Week 11: Single-cell genomics

- Introduction
- Dimensionality reduction
- Supervised machine learning

Single-cell RNA-seq

BREAKTHROUGH OF THE YEAR

Development cell by cell

With a trio of techniques, scientists are tracking embryo development in stunning detail

A zebrafish embryo at an early stage of development. Fluorescent markers highlight cells expressing genes that help determine the type of cell they will become. (JEFFREY FARRELL, SCHIER LAB/HARVARD UNIVERSITY)

The single-cell revolution is just starting.

— Elizabeth Pennisi

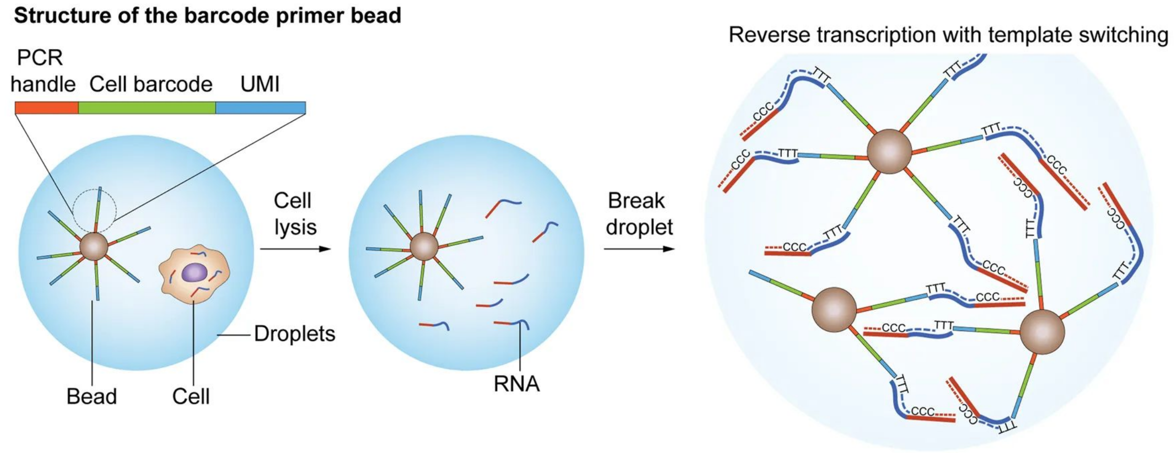
https://vis.sciencemag.org/breakthrough2018/finalists/#cell-development

Single-cell RNA-seq

	Bulk strategy	Single-cell-specific aspects
Alignment and generation of counts	Existing mapping tools (e.g. TopHat[42] or GSNAP[43]) and approaches for generating read counts (e.g. HTseq[45]) can be used	
Quality control	• Focuses on investigating library quality • PCA or clustering to confirm that samples of the same class (that is, biological or technical replicates) cluster together	In addition to the quality control strategies necessary for bulk RNA-seq, it is important to determine whether the RNA in each captured cell is degraded. Studying the total percentage of mapped reads and the proportion of reads mapped to the spiked-in molecules can be useful. Cells with aberrant patterns can be discarded from downstream analyses
Normalization	Standard approaches adjust for sequencing depth (e.g., using FPKM[10] or a scaling factor-based approach[50,51])	For scRNA-seq it is also necessary to decide how to normalize for differences in the mRNA content between cells. Moreover, 3′ biases in some experimental protocols[28] mean that normalization approaches that adjust for transcript length might be problematic
Confounding factors	PCA-regression-like approaches can be used to identify and account for latent structure between samples or cells that contributes to the expression landscape but that is independent of the biological quantity of interest. Specialized tools[61] that use such corrections automatically and that can handle multiple confounding factors can also be used	
Cell type identification	Clustering approaches based on latent-variable models (with some scRNA-seq modifications) can be used	
Cell type characterization	Differential expression tools[50] and transcript usage tools[74] built for bulk analyses can generally be applied	

Stegle (2015) Nat. Rev. Genet.

Single-cell isolation and library preparation



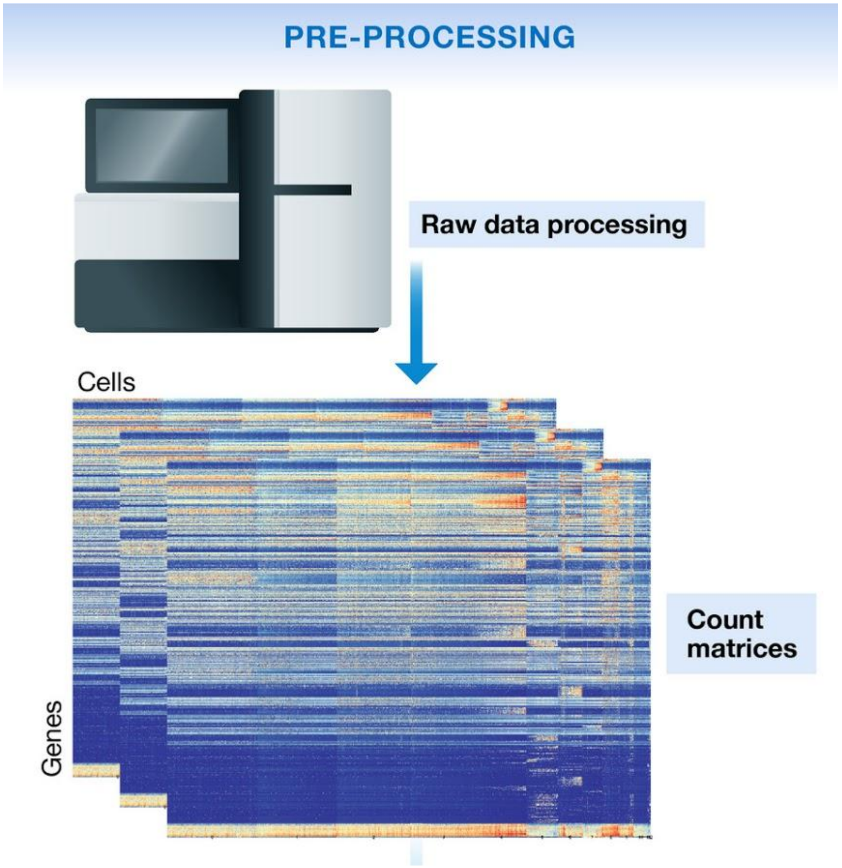
A schematic example of droplet-based library generation.

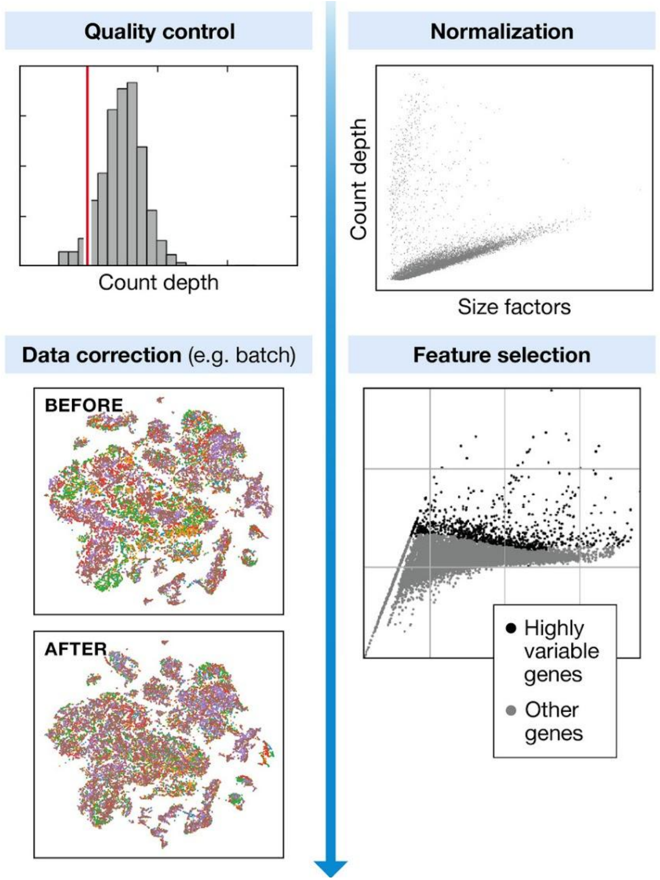
Libraries for scRNA-seq are typically generated via:

- Cell lysis
- Reverse transcription into first-strand cDNA using uniquely barcoded beads
- Second-strand synthesis, &
- cDNA amplification.

Hwang (2018) Exp. Mol. Med.

Pre-processing, QC, & normalization of scRNA-seq data





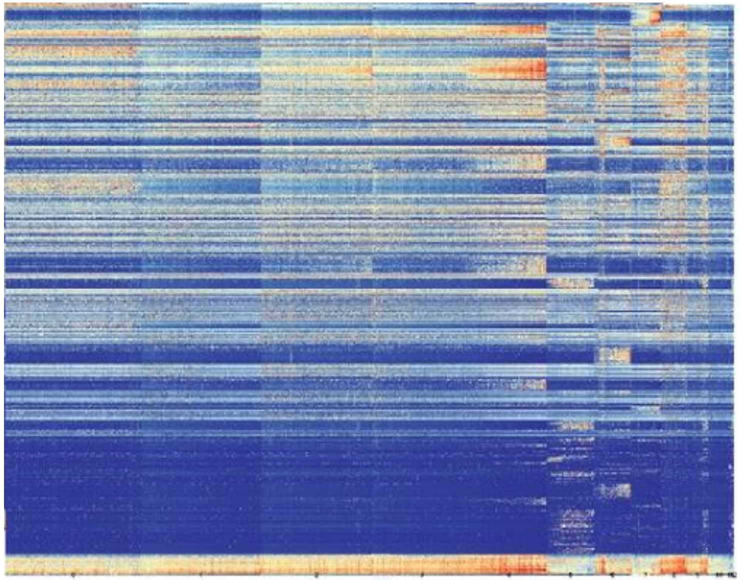
Luecken & Theis (2019) Mol. Sys. Biol.

Dimensionality reduction of scRNA-seq data for visualization

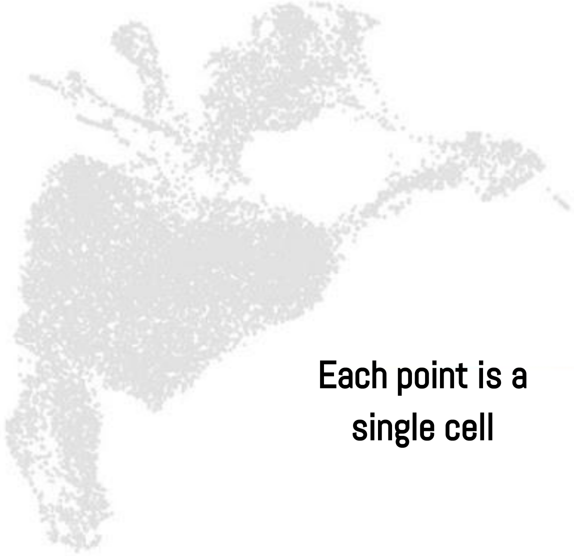
Count matrix

cells

genes

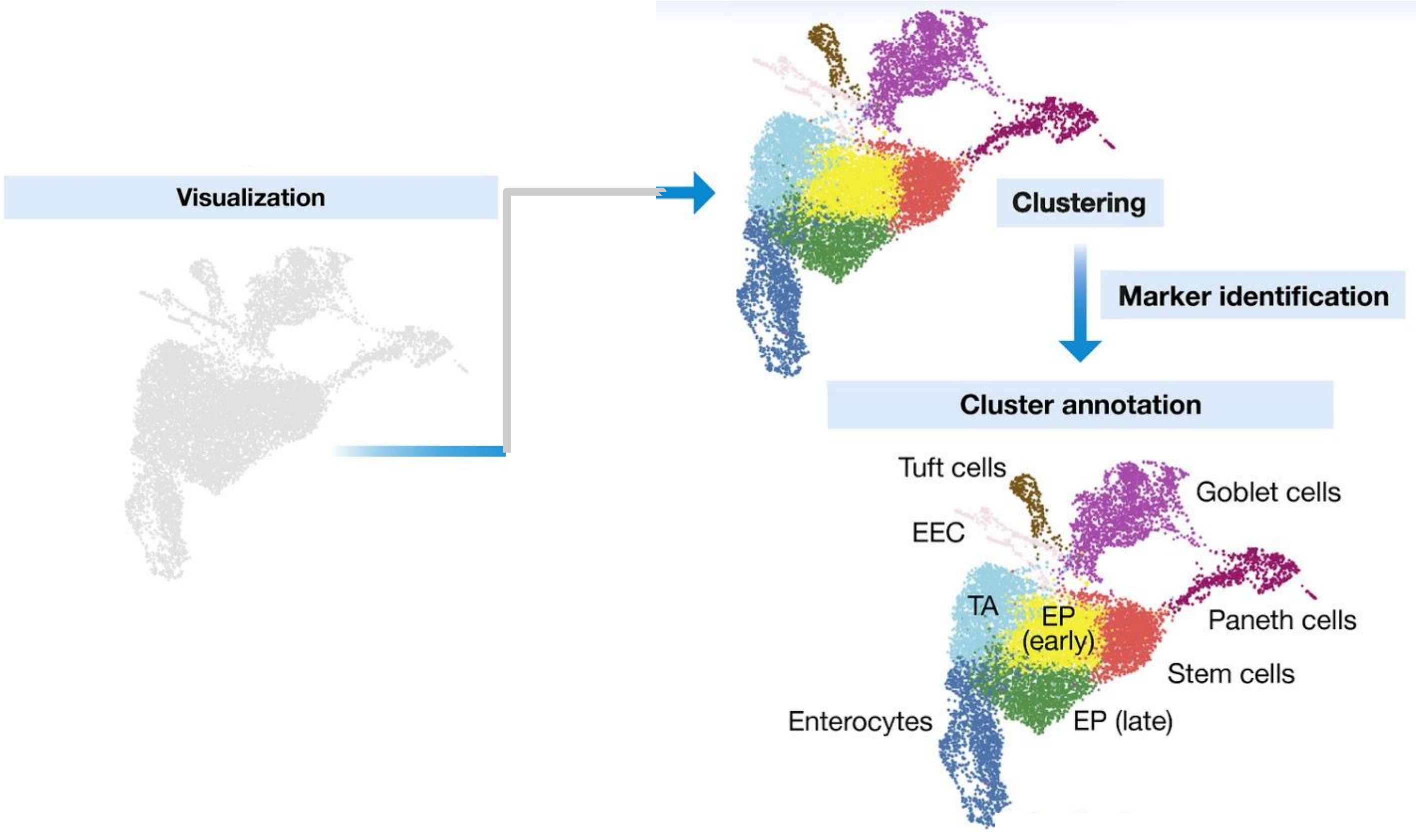


Visualization



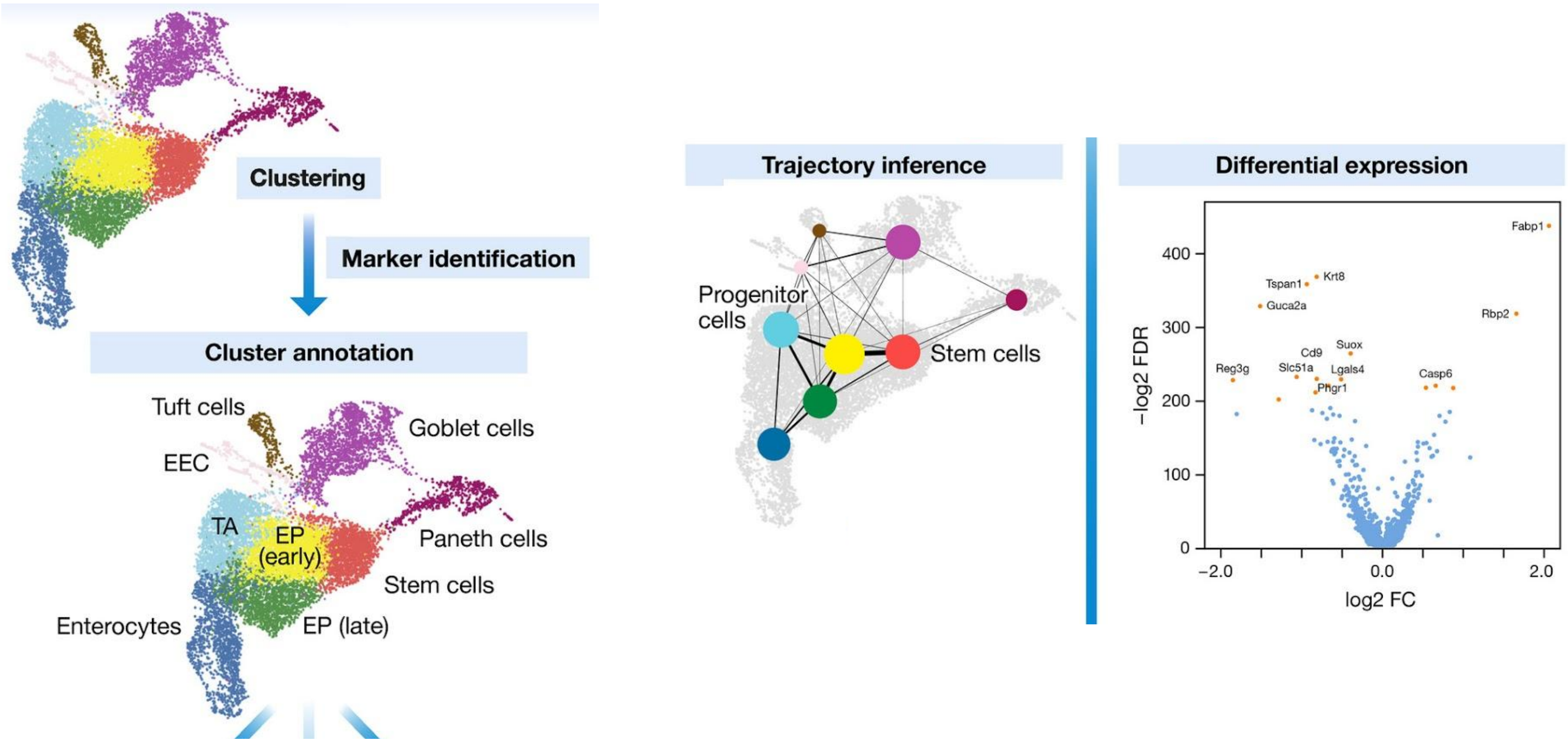
Luecken & Theis (2019) Mol. Sys. Biol.

Clustering scRNA-seq data to identify cell types/states



Luecken & Theis (2019) Mol. Sys. Biol.

Clustering scRNA-seq data to identify cell types/states



Luecken & Theis (2019) Mol. Sys. Biol.

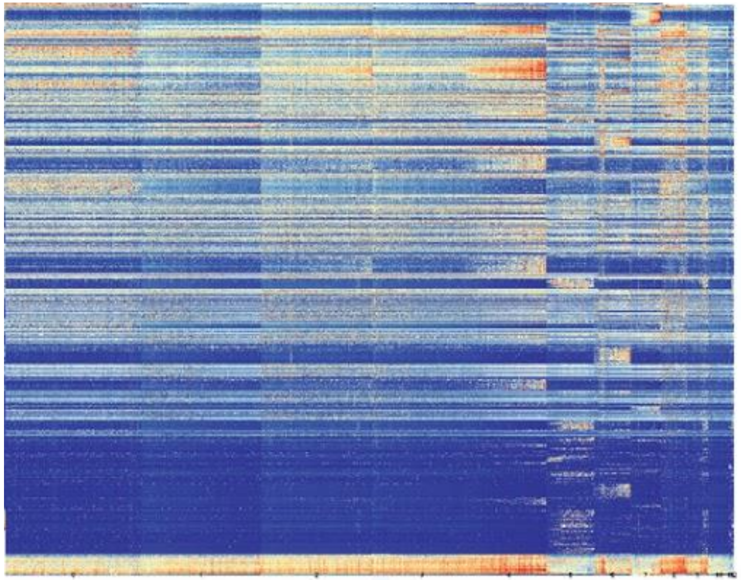
Dimensionality reduction of scRNA-seq data for visualization

Count matrix

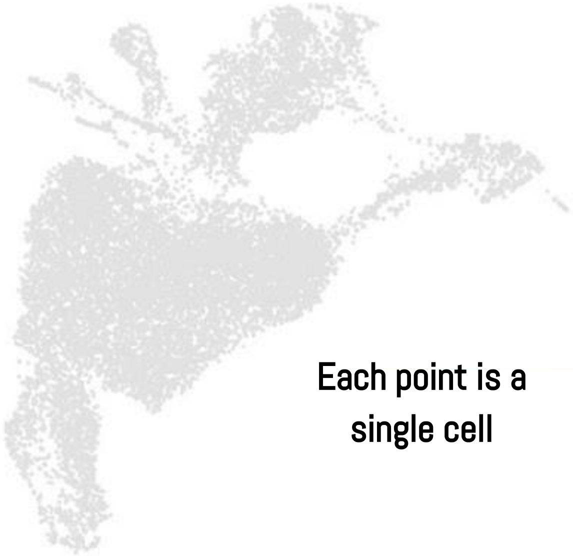
Visualization

cells

genes



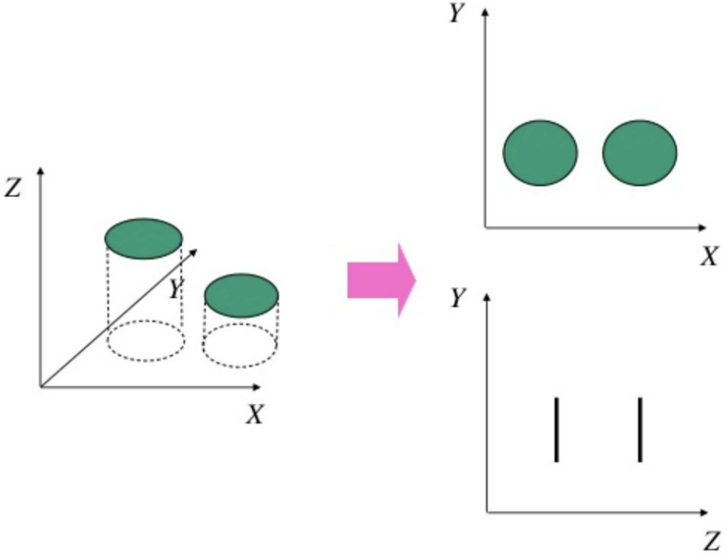
?



Each point is a single cell

Luecken & Theis (2019) Mol. Sys. Biol.

Dimensionality reduction – Projecting data into low-dim space

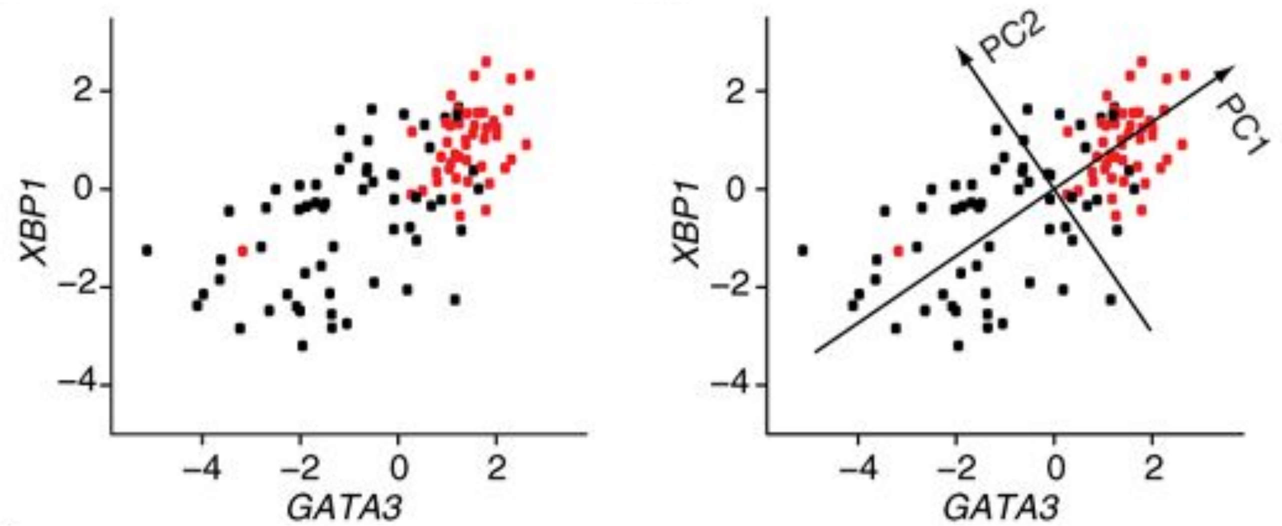


Lever, Krzywinski, & Altman (2017) Nat. Meth.

Dimensionality reduction using Principal Components Analysis

PCA geometrically projects data onto a lower-dimensional space

- Each lower dimension is a ‘linear’ combination of correlated original dimensions.
- The principal components (PCs) represent the directions of maximum variation.

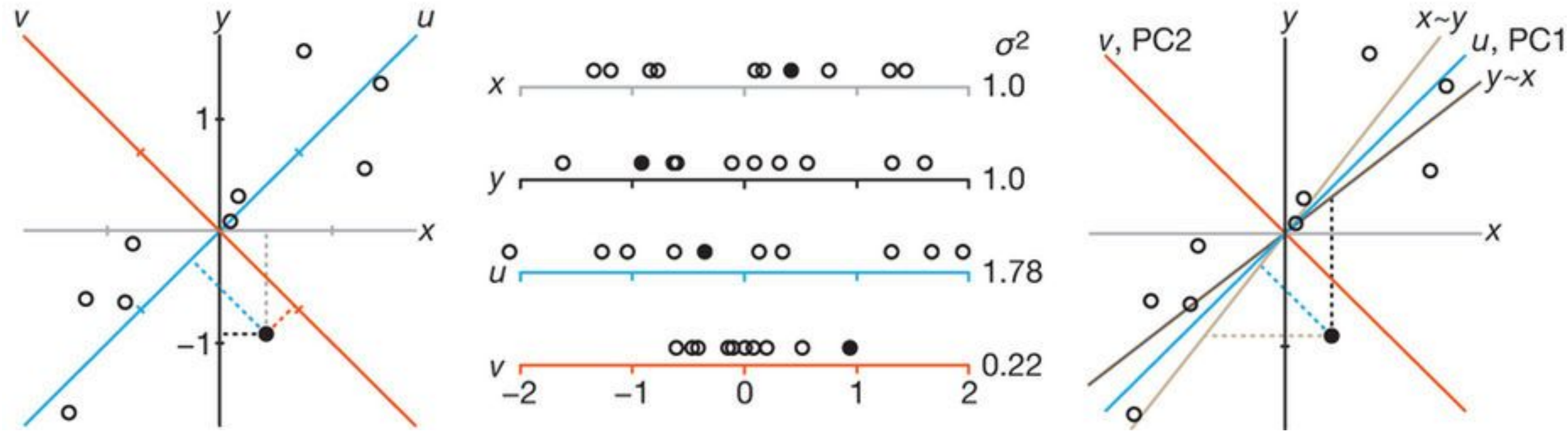


Ringnér (2008) Nat. Biotech

Dimensionality reduction using Principal Components Analysis

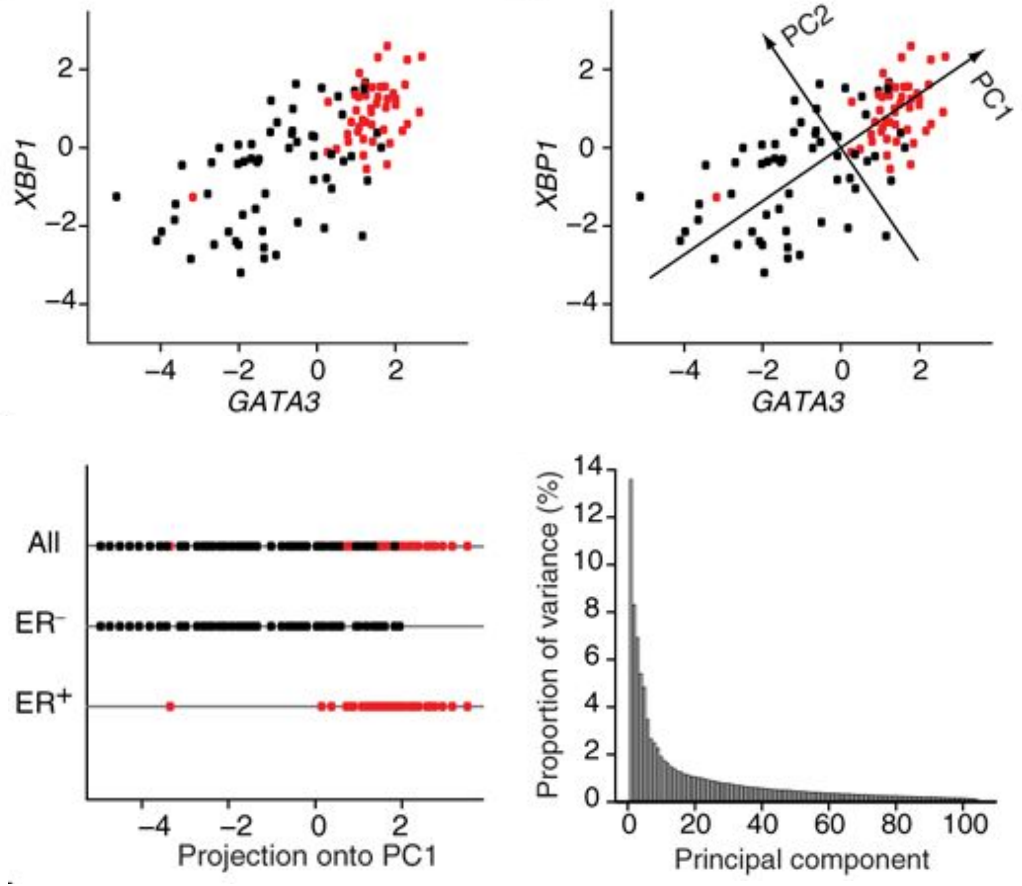
PCA geometrically projects data onto a lower-dimensional space

- Each lower dimension is a 'linear' combination of correlated original dimensions.
- The principal components (PCs) represent the directions of maximum variation.



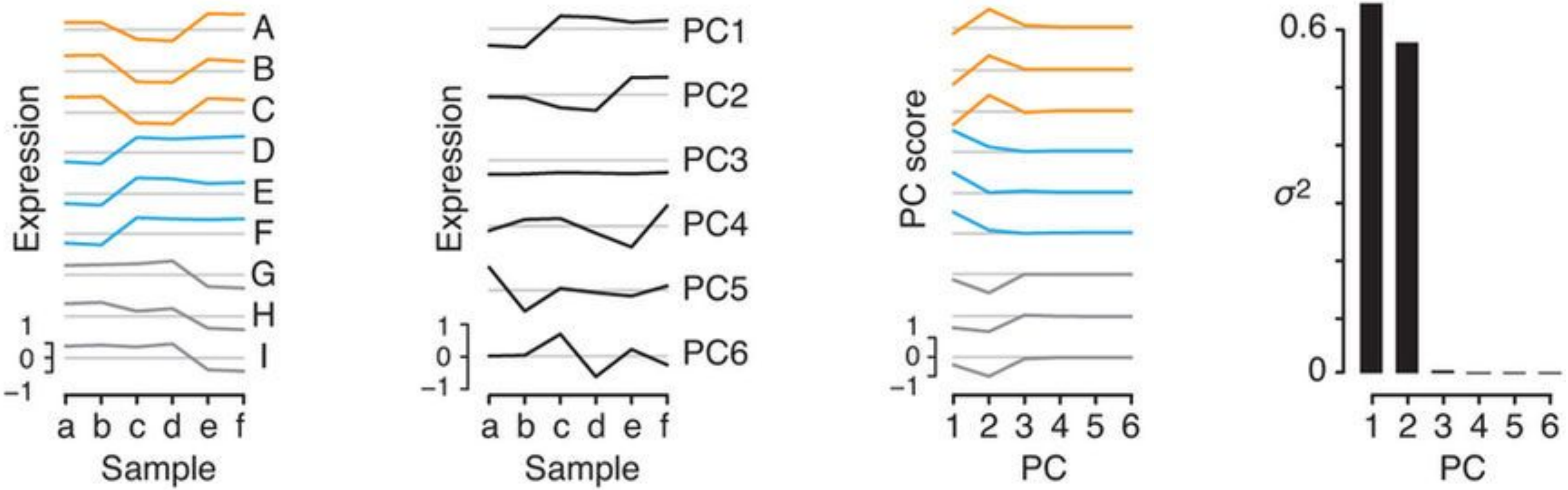
Lever, Krzywinski, & Altman (2017) Nat. Meth.

Dimensionality reduction by PCA



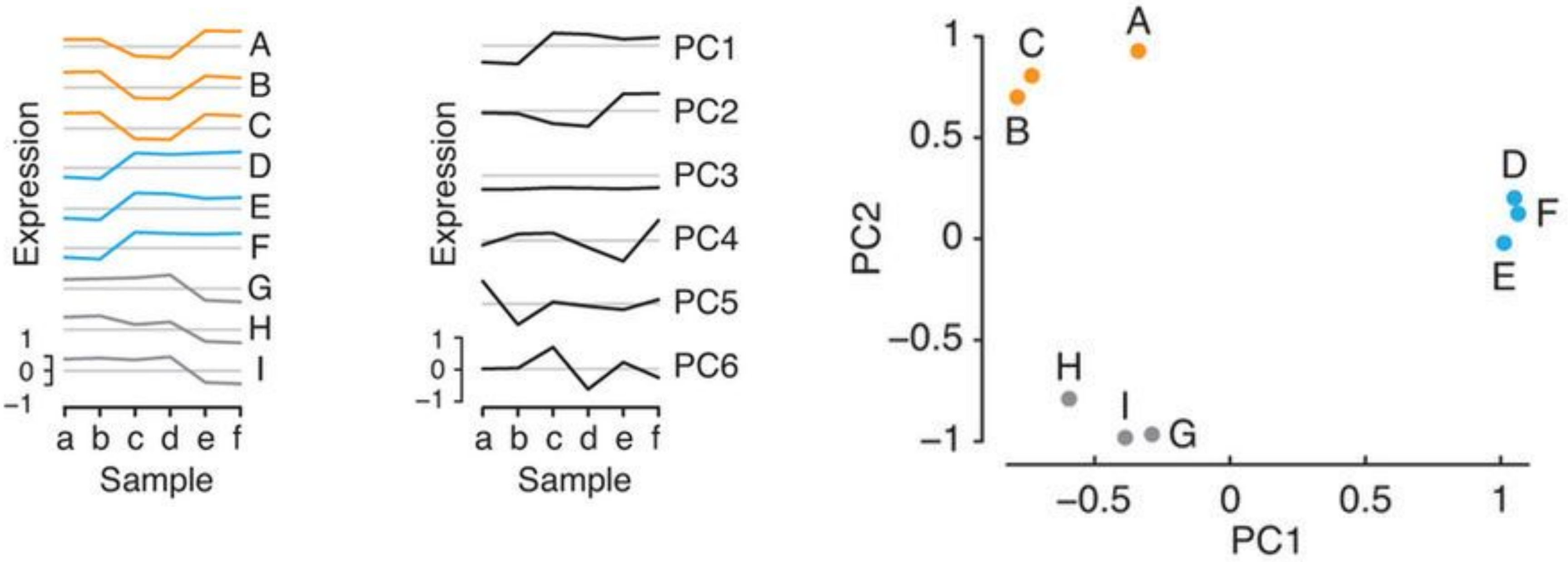
Ringnér (2008) Nat. Biotech

Dimensionality reduction by PCA



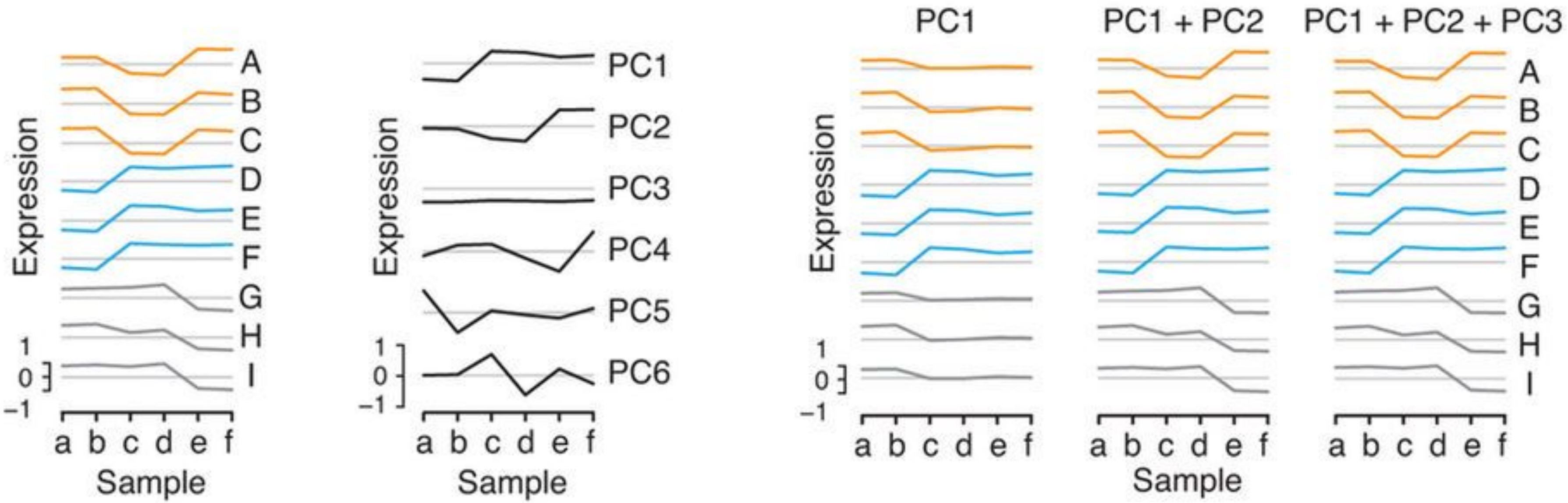
Lever, Krzywinski, & Altman (2017) Nat. Meth.

Dimensionality reduction by PCA



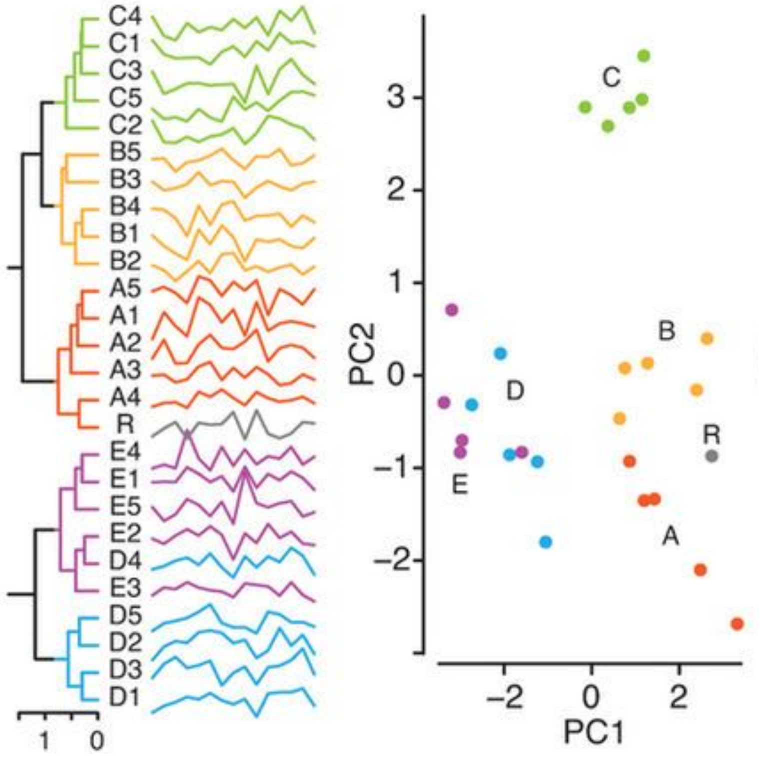
Lever, Krzywinski, & Altman (2017) Nat. Meth.

Dimensionality reduction by PCA



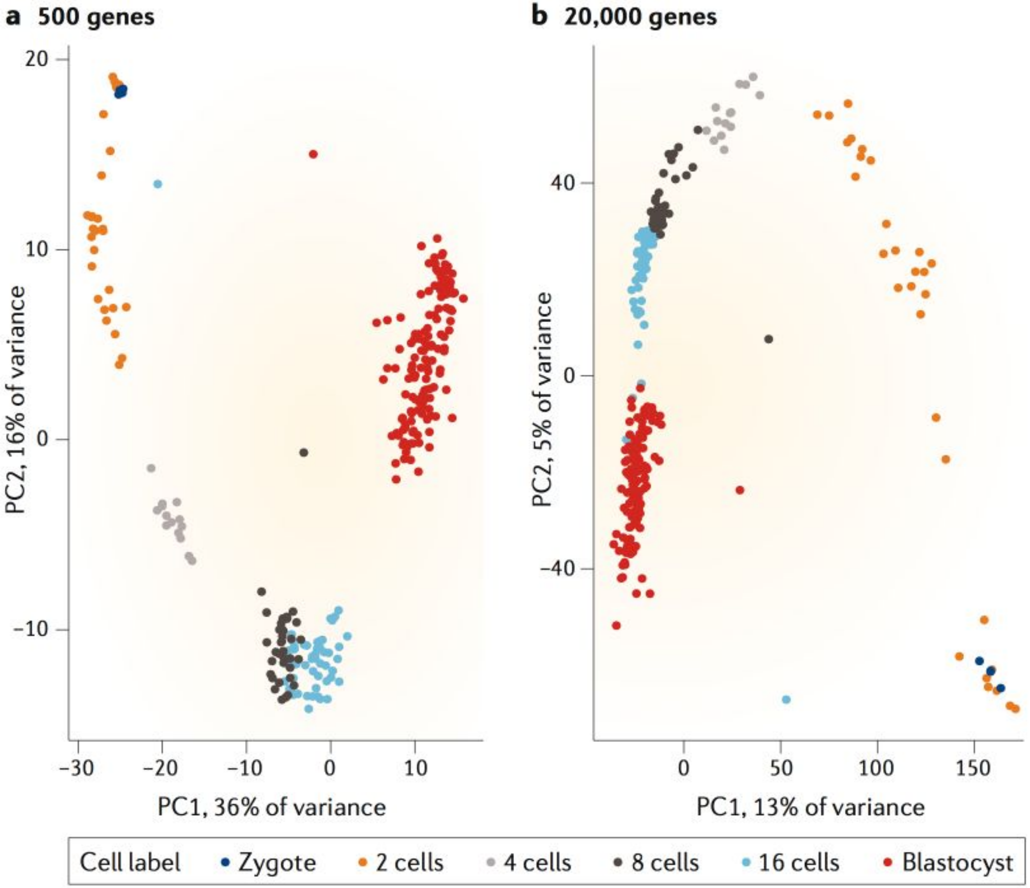
Lever, Krzywinski, & Altman (2017) Nat. Meth.

Dimensionality reduction by PCA



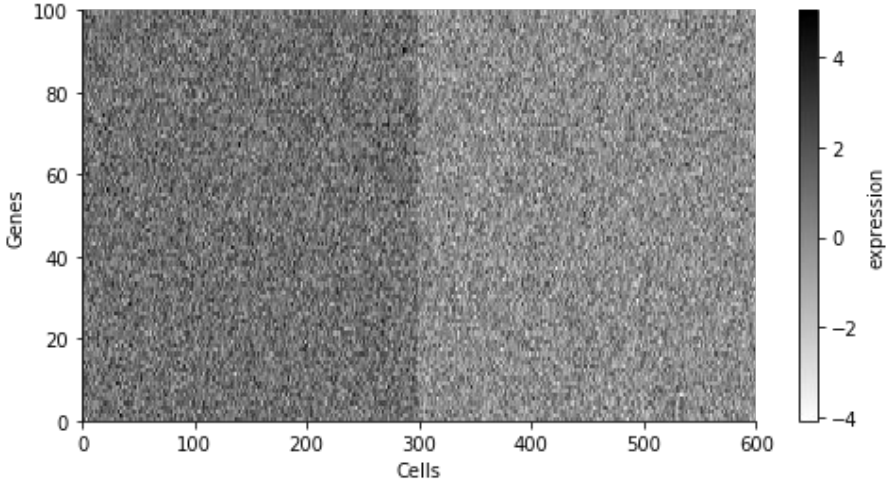
Lever, Krzywinski, & Altman (2017) Nat. Meth.

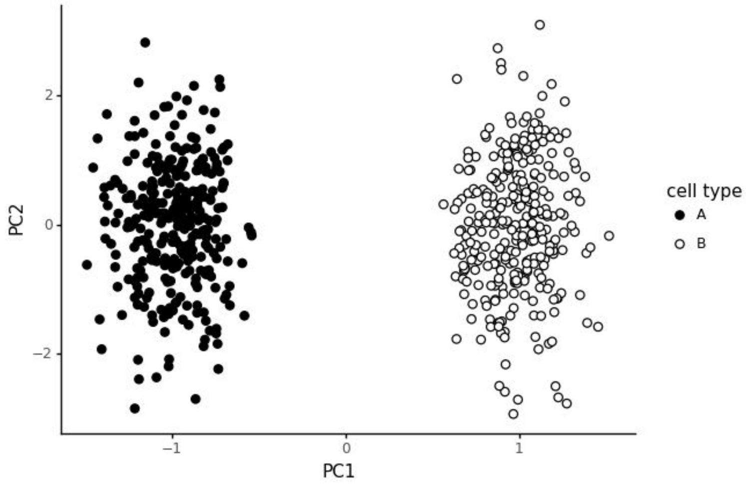
Dimensionality reduction by PCA



Lever, Krzywinski, & Altman (2017) Nat. Meth.

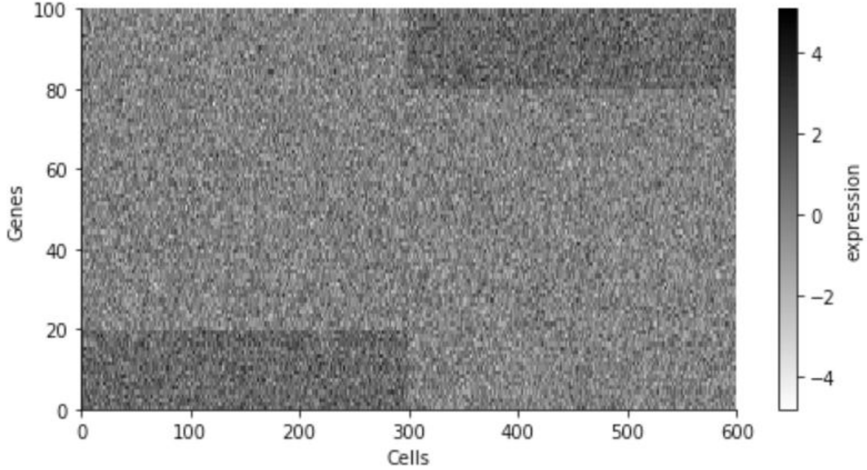
Dimensionality reduction by PCA

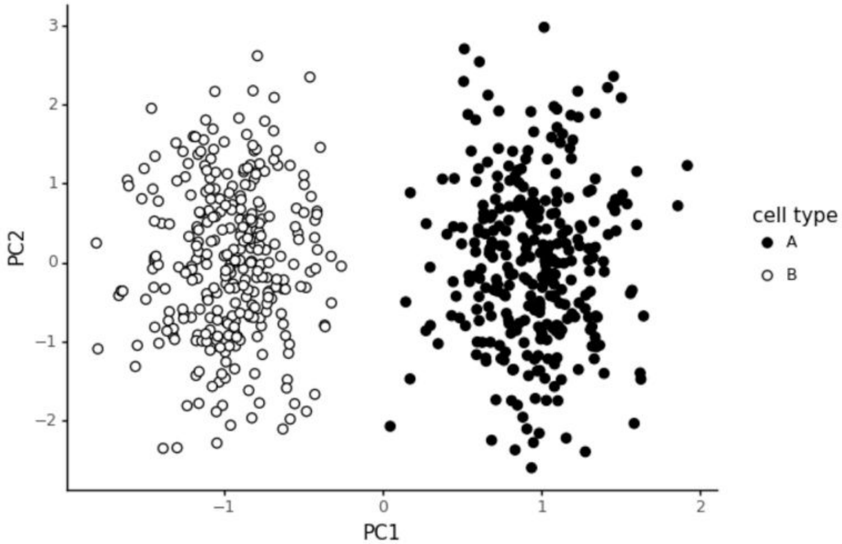




http://www.nxn.se/valent/2017/6/12/how-to-read-pca-plots

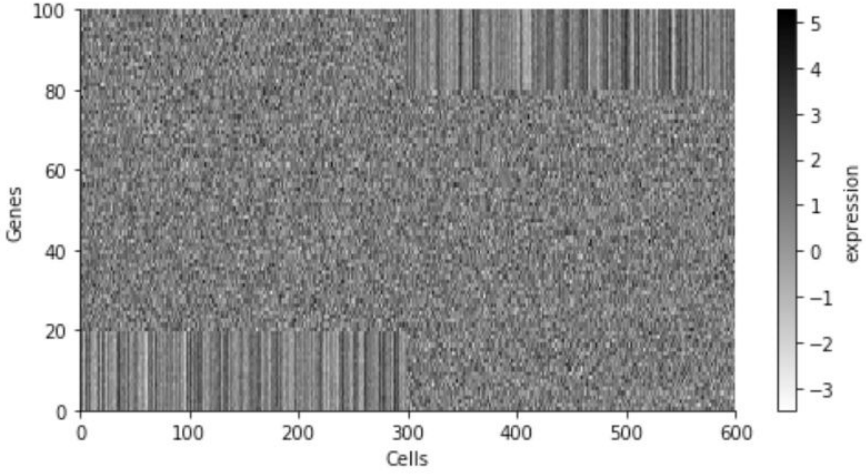
Dimensionality reduction by PCA

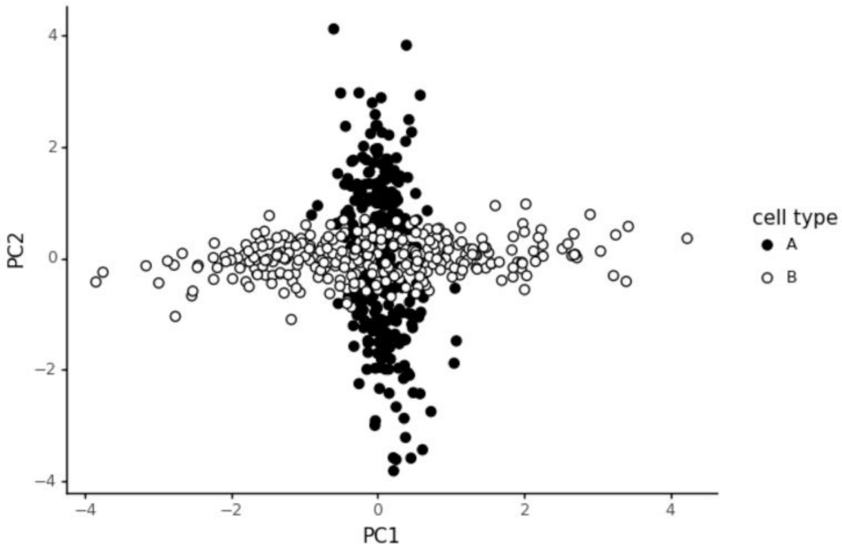




http://www.nxn.se/valent/2017/6/12/how-to-read-pca-plots

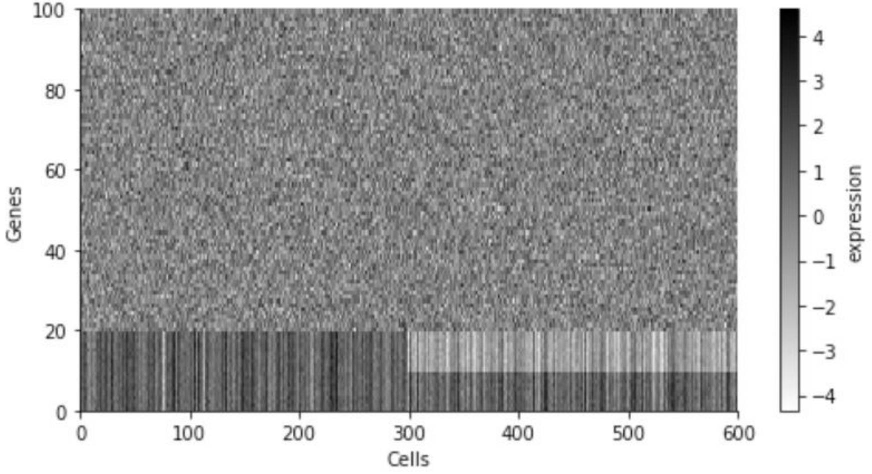
Dimensionality reduction by PCA

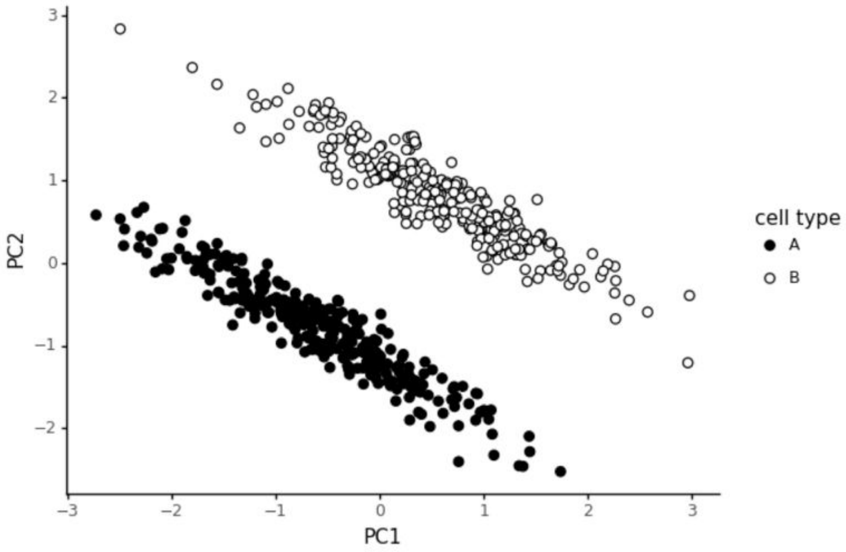




http://www.nxn.se/valent/2017/6/12/how-to-read-pca-plots

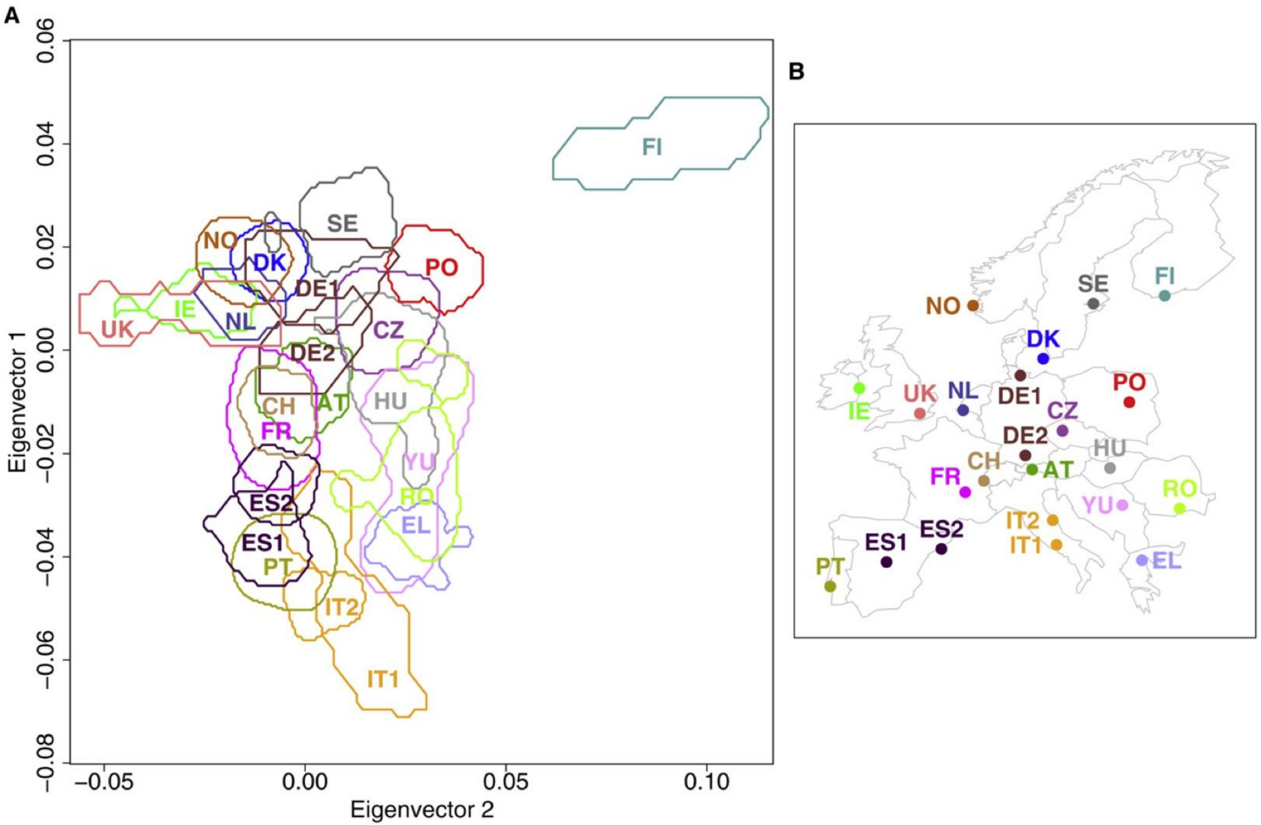
Dimensionality reduction by PCA





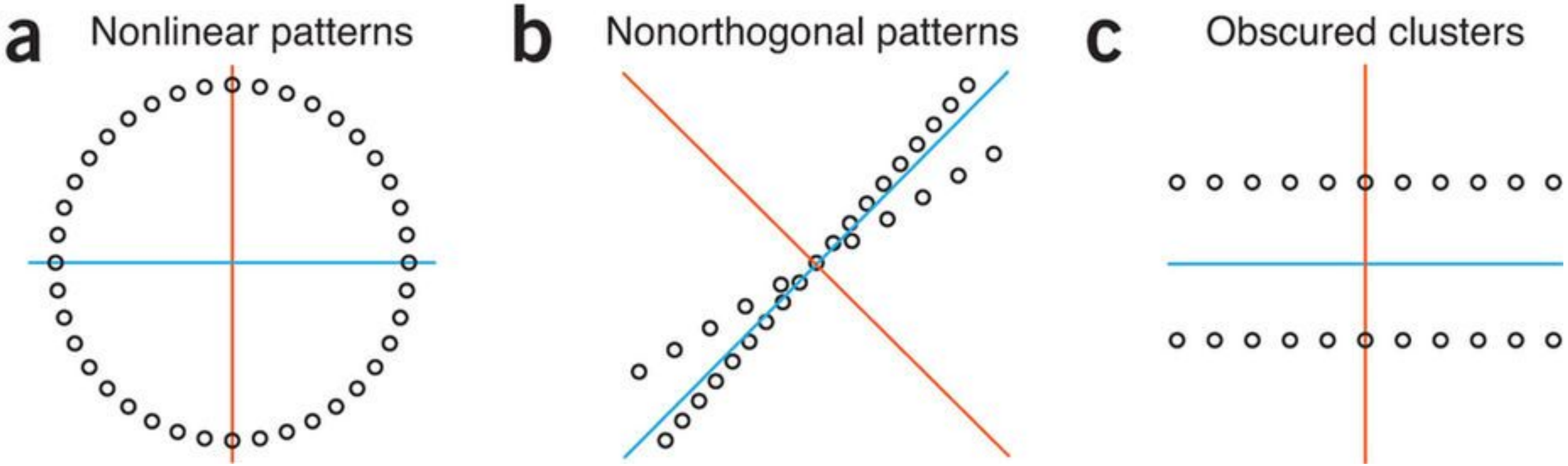
http://www.nxn.se/valent/2017/6/12/how-to-read-pca-plots

Dimensionality reduction



"Genes mirror geography within Europe." Novembre (2008) Nat.

Limitations of PCA, alternatives



Lever, Krzywinski, & Altman (2017) Nat. Meth.