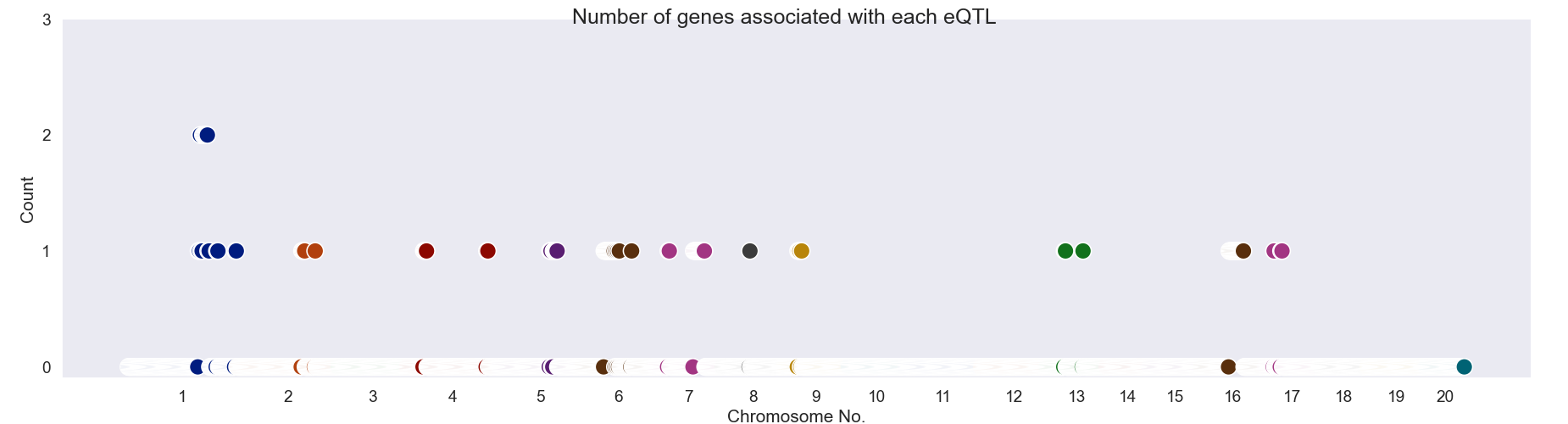
**מטלה 3 – גנטיקה בעידן הרפואה האישית**

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: