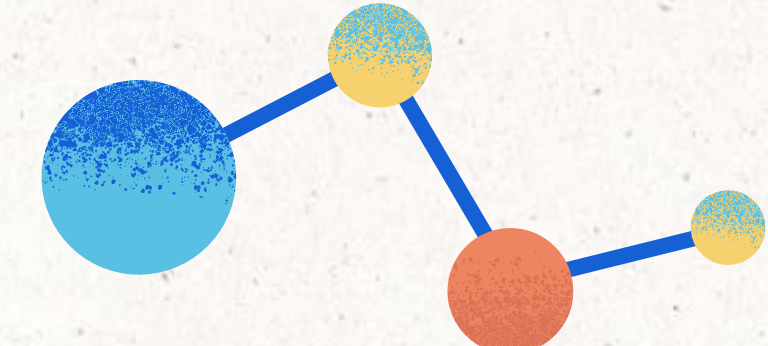
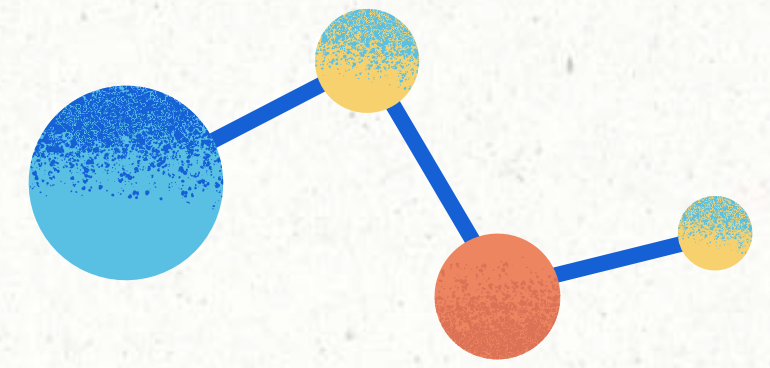




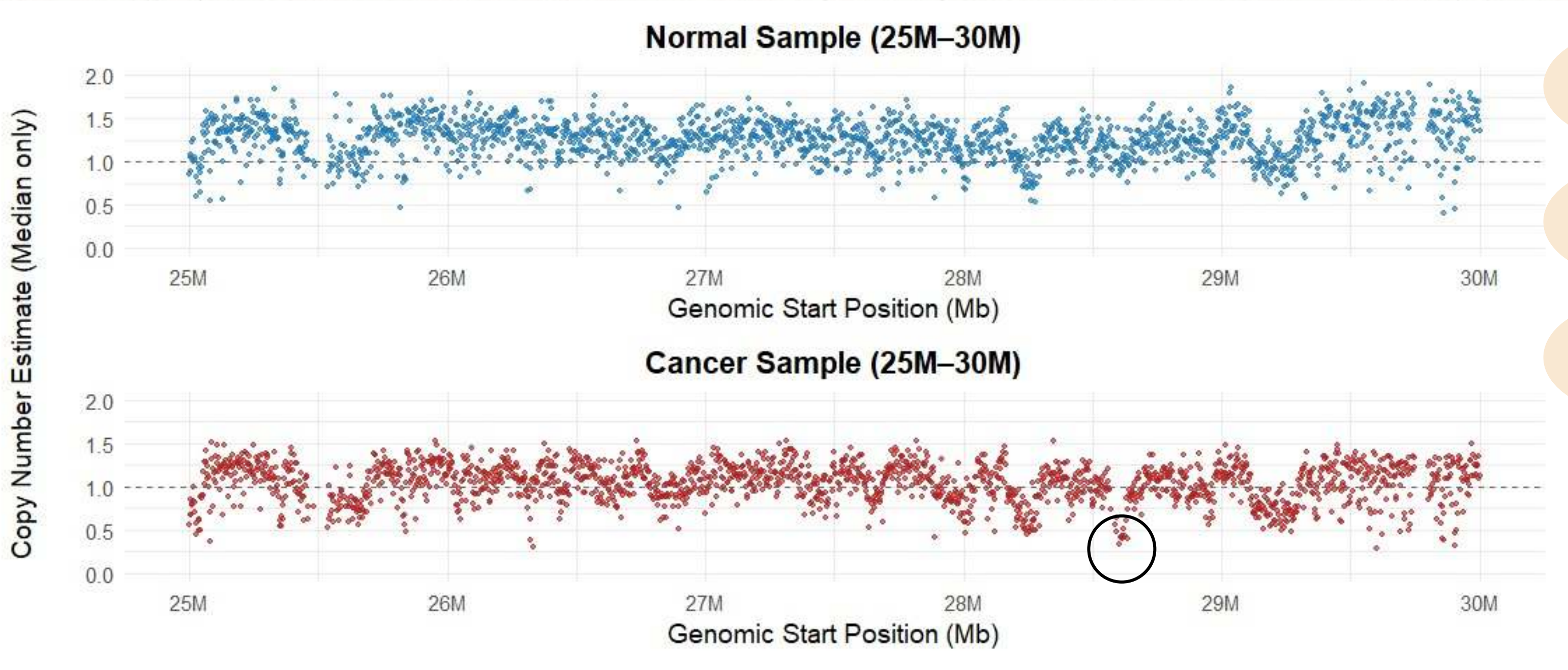
הצגה מסכמת

קבוצה B

מציגות: סלומה ברנס, עדן מלכה ועומר אבוחצירה



האם יש שינוי במספר העותקים סביב 28.5M בכרומוזום?



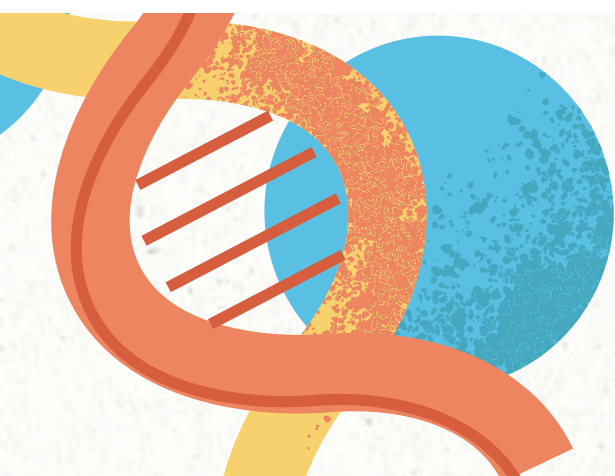
הבקשה

הנתונים

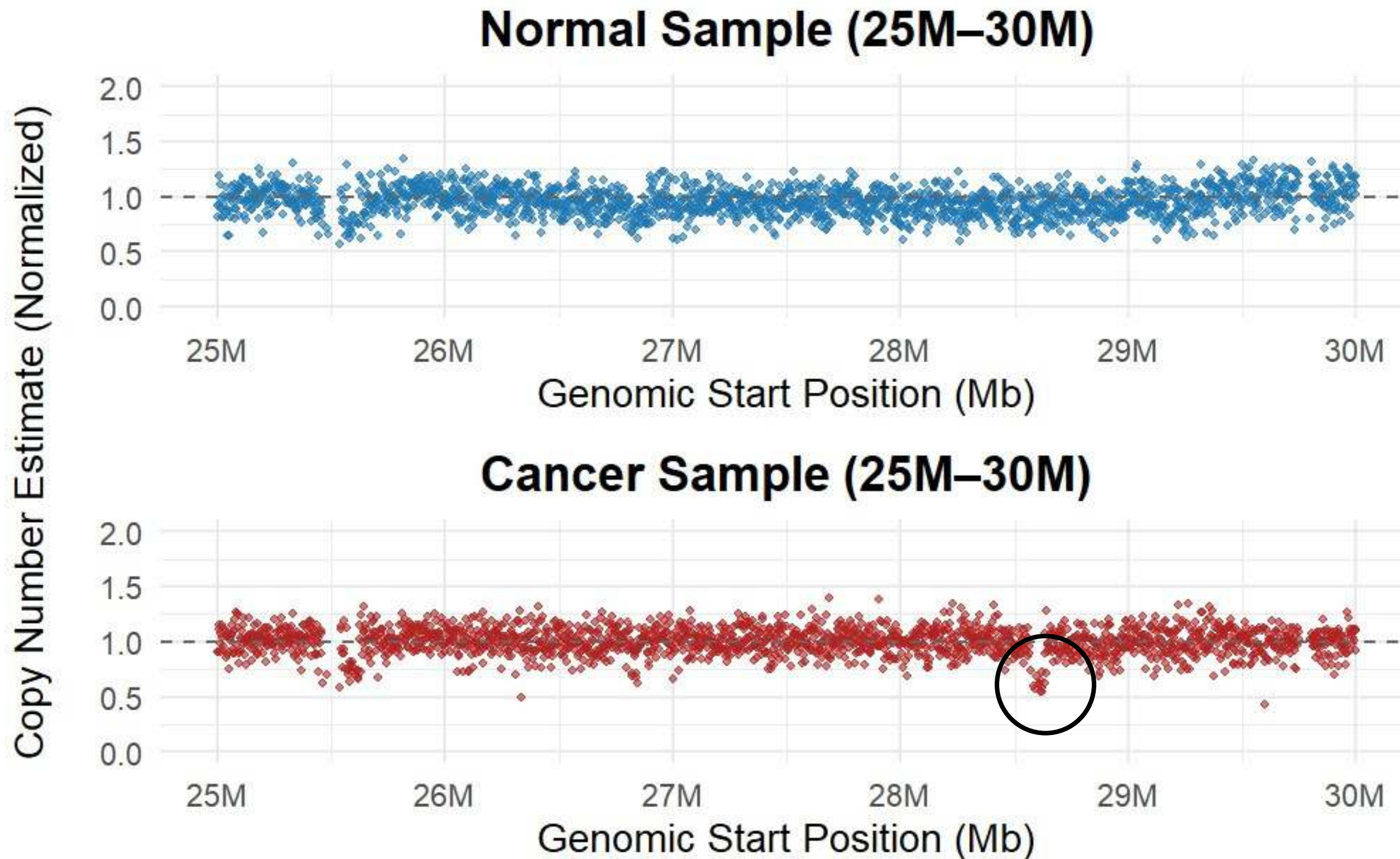
הבעיה



כולל תיקון חציון



האם יש שינוי במספר העותקים סביב 28.5M בכרומוזום?



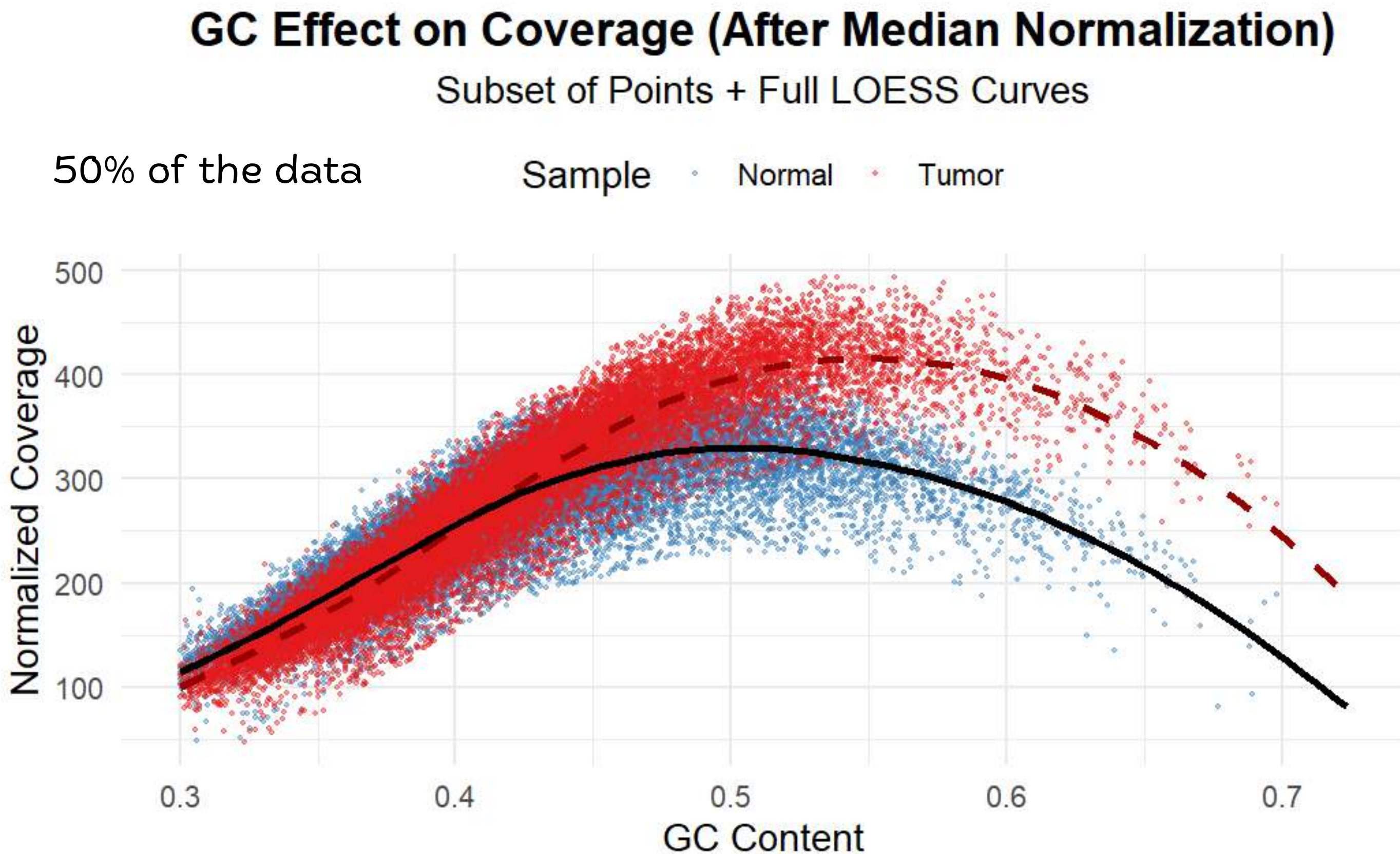
הבקשה

הנתונים

הבעיה

כולל תיקון חציון
ותיקון GC

דגימה בריאה למול סרטנית



המודל הסטטיסטי באמידת מס' העותקים

$$Y_k^t = a_k \cdot f^t(GC_k) \cdot \gamma_k \cdot \delta_k^t + \eta_k^{t,Pois}$$

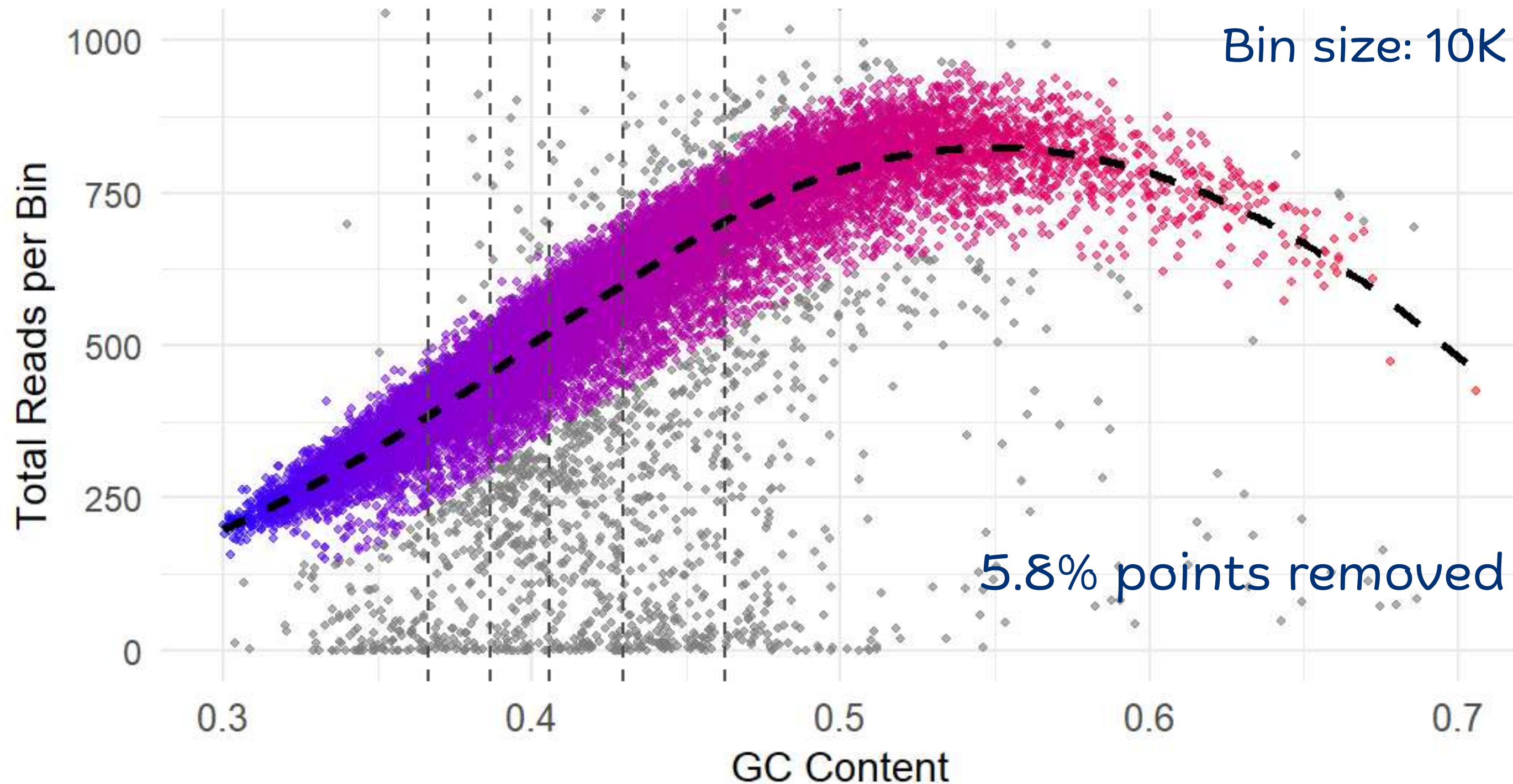
$$Y_k^n = f^n(GC_k) \cdot \gamma_k \cdot \delta_k^n + \eta_k^{n,Pois}$$

- Y_k^t, Y_k^n : observed read counts in bin k for tumor and normal samples
- $f^t(GC_k), f^n(GC_k)$: GC-bias correction functions (per sample)
- $a_k \in \{0.5, 1, 1.5, 2\}$ is the unknown copy number
- γ_k : shared signal across both samples
- δ_k^t, δ_k^n : sample-specific multiplicative effects
- $\eta_k^{t,Pois}, \eta_k^{n,Pois}$: Poisson noise

אומדן אפקט ה־GC ע"י רגרסיית SPLINE

Coverage vs GC Content

Cubic Spline with 5 Knots (Degree = 3), MAE = 35



Coverage < 5

LOESS ± 1.5 IQR

LOESS ± 3.25 IQR

Mean Average Error:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

בחירת אומדן למספר העותקים

Model B

$$\hat{a}_k = \frac{Y_k}{\hat{f}(GC_k)} + C$$

דגימה אחת עם תיקון להטיית GC בלבד

Model C

$$\hat{a}_k = \frac{Y_k^T}{Y_k^N} + C$$

יחס כיסויים בין דגימה סרטנית
לנורמלית : ללא תיקון GC

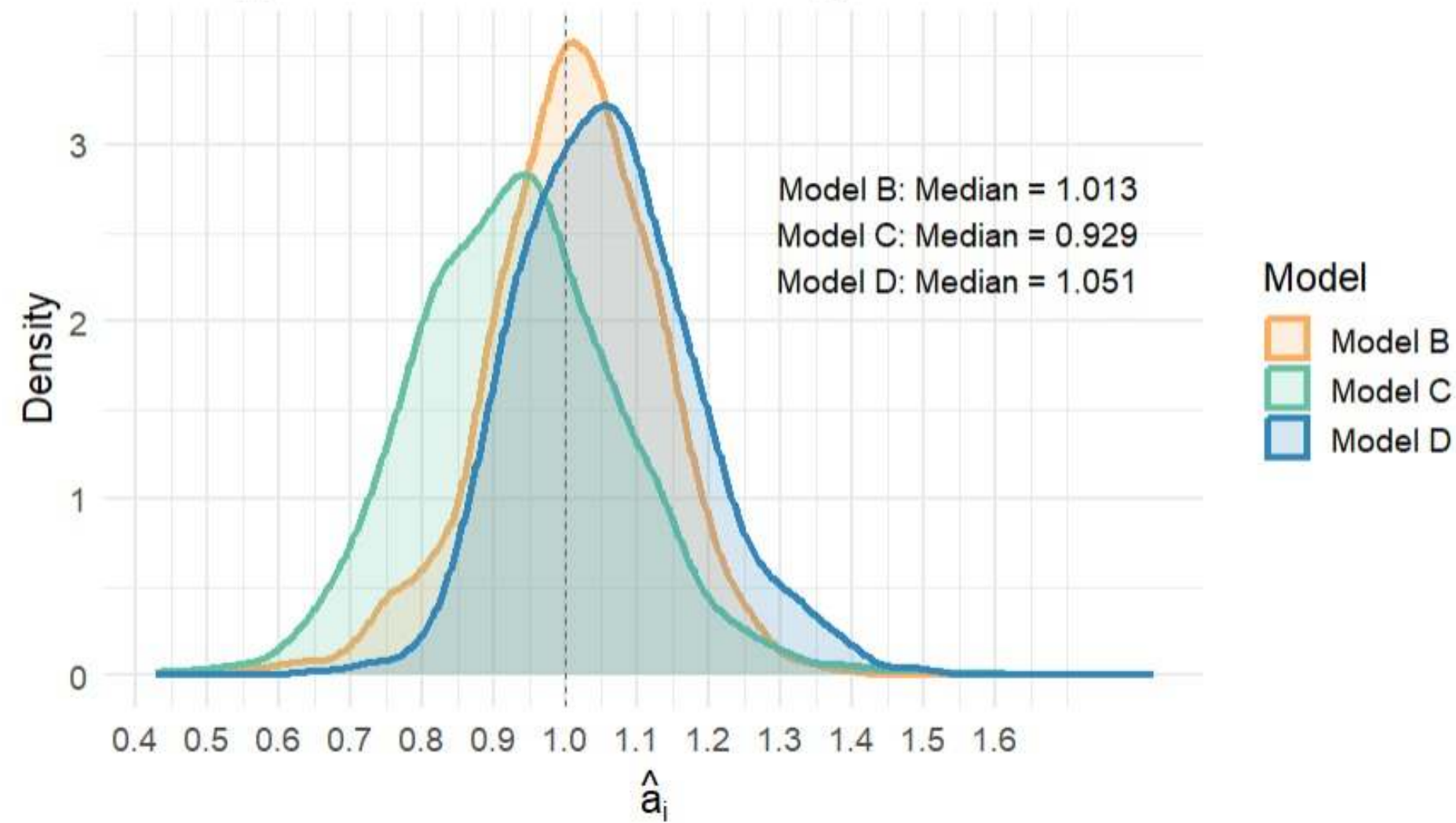
Model D

$$\hat{a}_k = \frac{\left(\frac{Y_k^{(T)}}{\hat{f}_{T(GC_k)}} + c \right)}{\left(\frac{Y_k^{(N)}}{\hat{f}_{N(GC_k)}} + c \right)}$$

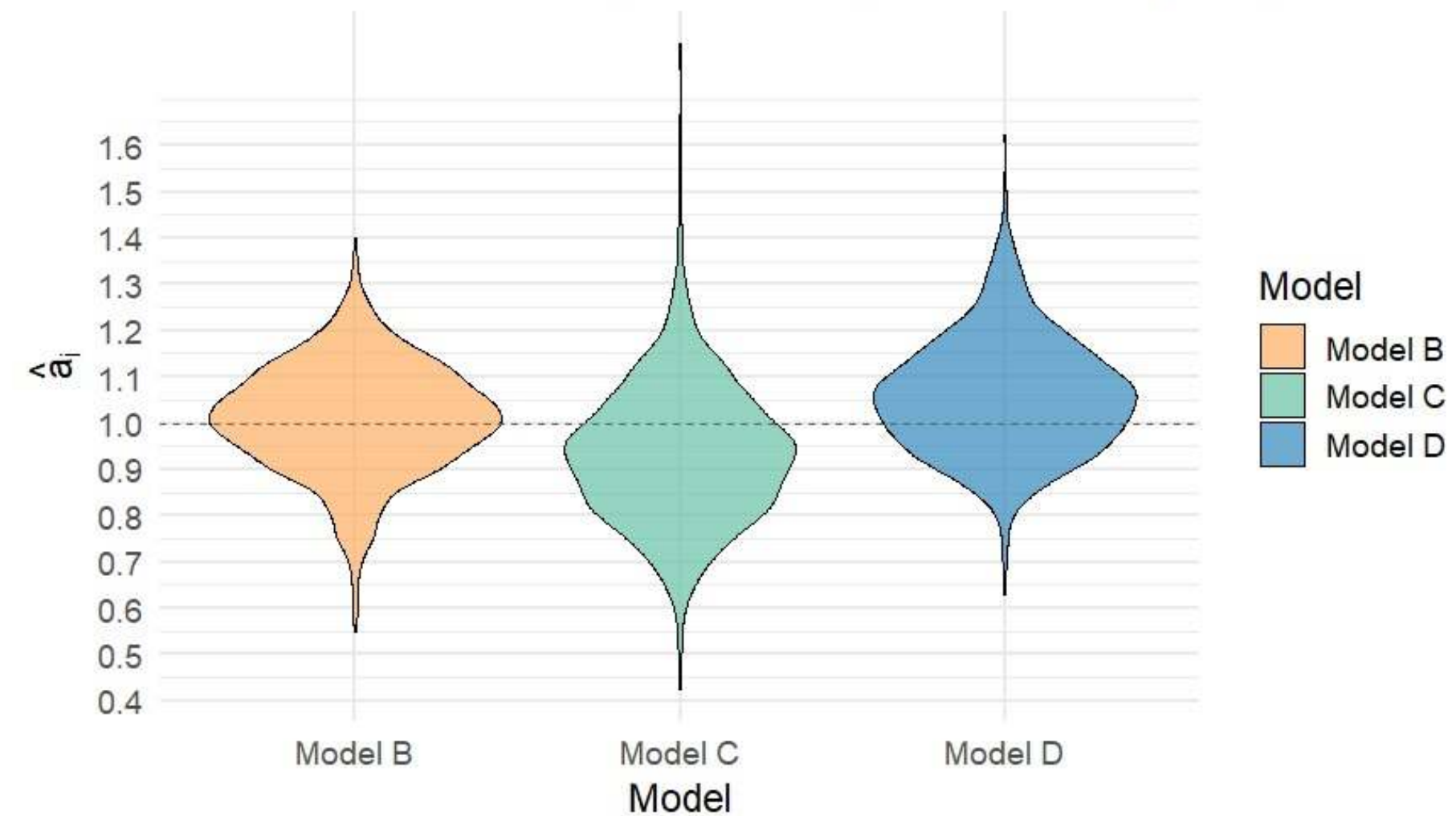
יחס מתוקן בין שתי הדגימות :
עם תיקון להטיית GC +
קבוע ייצוב $c = 0.1$

מדדי טיב התאמת המודל

Copy Number Estimation – Region 25M–30M



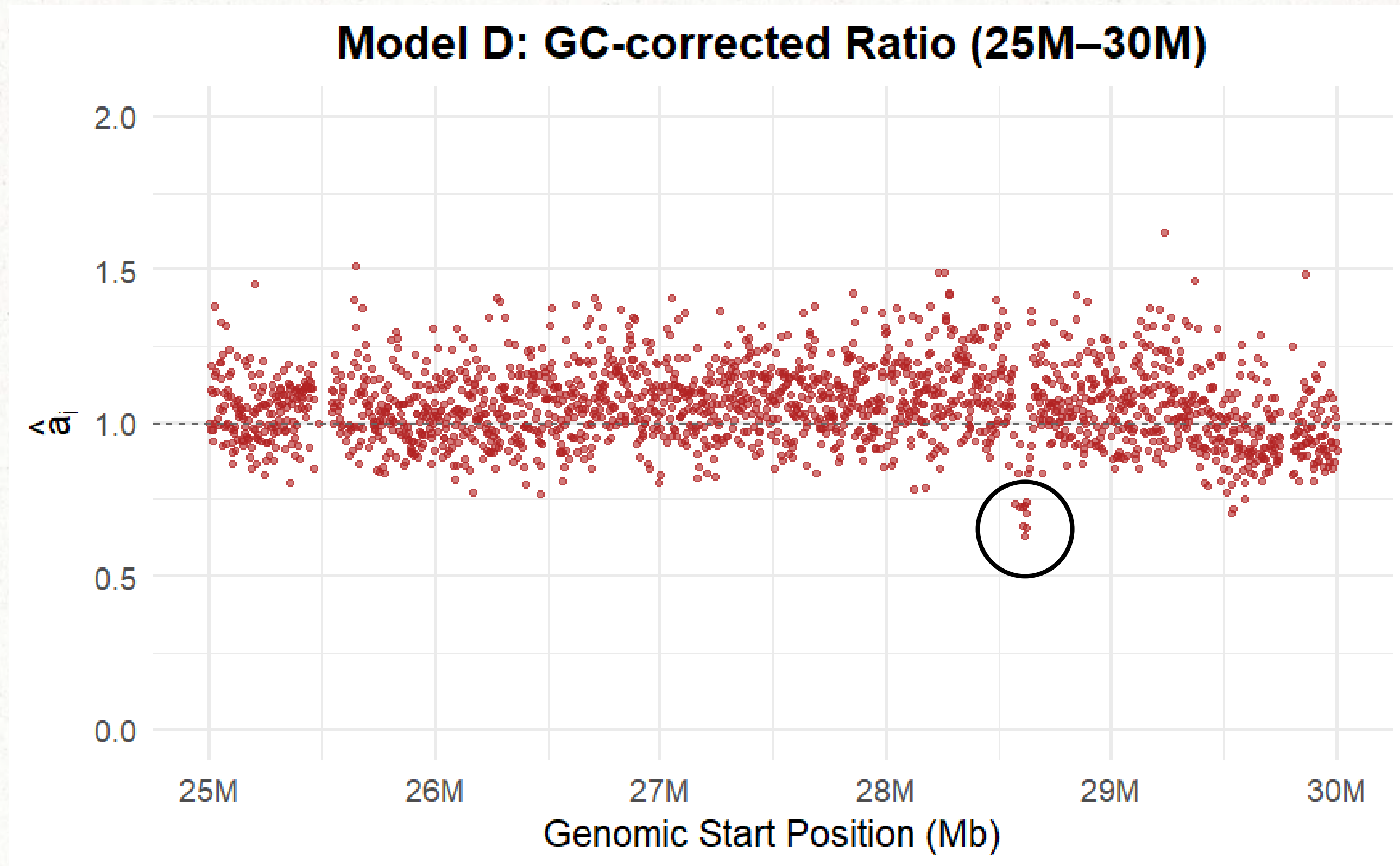
Violin Plot of Estimated $\hat{a}_i$ per Model (Combined Regions)



Model	$\hat{a}_i \in [0.9, 1.1]$
Model B	0.735
Model C	0.482
Model D	0.749

בדיקת פרופורציה באזורים
:26.5M–27M

הצגת האזור המתוקן



תשובה למתייעץ ומסקנות

נמצאו חריגות ברורות באזור 28.5M בסיסים לאחר תיקון מלא.



אפקטים מקומיים של GC מקשים על גילוי CNV.



מודל D סיפק את האומדן היציב והמדויק ביותר.

