה חזקה. ניתן לראות גם שרוב הזוגות המובהקים שבטראנס נמצאים גם הם על האלכסון.  
בנוסף, ניתן היה לצפות שנקבל ערכים היוצרים מעין "עמודות" על הגרף, שבתורן היו רומזות על קיומן של נקודות חמות. העובדה שאין כאלה מחזקת את המסקנות מסעיף 2 שאין נקודות חמות, קרי, שאין סניפים שנמצאים באסוציאציה עם מספר רב של גנים.