

تمرین سوم درس مبانی بیوانفورماتیک

دکتر فاطمه زارع میرک آباد

نیمسال اول سال تحصیلی ۱۴۰۳-۱۴۰۴

توضیح تمرین:

هدف از این تمرین آشنایی با تحلیل داده‌های بیان ژن با استفاده از ابزارهای GEO2R و Enrichr است. دانشجویان باید با استفاده از داده‌های موجود در پایگاه داده GEO، به شناسایی ژن‌های معنادار و تحلیل مسیرهای زیستی مرتبط بپردازند.

مراحل انجام تمرین:

بخش اول: تحلیل بیان ژن تفاضلی^۱ با استفاده از GEO2R

وارد پایگاه داده GEO شوید و دیتاست **GSE15852** را پیدا کنید. این دیتاست شامل ۴۳ نمونه سرطان سینه و بافت‌های سالم متناظر آن‌هاست. با استفاده از ابزار GEO2R نمونه‌های سرطانی را با سالم مقایسه کنید.

۱. توضیح دهید در بیان ژن تفاضلی مشکل استفاده از **p-value** چیست؟ آزمایشی طراحی کنید که نشان دهد احتمال رخ دادن خطای نوع اول با افزایش تعداد آزمون‌ها چگونه تغییر می‌کند. درباره روش‌های اصلاح **p-value** تحقیق کنید و دو روش **Benjamini-Hochberg (BH)** و **Bonferroni** را توضیح دهید و تفاوت آن‌ها را بیان کنید.

۲. ژن‌هایی را معنادار در نظر بگیرید که $\log FC > 1$ و $\text{adj } p\text{-value} < 0.001$ باشد. سپس نمودارهای **Volcano plot** و **Mean difference (MD) plot** را رسم کنید و هر کدام را توضیح دهید. ژن‌های معنا دار را ذخیره کنید. از آن‌ها برای بخش دوم تمرین استفاده خواهیم کرد.

۳. در انتها، نمونه‌ها را براساس **Grade** به چهار گروه تقسیم کنید. سپس مجدد آنالیز را انجام دهید. نمودارهای **UMAP**، **Venn diagram** و **Boxplot** را رسم کنید و آن‌ها را تحلیل کنید.

بخش دوم: غنی‌سازی مسیرهای زیستی^۲

در این بخش، ژن‌های معنادار شناسایی شده در مرحله قبل را با استفاده از ابزار **Enrichr** از نظر مسیرهای زیستی و عملکردهای ژنی تحلیل کنید. برای این منظور از پایگاه داده‌های **KEGG** و **GO** استفاده کنید.

۱. تحلیل مسیرهای زیستی در **KEGG**:

لیست ژن‌های معنادار را در ابزار **Enrichr** وارد کنید و مسیرهای زیستی معنادار در پایگاه داده **KEGG** را شناسایی کنید. توضیح دهید که این مسیرهای زیستی چه نقشی دارند و چگونه با پیشرفت یا شکل‌گیری سرطان مرتبط هستند. نتایج را با جداول و نمودارهای مناسب ارائه دهید و نکات برجسته هر مسیر زیستی را توضیح دهید.

¹ Differential gene expression (DGE)

² Pathway Enrichment

۲. تحلیل عملکرد ژن‌ها در GO :

دیتاست GO را بررسی کنید و ژن‌های معنادار را در سه دسته زیر تحلیل کنید. نتایج را با استفاده از نمودار و جداول مناسب نشان دهید.

- فرایندهای زیستی (Biological Process): بررسی کنید کدام فرایندهای زیستی معنادار هستند و نقش آن‌ها در سرطان را توضیح دهید.
- اجزای سلولی (Cellular Component): مشخص کنید ژن‌های معنادار در کدام بخش‌های سلول فعال هستند و اهمیت این بخش‌ها را توضیح دهید.
- عملکردهای مولکولی (Molecular Function): عملکردهای مولکولی معنادار را بررسی کرده و ارتباط آن‌ها با فرایندهای سرطان‌زا را توضیح دهید.

نکات تکمیلی:

- گزارش خود را در قالب فایل PDF تا تاریخ ۲۸ دی‌ماه در سامانه آپلود کنید.
- لینک‌های مفید:

<https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html> ○

[/https://maayanlab.cloud/Enrichr](https://maayanlab.cloud/Enrichr) ○

<https://www.ncbi.nlm.nih.gov/geo> ○

<https://geneontology.org> ○

<https://www.genome.jp/kegg/pathway.html> ○