

## פרויקט סופי – גנטיקה בעידן הרפואה האישית

יהלי בן דוד 315639898

ניר בורגר 313580920

### חלק 2 – עיבוד מקדים:

בחרנו לעבוד עם הפנוטיפים שעוסקים בהשפעת הזרקת מורפיום בטווחים של 45-60 דקות מההזרקה על זכרים, נקבות, וזכרים ונקבות:

Morphine response (50 mg/kg ip), locomotion (open field) from 45-60 min after injection in an activity chamber for males [cm] (970)

Morphine response (50 mg/kg ip), locomotion (open field) from 45-60 min after injection in an activity chamber for females [cm] (1224)

Morphine response (50 mg/kg ip), locomotion (open field) from 45-60 min after injection in an activity chamber for males and females [cm] (1478)

נבחן את הביטוי הפנוטיפי על מסדי נתונים הלקוחים מההיפותלמוס (מוח) והכבד, מהמאגרים שסופקו.

בחרנו בהיפותלמוס משום שלמורפיום יש השפעה ישירה על המוח, ובכבד משום שהוא מהווה גורם מרכזי בפירוק החומר בגוף.

נצפה לראות השפעה של האזורים הללו על הביטוי הפנוטיפי. כמו כן, מעניין יהיה לגלות אם יש קשר בין הביטוי הפנוטיפי לגנים למין הנבדק.

בתהליך העיבוד המקדים נירמלנו את הנתונים על ההיפותלמוס משום שלהם תוחלת 8 ושונות 4. החלטנו לנרמל להתפלגות נורמלית סטנדרטית.

את הטבלה בנינו כך ששורות הן גנים ועמודות הן BXD כנדרש והורדנו שורות ללא מזהה גנטי, כנדרש. החלטנו לסנן 50% מהגנים שיש להם ביטוי מקסימלי נמוך. ומתוכם נשארו עם 1500 גנים בעלי שונות מקסימלית. סה"כ נשארו עם 3000 גנים, 1500 מכל מסד נתונים.

מספר הגנים לאחר הסינון הראשוני	מספר הגנים שמהווה 50% מהגנים עם הביטוי המקסימלי הגבוה ביותר	מספר הגנים שנשארו לאחר הוצאת גנים עם שונות נמוכה	מספר ה-Strains שנשארו בסוף הסינון	
24197	12098	1500	46	היפותלמוס
17872	8936	1500	40	כבד

### חלק 3 – ניתוח eQTL:

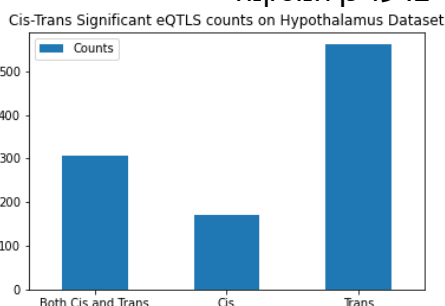
בדומה למטלה 3, ספרנו את מספר הזוגות של (גן, סניף) שלהם היה P-value מובהק (שאינו גדול מ-0.05). ביניהם ספרנו את אלה שמתנהגים כציס וכטראנס, כאשר ההגדרה לציס נשארה כשהייתה מהמטלה הקודמת – על אותו הכרומוזום ובמרחק שאינו עולה על 2Mbp.

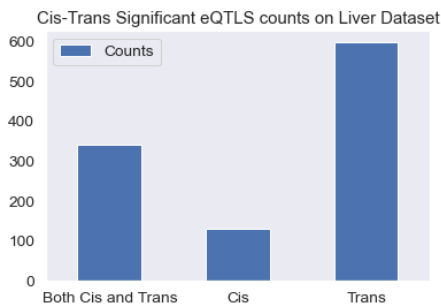
בהתאם לתוצאות שקיבלנו, ראינו כי יש יותר צמידים מובהקים מ-eQTLs ייחודיים. על כן המסקנה היא שיש כמה סניפים שמתנהגים גם כציס וגם כטראנס (להלן דואלים), כתלות בגן, ואין הצדקה להחשיבם כאחד או כשני בלבד.

#### בהיפותלמוס:

קיבלנו סה"כ 1041 eQTLs שונים, ו-3001 זוגות מובהקים. מתוך 1041 ה-eQTLs השונים, 172 מהם הם בצורת ציס, 562 בצורת טראנס, ו-307 בשתי הצורות.

התפלגות מובאת בגרף הבא:





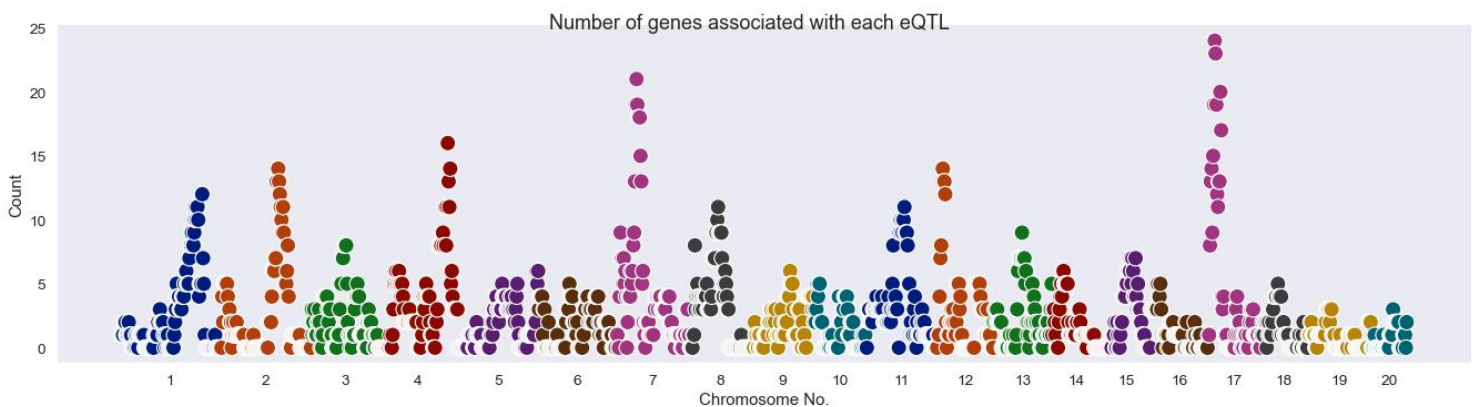
#### כבד:

קיבלנו סה"כ 1070 eQTLs שונים, ו-3473 זוגות מובהקים. מתוך 1070 ה- eQTLs השונים, 132 מהם הם בצורת ציס, 597 בצורת טראנס, ו-341 בשתי הצורות. התפלגות מובאת בגרף הבא:

ניתן לראות הבדלים בכמויות בין ציס, טראנס ודואלים. סיבה אפשרית לכך היא שלעיתים, סניפים הקשורים לגנים שרחוקים מהם (כלומר, זוג שבמצב טראנס) הם לעיתים סניפים שלמעשה משפיעים על חלבוני שיעתוק שמבקרים מספר גנים שעשויים להיות גם קרובים וגם רחוקים מאותו הסניף. על כן, נקבל יותר גנים שבאסוציאציה לאותו הסניף, חלקם קרובים (ציס) וחלקם רחוקים (טראנס), ומספר כולל גבוה יותר של eQTLs בסופו של דבר. בנוסף, נזכור שיש יותר גנים רחוקים מסניף נתון מאשר גנים קרובים, שכן יש מספר סופי ומוגבל מאוד של גנים הקרובים לאותו סניף נתון.

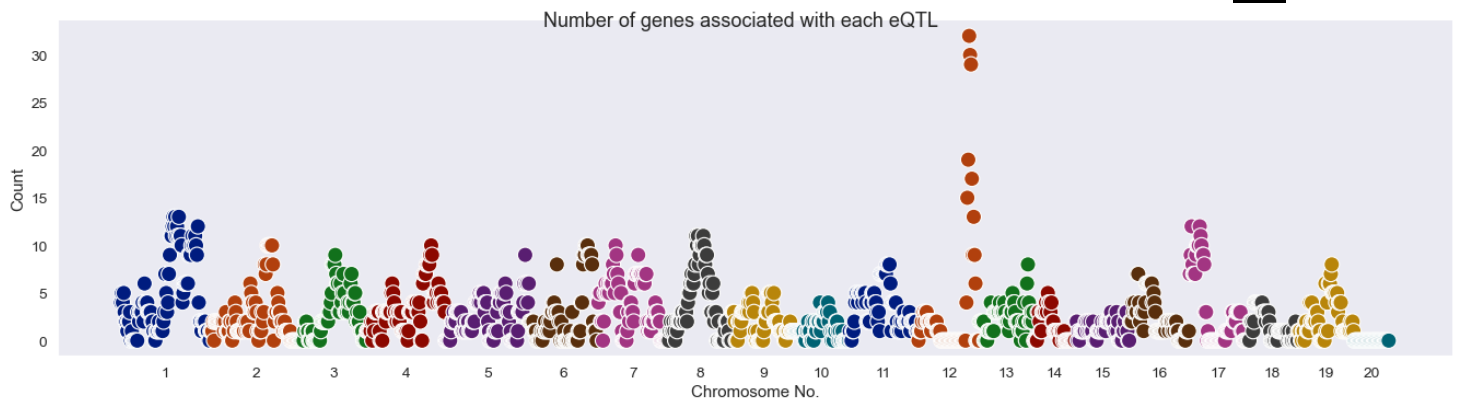
#### מספר הגנים שבאסוציאציה עם כל eQTL לאורך הגנום

#### היפותלמוס:



ניתן לראות שבכרומוזום 17 נמצא הסניף שבאסוציאציה עם הכי הרבה גנים (קרוב ל-25), כלומר ייתכן שבכרומוזום זה נמצא סניף המהווה נקודה חמה (hotspot). ניתן אמנם לראות סניפים שבאסוציאציה עם 15-20 גנים בכרומוזומים 2, 4, 7 ו-12, אך אלה לא בהכרח ייחשבו לנקודות חמות, משום שהן מצריכות אסוציאציה עם הרבה גנים – מעל 20 לפחות.

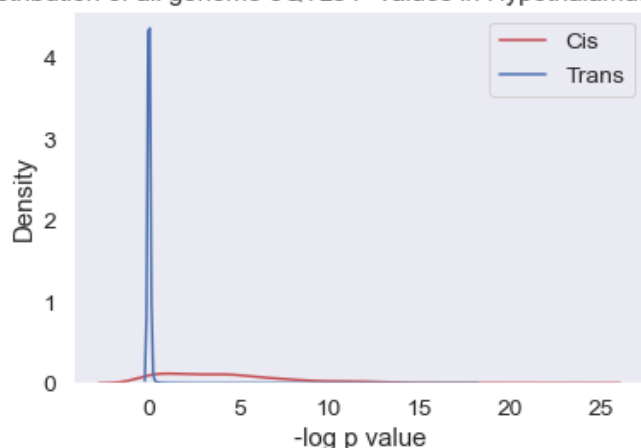
#### כבד:



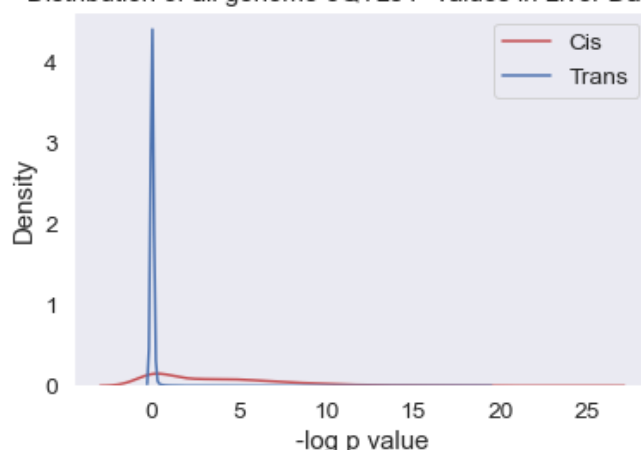
ניתן לראות שבכרומוזום 12 נמצא הסניף שבאסוציאציה עם הכי הרבה גנים (מעל 30), כלומר ייתכן שבכרומוזום זה נמצא סניף המהווה נקודה חמה (hotspot).  
ניתן אמנם לראות סניפים שבאסוציאציה עם 10-15 גנים במספר כרומוזומים (למשל 1, 8, 17 ועוד), אך אלה לא בהכרח ייחשבו לנקודות חמות, משום שהן מצריכות אסוציאציה עם הרבה גנים – מעל 20 לפחות.

### התפלגות כל ערכי P-values עבור eQTLs שבציס ובטראנס

Distribution of all genome eQTLs P-values in Hypothalamus Dataset



Distribution of all genome eQTLs P-values in Liver Dataset



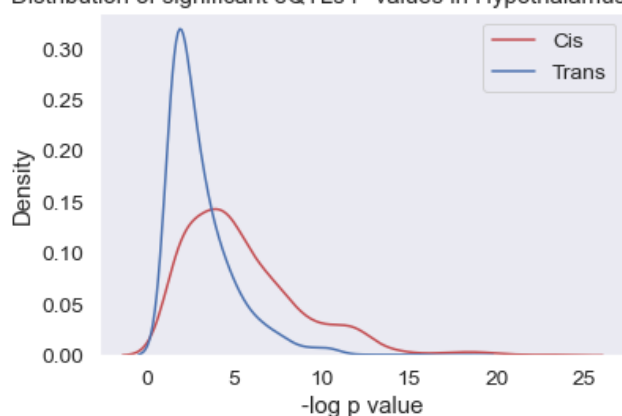
נדגיש כי הגרפים הבאים מציגים את ההתפלגות של ערכי P-value לאחר שהופעלה עליהם טרנספורמציה  $-\log$  בבסיס 10.

נבחין כי בשתי הסביבות, הן בהיפותלמוס והן בכבד, ל-eQTLs שבצורת טראנס יש ערכי  $-\log_{10}(P_{value})$  נמוכים יותר, כלומר ערכי P-value גבוהים יותר מאשר אלה שבציס. יתרה מזו, לרוב ה-eQTLs שבצורת טראנס יש ערכי P-value גבוהים (ערכי ה- $-\log_{10}(P_{value})$  – קרובים מאוד ל-0), בעוד שעבור eQTLs שבצורת ציס יש ערכי P-value נמוכים (ערכי  $-\log_{10}(P_{value})$  – קרובים ל-0).

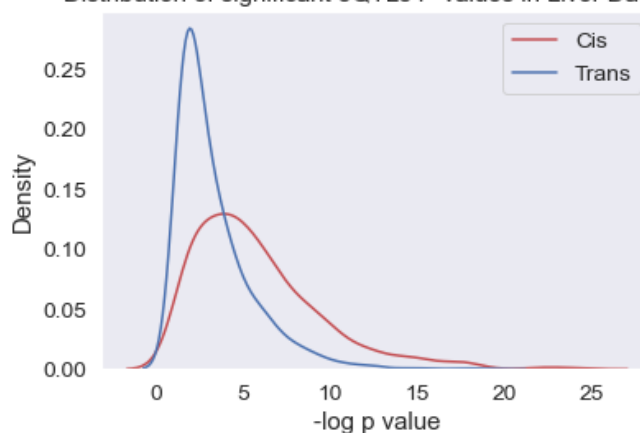
בנוסף, אחוז ה-eQTLs המובהקים (שעבורם מתקיים  $P_{value} \leq 0.05$ ) גבוה יותר בקרב אלה שבצורת ציס.

### התפלגות ערכי P-values המובהקים עבור eQTLs שבציס ובטראנס

Distribution of significant eQTLs P-values in Hypothalamus Dataset



Distribution of significant eQTLs P-values in Liver Dataset



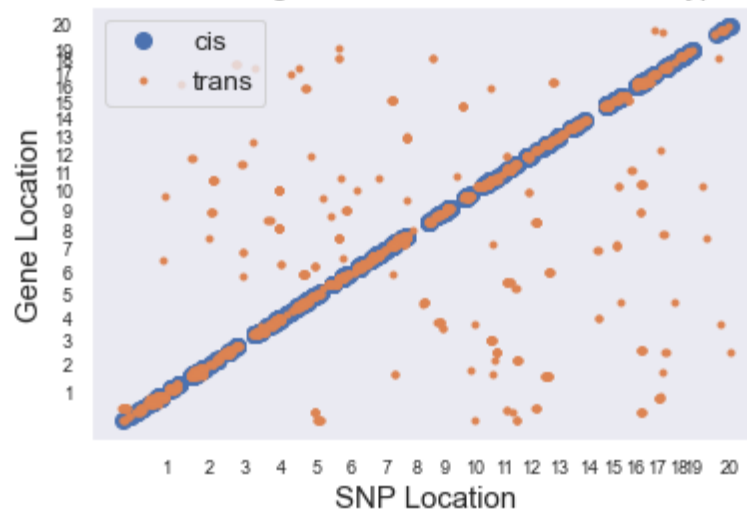
נדגיש כי הגרפים הבאים מציגים את ההתפלגות של ערכי P-value לאחר שהופעלה עליהם טרנספורמציה  $-\log$  בבסיס 10. נבחין כי בשתי הסביבות, ההיפותלמוס והכבד, בדומה להתפלגות על פני כל הצימודים של גן-סניף (משמעותיים או לא), גם עבור הצימודים המשמעותיים נבחין כי בשתי הסביבות, הן בהיפותלמוס והן בכבד, ל-eQTLs שבצורת טראנס יש ערכי  $-\log_{10}(P_{value})$  גבוהים יותר.

נמוכים יותר, כלומר ערכי P-value גבוהים יותר מאשר אלה שבציס. עם זאת עבור הצימודים המשמעותיים נבחין כי קיים אחוז מסוים של צימודים עבורם ה-p-value נמוך יחסית.

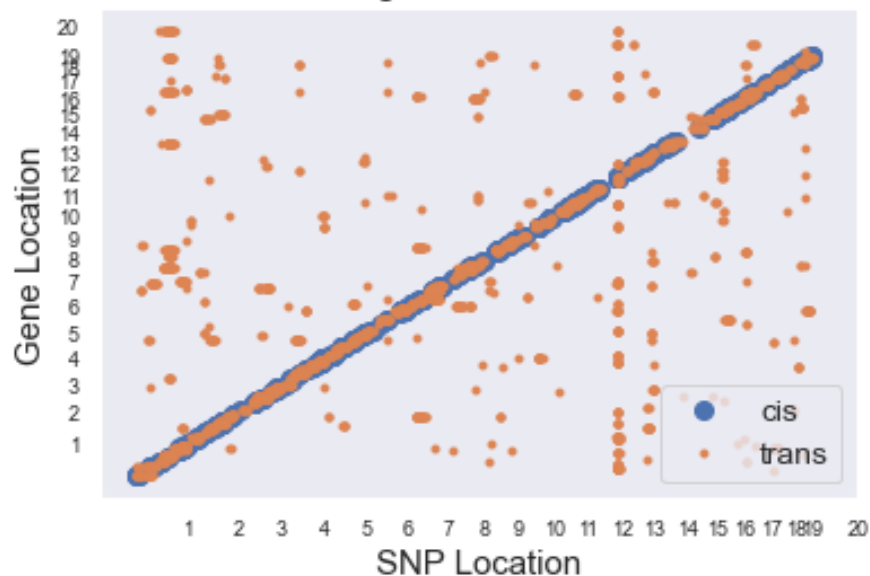
#### פיזור של תוצאות משמעותיות לפי מיקום ה-SNP ומיקום הגן

כל נקודה מייצגת eQTL משמעותי, כאשר נקודות כתומות מייצגות trans ונקודות כחולות מייצגות cis. כדי שהמיקום יהיה רציף, הגדרנו את המיקום בתור המיקום על הכרומוזום הרלוונטי בתוספת המיקום האחרון על הכרומוזום הקודם.

Visualization of cis and trans genes and SNPs locations in Hypothalamus Dataset



Visualization of cis and trans genes and SNPs locations in Liver Dataset



נשים לב לאלכסונים ברורים בשני הגרפים, שמייצגים את הצימודים בציס או הצימודים בטרנס, עבורם ה-eQTL וה-SNP קרובים יחסית. תוצאות אלה תואמות את הממצא לפיו מרבית ה-eQTLים המשמעותיים היו cis. בנוסף לאלכסון נחפשו Hotspots כלומר SNPs בצימוד טראנס למספר גנים גדול שמיוצגים בגרף ע"י "עמודה מלאה". עבור ההיפותלמוס אין "עמודות" מלאות, בכרומוזום 17 נבחין בכמה נקודות, ואכן בגרף קודם הראנו שמספר הגנים המקסימלי שהיו באסוציאציה משמעותית אתו היה בערך 25, כלומר לא מעט אבל לא מוגדר כ-Hotspot. לעומת זאת בכבד נראה עמודה מלאה יחסית בכרומוזום 12. בנוסף קיימות נקודות גם לאורך כרומוזום 1 ו-8. התוצאות הנ"ל הן

בהלימה לגרף הקודם בו ראינו בכרומוזום 12 Hotspot של מעל 30 גנים, וכ-15 גנים באסוציאציה לכרומוזומים 1,8.

חלק 4 – ניתוח QTL:

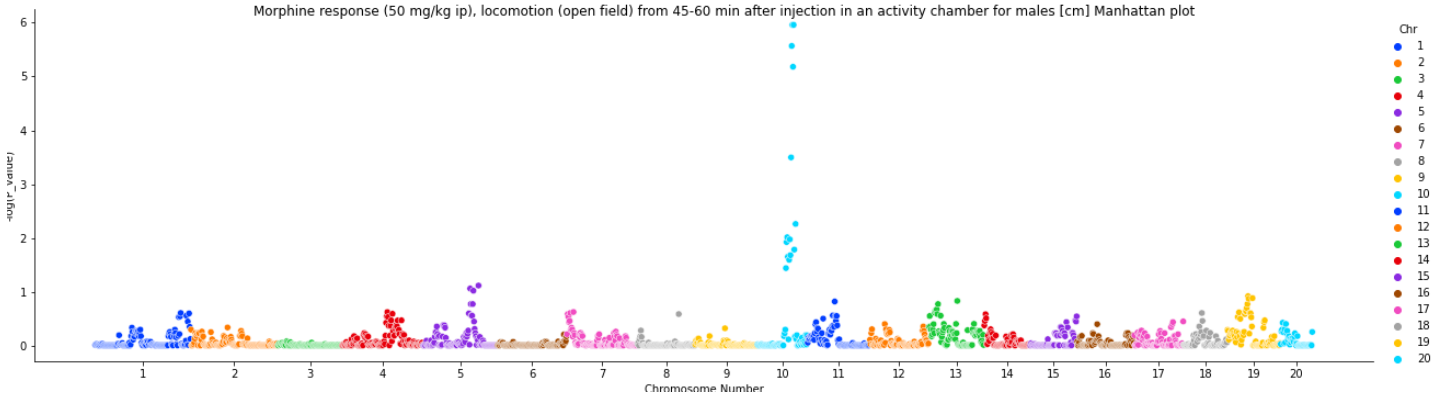
נסכם את ממצאי הניתוח:

פנוטיפ	מספר ה-QTLים	מספר הכרומוזום	כמות
Morphine response (50 mg/kg ip), locomotion (open field) from 45-60 min after injection in an activity chamber for males [cm]	14	10	14
Morphine response (50 mg/kg ip), locomotion (open field) from 45-60 min after injection in an activity chamber for females [cm]	16	5	1
		10	14
		11	1
Morphine response (50 mg/kg ip), locomotion (open field) from 45-60 min after injection in an activity chamber for males and females [cm]	19	5	4
		10	14
		11	1

ניתן לראות שרוב ה-QTL נמצאים על כרומוזום מספר 10.

נבחן את תרשימי מנהטן של כל אחד מהפנוטיפים:

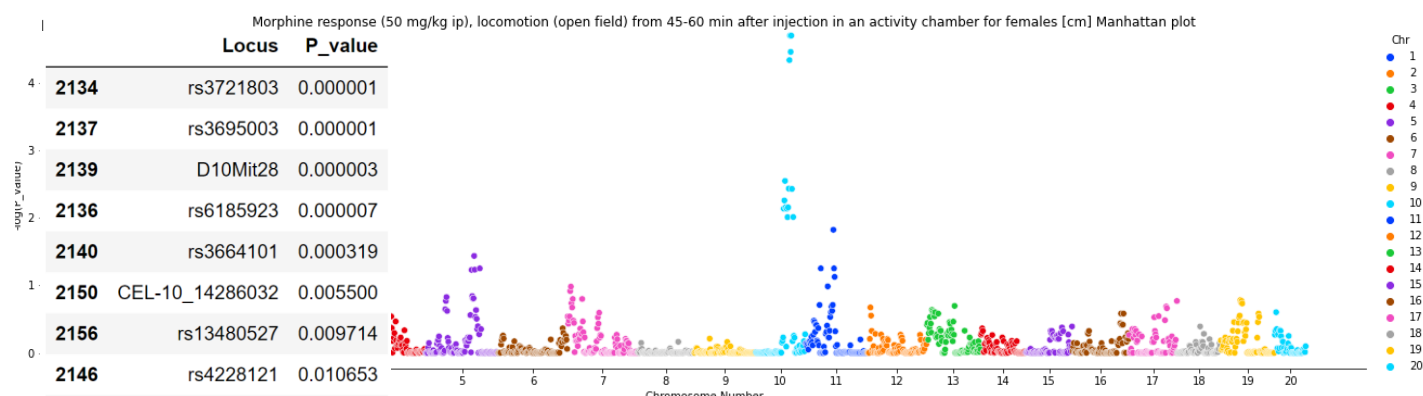
בזכרים - Morphine response (50 mg/kg ip), locomotion (open field) from 45-60 min after injection in an activity chamber for males [cm]



ניתן לראות שבדומה לניתוח מהטבלה, בכרומוזום 10 קיימים סניפים שעבורם ערך ה-P-value נמוך (משום ש— $\log(P\text{-value})$  גבוה מאוד), בניגוד לשאר הכרומוזומים, בהם הערכים הנ"ל מראים על חוסר מובהקות. המסקנה היא שכנראה יש קשר בין קבוצת סניפים שנמצאת על כרומוזום 10 ובין הפנוטיפ שבדקנו – השפעה של מורפיום על עכברים זכרים 45-60 דקות מרגע ההזרקה.

בטבלה משמאל מובאים כל הלוקוסים שקיבלו ערכים מובהקים, ממוינים בסדר עולה.

### בנקבות - Morphine response (50 mg/kg ip), locomotion (open field) from 45-60 min after injection in an activity chamber for females [cm]

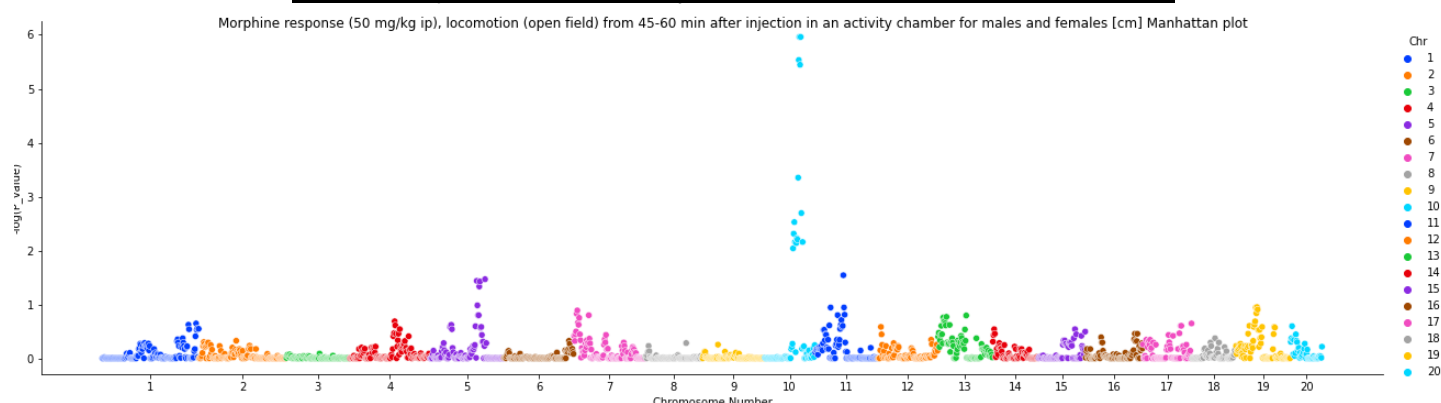


בדומה לזכרים, גם כאן אין שוני מהטבלה – הסניפים שמראים מובהקות נמצאים בעיקר בכרומוזום 10, אך יש ניצנים של מובהקות בכרומוזומים 11 ו-5.

בטבלה משמאל מוצגים כל הלוקוסים שקיבלו ערכים מובהקים, ממוינים מהנמוך לגבוה.

	Locus	P_value
5		
2	2134 rs3721803	0.000020
5	2137 rs3695003	0.000020
7	2136 rs6185923	0.000035
	2139 D10Mit28	0.000046
	2156 rs13480527	0.002821
	2140 rs3664101	0.003669
	2150 CEL-10_14286032	0.003707
	2157 rs3688363	0.005500
	2144 rs4228112	0.006957
	2155 rs13480526	0.006957
	2154 rs13480525	0.007272
	2159 rs13459119	0.007272
	2152 rs3712394	0.009674
	2146 rs4228121	0.009714
	2442 rs13481186	0.014941
	1090 rs6404215	0.036457

# בזכרים ונקבות - Morphine response (50 mg/kg ip), locomotion (open field) from 45-60 min after injection in an activity chamber for males and females [cm]



	Locus	P_value
2134	rs3721803	0.000001
2137	rs3695003	0.000001
2139	D10Mit28	0.000003
2136	rs6185923	0.000004
2140	rs3664101	0.000445
2150	CEL-10_14286032	0.002021
2156	rs13480527	0.002973
2157	rs3688363	0.004875
2146	rs4228121	0.005939
2144	rs4228112	0.006259
2155	rs13480526	0.006957
2152	rs3712394	0.006957
2154	rs13480525	0.007272
2159	rs13459119	0.009130
2442	rs13481186	0.028714
1098	rs13478117	0.034030
1099	mCV22832219	0.036457
1089	mCV23582150	0.037544
1090	rs6404215	0.046437

בדומה לזכרים ולנקבות – גם כאן, בזכרים ובנקבות יחד, יש מובהקות גדולה בכרומוזום מספר 10, וכן סימנים של מובהקות בכרומוזומים 11 ו-5, כאשר יש יותר ב-5 מאשר בשני הפנוטיפים האחרים בנפרד.

בטבלה משמאל מוצגים הלוקוסים שקיבלו ערכים מובהקים, ממוינים בסדר עולה מהנמוך לגבוה.

המסקנה המתבקשת היא שיש קשר כלשהו בין הלוקוסים שבכרומוזומים 5, 10 ו-11 לבין הפנוטיפים שבחרנו. זאת ועוד, ובמיוחד בכרומוזום 10, יש קשר בין הלוקוסים שבו לבין עיבוד ופירוק מורפיום בעכברים משני המינים. אפשר גם לשער שבעיקר בנקבות יש השפעה גם ללוקוסים בכרומוזומים 5 ו-11.

## חלק 5 – שילוב התוצאות

בחלק זה נשווה בין תוצאות ה-eQTL לתוצאות ה-QTL. נבדוק אילו סניפים נמצאים באסוציאציה משמעותית גם עם הפנוטיפים וגם עם הביטוי הגנטי. באמצעות הסניפים הללו נייצר שלשות של (גן, סניף וביטוי פנוטיפי). בנוסף נתייחס אך ורק לשלשות בהן הגן וה-SNP נמצאות על אותו כרומוזום. סידרנו את השלשות בטבלאות לכל רקמה והן מובאות מטה. ננתח אותן כעת.

כבד: נבחין כי קיים קשר בין הגן Aig1 שנמצא על כרומוזום 10 לבין כל שלושת הפנוטיפים. בנוסף נבחין כי קיימים קשרים בין גנים על כרומוזום 11 ובין הפנוטיפים 1224 (נקבות) ו-1478 (זכרים ונקבות), ובין גנים על כרומוזום 5 לפנוטיפ 1478 (זכרים ונקבות) מה שתואם את הממצאים בחלק 4. לעומת זאת נשים לב כי לא התקבלו כלל שלשות של גנים מכרומוזום 12, על אף שבחלק שלוש ממצאנו כי כרומוזום נמצא ה-SNP עם מספר ה-eQTLים הגבוה ביותר.

היפותלמוס: נבחין כי קיימים קשר בין קבוצת גנים על כרומוזום 10 ובין שלושת הפנוטיפים. בנוסף כמעט לא קיימים קשרים בין הפנוטיפים הללו לגנים ה-SNP על כרומוזומים אחרים (שלשות בודדות על כרומוזום 11).

ניתן לשער כי אם היינו מגבילים את בדיקת ה-QTLים רק עבור גנוטיפים שקיימת אסוציאציה שלהם עם גן אחד לפחות, היינו מקטינים משמעותית את מספר ה-SNPים. כך התיקון היה פחות "דרסטי" והיינו מקבלים יותר QTLים ייחודיים, ואסוציאציות לגנים על כרומוזומים נוספים (למשל כרומוזום 12).



שלישית כבד:

Locus	gene name	snp name	Chr	phenotype	genotype na	P_value	-log(P_value)
D10Mit28	Aig1	2139	10	970	M	2.695E-06	5.56950561
rs3664101	Aig1	2140	10	970	M	0.0003187	3.496661343
rs4228112	Aig1	2144	10	970	M	0.0209649	1.678507087
rs4228121	Aig1	2146	10	970	M	0.0106533	1.972515388
CEL-10_1	Aig1	2150	10	970	M	0.0055001	2.259630964
rs3712394	Aig1	2152	10	970	M	0.0164914	1.782742869
rs1348052	Aig1	2154	10	970	M	0.0255251	1.593032858
rs1348052	Aig1	2155	10	970	M	0.0224222	1.649321598
rs3688363	Aig1	2157	10	970	M	0.0119482	1.922698627
rs1345911	Aig1	2159	10	970	M	0.0364569	1.438220288
rs1348118	Med1	2442	11	1224	F	0.0149408	1.825625307
rs1348118	Lsm12	2442	11	1224	F	0.0149408	1.825625307
D10Mit28	Aig1	2139	10	1224	F	4.576E-05	4.339484075
rs3664101	Aig1	2140	10	1224	F	0.0036689	2.435463847
rs4228112	Aig1	2144	10	1224	F	0.0069567	2.157598644
rs4228121	Aig1	2146	10	1224	F	0.0097141	2.012595348
CEL-10_1	Aig1	2150	10	1224	F	0.0037066	2.431022152
rs3712394	Aig1	2152	10	1224	F	0.0096736	2.01440983
rs1348052	Aig1	2154	10	1224	F	0.0072717	2.13836415
rs1348052	Aig1	2155	10	1224	F	0.0069567	2.157598644
rs3688363	Aig1	2157	10	1224	F	0.0055001	2.259630964
rs1345911	Aig1	2159	10	1224	F	0.0072717	2.13836415
rs1348118	Med1	2442	11	1478	M & F	0.028714	1.541905592
rs1348118	Lsm12	2442	11	1478	M & F	0.028714	1.541905592
D10Mit28	Aig1	2139	10	1478	M & F	2.931E-06	5.532925562
rs3664101	Aig1	2140	10	1478	M & F	0.0004454	3.35128877
rs4228112	Aig1	2144	10	1478	M & F	0.0062594	2.203470054
rs4228121	Aig1	2146	10	1478	M & F	0.0059391	2.226283003
CEL-10_1	Aig1	2150	10	1478	M & F	0.0020208	2.694487287
rs3712394	Aig1	2152	10	1478	M & F	0.0069567	2.157598644
rs1348052	Aig1	2154	10	1478	M & F	0.0072717	2.13836415
rs1348052	Aig1	2155	10	1478	M & F	0.0069567	2.157598644
rs3688363	Aig1	2157	10	1478	M & F	0.0048755	2.311982198
rs1345911	Aig1	2159	10	1478	M & F	0.0091297	2.039541532
rs1347811	Wdr86	1098	5	1478	M & F	0.0340301	1.468136211
rs1347811	Pion	1098	5	1478	M & F	0.0340301	1.468136211
mCV22832	Wdr86	1099	5	1478	M & F	0.0364569	1.438220288
mCV22832	Pion	1099	5	1478	M & F	0.0364569	1.438220288

Locus	gene name	snp name	Chr	phenotype	phenotype name	P_value	-log(P_value)
rs3664101	Nhs1	2140	10	970 M		0.0003187	3.496661343
rs3664101	Nmbr	2140	10	970 M		0.0003187	3.496661343
rs3664101	Hebp2	2140	10	970 M		0.0003187	3.496661343
rs3664101	Adat2	2140	10	970 M		0.0003187	3.496661343
rs3664101	Ppil4	2140	10	970 M		0.0003187	3.496661343
rs4228121	Nhs1	2146	10	970 M		0.0106533	1.972515388
rs4228121	Nmbr	2146	10	970 M		0.0106533	1.972515388
rs4228121	Hebp2	2146	10	970 M		0.0106533	1.972515388
rs4228121	Adat2	2146	10	970 M		0.0106533	1.972515388
rs4228121	Ppil4	2146	10	970 M		0.0106533	1.972515388
CEL-10_1	Nhs1	2150	10	970 M		0.0055001	2.259630964
CEL-10_1	Nmbr	2150	10	970 M		0.0055001	2.259630964
CEL-10_1	Hebp2	2150	10	970 M		0.0055001	2.259630964
CEL-10_1	Adat2	2150	10	970 M		0.0055001	2.259630964
CEL-10_1	Ppil4	2150	10	970 M		0.0055001	2.259630964
rs3721803	Nmbr	2134	10	970 M		1.095E-06	5.960620122
rs3721803	Adat2	2134	10	970 M		1.095E-06	5.960620122
rs3721803	Ppil4	2134	10	970 M		1.095E-06	5.960620122
rs3695003	Nmbr	2137	10	970 M		1.095E-06	5.960620122
rs3695003	Adat2	2137	10	970 M		1.095E-06	5.960620122
rs3695003	Ppil4	2137	10	970 M		1.095E-06	5.960620122
rs4228112	Nmbr	2144	10	970 M		0.0209649	1.678507087
rs4228112	Hebp2	2144	10	970 M		0.0209649	1.678507087
rs4228112	Adat2	2144	10	970 M		0.0209649	1.678507087
rs4228112	Ppil4	2144	10	970 M		0.0209649	1.678507087
rs1348052	Hebp2	2154	10	970 M		0.0255251	1.593032858
rs1348052	Adat2	2154	10	970 M		0.0255251	1.593032858
rs1348052	Hebp2	2155	10	970 M		0.0224222	1.649321598
rs1348052	Adat2	2155	10	970 M		0.0224222	1.649321598
rs1345911	Hebp2	2159	10	970 M		0.0364569	1.438220288
rs1345911	Adat2	2159	10	970 M		0.0364569	1.438220288
rs6185923	Adat2	2136	10	970 M		6.583E-06	5.181603074
rs6185923	Ppil4	2136	10	970 M		6.583E-06	5.181603074
rs3664101	Nhs1	2140	10	1224 F		0.0036689	2.435463847
rs3664101	Nmbr	2140	10	1224 F		0.0036689	2.435463847
rs3664101	Hebp2	2140	10	1224 F		0.0036689	2.435463847
rs3664101	Adat2	2140	10	1224 F		0.0036689	2.435463847
rs3664101	Ppil4	2140	10	1224 F		0.0036689	2.435463847
rs4228121	Nhs1	2146	10	1224 F		0.0097141	2.012595348
rs4228121	Nmbr	2146	10	1224 F		0.0097141	2.012595348
rs4228121	Hebp2	2146	10	1224 F		0.0097141	2.012595348
rs4228121	Adat2	2146	10	1224 F		0.0097141	2.012595348
rs4228121	Ppil4	2146	10	1224 F		0.0097141	2.012595348
CEL-10_1	Nhs1	2150	10	1224 F		0.0037066	2.431022152
CEL-10_1	Nmbr	2150	10	1224 F		0.0037066	2.431022152
CEL-10_1	Hebp2	2150	10	1224 F		0.0037066	2.431022152
CEL-10_1	Adat2	2150	10	1224 F		0.0037066	2.431022152
CEL-10_1	Ppil4	2150	10	1224 F		0.0037066	2.431022152
rs3721803	Nmbr	2134	10	1224 F		1.982E-05	4.70278817
rs3721803	Adat2	2134	10	1224 F		1.982E-05	4.70278817
rs3721803	Ppil4	2134	10	1224 F		1.982E-05	4.70278817
rs3695003	Nmbr	2137	10	1224 F		1.982E-05	4.70278817
rs3695003	Adat2	2137	10	1224 F		1.982E-05	4.70278817
rs3695003	Ppil4	2137	10	1224 F		1.982E-05	4.70278817
rs4228112	Nmbr	2144	10	1224 F		0.0069567	2.157598644
rs4228112	Hebp2	2144	10	1224 F		0.0069567	2.157598644
rs4228112	Adat2	2144	10	1224 F		0.0069567	2.157598644
rs4228112	Ppil4	2144	10	1224 F		0.0069567	2.157598644
rs1348052	Hebp2	2154	10	1224 F		0.0072717	2.13836415
rs1348052	Adat2	2154	10	1224 F		0.0072717	2.13836415
rs1348052	Hebp2	2155	10	1224 F		0.0069567	2.157598644
rs1348052	Adat2	2155	10	1224 F		0.0069567	2.157598644
rs1345911	Hebp2	2159	10	1224 F		0.0072717	2.13836415
rs1345911	Adat2	2159	10	1224 F		0.0072717	2.13836415
rs6185923	Adat2	2136	10	1224 F		3.456E-05	4.46143574
rs6185923	Ppil4	2136	10	1224 F		3.456E-05	4.46143574
rs1348118	Stac2	2442	11	1224 F		0.0149408	1.825625307
rs3664101	Nhs1	2140	10	1478 M & F		0.0004454	3.35128877
rs3664101	Nmbr	2140	10	1478 M & F		0.0004454	3.35128877
rs3664101	Hebp2	2140	10	1478 M & F		0.0004454	3.35128877
rs3664101	Adat2	2140	10	1478 M & F		0.0004454	3.35128877
rs3664101	Ppil4	2140	10	1478 M & F		0.0004454	3.35128877
rs4228121	Nhs1	2146	10	1478 M & F		0.0059391	2.226283003
rs4228121	Nmbr	2146	10	1478 M & F		0.0059391	2.226283003
rs4228121	Hebp2	2146	10	1478 M & F		0.0059391	2.226283003
rs4228121	Adat2	2146	10	1478 M & F		0.0059391	2.226283003
rs4228121	Ppil4	2146	10	1478 M & F		0.0059391	2.226283003
CEL-10_1	Nhs1	2150	10	1478 M & F		0.0020208	2.694487287
CEL-10_1	Nmbr	2150	10	1478 M & F		0.0020208	2.694487287
CEL-10_1	Hebp2	2150	10	1478 M & F		0.0020208	2.694487287
CEL-10_1	Adat2	2150	10	1478 M & F		0.0020208	2.694487287
CEL-10_1	Ppil4	2150	10	1478 M & F		0.0020208	2.694487287
rs3721803	Nmbr	2134	10	1478 M & F		1.095E-06	5.960620122
rs3721803	Adat2	2134	10	1478 M & F		1.095E-06	5.960620122
rs3721803	Ppil4	2134	10	1478 M & F		1.095E-06	5.960620122
rs3695003	Nmbr	2137	10	1478 M & F		1.095E-06	5.960620122
rs3695003	Adat2	2137	10	1478 M & F		1.095E-06	5.960620122
rs3695003	Ppil4	2137	10	1478 M & F		1.095E-06	5.960620122
rs4228112	Nmbr	2144	10	1478 M & F		0.0062594	2.203470054
rs4228112	Hebp2	2144	10	1478 M & F		0.0062594	2.203470054
rs4228112	Adat2	2144	10	1478 M & F		0.0062594	2.203470054
rs4228112	Ppil4	2144	10	1478 M & F		0.0062594	2.203470054
rs1348052	Hebp2	2154	10	1478 M & F		0.0072717	2.13836415
rs1348052	Adat2	2154	10	1478 M & F		0.0072717	2.13836415
rs1348052	Hebp2	2155	10	1478 M & F		0.0069567	2.157598644
rs1348052	Adat2	2155	10	1478 M & F		0.0069567	2.157598644
rs1345911	Hebp2	2159	10	1478 M & F		0.0091297	2.039541532
rs1345911	Adat2	2159	10	1478 M & F		0.0091297	2.039541532
rs6185923	Adat2	2136	10	1478 M & F		3.587E-06	5.445246636
rs6185923	Ppil4	2136	10	1478 M & F		3.587E-06	5.445246636
rs1348118	Stac2	2442	11	1478 M & F		0.028714	1.541905592

שלוש היפותלמוס:

## חלק 6: ניתוח סיבתיות

את מבחן הסיבתיות נבצע עבור 10 שלשות מתוך השלשות שבחרנו.

להלן השלשות:

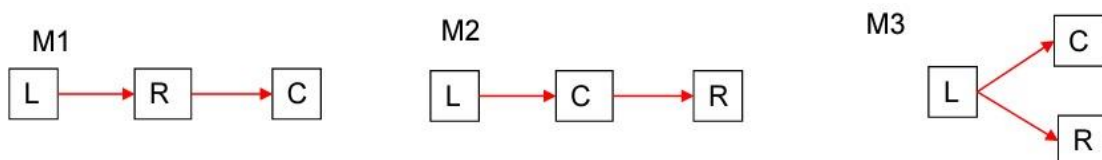
כבד:

היפותלמוס:

Locus	Gene	Phenotype
rs4228121	Nmbr	970
rs3721803	Nmbr	970
rs3721803	Adat2	970
rs3721803	Adat2	1224
rs3721803	Ppil4	1224
rs4228112	Adat2	1224
rs4228121	Ppil4	1478
rs3721803	Ppil4	1478
rs1345911	Adat2	1478
rs3664101	Nhsl1	1478

Locus	Gene	Phenotype
D10Mit28	Aig1	970
rs4228121	Aig1	970
rs1348052	Aig1	970
rs1348118	Med1	1224
rs3712394	Aig1	1224
rs1348052	Aig1	1224
rs1348118	Lsm12	1478
rs3664101	Aig1	1478
rs4228112	Aig1	1478
rs1348052	Aig1	1478

לכל שלשה נבחנו 3 המודלים שראינו בכיתה:



כאשר L – לוקוס, R – גן, C – תכונה מורכבת (=פנוטיפ).

נבחר את המודל הטוב ביותר באמצעות חישוב הנראות של כל מודל. במקרה שלנו חישובנו את  $\log$  הנראות, וכיוון שזו פונקציה עולה, המודל שעבורו "לוג" הנראות מקסימלי הוא אותו המודל שעבורו הנראות מקסימלית.

לאחר מכן נחשב את יחס הנראות. נניח בה"כ  $L(M_1) \geq L(M_2), L(M_3)$ .

$$LR = \frac{L(M_1)}{\max(L(M_2), L(M_3))} \rightarrow \log(LR) = \log(L(M_1)) - \max(\log(L(M_2)), \log(L(M_3)))$$

נגדיר את השערת האפס ואת האלטרנטיבה:

נניח כי  $L(M_i) \geq L(M_j, M_k), i, j, k \in \{1, 2, 3\}$

$$H_0: \log L(M_1) - \max(\log L(M_j), \log L(M_k)) \leq 0$$

$$H_1: \log L(M_1) - \max(\log L(M_j), \log L(M_k)) > 0$$

כדי להעריך את המובהקות של כל תוצאה נבצע מבחן פרמוטציות. במבחן נייצר 100 פרמוטציות ונחשב מחדש את כל עמודות הטבלה הנוספות. לאחר מכן נספור את הפרמוטציות שעבורן יחס הנראות (LR) גבוה מהערך שמצאנו ונחלק במספר הפרמוטציות. מטרת המבחן היא "לנתק" את הקשרים שעבורם אחד המודלים הוא הטוב ביותר (האלטרנטיבה מתקיימת). לצורך כך נבצע פרמוטציות של עמודות הגנים והפנוטיפים. כך השלשות האמיתיות ינותקו והמידע עצמו יהיה רנדומלי בפועל.

## סיכום התוצאות:

כבד:

Locus	Gene	Phenotype	Model	p_val
D10Mit28	Aig1	970	3	0
rs4228121	Aig1	970	3	0.37
rs1348052	Aig1	970	1	0.01
rs1348118	Med1	1224	3	0
rs3712394	Aig1	1224	1	0.01
rs1348052	Aig1	1224	1	0.04
rs1348118	Lsm12	1478	3	0.01
rs3664101	Aig1	1478	3	0.09
rs4228112	Aig1	1478	1	0.05
rs1348052	Aig1	1478	1	0.01

היפותלמוס:

Locus	Gene	Phenotype	Model	p_val
rs4228121	Nmbr	970	3	0.05
rs3721803	Nmbr	970	3	0
rs3721803	Adat2	970	3	0
rs3721803	Adat2	1224	3	0
rs3721803	Ppil4	1224	3	0
rs4228112	Adat2	1224	3	0
rs4228121	Ppil4	1478	3	0.13
rs3721803	Ppil4	1478	3	0
rs1345911	Adat2	1478	3	0
rs3664101	Nhs1	1478	3	0

ראשית נציין כי הערך  $P_{value} = 0$  הוא לא אפשרי תיאורטית. בפועל, תוצאה זו התקבלה עבור שלישיות שהערך שהתקבל עבורן הוא נמוך מאוד (כלומר קרוב ל-0), וכיוון שמספר הפרמוטציות מוגבל ל-100 לא נמצאה שלשה עם יחס נראות גבוה יותר מאשר היחס שהתקבל עבורן, ולכן ערך  $P_{value}$  שמופיע עבורן הוא 0. אילו היינו מבצעים פרמוטציות נוספות (מעל 10,000), היינו עשויים לקבל ערכים שאינם 0 במקרים אלה.

בחרנו להציג את עשר השלשות הנ"ל לכל רקמה כמדגם מייצג של תוצאות מבחן הסיביות עבור כל רקמה. נבחין כי עבור הכבד יש חלוקה כמעט שווה בין מודל 1 למודל 3, כלומר חלק מהתוצאות ניתן להסביר באמצעות קשר סיבתי ישיר בין הגנוטיפ ל-SNP ובין ה-SNP לפנוטיפ, ואת חלקן באמצעות קשר בין הגנוטיפ ל-SNP ובין הגנוטיפ לפנוטיפ, אך ללא קשר מוכח בין ה-SNP לפנוטיפ. לעומת זאת, עבור ההיפותלמוס עבור רוב מוחלט של השלשות מודל 3 היה "המוצלח" ביותר, כלומר ניתן להסביר את הרוב המוחלט באמצעות קשר ישיר בין הגנוטיפ ל-SNP ובין הגנוטיפ לפנוטיפ אך ללא קשר מוכח בין ה-SNP לפנוטיפ.

כמו כן נבחין כי אכן קיבלנו תוצאות מובהקות של קשר סיבתי תחת מודלים 1 ו-3 בין הגן Aig1 ובין הפנוטיפים שבחנו, וכן בין רוב מוחלט של הגנים שנבדקו בהיפותלמוס (שרובם בכרומוזום 10) ובין הפנוטיפים שנבדקו. ממצא זה תואם את התוצאות בחלק 4 ומהווה אישוש להשערה כי קיים קשר בין גנוטיפים בכרומוזום 10 ובין הפנוטיפים שנבדקו.