

به نام خدا



# تحلیل دادگان ریزآرایی لوکمی حاد مغز استخوان

مقدمه‌ای بر بیوانفورماتیک

پروژه‌ی نهایی

گروه ۲

دکتر علی شریفی زارچی

دکتر سمیه کوهی

گردآورنده: مهدی غزنوی

هدف از این پروژه، تجربه‌ای اولیه در تحلیل داده‌های بیوانفورماتیک می‌باشد. شما بایستی داده‌های ریزآرایه<sup>۱</sup> مربوط به سرطال خون (لوکمیا) را تحلیل کنید. صفحه‌ی مربوط به دادگان و سایر اطلاعات مجموعه داده را می‌توانید در [این لینک](#) مشاهده کنید.

لوکمیا انواع مختلفی دارد. یکی از انواع این سرطان، لوکمی حاد مغز استخوان<sup>۲</sup> (AML) است. شما بایستی با تحلیل دادگانی که در اختیار دارید، ژن‌هایی را که در این نوع سرطان نقش موثرتری دارند به دست آورید. به این منظور در مجموعه دادگان، داده‌هایی را که Phenotype آن‌ها Normal است را گروه نرمال و نمونه‌هایی را که Source name آن‌ها AML patient است را گروه تست در نظر بگیرید.

موارد زیر از شما خواسته می‌شود:

#### • کنترل کیفیت داده (۱۵ نمره)

ممکن است داده‌های اولیه آماده‌ی تحلیل نباشند و نیاز به تغییراتی داشته باشند. در صورت لزوم، تغییراتی بر روی دادگان بدهید. نمودار دادگان را پیش و پس از تغییراتی که اعمال می‌کنید نمایش دهید. تغییرات مذکور می‌توانند موارد مختلفی (مثلاً نرمال‌سازی در صورت نیاز) باشند.

#### • کاهش ابعاد داده (۱۵ نمره)

برخی ژن‌ها در نمونه‌ها میزان بیان تقریباً برابری دارند. در این قسمت با استفاده از روش کاهش ابعاد، تاثیر ژن‌هایی را که میزان بیان تقریباً برابر در نمونه‌ها دارند را کم کنید. داده‌هایی با بعد کمتر به دست آوردید.

#### • بررسی همبستگی بین نمونه‌ها (۱۵ نمره)

Source name در نمونه‌های نرمال متفاوت است. پس از کاهش ابعاد، رابطه‌ی همبستگی بین نمونه‌ها را براساس Source name بررسی کرده و خروجی را به شکل یک نمودار نمایش دهید.

#### • بررسی تمایز در بیان ژن‌ها (۲۵ نمره)

دسته نمونه‌های سالمی را که میزان همبستگی‌شان با نمونه‌ی تست بیشتر است را بیابید. سپس میزان بیان ژن‌ها در دسته‌ی مذکور را با میزان بیان ژن‌ها در افراد بیمار مقایسه کنید. برخی از این ژن‌ها در افراد بیمار به صورت معنی‌داری بیان کمتری دارند و برخی دیگر به صورت معنی‌دار بیان بیشتری دارند. ژن‌های هر دسته را گزارش کنید (سطح معنی داری را ۰,۰۵ در نظر بگیرید).

#### • آنالیز gene anthology و pathwayها (امتیازی، ۲۵ نمره)

برای هر کدام از دسته ژن‌هایی که افزایش یا کاهش بیان در افراد بیمار دارند، gene anthology و pathway مرتبط را گزارش کنید. (برای این کار می‌توانید از Enrichr کمک بگیرید).

#### • موارد امتیازی دیگر (موارد این قسمت بسته به میزان و کیفیت توضیحات ارائه شده، تا سقف ۳۵ نمره امتیازی دارند).

- توضیح دهید موارد فوق چگونه در AML تاثیرگذارند.
- نتایج خود را با مقالات مرتبط در این زمینه مقایسه کنید.
- دیگر داده‌های مرتبط با این بیماری را تحلیل کنید.
- هر موردی که به نظرتان در بررسی AML مفید و جذاب است ذکر و تحلیل کنید.

#### نکات پایانی

- گزارش پایانی ۳۰ نمره دارد. نتایج و تحلیل‌هایتان را به طور کامل توضیح دهید.
- موارد ارسالی بایستی شامل کدهای زده شده، گزارش، نمودارها و نتایج نهایی باشند.

<sup>1</sup> Microarray

<sup>2</sup> Leukemia

<sup>3</sup> Acute Myeloid Leukemia

- از کتابخانه‌ی GEOquery می‌توانید برای خواندن داده‌گان استفاده کنید.
- برای انجام پروژه می‌توانید از [فیلم‌های درس بیوانفورماتیک پیشرفته در مکتب‌خونه](#) (جلسات ۹ تا ۱۲ در تحلیل داده‌های ریزآرایه) استفاده کنید.
- سوالات خود را در صفحه‌ی پیازای درس مطرح کنید.
- مهلت تحویل پروژه ساعت ۲۳:۵۹ روز ۱۵ بهمن ماه است.

موفق باشید 😊