

生成AIを活用したTCGAデータ処理 膀胱がん予後関連遺伝子の探索

上條千太

北海道大学 大学院 医学研究院 腎泌尿器外科学教室
北海道大学 遺伝子病制御研究所 がん制御学分野

【背景①】

泌尿器科医として数年間勤務した後、筋層浸潤膀胱がんの新規治療法開発に関する研究を志し大学院に進学した。現在、遺伝子病制御研究所 がん制御学分野 園下研究室で研究を行っている。

プログラミングに興味はあったものの、Web上の無料講座で基礎を学習した程度であった。

「R7年度データハンドリング演習 生成AIを活用したプログラミングの基礎」を受講し、TCGAデータベースで、膀胱がんの予後に関連する遺伝子を探索するRコード作成に取り組んだため、その結果を報告する。

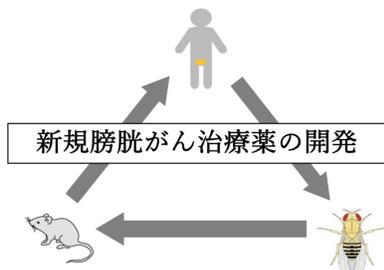
【背景②】

がん制御学分野 園下研究室では、当研究室では、細胞やマウス実験に加え

ショウジョウバエを相補的に用い、新規治療標的遺伝子の探索を行っている。

膀胱がん遺伝子型モデルハエと1系統あたり20-90遺伝子が欠失している遺伝子欠失システムを活用し、膀胱がん遺伝子型モデルハエの生存率に影響を与えた遺伝子群を同定した。

同定した遺伝子のヒトオルソログ（相同遺伝子）と、膀胱がん患者の予後の関係を調査した。



【対象と方法】

TCGA-BLCAコホートのpT2-4膀胱がん369例
臨床情報、生存情報、mRNA発現データ

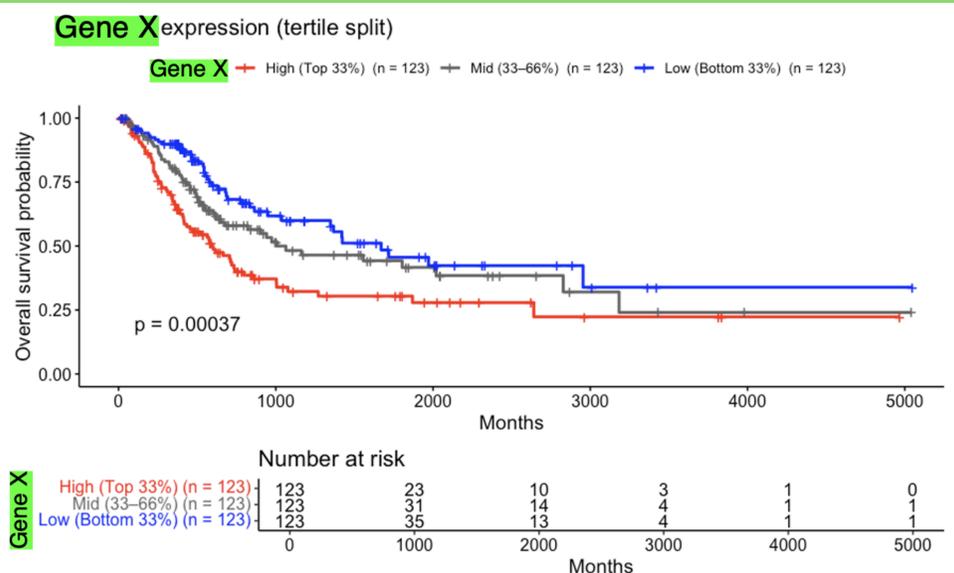
ハエ実験で同定された215遺伝子のヒトオルソログ

各候補遺伝子の発現量に基づいて症例を3群（high, middle, low）に分類
発現量に基づき症例を3群（High：上位33%、Middle：中間33%、Low：下位33%）に分類した。
全生存期間（OS）はKaplan-Meier法により解析、
群間比較にはlog-rank検定を用いた
p < 0.05を統計学的有意差ありとした

解析に用いたRコードはChatGPT（OpenAI, San Francisco, CA, USA; GPT-5 model）を用いて作成し、著者が内容を確認・修正した。

【結果】

遺伝子名を入力すると、mRNA発現量に基づき3群に分類したKaplan-Meier曲線を作成するコードを作成した。
plot_km("GeneX", split = "tertile")



このコードを活用し、215遺伝子のヒトオルソログを3群に分け解析したところ、Low vs Middle、またはMiddle vs High で有意差を認める25遺伝子を同定した。

今回の検討で同定した25遺伝子について、ヒト遺伝子-ハエ遺伝子間の相同性の高さ、新規性、阻害剤の有無を検討し、細胞実験・マウス実験へと進める遺伝子を決定するプロセスを現在進めている。

【結語】

プログラミング初心者が「生成AIを活用したプログラミングの基礎」を受講し、ChatGPTとR studioを活用し、TCGAデータの解析を行った。

当研究室では、モデルの生存率を、羽化した個体と羽化しなかった個体の割合を手作業でカウントして算出している。今後は画像解析を用いて生存率を自動算出するツールの開発を目指す。



$$\text{生存率 (\%)} = \frac{\text{羽化した蛹 (○)} }{\text{すべての蛹 (○+●)} } \times 100$$



本研究は、
JST 次世代研究者挑戦的研究プログラム
JPMJSP2119の支援を受けたものです。

