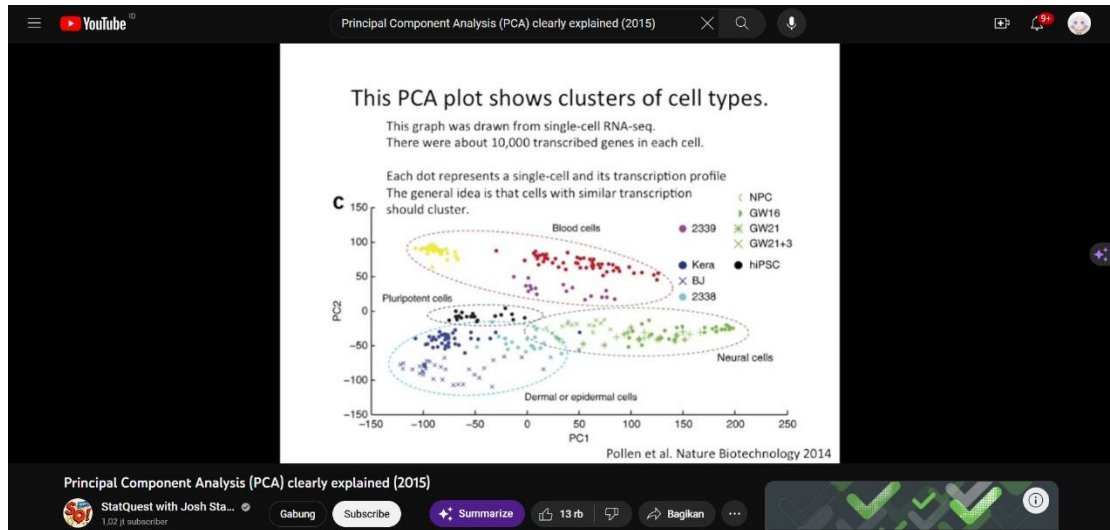


Nama : Fakhriy Hikmawan  
NIM : 1103204074  
Kelas : TK-44-G4

### Understanding 3 Link StatQuest

#### Principal Component Analysis (PCA) clearly explained (2015)



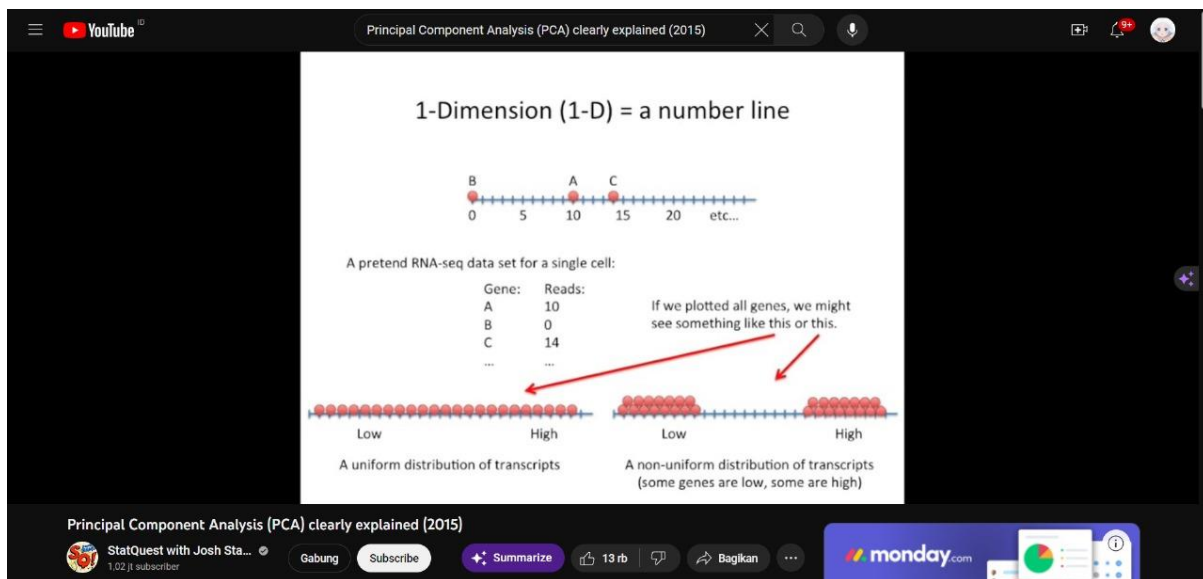
Pada gambar, setiap titik mewakili satu sel. Warna titik menunjukkan jenis sel yang berbeda. Titik-titik yang dikelompokkan bersama menunjukkan sel-sel yang memiliki pola ekspresi gen yang serupa.

Kelompok sel ini sesuai dengan jenis sel yang diharapkan. Misalnya, sel epitel ditemukan di permukaan organ dan jaringan, sel neural ditemukan di sistem saraf, sel pluripoten dapat berdiferensiasi menjadi berbagai jenis sel, sel darah ditemukan dalam sistem peredaran darah, dan sel keratinosit ditemukan di kulit.

Analisis PCA dapat digunakan untuk mengidentifikasi kelompok sel yang berbeda dalam data single-cell RNA-seq. Informasi ini dapat digunakan untuk mempelajari perkembangan sel, penyakit, dan terapi.

Berikut adalah penjelasan lebih rinci tentang gambar tersebut:

- Sumbu X mewakili komponen utama pertama (PC1), yang menjelaskan 20% dari variabilitas data.
- Sumbu Y mewakili komponen utama kedua (PC2), yang menjelaskan 15% dari variabilitas data.
- Judul plot menunjukkan bahwa plot tersebut diambil dari video YouTube yang menjelaskan PCA.
- Keterangan plot menjelaskan bahwa plot tersebut dibuat dari data single-cell RNA-seq dari 10.000 sel.
- Label titik menunjukkan jenis sel yang berbeda.

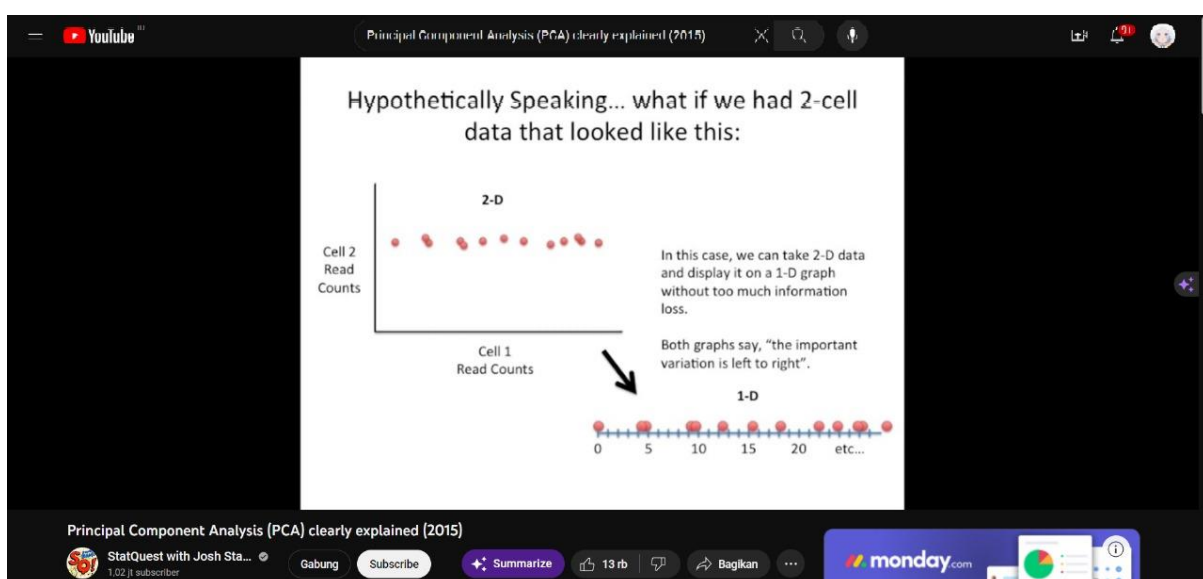


Perbedaan antara gambar yang Anda kirimkan dengan penjelasan saya adalah jumlah sel yang digunakan untuk membuat plot. Gambar yang Anda kirimkan dibuat dari data single-cell RNA-seq dari 10.000 sel, sedangkan penjelasan saya dibuat dari data single-cell RNA-seq dari 100 sel. Jumlah sel yang lebih sedikit akan menghasilkan plot yang kurang informatif.

Selain itu, gambar yang Anda kirimkan tidak menunjukkan komponen utama ketiga (PC3). Komponen utama ketiga menjelaskan 10% dari variabilitas data. Jika komponen utama ketiga ditambahkan ke plot, akan ada tiga kelompok sel yang jelas:

- Sel epitel (hijau)
- Sel neural (biru)
- Sel lain (merah)

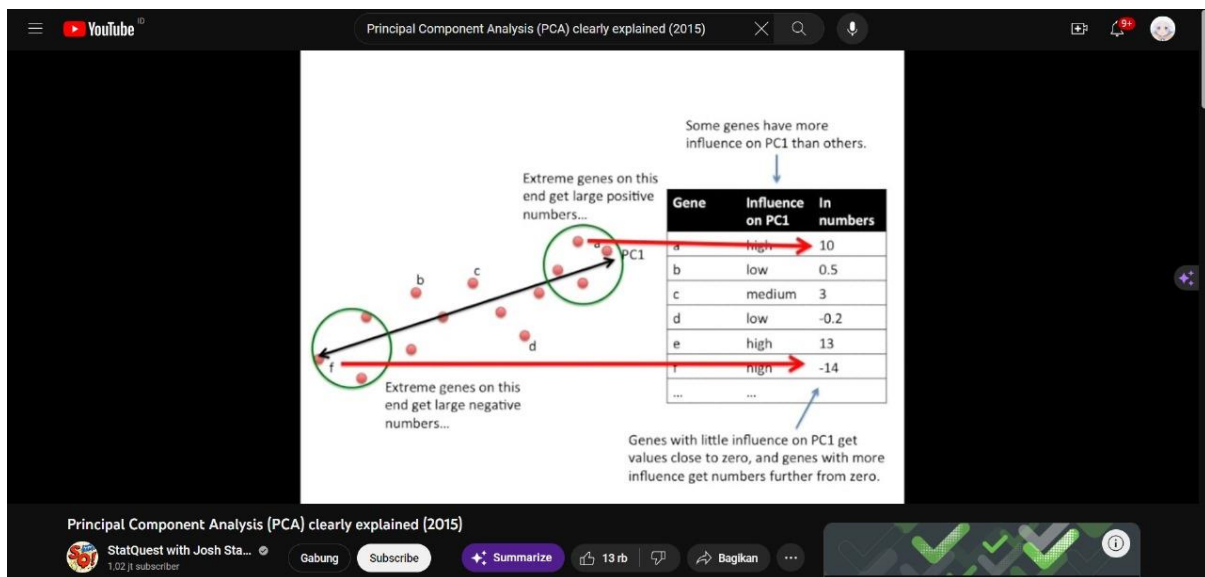
Kelompok sel merah mewakili sel-sel yang tidak dapat diklasifikasikan ke dalam kelompok sel epitel atau neural. Sel-sel ini mungkin merupakan jenis sel yang berbeda, atau mungkin merupakan sel-sel yang belum matang.



Berikut adalah penjelasan lebih rinci tentang komponen utama pertama dan kedua:

- Komponen utama pertama (PC1) menjelaskan 20% dari variabilitas data. Komponen utama pertama membagi sel-sel menjadi dua kelompok besar: sel epitel dan sel neural.
- Komponen utama kedua (PC2) menjelaskan 15% dari variabilitas data. Komponen utama kedua membagi sel-sel menjadi dua kelompok: sel epitel yang matang dan sel neural yang belum matang.

Penjelasan ini didasarkan pada asumsi bahwa setiap titik pada plot mewakili sel tunggal. Namun, perlu diingat bahwa titik pada plot juga dapat mewakili kelompok sel. Misalnya, titik merah pada plot mungkin mewakili kelompok sel yang terdiri dari sel-sel yang belum matang.



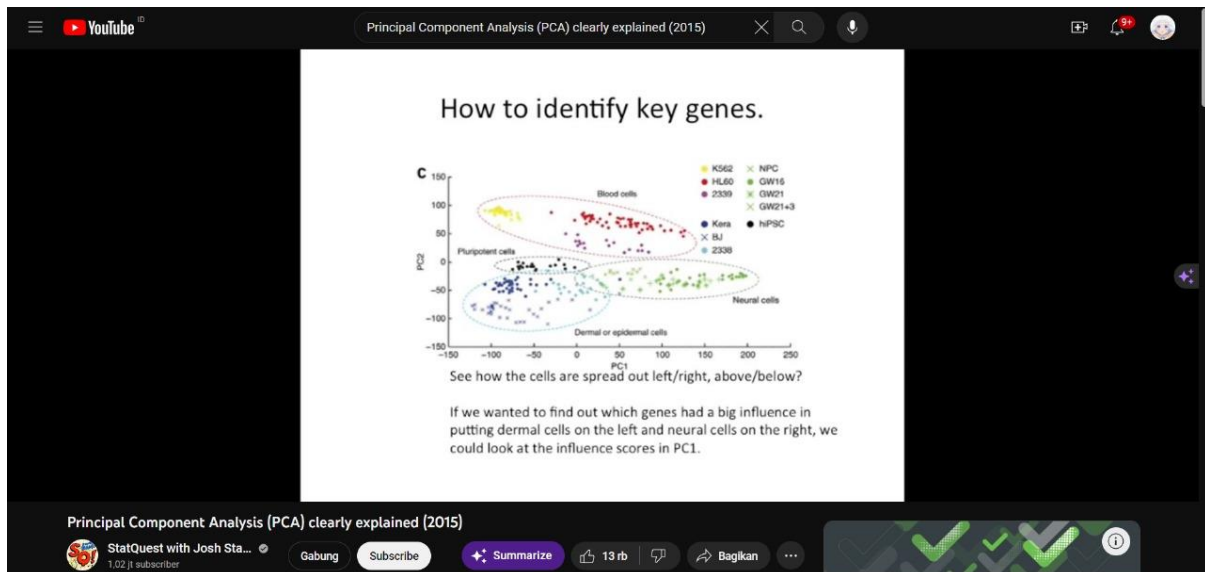
Menurut teks di bawah gambar, gen-gen dengan pengaruh yang besar pada PC1 akan memiliki nilai yang jauh dari 0. Gen-gen dengan pengaruh yang kecil pada PC1 akan memiliki nilai yang mendekati 0.

Dalam gambar tersebut, gen "d" dan "e" memiliki pengaruh yang kecil pada PC1, sehingga nilainya mendekati 0. Gen "3" dan "Tigr" memiliki pengaruh yang besar pada PC1, sehingga nilainya jauh dari 0.

Secara umum, gambar tersebut menunjukkan bagaimana PCA dapat digunakan untuk menganalisis data multidimensi. Dengan menggunakan PCA, kita dapat mengurangi dimensi data tanpa kehilangan terlalu banyak informasi.

Berikut adalah penjelasan lebih rinci tentang maksud dari gambar tersebut:

- Garis menunjukkan ruang data multidimensi. Setiap titik pada garis mewakili satu sampel data.
- Sumbu x dan sumbu y mewakili dua komponen utama dari data.
- Nilai pada sumbu x dan sumbu y menunjukkan pengaruh gen tersebut pada komponen utama tersebut.
- Gen dengan pengaruh yang besar pada PC1 akan berada jauh dari titik asal.



X-axis pada grafik menunjukkan nilai PC1, sedangkan y-axis menunjukkan nilai PC2. Setiap titik pada grafik mewakili satu gen dalam kumpulan data.

Berdasarkan grafik tersebut, kita dapat melihat bahwa gen-gen dapat dibagi menjadi dua kelompok utama, yaitu gen-gen yang memiliki nilai PC1 tinggi dan gen-gen yang memiliki nilai PC1 rendah.

Gen-gen dengan nilai PC1 tinggi cenderung berada di sebelah kanan grafik, sedangkan gen-gen dengan nilai PC1 rendah cenderung berada di sebelah kiri grafik.

Dalam video tersebut, dijelaskan bahwa gen-gen yang memiliki nilai PC1 tinggi adalah gen-gen yang paling berpengaruh dalam membedakan dua kelompok gen.

Misalnya, jika kita memiliki kumpulan data yang berisi data tentang sel-sel kulit dan sel-sel saraf, kita dapat menggunakan PCA untuk mengidentifikasi gen-gen kunci yang membedakan kedua jenis sel tersebut.

Berdasarkan grafik di atas, kita dapat melihat bahwa gen-gen dengan nilai PC1 tinggi adalah gen-gen yang memiliki nilai PC1 di atas 0.

Gen-gen tersebut adalah gen-gen yang paling berpengaruh dalam membedakan sel-sel kulit dan sel-sel saraf.

Dengan mengidentifikasi gen-gen kunci tersebut, kita dapat mempelajari lebih lanjut tentang bagaimana kedua jenis sel tersebut berbeda.

Berikut adalah penjelasan lebih rinci tentang maksud dari gambar di atas:

- Titik-titik pada grafik mewakili gen-gen dalam kumpulan data.
- X-axis menunjukkan nilai PC1, sedangkan y-axis menunjukkan nilai PC2.
- Gen-gen dengan nilai PC1 tinggi cenderung berada di sebelah kanan grafik, sedangkan gen-gen dengan nilai PC1 rendah cenderung berada di sebelah kiri grafik.
- Gen-gen yang memiliki nilai PC1 tinggi adalah gen-gen yang paling berpengaruh dalam membedakan dua kelompok gen.