به نام خدا



گزارش فاز اول پروژه مقدمه‌ای بر بیوانفورماتیک

استاد

**دکتر شریفی و کوهی**

نویسندگان

**امیرحسین عابدی، امیرمحمد ایمانی، سیاوش رحیمی شاطرانلو**

**( به ترتیب الفبا )**

محتواها

[1 Micro-Array 2](#_Toc121828900)

[2 تغییرات بر روی داده‌ها 2](#_Toc121828901)

[3 کاهش ابعاد 2](#_Toc121828902)

[4 تحلیل نمونه‌ها 3](#_Toc121828903)

# Micro-Array

در ابتدا لازم است درمورد داده‌های بدست آمده از micro-array توضیح دهیم.

Micro-array یک لوازم آزمایشگاهی است که با استفاده از آن میتوانیم میزان بیان تعداد زیادی از ژن‌ها را برای یک DNA خاص مشخص کنیم. این لوازم صفحاتی دارند که با قرار گرفتن cDNA بر روی این صفحات مقدار بیان یک ژن در آنها مشخص میشود.

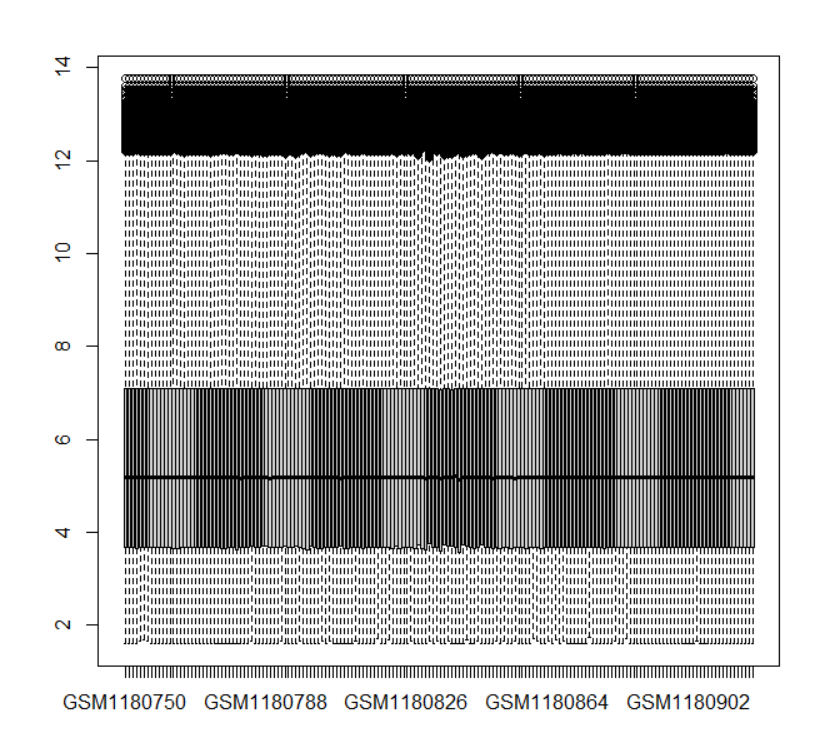
اینکار را برای چند سمپل انجام میدهیم. به طور مثال انسان‌هایی که سالم هستند و ( در این مسئله ) نمونه‌هایی که سرطان AML دارند.

داده خروجی آن که قرار است در ادامه بر روی آن کار کنیم به صورت یک ماتریس است که آنرا در شکل زیر میبینیم :



# بررسی کیفیت داده‌ها

در ابتدا چندمعیار برای درست بودن داده‌هایی که در دست داریم چک میکنیم. در ابتدا میتوانیم از خروجی boxplot برای این داده‌ها استفاده کنیم.



همانطور که مشاهده میشود خط مشکی وسط برای تمامی نمونه‌ها یکی میباشد. یعنی میانه میزان بیان ژن‌ها برای تمامی آنها به طور تقریبی برابر میشود. این یکی از معیارهایی است که به ما نشان میدهد که داده‌هایی که در دست داریم داده‌های خوبی هستند و نیاز به Augment کردن آنها نیست. در صورتی که این اتفاق نمی‌افتد میتوانستیم نتیجه بگیریم که داده‌هایی که بدست آوردیم میتواند اشتباه باشد و DNA هایی که داریم مربوط به نمونه‌های ما نیستند.

یکی دیگر از روش‌هایی که میتوانیم از صحت داده‌ها ( و تست ) مطمئن شویم انجام آزمون‌های فرض بر روی داده‌ها است. توجه کنید که از بین ژن هایی که میزان بیان آنهارا داریم حداقل میزان بیان یکی از آنها باید بر روی نمونه‌های مختلف اثر گذاشته باشد. اگر اینطور نباشد داده‌ها درست نیستند و جایی در روند استخراج داده‌ها اشتباهی رخ داده است.

# کاهش ابعاد

یکی از روش‌های دیگری که میتوانیم از صحت تست خود اطمینان حاصل کنیم این است که داده‌های خود را به بعد 2 ببریم و پراکندگی آنهارا چک کنیم. تستی که درست انجام شده باشد باید به درستی نمونه‌های مختلف را از هم جدا کند. در اینجا از روش های ذکر شده در شرح پروژه استفاده میکنیم.

# تحلیل نمونه‌ها

شکسمنیبتکمشنستکیمبتشسکیب