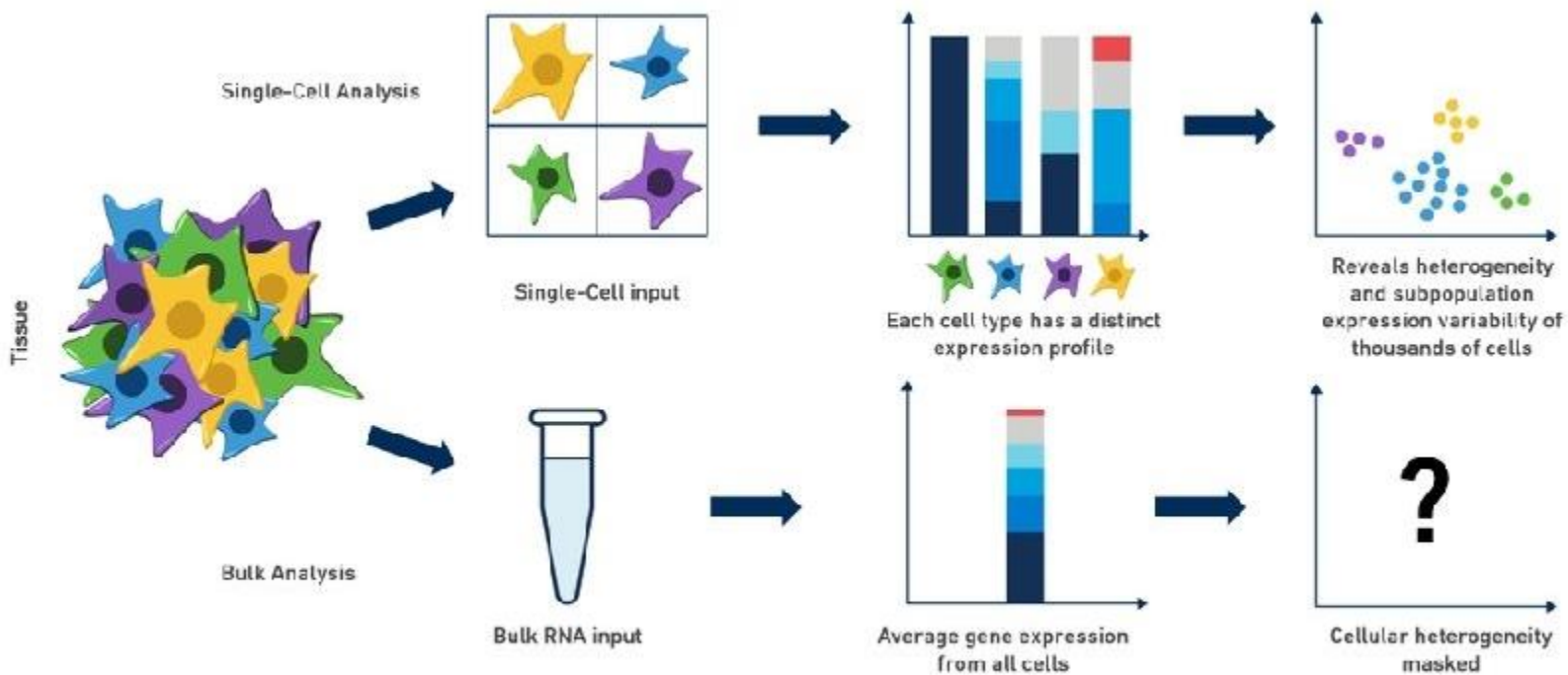
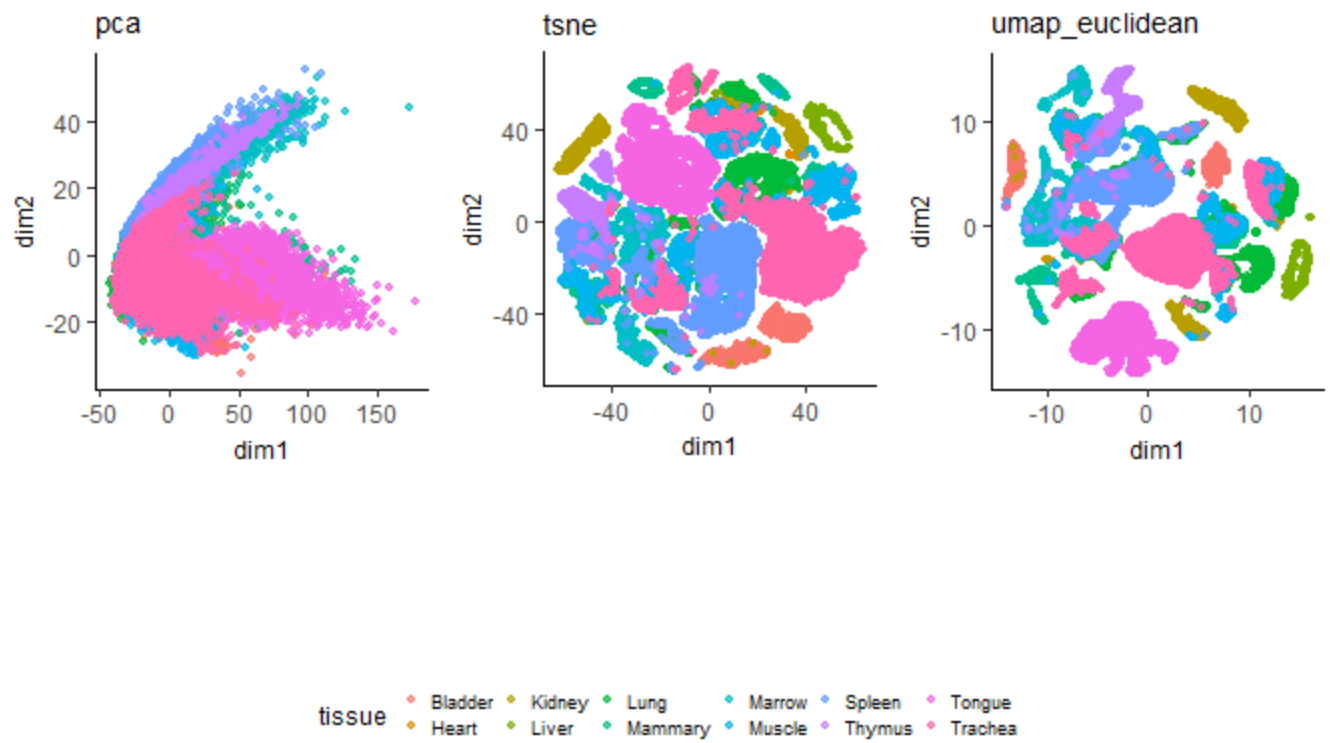
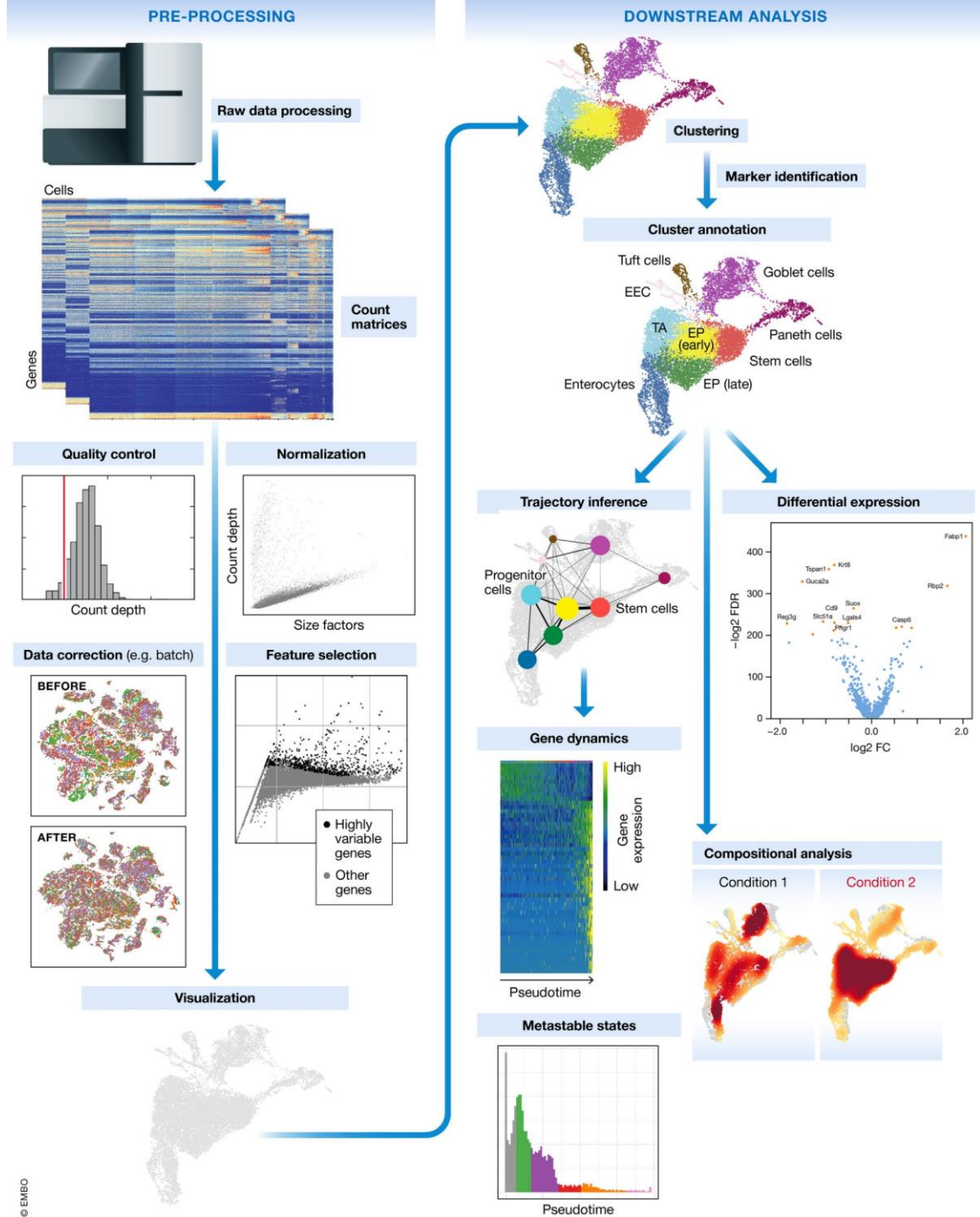
Para llevar a casa

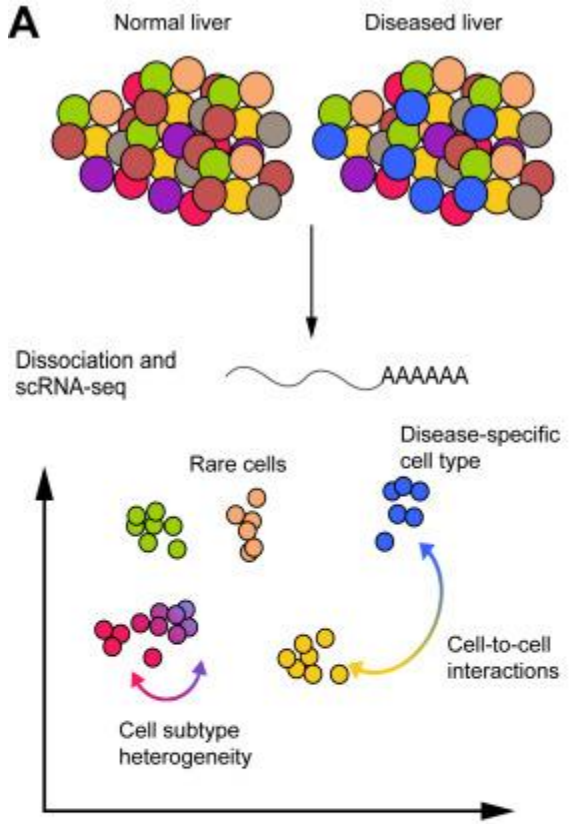
- Tres o más réplicas biológicas
- Haced réplicas técnicas (¡háh!)
- Las muestras control no son más simples, así que no uses menos muestras en ese grupo
- Planifica con tiempo. Determina bien qué quieres comparar y qué necesitarás.
- Idealmente iguala el número de sujetos en cada grupo
- Cuidado al analizar grupos pequeños. No vale comparar 20000 genes si tienes 5 muestras.
- Evita que el efecto batch afecte sólo a un grupo.
- Normaliza todo, batches incluidos, si puedes
- Corrige el p-valor
- No te lances, revisa los datos, cómo se comportan
- Prueba, cambia, revisa, tunea.
- Practica Python!

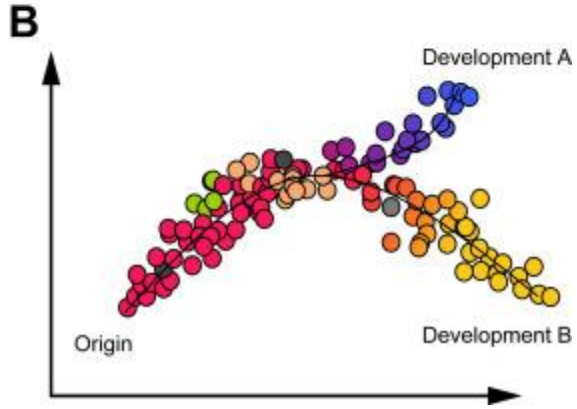

Single-Cell Analysis
Tissue
Single-Cell input
Each cell type has a distinct expression profile
Reveals heterogeneity and subpopulation expression variability of thousands of cells
Bulk Analysis
Bulk RNA input
Average gene expression from all cells
?
Cellular heterogeneity masked

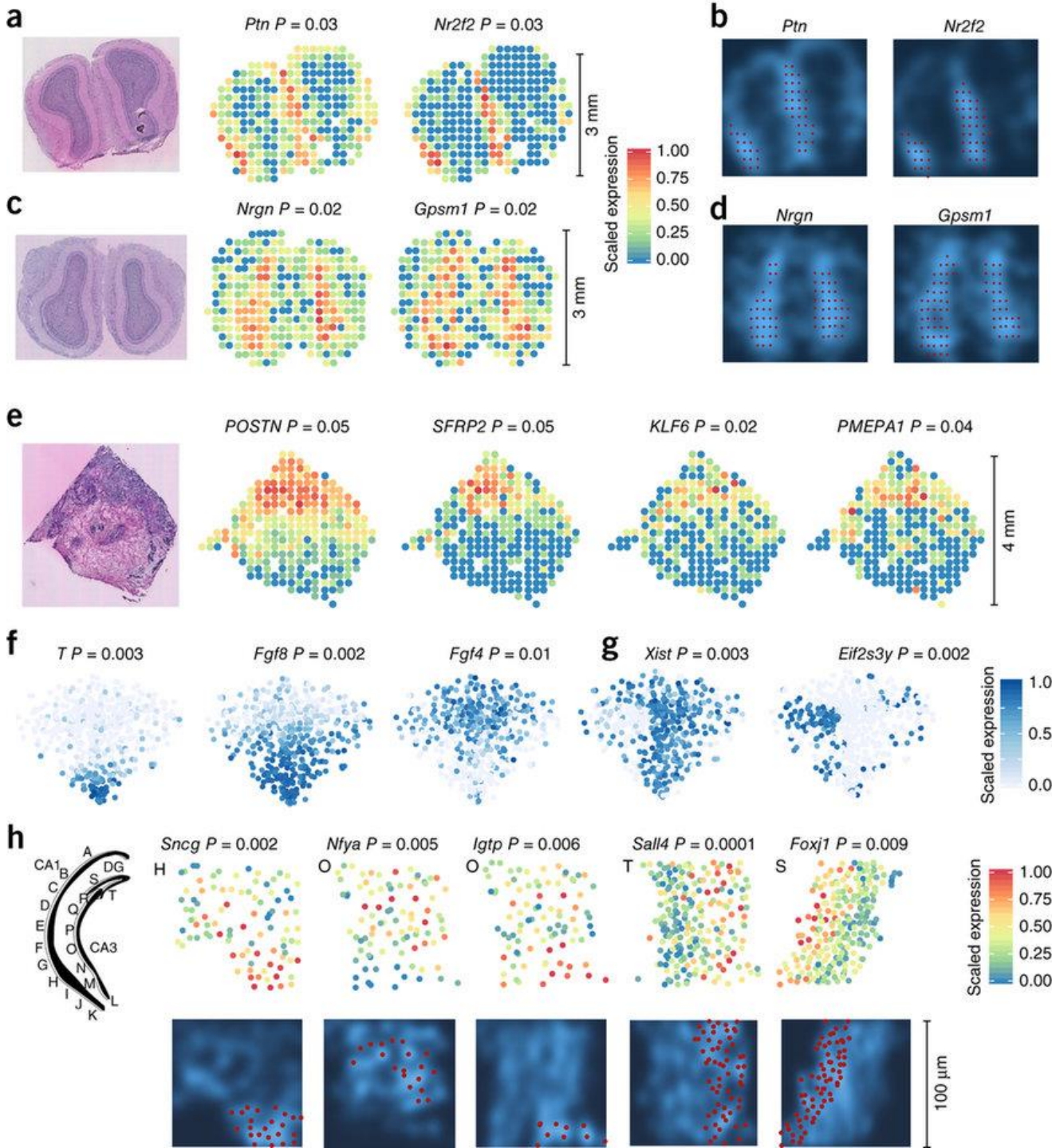
	Goal	Protocol	Quality control	Normalization	Analyses
Bulk RNA-seq	• Measure the average gene expression across the population of cells in a sample • To identify differences between sample conditions	• RNA is extracted from all cells in the sample • Reverse transcription converts RNA to cDNA, facilitates ligation of sequencing adaptors • Amplification	• GC content, presence of adaptors, overrepresented k-mers, duplicated reads • Percentage of reads that map to reference • Reproducibility between replicates	• Batch effect • Between-sample variability: sequencing depth Quantile normalization, spike-ins • Within-sample variability: feature length, library size effects RPKM, FPKM, TPM	• Estimate gene and transcript expression • Differential expression analysis • Alternative splicing
scRNA-seq	• Measure the gene expression of individual cells in a sample • To identify differences between cell types/states	• RNA is extracted from isolated cells, labeled with cell specific identifier • UMIs, spike-ins often included, to account for higher levels of noise • Reverse transcription, amplification similar to bulk protocol	• Reads, number of genes per cell • Percentage of reads that map to spike-ins (if used), percentage of reads that map to mitochondria • QC metrics used in bulk RNA-seq are also examined	• Batch effect and within-sample variability are corrected for similarly to bulk RNA-seq • Between-sample variability methods must additionally account for capture efficiency and dropout sources of noise	• Dimensionality reduction • Identify cell subpopulations • Differential expression • Pseudotime/ trajectory analysis

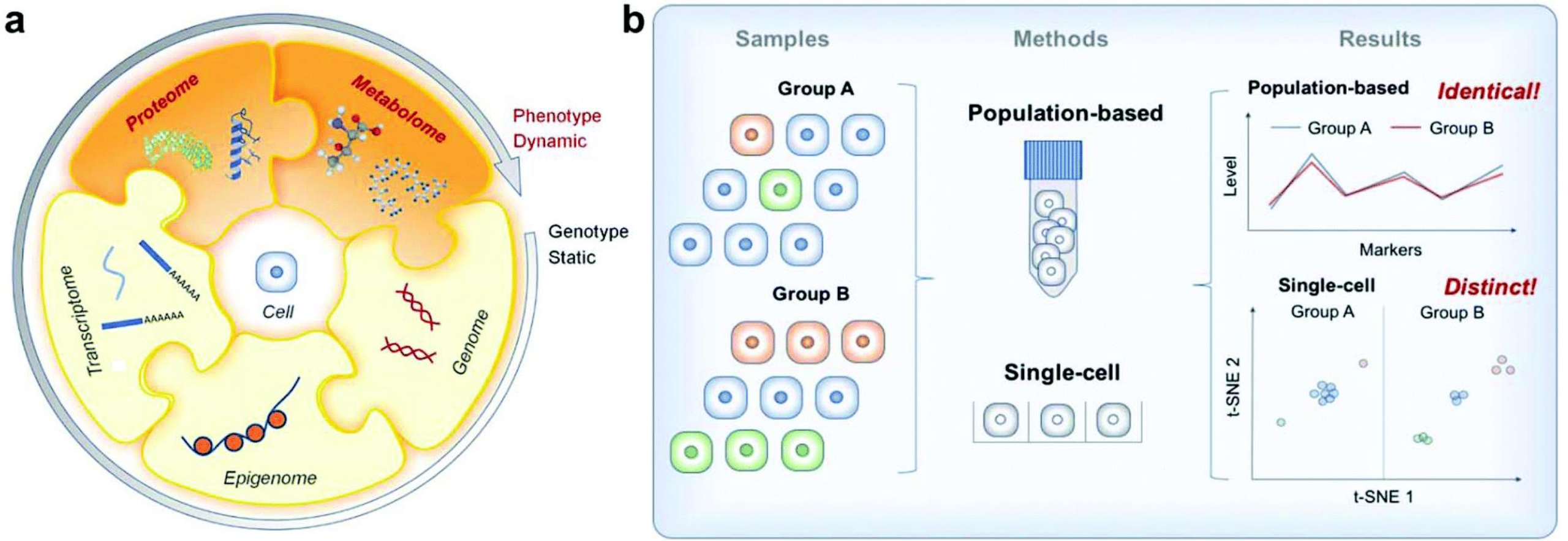

pca
tsne
umap_euclidean
dim1
dim2
tissue
Bladder
Heart
Kidney
Liver
Lung
Mammary
Marrow
Muscle
Spleen
Thymus
Tongue
Trachea


PRE-PROCESSING
DOWNSTREAM ANALYSIS
Raw data processing
Cells
Genes
Count matrices
Quality control
Count depth
Normalization
Count depth
Size factors
Data correction (e.g. batch)
BEFORE
AFTER
Feature selection
Highly variable genes
Other genes
Visualization
Clustering
Marker identification
Cluster annotation
Tuft cells
Goblet cells
EEC
TA
EP (early)
Paneth cells
Stem cells
Enterocytes
EP (late)
Trajectory inference
Progenitor cells
Stem cells
Differential expression
−log2 FDR
log2 FC
Gene dynamics
High
Gene expression
Low
Pseudotime
Compositional analysis
Condition 1
Condition 2
Metastable states
Pseudotime


A
Normal liver
Diseased liver
Dissociation and scRNA-seq
AAAAAA
Rare cells
Disease-specific cell type
Cell-to-cell interactions
Cell subtype heterogeneity


B
Development A
Origin
Development B


a
Ptn P = 0.03
Nr2f2 P = 0.03
3 mm
Scaled expression
1.00
0.75
0.50
0.25
0.00
b
Ptn
Nr2f2
c
Nrgn P = 0.02
Gpsm1 P = 0.02
3 mm
d
Nrgn
Gpsm1
e
POSTN P = 0.05
SFRP2 P = 0.05
KLF6 P = 0.02
PMEPA1 P = 0.04
4 mm
f
T P = 0.003
Fgf8 P = 0.002
Fgf4 P = 0.01
g
Xist P = 0.003
Eif2s3y P = 0.002
Scaled expression
1.0
0.5
0.0
h
Sncg P = 0.002
Nfya P = 0.005
Igtp P = 0.006
Sall4 P = 0.0001
Foxj1 P = 0.009
H
O
O
T
S
CA1
DG
CA3
Scaled expression
1.00
0.75
0.50
0.25
0.00
100 µm


a
Proteome
Metabolome
Phenotype
Dynamic
Genotype
Static
Transcriptome
AAAAAA
AAAAAA
Cell
Genome
Epigenome
b
Samples
Methods
Results
Group A
Group B
Population-based
Single-cell
Population-based
Identical!
Group A
Group B
Level
Markers
Single-cell
Distinct!
Group A
Group B
t-SNE 2
t-SNE 1


#becurious


bac
ZARAGOZA 2021
Online
thank you