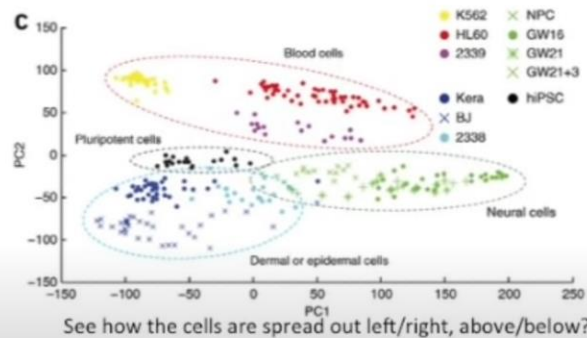


How to identify key genes.



See how the cells are spread out left/right, above/below?

If we wanted to find out which genes had a big influence in putting dermal cells on the left and neural cells on the right, we could look at the influence scores in PC1.

Pencarian langkah pencarian langkah pencarian halo dan selamat datang di pencarian stat pencarian stat dipersembahkan oleh orang-orang yang ramah di departemen genetika di Universitas North Carolina di Chapel Hill hari ini kita akan berbicara tentang analisis komponen utama atau singkatnya PCA mari kita mulai dengan contoh aksi analisis komponen prinsip dia adalah contoh plot PCA yang saya dapatkan dari artikel yang baru saja saya baca menunjukkan kelompok jenis sel grafik ini diambil dari data sekuensing RNA sel tunggal ada sekitar 10.000 gen yang ditranskripsi di setiap sel dan setiap titik dalam grafik ini mewakili satu sel dan profil transkripsinya. Ide umumnya adalah bahwa sel dengan profil transkripsi yang serupa harus berkelompok dan kita melihat bahwa dalam grafik ini kita melihat bahwa sel darah membentuk satu kelompok yang berbeda dari sel pluripoten yang berbeda dengan sel saraf dan sel dermal atau epidermal jadi pertanyaan besarnya adalah bagaimana transkripsi dari 10.000 gen dikompresi menjadi satu titik pada grafik jawabannya adalah PCA PCA adalah metode untuk mengompresi banyak data menjadi sesuatu yang menangkap inti dari data asli dalam pencarian stat ini kita akan mempelajari semua tentang bagaimana PCA melakukan kompresi ini dan kita juga akan mencari tahu apa yang dirujuk oleh label akses ini sebelum kita mendalami seluk beluk PCA. akan membahas sedikit materi latar belakang. Kami akan memperkenalkan dimensi hanya untuk memperingatkan Anda bahwa ini akan tampak sangat sederhana, tetapi tunggu saja Anda akan senang kami melakukan ini.

1) Tujuan Utama PCA:

- PCA bertujuan untuk mereduksi dimensi dalam dataset sambil mempertahankan sebanyak mungkin informasi asli.
- Hal ini mencapai dengan mengubah ruang koordinat data ke ruang yang lebih kecil, yang didefinisikan oleh "komponen-komponen utama" yang memaksimalkan variasi data.

2) Langkah-langkah PCA:

- Standarisasi Data: Langkah pertama adalah untuk mengstandarisasi data dengan mengurangi rata-rata dari setiap fitur (kolom) dan mengukur variabilitas dengan menghitung deviasi standar. Ini penting untuk memastikan bahwa semua fitur memiliki pengaruh yang seimbang dalam analisis.

- Matriks Kovariansi: PCA melibatkan perhitungan matriks kovariansi antara semua pasangan variabel (fitur) dalam dataset. Matriks kovariansi mengukur hubungan antara setiap pasang variabel dan memberi tahu seberapa bersamaan atau berbedanya mereka.

3) Perhitungan Komponen Utama:

- Setelah mendapatkan matriks kovariansi, langkah selanjutnya adalah menghitung "komponen-komponen utama" atau "vektor-vektor eigen." Komponen-komponen ini adalah kombinasi linear dari variabel asli dan merupakan arah di mana data memiliki variasi maksimum.
- Komponen pertama adalah yang paling signifikan karena mencakup sebagian besar variasi data. Komponen kedua adalah yang berikutnya, dan seterusnya.

4) Proyeksi Data: PCA memungkinkan data diproyeksikan ke dalam ruang komponen utama, yang memungkinkan representasi data dengan dimensi yang lebih rendah daripada data asli.

5) Pemilihan Komponen Utama: Dalam prakteknya, kita biasanya memilih sejumlah komponen utama yang mewakili sebagian besar variasi dalam data. Namun, jumlah komponen ini dapat dipilih berdasarkan aturan baku, seperti mengambil komponen hingga sejumlah tertentu atau dengan merujuk pada rasio varians yang dijelaskan.

6) Keuntungan PCA:

- Mereduksi dimensi dataset, yang mengurangi kompleksitas dan kebisingan.
- Mengungkap struktur dan pola dalam data yang mungkin sulit terlihat dalam dimensi asli.
- Membantu mengidentifikasi fitur-fitur yang paling penting dalam analisis data.
- Berguna untuk kompresi data.

7) Keterbatasan PCA:

- Kehilangan interpretasi fitur asli karena transformasi ke dimensi yang lebih rendah.
- Komponen utama harus linear, yang berarti PCA kurang efektif dalam mengidentifikasi pola non-linear dalam data.