

بخش ۱.

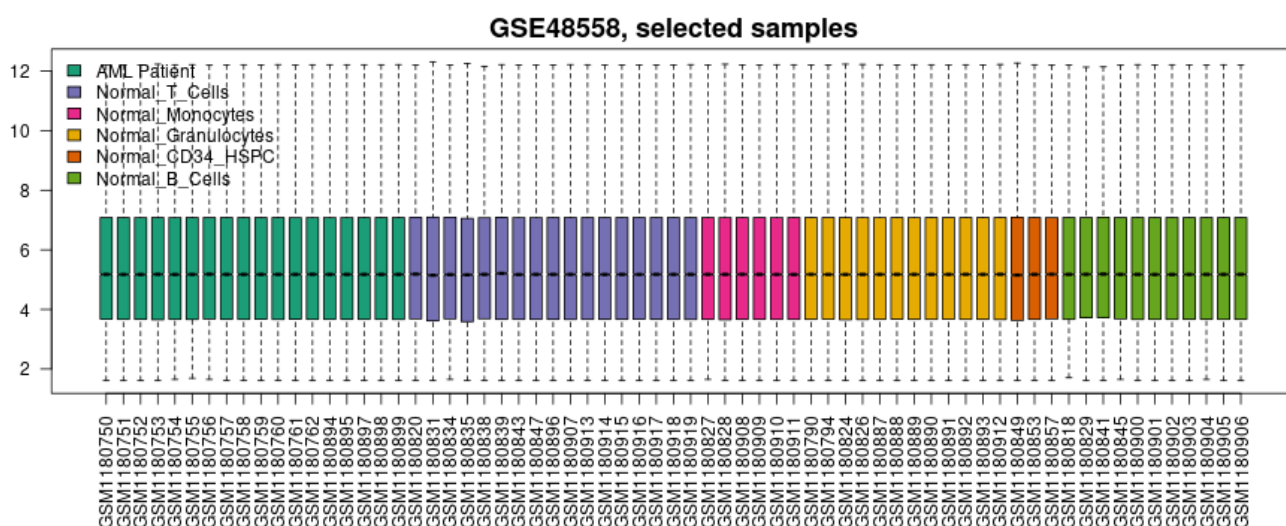
کنترل کیفیت داده

گام ۱ - Importing Libraries

ابتدا کتابخانه‌های لازم را import می‌کنیم. این قسمت کپی شده‌ی خود سایت ncbi است که یک کد R به صورت ابتدایی قرار داده است.

گام ۲ - Analysis

همانطور که در صورت پروژه گفته شده می‌بایست داده‌های Normal را به عنوان پایه و AML Patient را تست در نظر بگیریم. در این بخش، با کمک برچسب‌گذاری انجام شده در دیتاست، این دو گروه را جدا می‌کنیم. به کمک خود سایت از قسمت داده‌های خود سایت به خوبی نرمالایز شده‌اند. با توجه به نمودار زیر می‌توان دید که میانگین‌ها به اندازه‌ی کافی نزدیک به هم هستند. البته برای آنکه نرمال‌های متفاوت را داشته باشیم، بر اساس جفت Normal و Source name آن‌ها را جدا می‌کنیم.



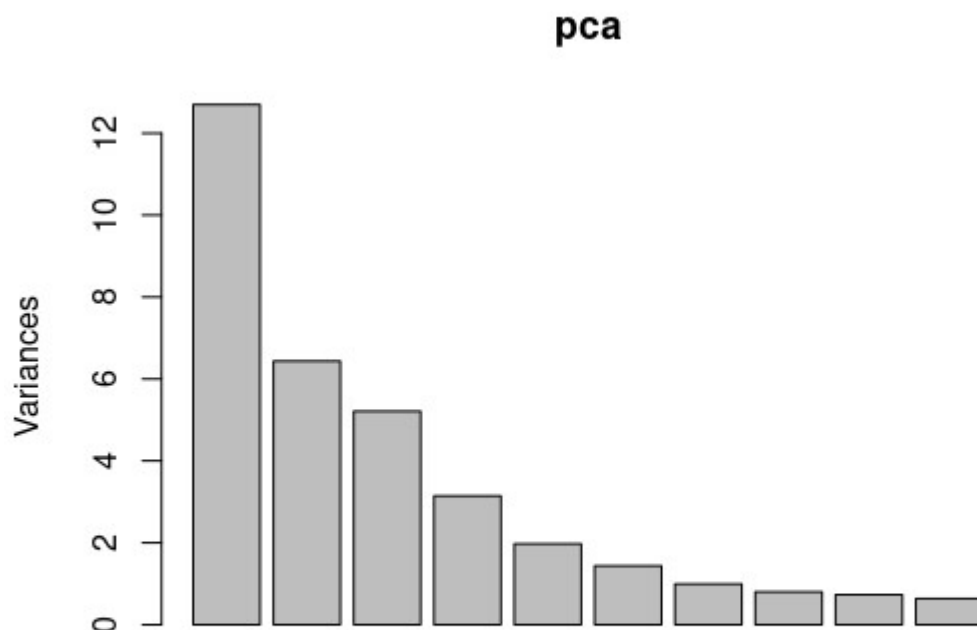
بخش ۲.

کاهش ابعاد داده

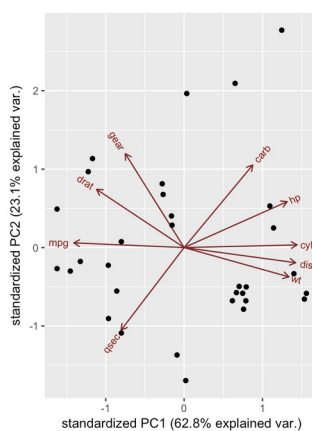
گام ۳ - Dimension reduction

برای کاهش بعد دو راه PCA و t-SNE وجود دارد. در اینجا من PCA را استفاده کرده‌ام.

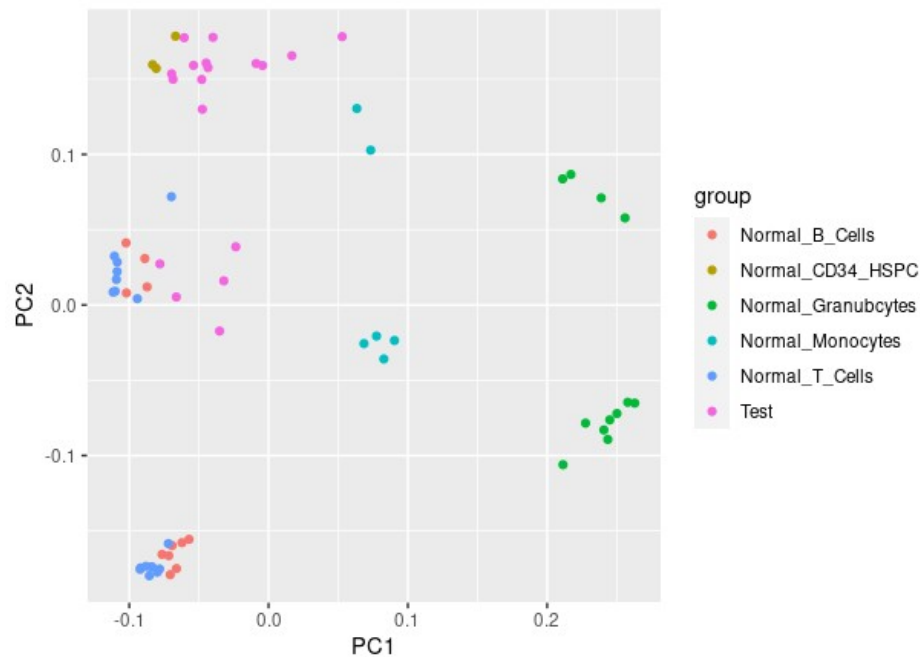
حالتی که PCA را بررسی می‌کنیم می‌بینیم که یک سری از ژن‌ها بیان بالایی دارند و باقی خیر. به صورت کلی، ژن‌هایی که مسئول بقای سلول هستند به صورت کلی بیانشان بالاست. یعنی به هر حال این ژن‌ها متمایزکننده‌ی خوبی نیستند. یک راه این است که داده‌ها را مقیاس کنیم، سپس میانگین را صفر کنیم.



از طرفی ترسیم خروجی PCA اساس چیزی شبیه به این است:



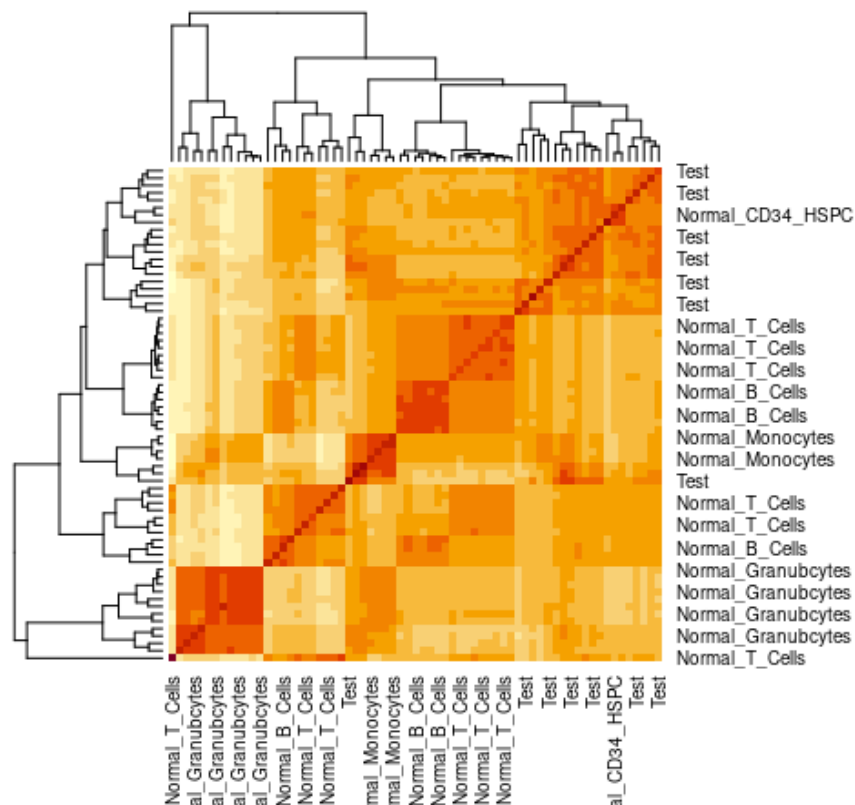
در خود سایت از umap استفاده شده ولی ما اینجا با کمک زاویه و طول بردار می توان PCA را نمایان کرده ایم..



همانگونه که می بینید، داده های تست به Normal_CD34_HSPC بسیار نزدیک هستند (گوشه بالا سمت چپ). این یعنی اینکه در بخش چهارم، این گروه از داده های تست را جدا می کنیم.
بخش ۳.

خروجی Heatmap

گام ۴ - Correlation



بخش ۴.

بررسی تمایز ژن‌ها

گام ۵ - جداسازی.

همانطور که در انتهای بخش ۲ گفته شد، گروهی از داده‌های تست که PC2 بیشتر ۰.۱۳ و PC1 کمتر از ۰.۰۲۵- دارند نزدیک به CD34 هستند. همچنین در نمودار همبستگی نیز نواحی تست با این گروه همبستگی بالاتری دارند.

گام ۵ - مرتب‌سازی.

اکنون (مانند کلاس آقای سلیمی) این گروه Near را به گروه‌های gset اضافه می‌کنیم و نهایتاً آن را به صورت one-hot در می‌آوریم. این برای این است که از حالت categorical خارج شود.

گام ۶ - اعمال مدل‌ها.

در این قسمت با کمک کتابخانه‌ی Limma از یک مدل خطی استفاده می‌کنیم. پس از fit کردن آن، می‌توان اختلاف بین بیانگرها را برای حالت Test و CD34_HSPC به دست آورد. سپس با فرض اینکه آستانه‌ی تفاوت ۰/۰۵ است، توزیع پیدا می‌شود. مشابه خود سایت،

logFC و adj.P_value را جدولی نگه می‌داریم.

گام ۷ - نتایج.

با توجه به شرط کمتر ۰.۰۵ بودن adj.P_value و همچنین logFC (بسته به اینکه بالا را بخواهیم یا پایین را) ژن‌ها را جدا می‌کنیم.

بخش ۵ - آنالیز Gene Ontology و Pathway

گام ۸ - Gene Ontology

برای این قسمت از سایت <https://maayanlab.cloud/Enrichr/enrichr> استفاده کرده‌ام.

برای این قسمت، ۳ نوع امتیاز حساب می‌شود.

ژن‌های بیان بالا:

Biological Process:

neutrophil activation involved in immune response (GO:0002283) – Adj. P_val 2.488e-17

neutrophil degranulation (GO:0043312) – Adj. P_val 3.704e-17

neutrophil mediated immunity (GO:0002446) – Adj.P_val 5.301e-17

Molecular Function:

kinase binding (GO:0019900) – Adj.P_val 0.000001018

protein kinase binding (GO:0019901) – Adj.P_val 0.000002972

cyclin-dependent protein serine/threonine kinase regulator activity (GO:0016538) – Adj.P_val 0.000004392

Cellular component:

specific granule (GO:0042581) – 1.535e-8

specific granule membrane (GO:0035579) – 3.618e-8

secretory granule lumen (GO:0034774) – 1.548e-7

پایین:

Biological Process:

secondary alcohol biosynthetic process (GO:1902653) – 0.1596

cholesterol biosynthetic process (GO:0006695) – 0.1596

sterol biosynthetic process (GO:0016126) - 0.1596

Molecular Function:

bis(5'-adenosyl)-hexaphosphatase activity (GO:0034431) – 0.6603

bis(5'-adenosyl)-pentaphosphatase activity (GO:0034432) – 0.6603

cell adhesive protein binding involved in bundle of His cell-Purkinje myocyte communication
(GO:0086083) - 0.6603

Cellular Component:

MHC class II protein complex (GO:0042613) – 0.6846

platelet alpha granule membrane (GO:0031092) – 0.6846

integral component of luminal side of endoplasmic reticulum membrane (GO:0071556) - 0.6846

امتیازی:

درمان AML.

بسیاری از روش‌های درمان مبتنی بر شیمی درمانی هستند. به صورت معمول از ترکیب Cytarabine و daunorubicin استفاده می‌شود اما داروهای دیگری نظیر Hydroxyurea، Cladribine و Azacitidine نیز به کار می‌آیند.