

به نام خدا



مقدمه ای بر بیوانفورماتیک

نیمسال دوم ۹۷-۹۸

دانشکده مهندسی کامپیوتر

دانشگاه صنعتی شریف

دکتر علی شریفی زارچی
دکتر سمیه کوهی
پروژه‌ی نهایی درس
آنالیز داده‌های میکرواری
عذری راستی

مدرسین

موضوع

مبحث

گردآورنده

لوکمیا^۱ یا سرطان خون، یکی از انواع سرطان است که معمولاً از مغز استخوان آغاز می‌شود و تعداد زیادی سلول خونی غیرعادی و نابالغ تولید می‌کند. این سلول‌های خونی به طور کامل تکامل نیافته‌اند و به آنها بلاست (blast) و یا سلول‌های لوکمی گفته می‌شود. علت بروز این سرطان هنوز ناشناخته است اما ترکیبی از عوامل ژنتیکی و محیطی (غیر ارثی) به عنوان عوامل موثر در نظر گرفته می‌شوند.

لوکمیا دارای انواع مختلفی است که یکی از آن‌ها لوکمی حاد مغز استخوان^۲ یا به اختصار AML است.^۳ AML دومین سرطان خون شایع در کودکان است. معمولاً در اثر بروز جهشی در دی‌ان‌ای سلول‌های پیش‌ساز خون که از تمایز کامل این سلول‌ها جلوگیری می‌کند و جهشی دیگر که موجب تقسیم و تکثیر غیرقابل کنترل سلول‌ها می‌شود، رخ می‌دهد.

در سال‌های اخیر از آنالیز داده‌های میکروآرای^۴ برای تشخیص این بیماری و بیماری‌های مشابه که در اثر بروز جهش و تغییر در بیان ژن‌ها به وجود می‌آیند استفاده می‌شود. در این پروژه شما قرار است از داده‌های GSE48558 که شامل تعدادی نمونه‌ی سرطانی و تعدادی نمونه‌ی سالم است استفاده کنید. صفحه مربوط به این داده در این [لینک](#) قرار دارد. مطالعات انجام شده در زمینه این داده در مقاله با PMID به شماره ۲۳۸۳۶۵۶۰ موجود است.

هدف این پروژه آن است که شما این داده را تحلیل کنید و با تمرکز بر روی تحلیل داده‌های میکروآرای، ژن‌هایی که در بروز این نوع سرطان نقش موثری دارند را پیدا کنید.

در این زمینه، شما می‌توانید از آنچه در فیلم‌های آموزشی کارگاه‌ها آموخته‌اید و هرگونه تکنیک دیگری که برای تحلیل این نوع داده‌های زیستی مناسب می‌دانید استفاده کنید. در این مطالعه دو گروه داده سالم و سرطانی وجود دارد که تفاوت‌های آنها از هر جهت می‌تواند مورد توجه باشد. می‌توانید نمونه‌هایی که source name آنها AML patient است را گروه بیمار و نمونه‌های با عنوان normal را گروه سالم در نظر بگیرید. بعد از تحلیل داده باید نتایج خود را از لحاظ زیستی تحلیل کنید. سعی کنید ارتباطی بین ژن‌ها و هدف مطالعه بیابید. در این راه، می‌توانید از مقالات و داده‌هایی که مرتبط با این نوع سرطان هستند نیز استفاده کنید. سعی کنید ارتباط یافته‌های خود و نتایج مقالات به روز این حیطه را پیدا کنید.

¹ Leukemia

² Acute myeloid leukemia

³ [Wikipedia](#)

⁴ microarray

به طور کلی تحلیل شما باید موارد زیر را شامل شود:

- کنترل کیفیت داده (۱۰ نمره)
- کاهش ابعاد داده (۱۰ نمره)
- بررسی همبستگی بین نمونه‌ها (۱۰ نمره)
- بررسی تمایز در بیان ژن‌های نمونه‌ها (۲۰ نمره)
- آنالیز gene anthology و pathwayها (۲۰ نمره)

بررسی تفاوت‌ها بین زیرگروه‌های داده، تحلیل داده‌های دیگر مرتبط با همین بیماری و مقایسه‌ی نتایج خود با سایر مقالات زیستی، دارای نمره اضافه خواهد بود. در صورت آوردن موارد بیشتر هم نمره اضافی به ازای هر مورد به شما تعلق میگیرد.

بعد از تحلیل داده‌ها گزارشی کامل از روش‌ها، تحلیل‌ها و نتایج خود ارائه دهید. گزارش شما حتما باید بخش مباحثات آینده داشته باشد. سعی کنید چالش‌ها را پیدا کنید و راهکار ارائه کنید. آوردن نمودارهای مرتبط در گزارش و استفاده از نکات گزارش نویسی نیز دارای اهمیت می‌باشد (نمره گزارش: ۳۰).