**דוח פרויקט גמר**

**מגישים:**

**בר אליחי 322894502**

**עוז אשר 213497365**

**הקדמה**

הפרויקט שלנו מתמקד בסוגים שונים של סרטן ואללים המצויים בהם, כך ששאלת המחקר העיקרית שלנו היא למצוא איזה גנים מעורבים בכל סוג גידול של סרטן מוחי, ומהו אחוז הביטוי של סוג גן מסוים בכל סוג גידול (אם הוא מפגין ביטוי עם אפקט משמעותי לעומת המצב הבריא), יחד עם מידע סטטיסטי נוסף עליו. החלק "המשני" הקיים בקוד הוא אפשרות למשתמש להזין שם של אלל או סוג סרטן ולקבל מידע עליהם כפלט, למשל כאשר יזין שם של אלל - אז הוא יקבל כפלט את סוגי הגידול בהם הוא מתבטא ברמה ששונה מרקמה בריאה, ובכמה אחוזים. אם המשתמש יזין שם של סרטן הפלט יהיה מידע על סוג הסרטן וסוגי האללים הדומיננטים בו,  וכמו כן גם מטריצה של קורלציה בין האללים עם הביטוי הכי משמעותיים בו, בהתאם לניתוח הסטטיסטי.

**שיטות**

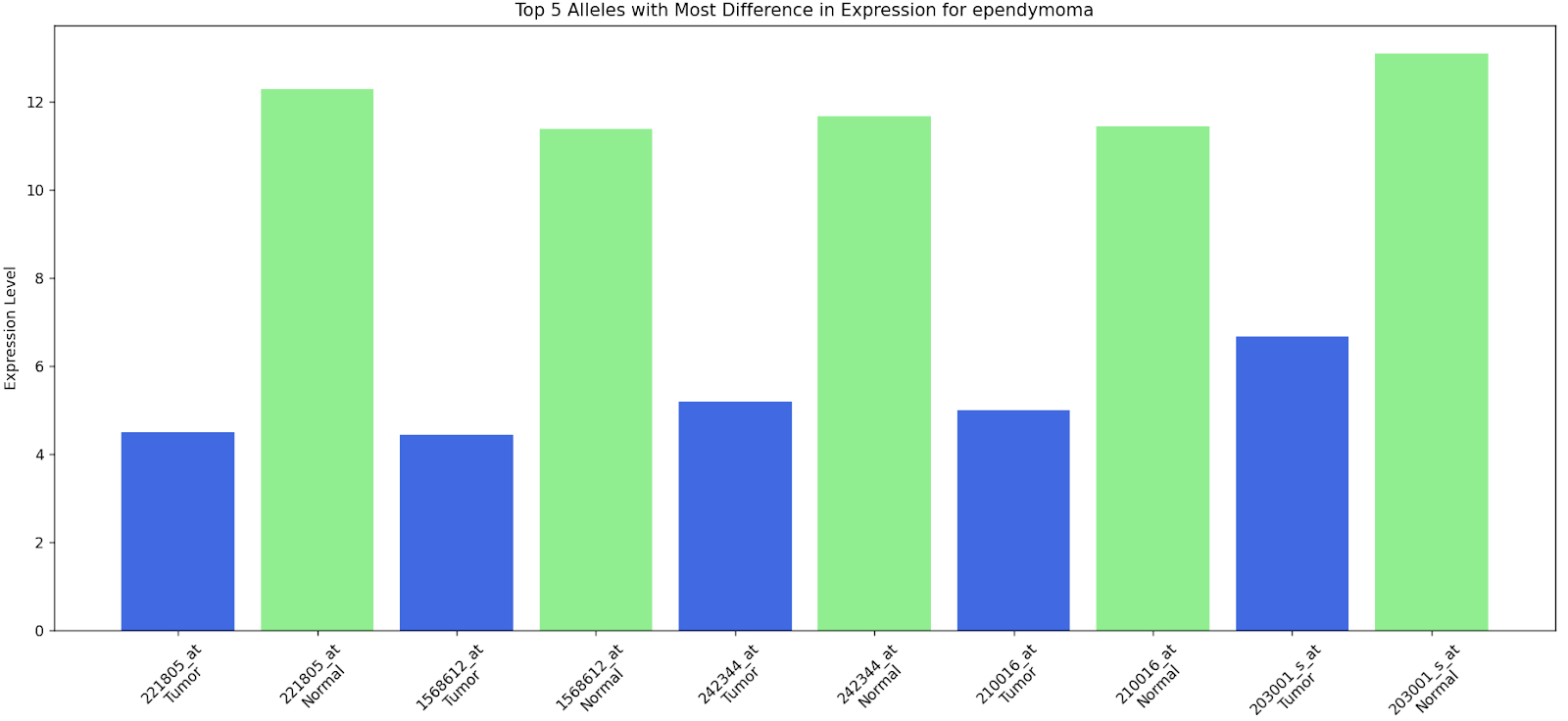
המידע שלנו מתבסס על טבלת אקסל ב- [Brain cancer gene expression - CuMiDa](https://www.kaggle.com/datasets/brunogrisci/brain-cancer-gene-expression-cumida?resource=download). הטבלה מכילה מידע על ביטוי הגנים של ארבעה סוגי סרטן ורקמה בריאה (עם מספר דגימות לכל אחד).

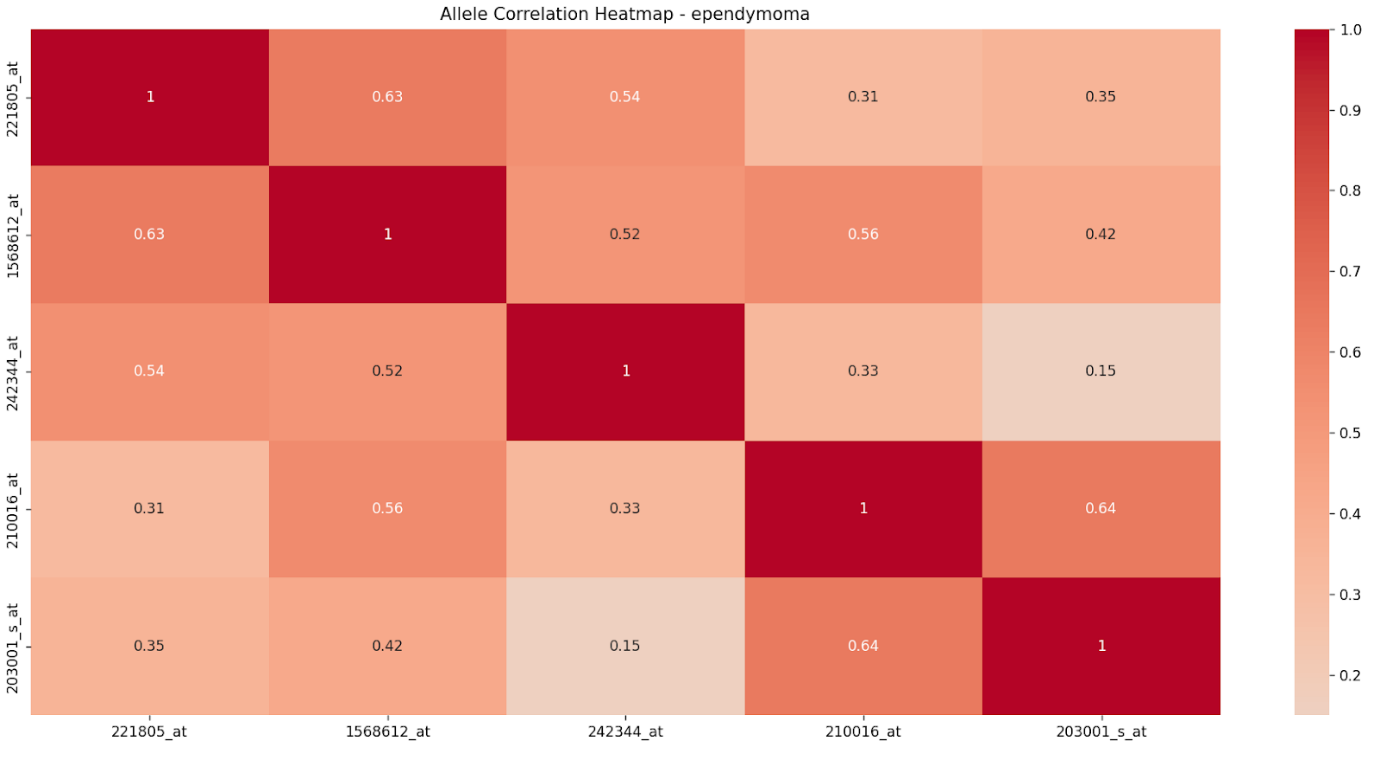
השתמשנו בפונקציות של ניקוי הקוד על מנת לסנן רק את חמשת הגנים הבולטים לכל סוג סרטן מבין כל 50 אלף האפשרויות בטבלת הנתונים. בנוסף לכך השתמשנו בפונקציות שונות על מנת למצוא ערכים סטטיסטיים כמו ערכי קיצון וחציונים, ופונקציות שנועדו לביצוע אינטראקציות עם המשתמש: פונקציות המאתרות את שם הגן או הסרטן הנדרש, פונקציה הפולטת למשתמש את האופציות להזנה וקולטת את המידע שהזין כך שתוכל לפלוט את המידע המתאים כתגובה לכך, פונקציות המחשבות ומציגות את הנתונים הרלוונטים על הדבר שהמשתמש הזין כמו רמת הביטוי של הגנים בכל סרטן וכו. לכל הפונקציות הללו הוספנו בדיקות (פונקציות testing) על מנת לוודא את תקינותן.

**תוצאות**

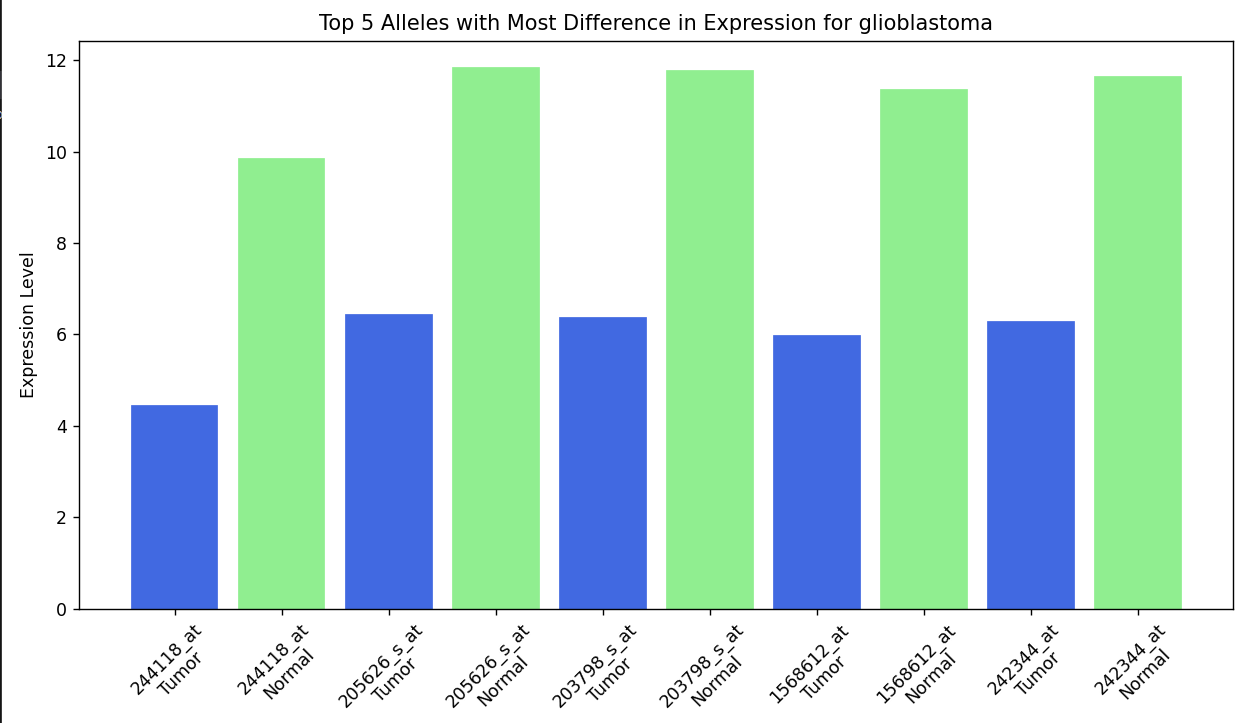
הצלחנו למצוא את הגנים הדומיננטים לכל סוג סרטן, והשגנו גם נתונים על רמת הביטוי שלהם ביחס לרמת הביטוי שלהם ברקמה בריאה. יש גנים למשל אשר מתבטאים פחות בסוגים מסוימים של סרטן, מה שדווקא מצביע על הטענה של חוסר של הביטוי שלהם הוא אחד הגורמים שהובילו להתפתחות סוגי סרטן מסוג זה. בנוסף ראינו גם את הקורלציה בין מספר סוגי גנים המשחקים את התפקיד הכי חשוב מה שמצביע על עד כמה הם משפיעים אחד על השני.

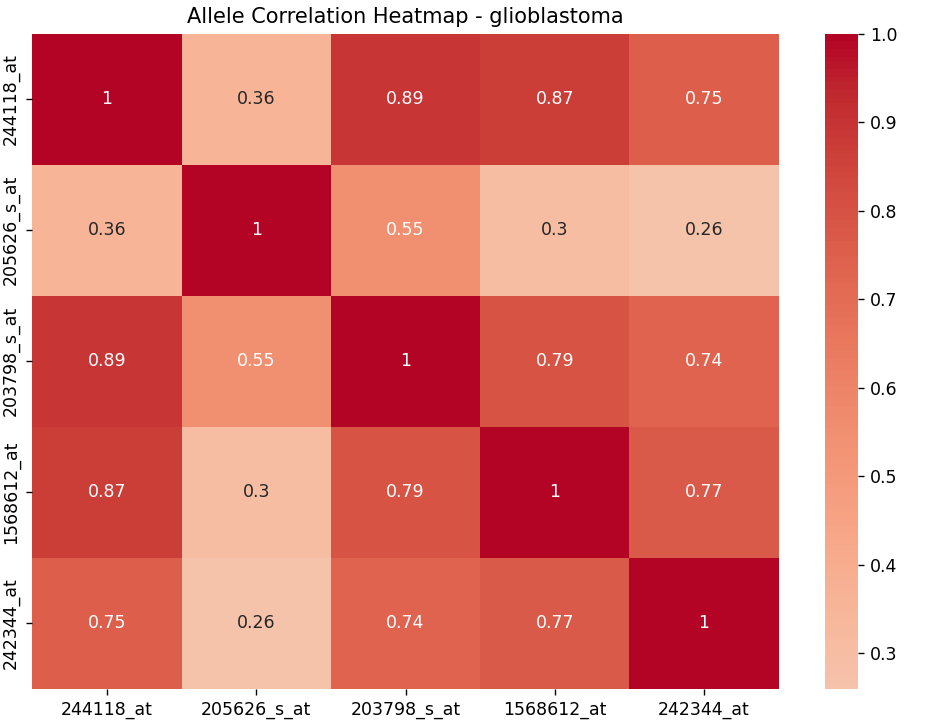
תוצאות עבור Ependymoma:



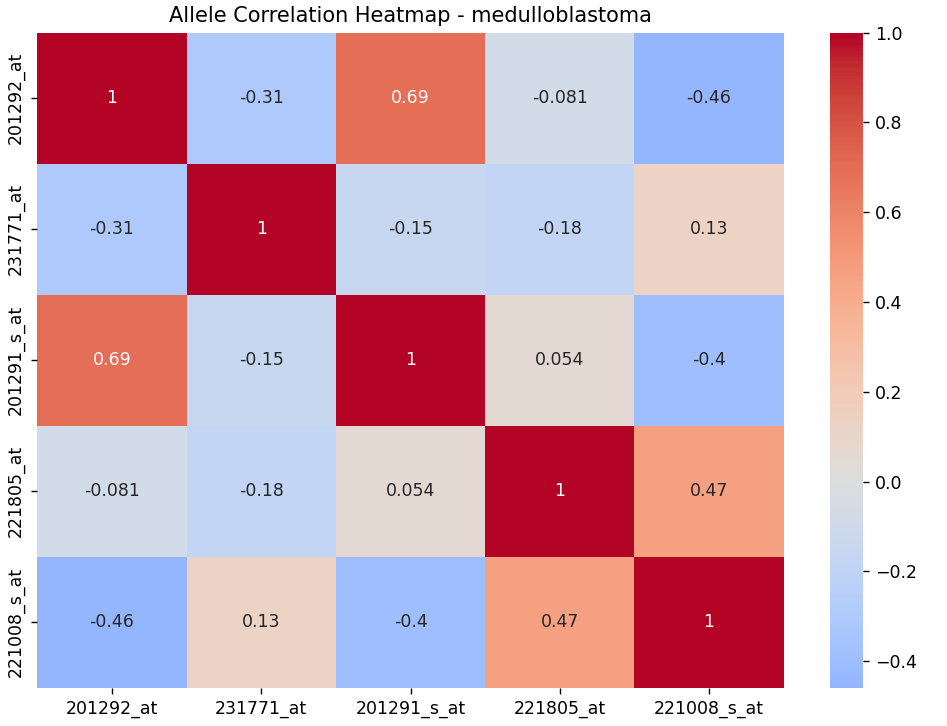
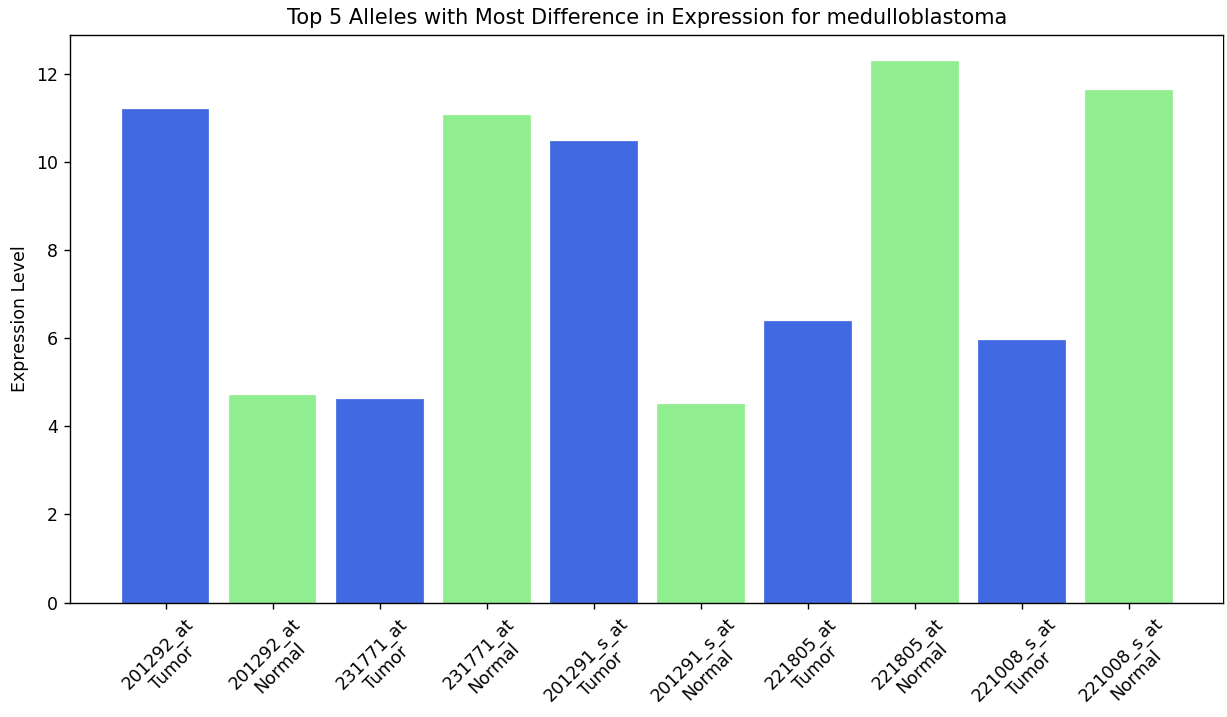


תוצאות עבור Giloblastoma:

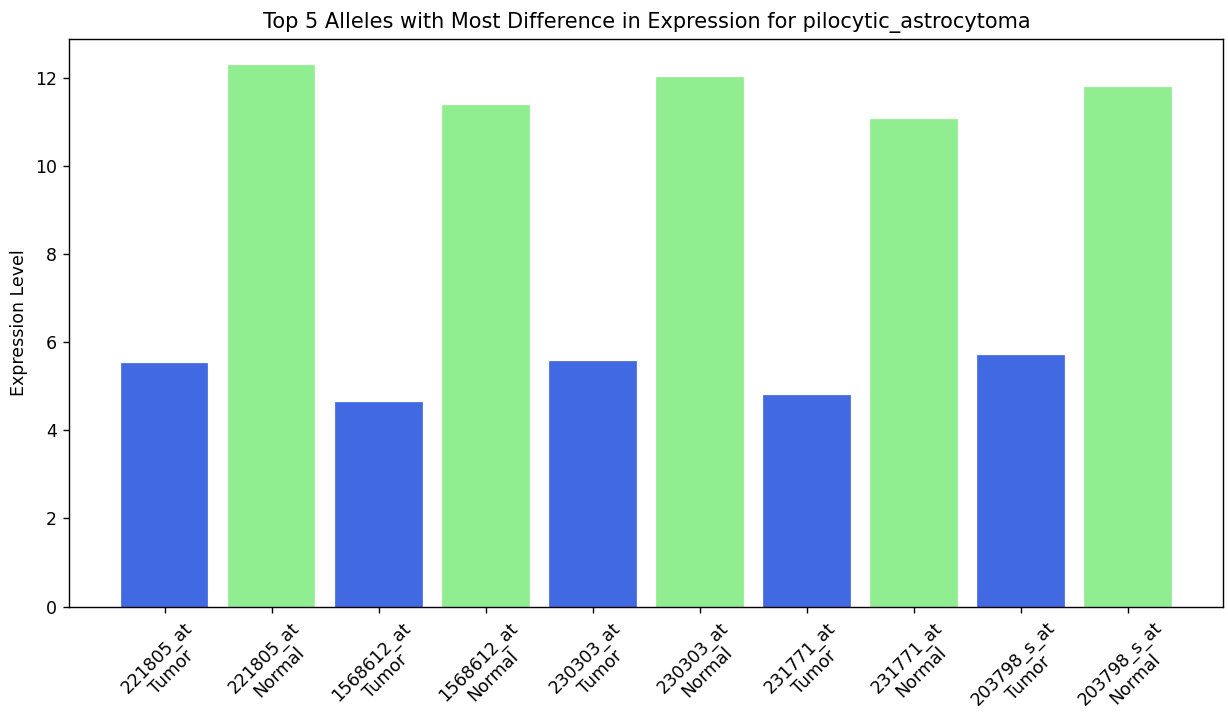


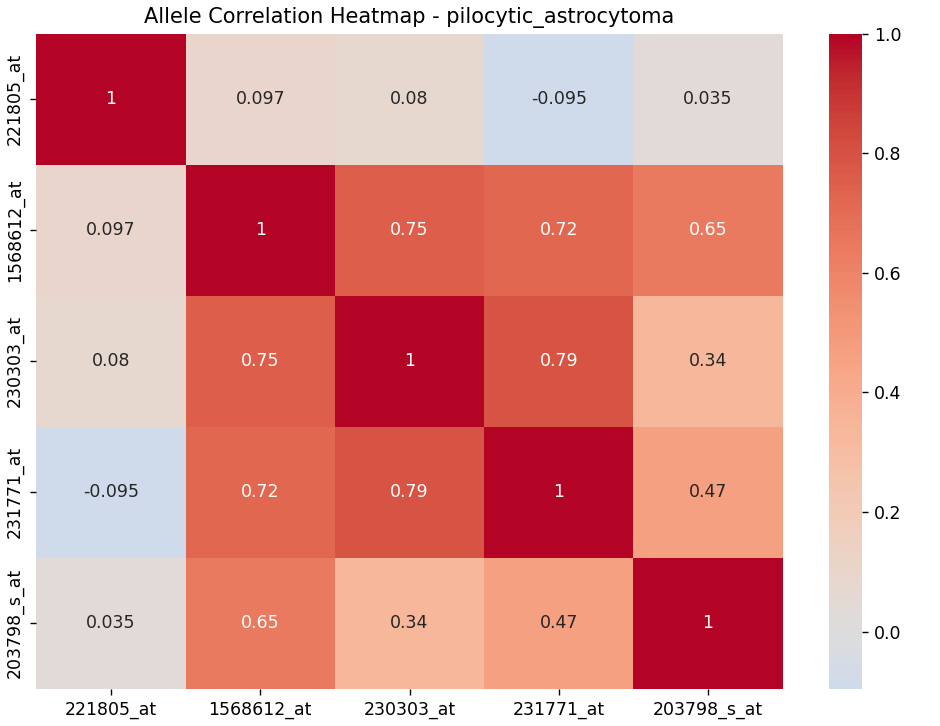


תוצאות עבור Medulloblastoma:



תוצאות עבור Pilocytic Astrocytoma:





דוגמא לתוצאה של הנתונים על אלל 1316\_at

