

Single-cell multiome technology and analysis methods

Elizabeth Dorans

MPG Primer | January 16, 2025

Outline



Technology

What is single-cell technology?

What is single-cell multiome technology?



Analysis methods

What questions can single-cell multiome data answer?

Technical challenges in single-cell multiome data analysis

Example: linking enhancers to genes

Outline

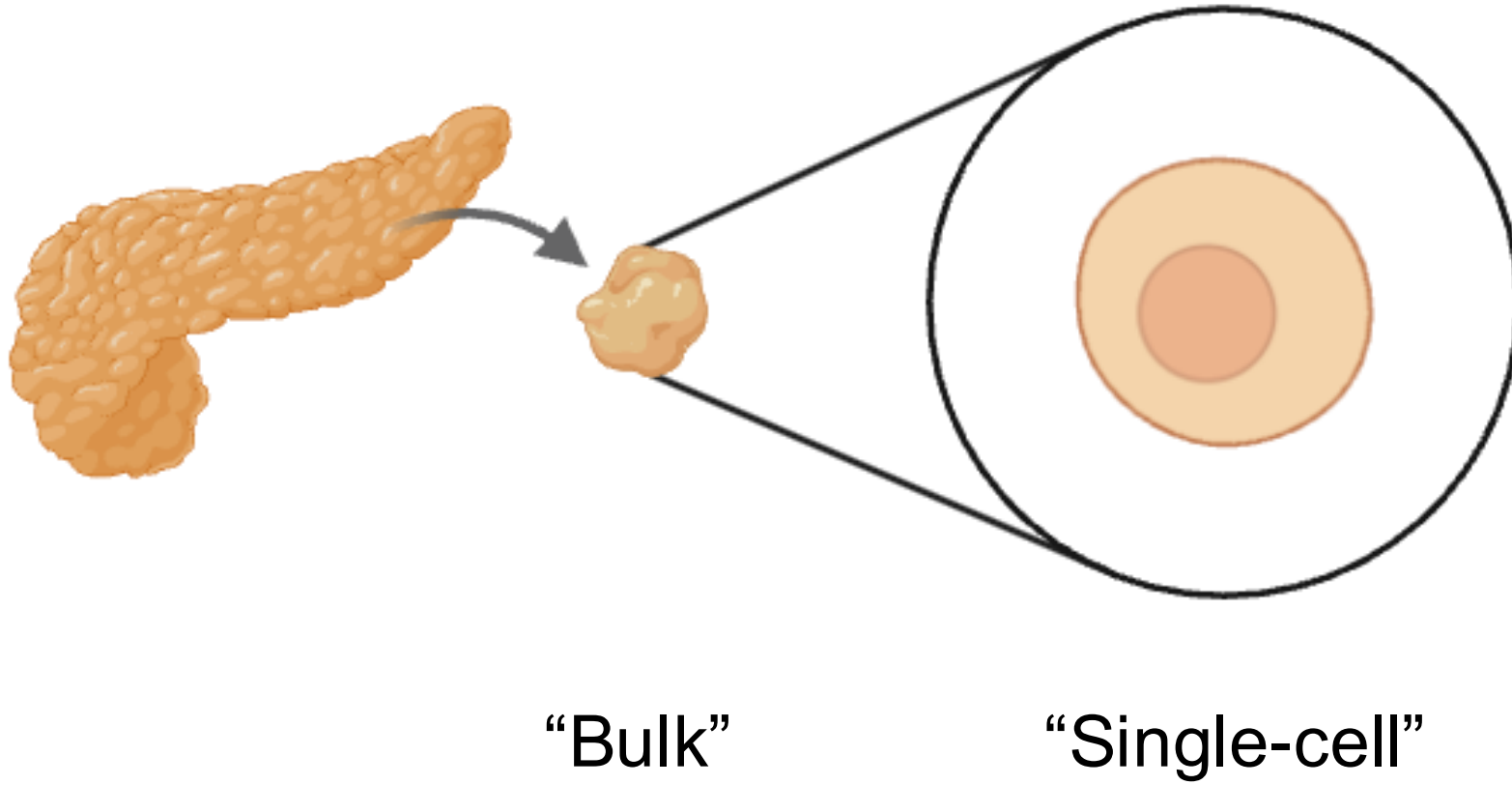


Technology

What is single-cell technology?

Analysis
methods

Single-cell technologies measure biology at the level of individual cells



Why single-cell?

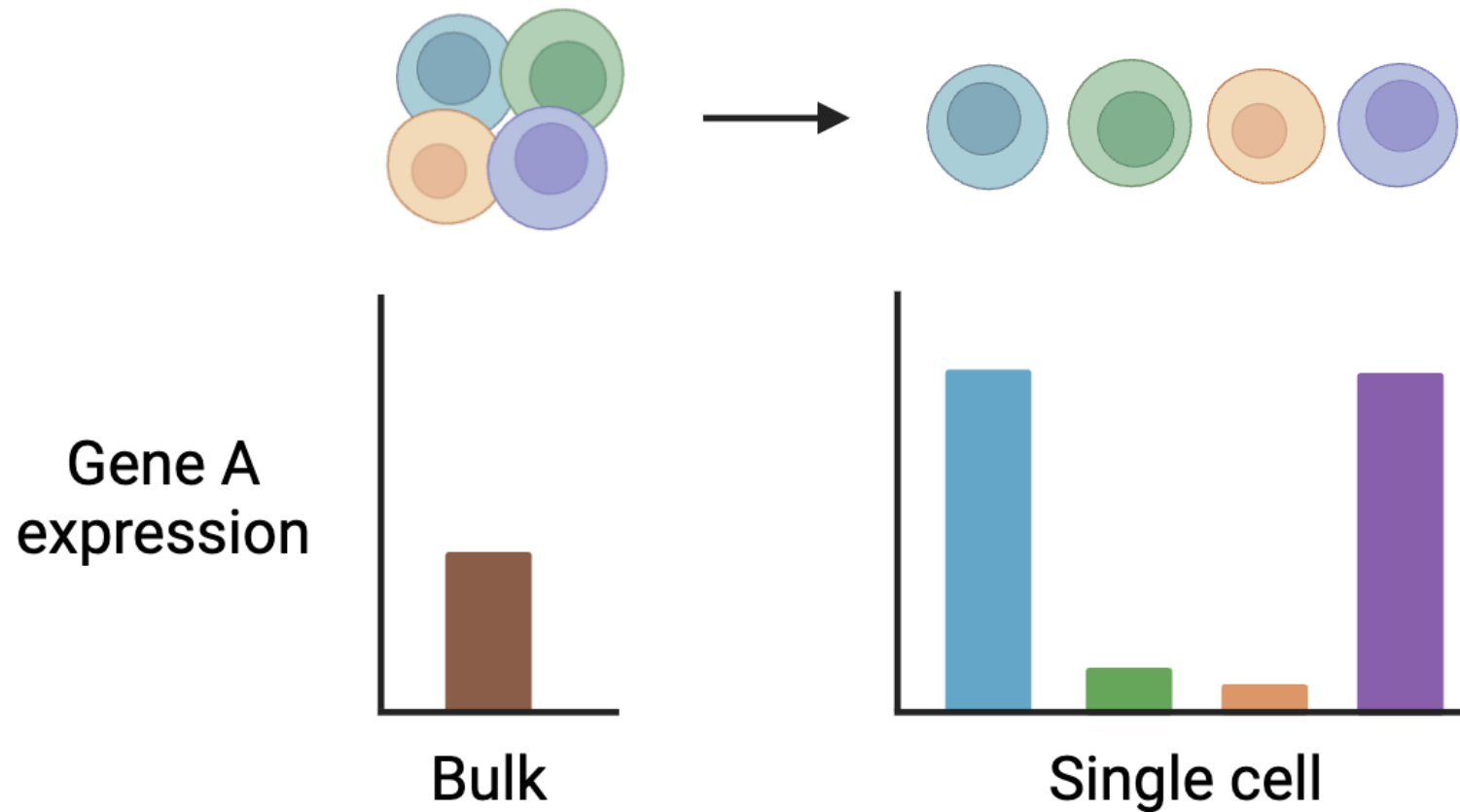


“Bulk” data = fruit smoothie



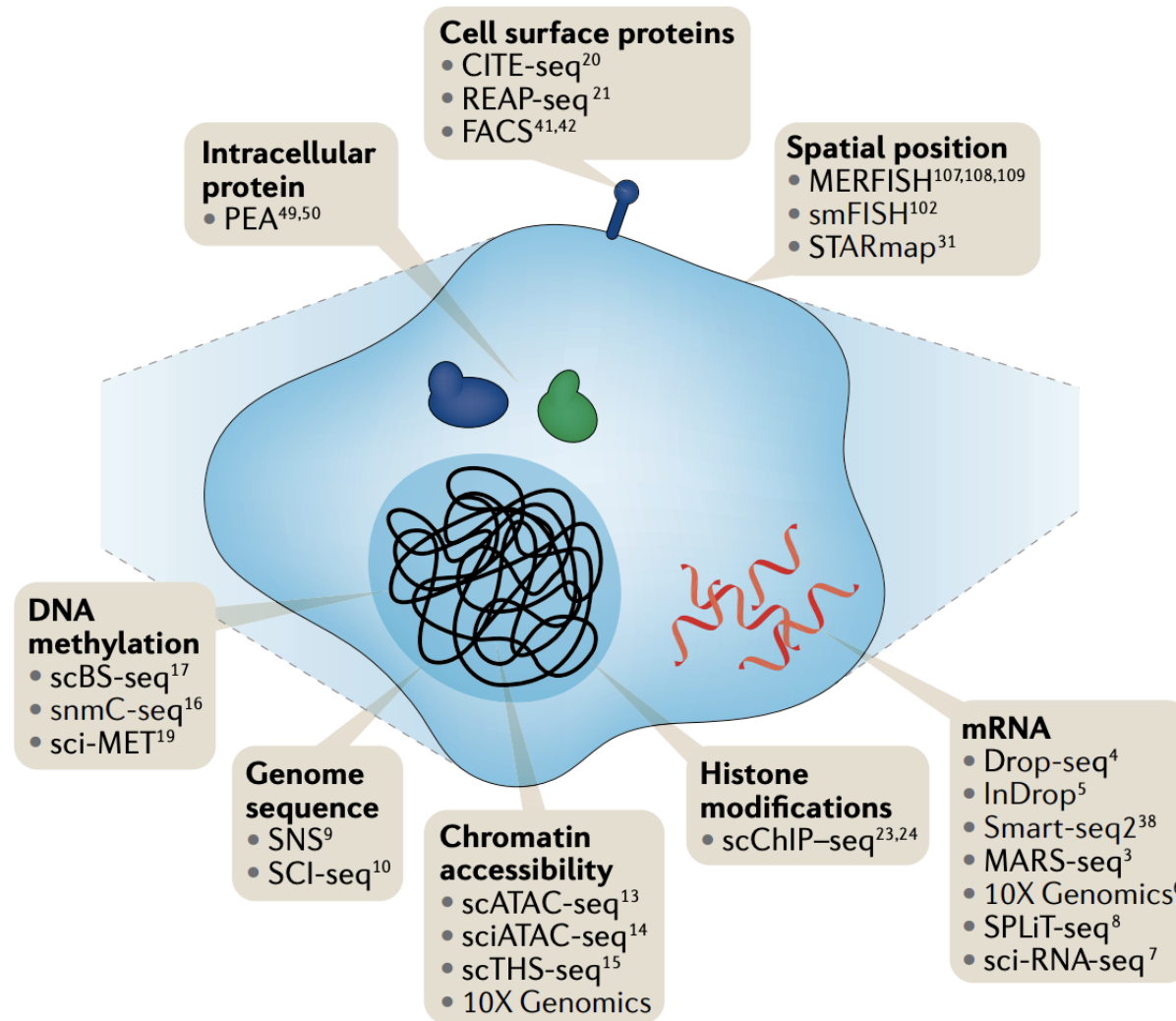
Single-cell data = fruit salad

Why single-cell?



Single-cell data retains cellular heterogeneity

Single-cell genomic technologies measure a variety of biological modalities



Outline



Technology

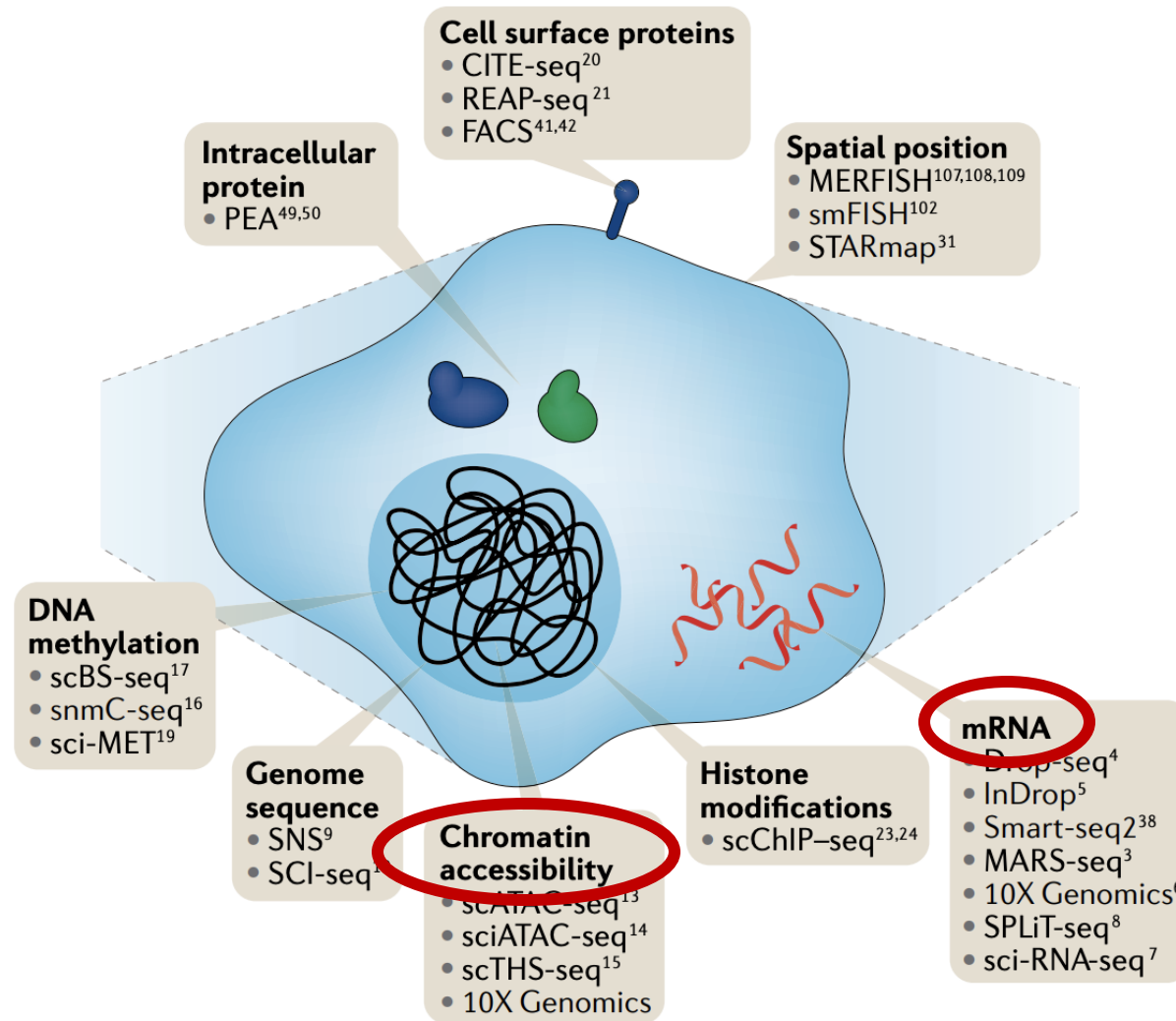
What is single-cell technology?

What is single-cell multiome technology?



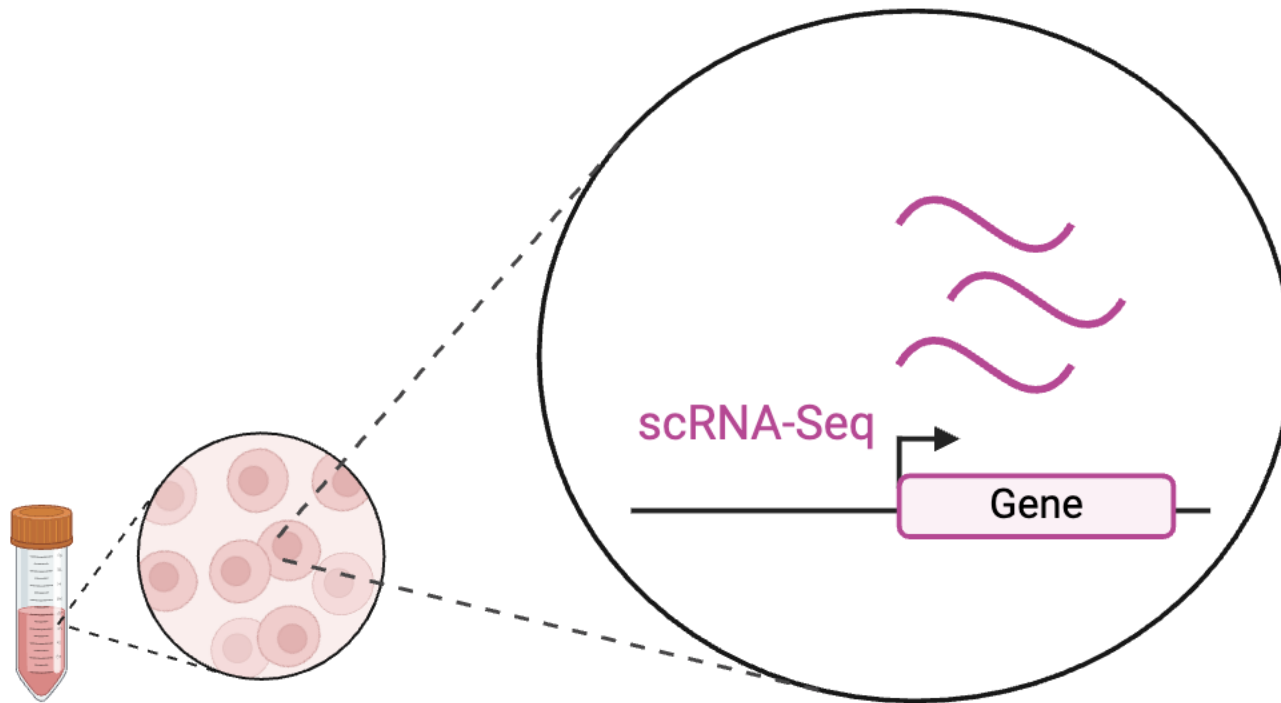
Analysis methods

Single-cell multiome technologies jointly measure multiple modalities

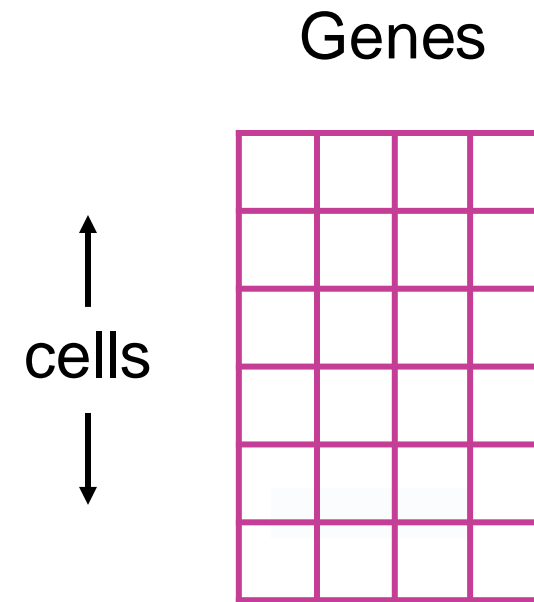


scRNA-Seq measures gene expression

Primers target mRNA for sequencing

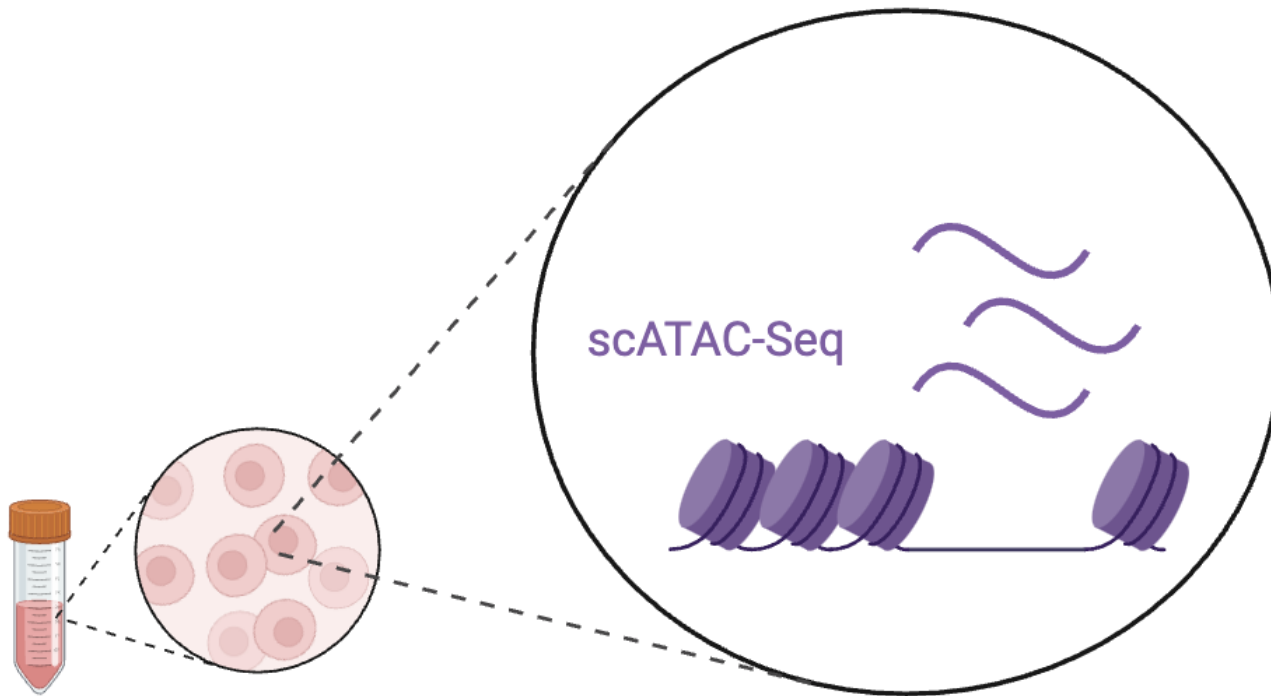


Data: expression matrix



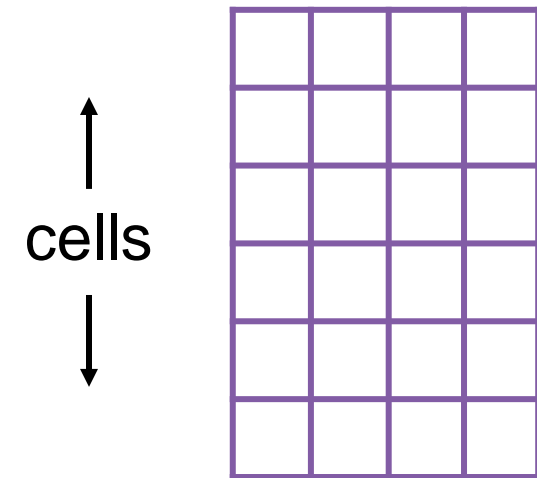
scATAC-Seq measures chromatin accessibility

Transposase enzyme targets regions of open chromatin for sequencing

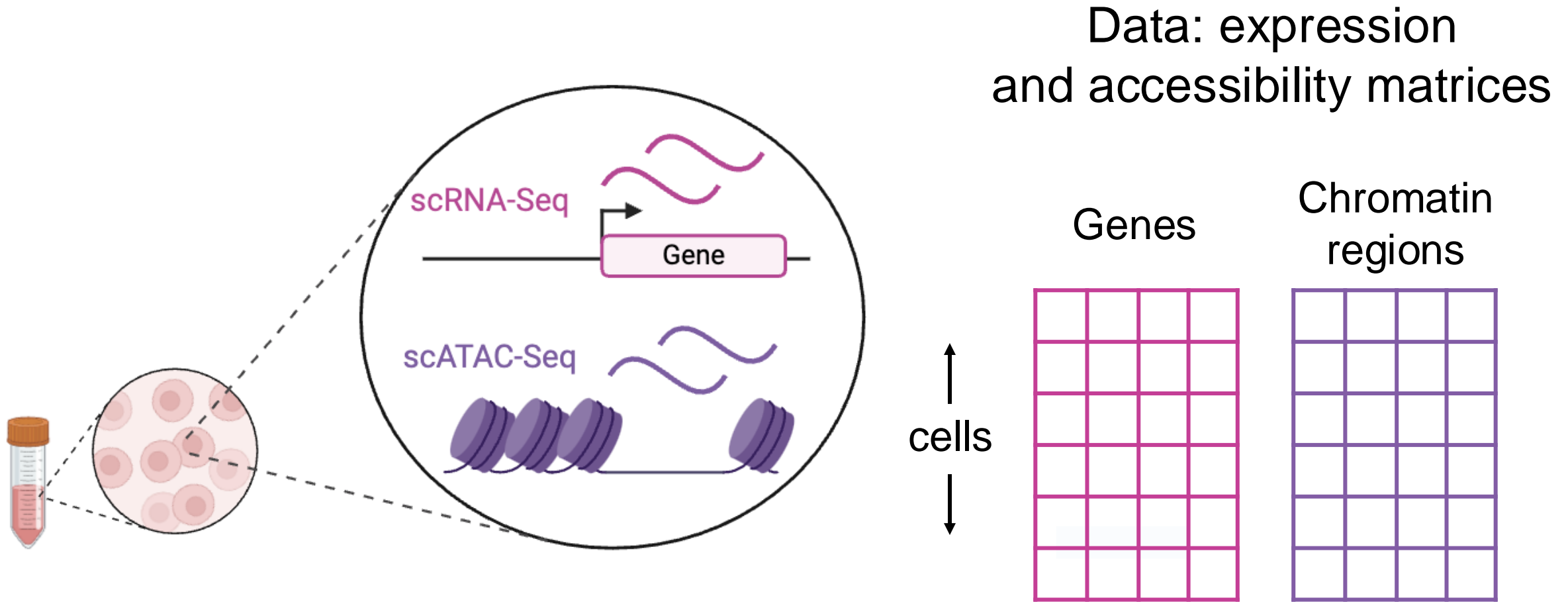


Data: accessibility matrix

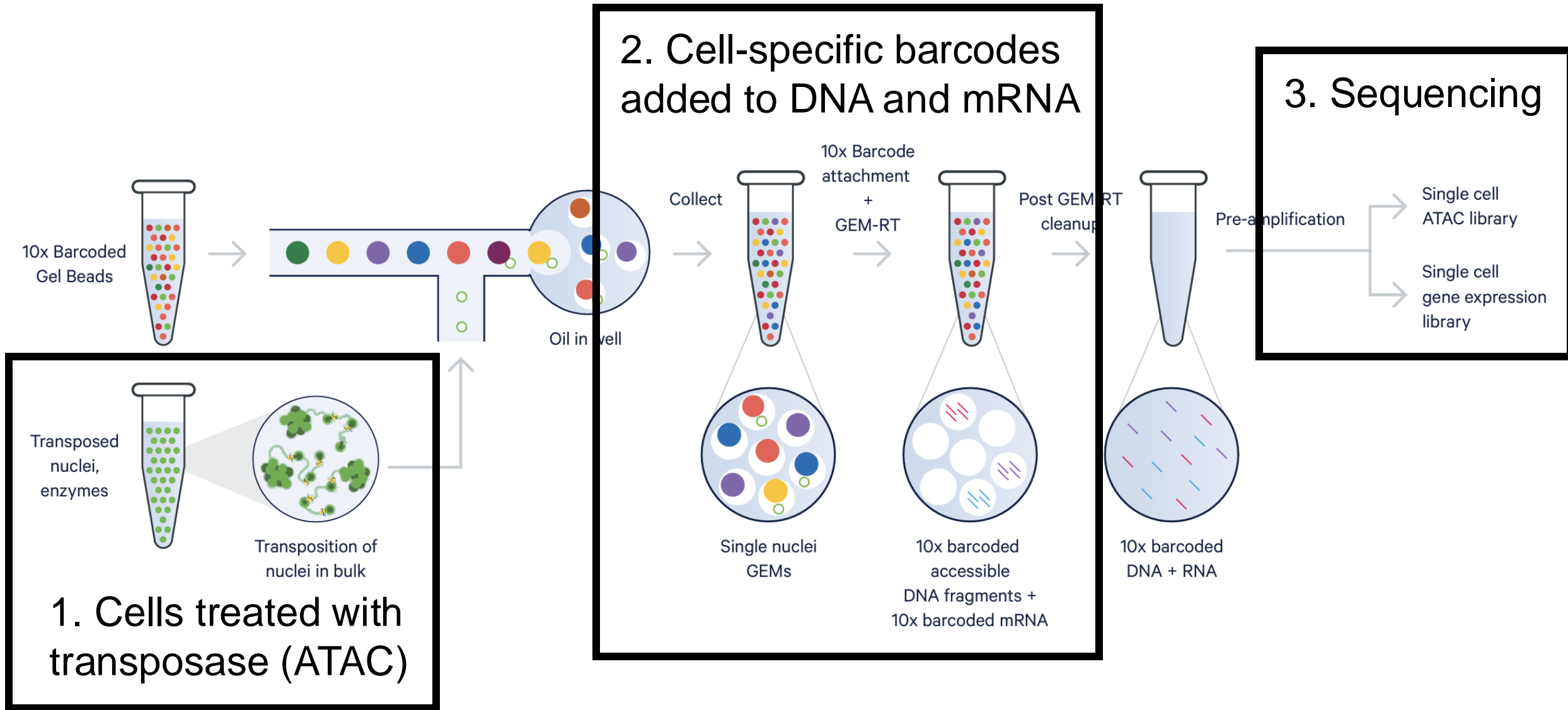
Chromatin regions
(e.g. peaks, tiles)



scRNA/ATAC-Seq jointly measures RNA + ATAC



10X multiome technology jointly profiles single-cell RNA + ATAC



Outline



Technology

What is single-cell technology?

What is single-cell multiome technology?



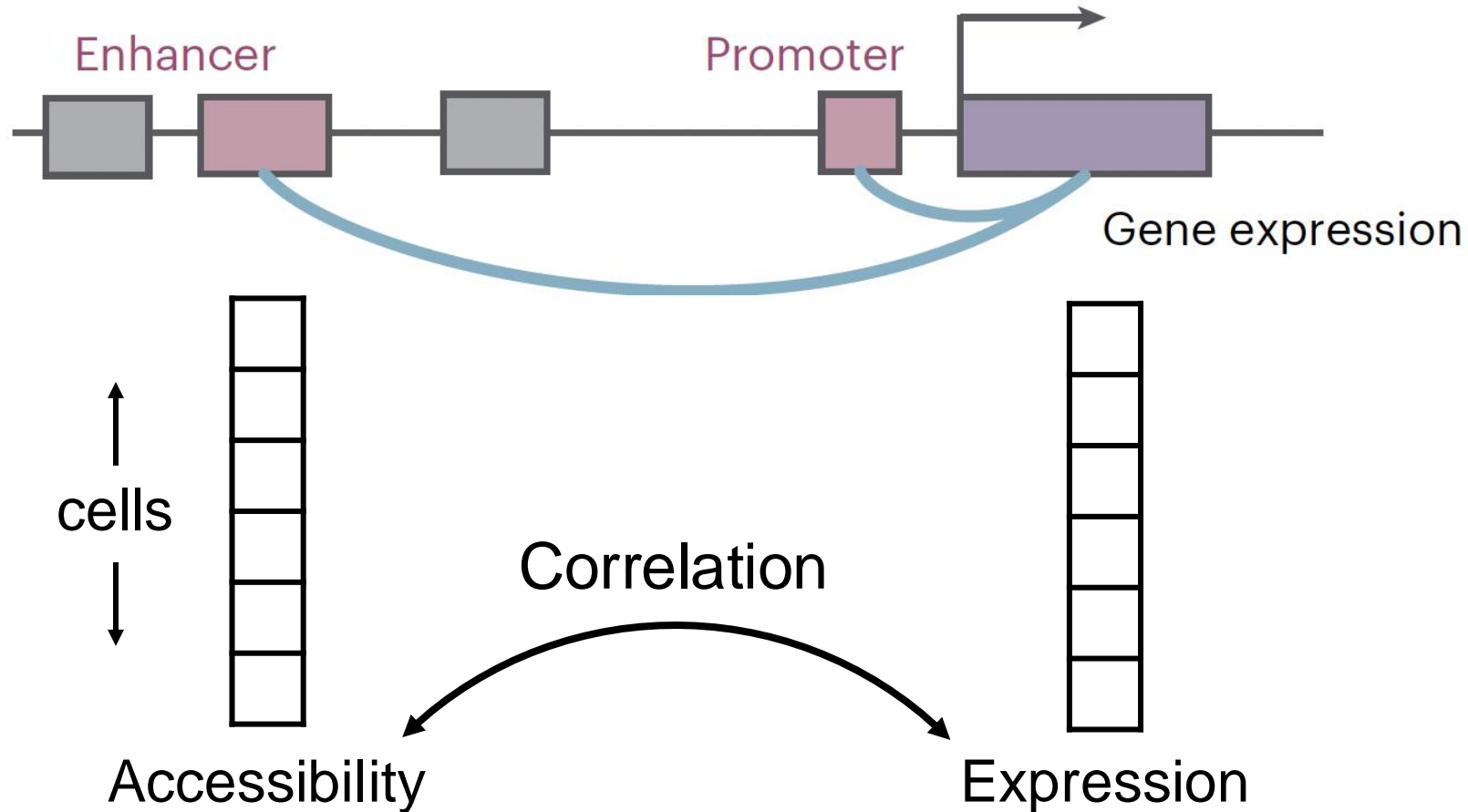
**Analysis
methods**

What questions can single-cell multiome data answer?

Biological applications

- Linking regulatory elements to target genes (and upstream TFs)
- Inferring developmental trajectories
- Discovery of rare or transient cell types / states
- Using one modality to predict another

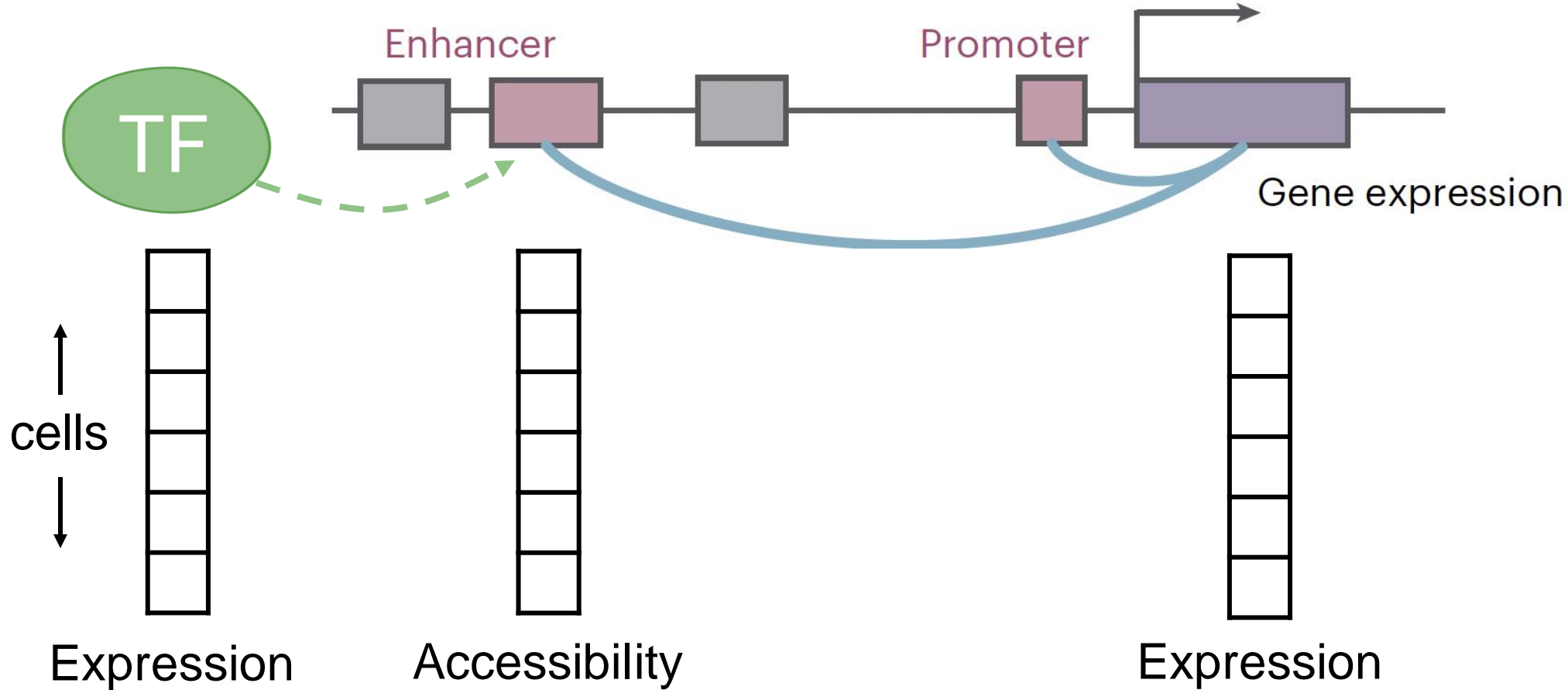
Linking regulatory elements to their target genes



Examples: SCENT (Sakaue 2024 Nat Genet), SCARlink (Mitra 2024 Nat Genet), pgBoost (Dorans *In Press* Nat Genet), Signac (Stuart 2021 Nat Methods), ArchR (Granja 2021 Nat Genet)

adapted from Sakaue 2024 Nat Genet (Fig. 1a)

Linking regulatory elements to their target genes and upstream TFs




Examples: SCENIC+ (González-Blas 2023 Nat Methods), REUNION (Yang Bioinformatics 2024)

adapted from Sakaue 2024 Nat Genet (Fig. 1a)

Inferring developmental trajectories

Chromatin potential / lineage priming: increase in accessibility of regulatory chromatin prior to a gene being “turned on” during development

Wnt3

norm. counts
0  1



introns
exons

Gene expression

Early development



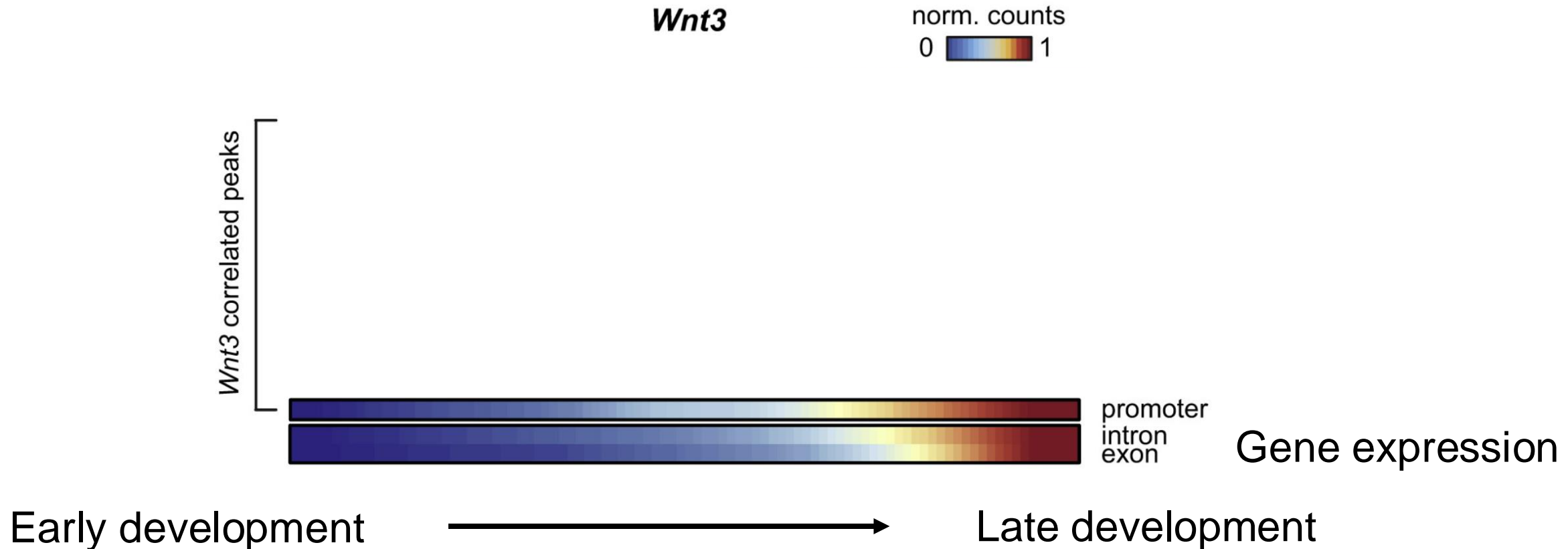
Late development

Examples: Ma 2020 Cell, Mitra 2024 Nat Genet

adapted from Ma 2020 Cell (Fig. 4E)

Inferring developmental trajectories

Chromatin potential / lineage priming: increase in accessibility of regulatory chromatin prior to a gene being “turned on” during development

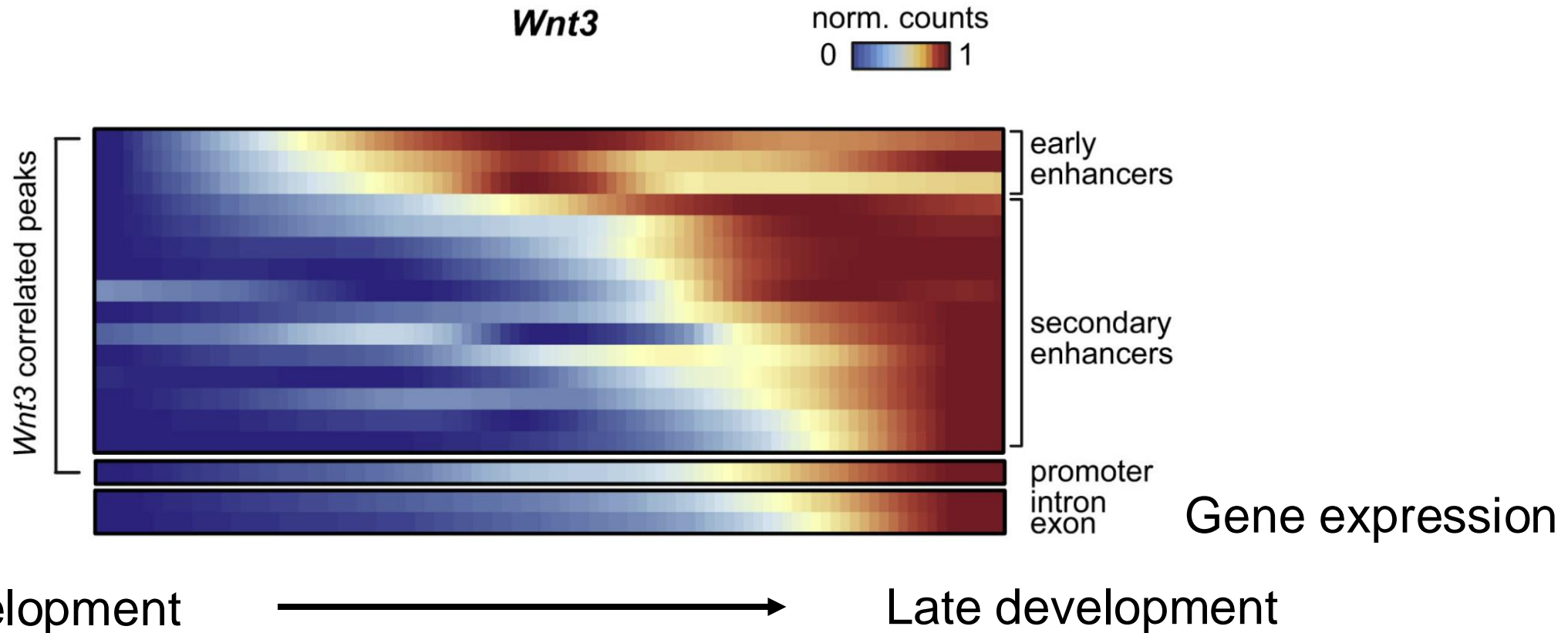


Examples: Ma 2020 Cell, Mitra 2024 Nat Genet

adapted from Ma 2020 Cell (Fig. 4E)

Inferring developmental trajectories

Chromatin potential / lineage priming: increase in accessibility of regulatory chromatin prior to a gene being “turned on” during development

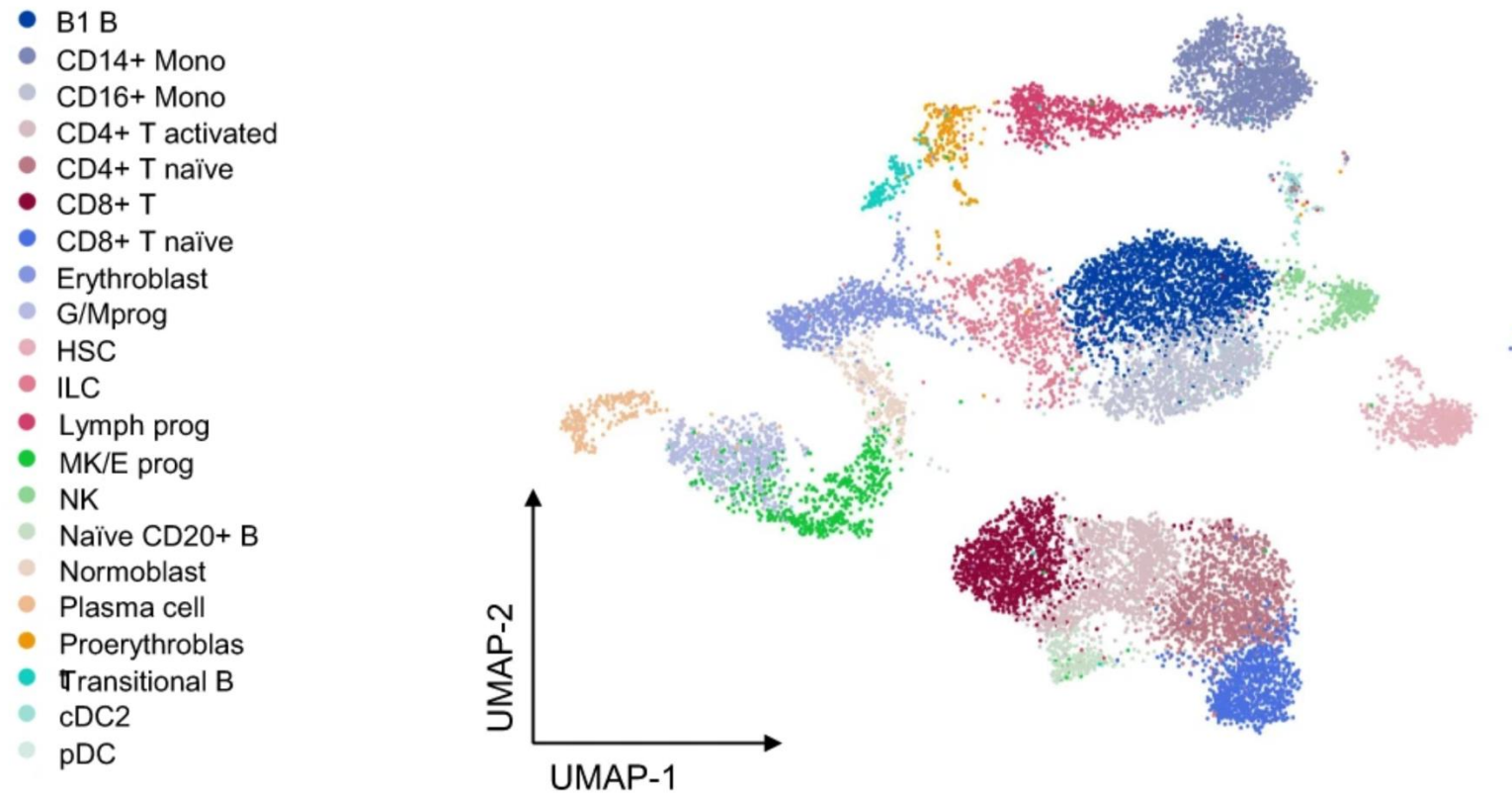


Examples: Ma 2020 Cell, Mitra 2024 Nat Genet

adapted from Ma 2020 Cell (Fig. 4E)

Discovery of rare or transient cell types / states

Dimension reduction + clustering used to define cell types / states from unimodal data
Multimodal data increases biological resolution, enables joint dimension reduction

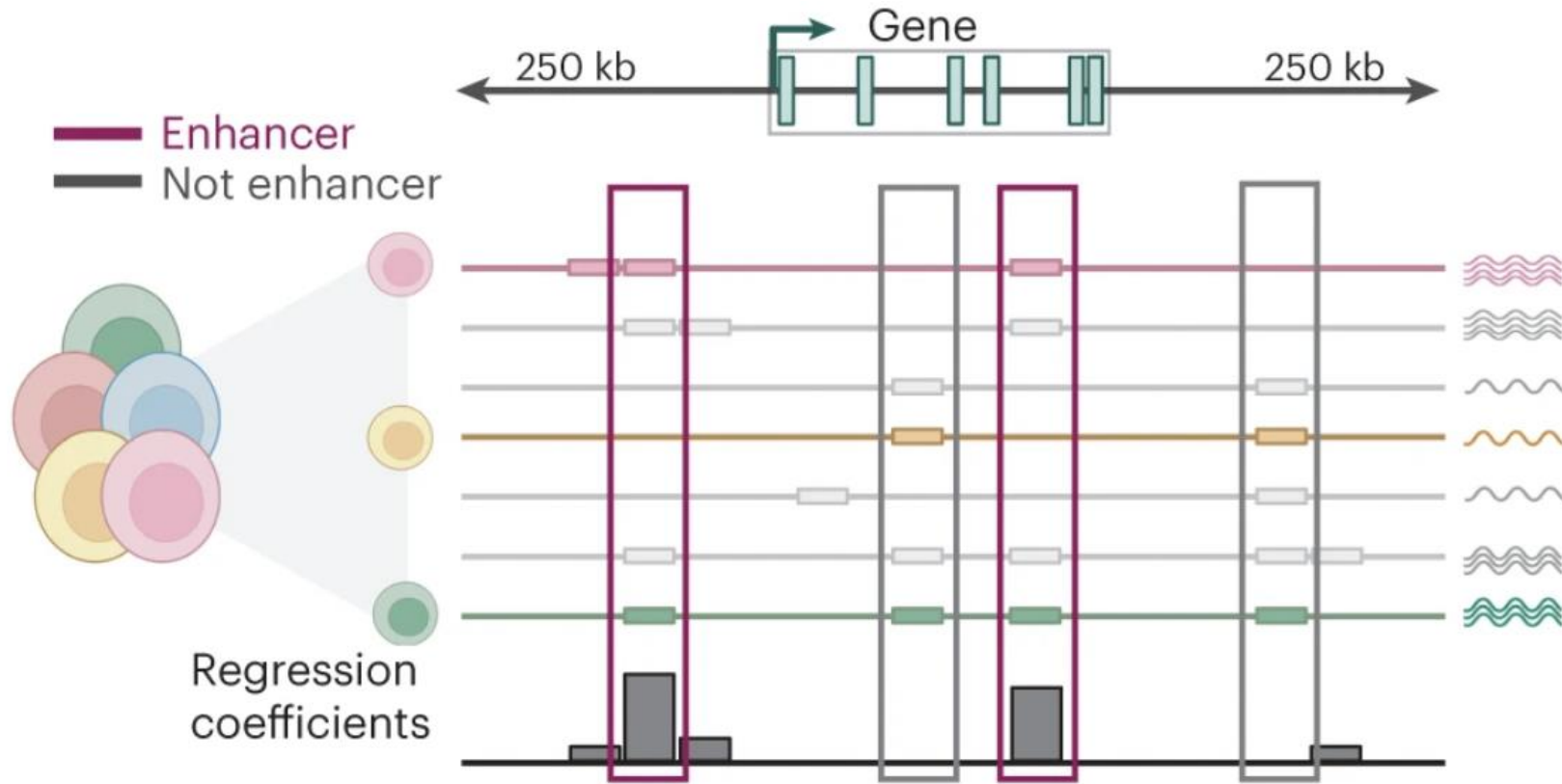


Examples: MarsGT (Wang 2024 Nat Comm), ArchR (Granja 2021 Nat Genet)

Adapted from Wang 2024 Nat Comm (Fig. 2d)

Using one modality to predict another

Model gene expression as a function of *cis*-chromatin accessibility
→ predict gene expression from chromatin accessibility alone



Example: SCARlink (Mitra 2024 Nat Genet)

Mitra 2024 Nat Genet (Fig. 1a)

Outline



Technology

What is single-cell technology?

What is single-cell multiome technology?



Analysis methods

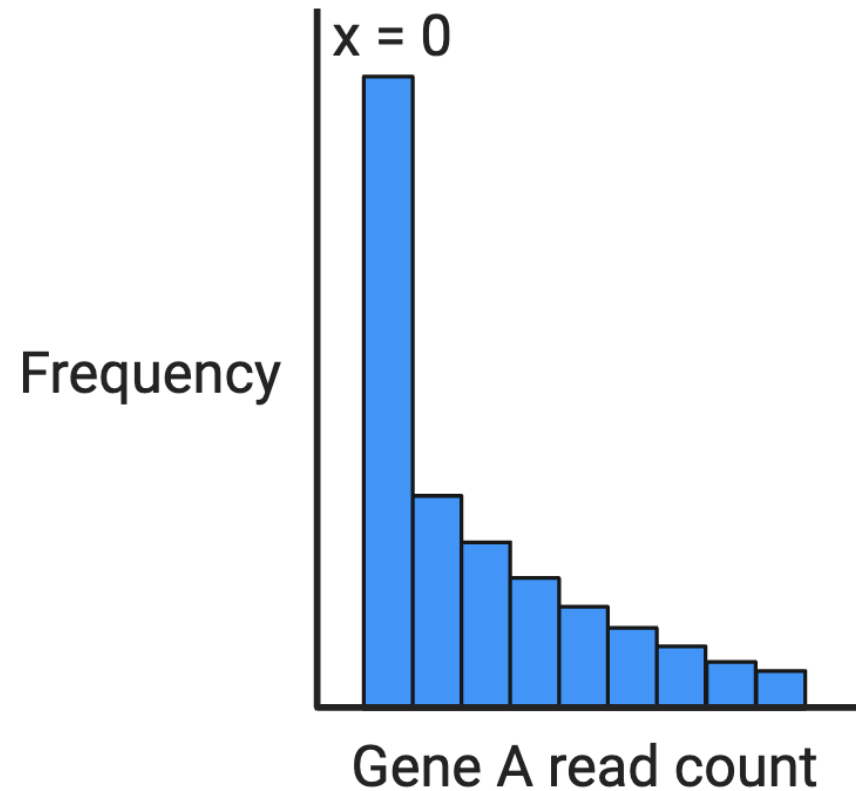
What questions can single-cell multiome data answer?

Technical challenges in single-cell multiome data analysis

Sparsity and noise

Challenge:

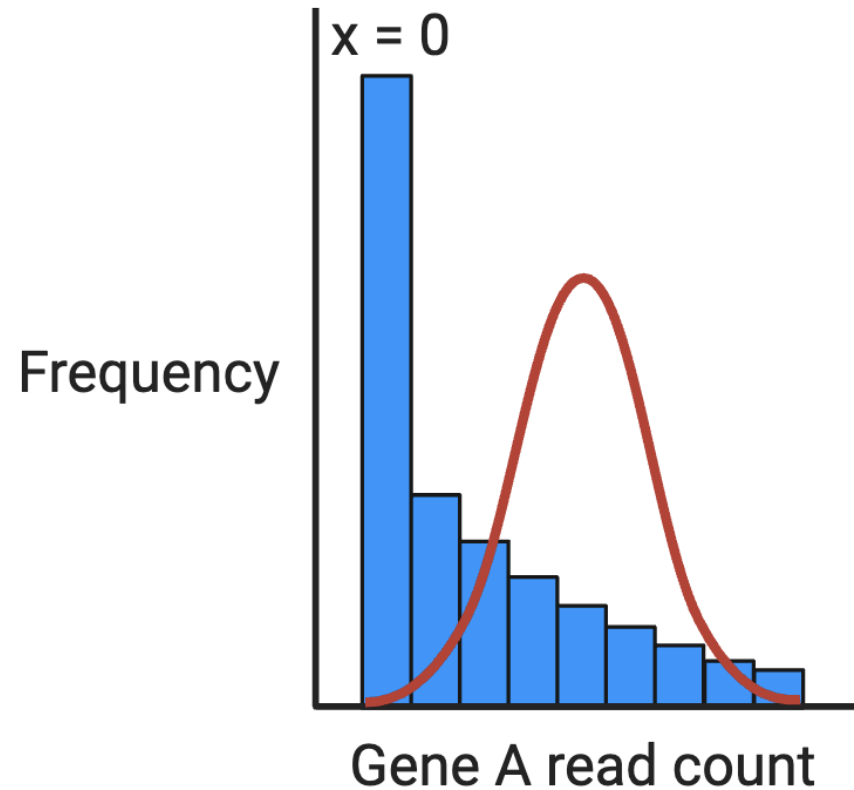
- Single-cell data is sparse (many zeros) and noisy
 - Due to biological variation and incomplete sampling (“drop out”)



Sparsity and noise

Challenge:

- Single-cell data is sparse (many zeros) and noisy
 - Due to biological variation and incomplete sampling (“drop out”)
- Many regression-based models assume normality



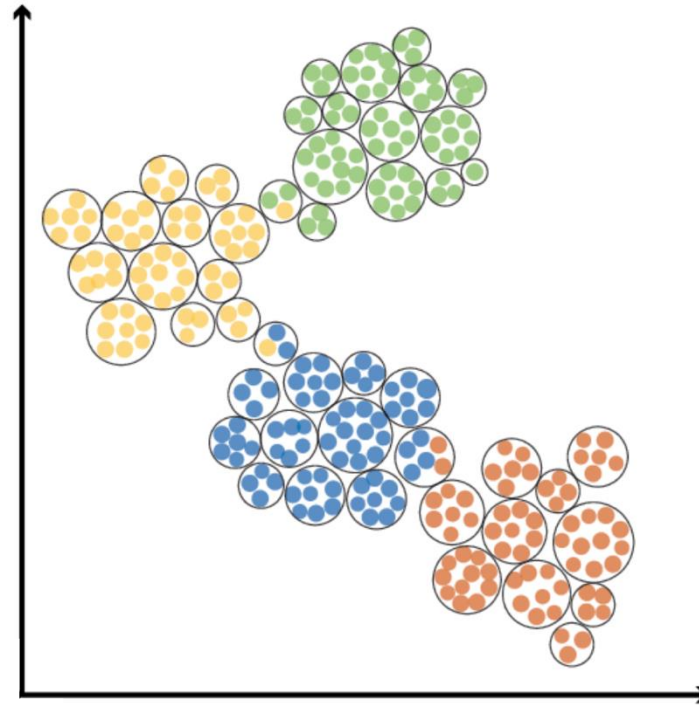
Sparsity and noise

Challenge:

- Single-cell data is sparse (many zeros) and noisy
 - Due to biological variation and incomplete sampling (“drop out”)
- Many regression-based models assume normality

Solutions:

- Normal model
 - Aggregate single cells into metacells
 - ArchR (Granja 2021 Nat Genet)
 - SEACells (Persad 2023 Nat Biotech)
 - SuperCell (Bilous 2022 Bioinformatics)



adapted from Bilous 2024 Mol Systems Bio (Fig. 1)

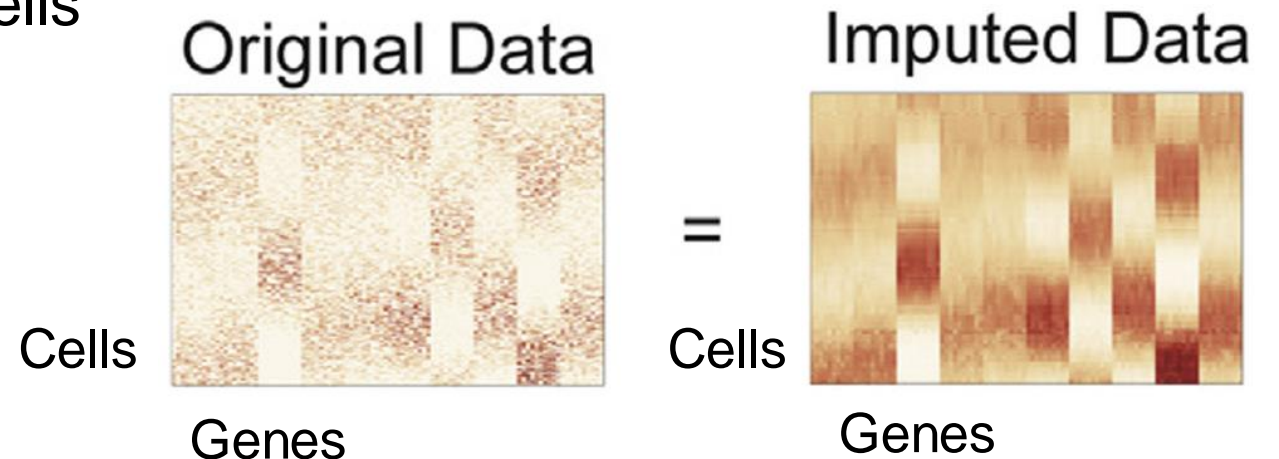
Sparsity and noise

Challenge:

- Single-cell data is sparse (many zeros) and noisy
 - Due to biological variation and incomplete sampling (“drop out”)
- Many regression-based models assume normality

Solutions:

- Normal model
 - Aggregate single cells into metacells
 - Impute “missing” counts
MAGIC (van Dijk 2018 Cell)



adapted from van Dijk 2018 Cell (Fig. 1)

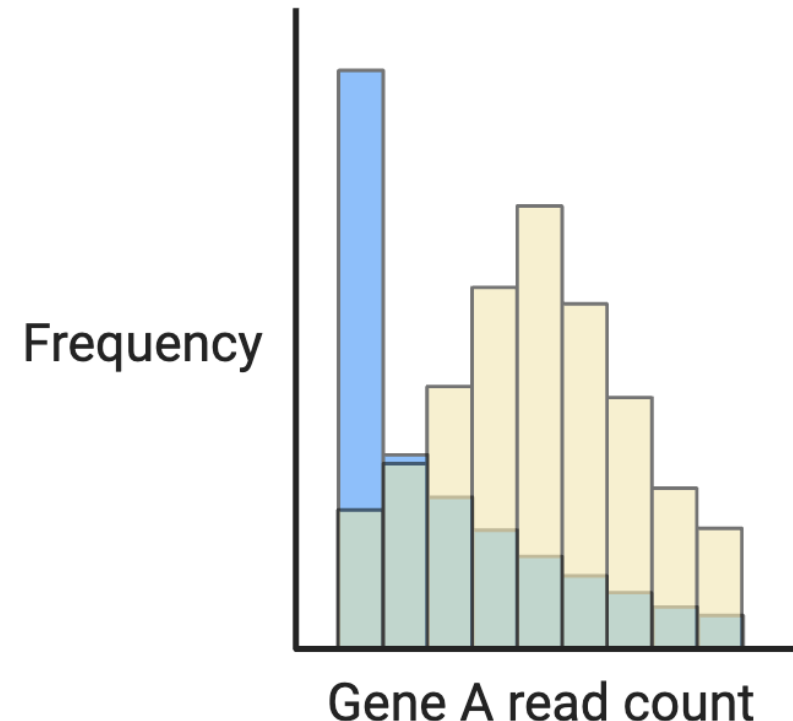
Sparsity and noise

Challenge:

- Single-cell data is sparse (many zeros) and noisy
 - Due to biological variation and incomplete sampling (“drop out”)
- Many regression-based models assume normality

Solutions:

- Normal model
 - Aggregate single cells into metacells
 - Impute “missing” counts
 - Normalize / transform counts
 - Signac (Stuart 2021 Nat Methods)
 - ArchR (Granja 2021 Nat Genet)



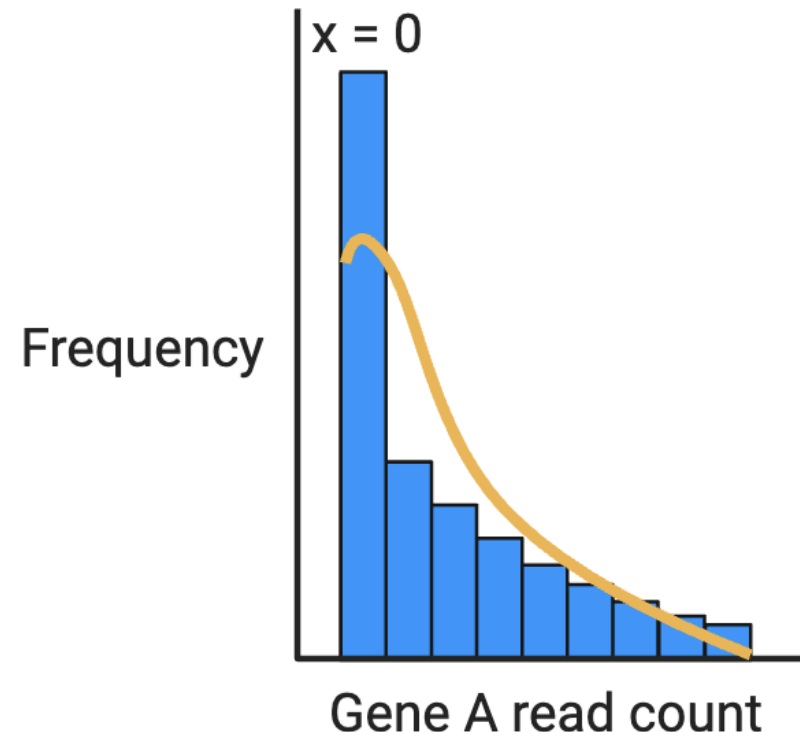
Sparsity and noise

Challenge:

- Single-cell data is sparse (many zeros) and noisy
 - Due to biological variation and incomplete sampling (“drop out”)
- Many regression-based models assume normality

Solutions:

- Normal model
 - Aggregate single cells into metacells
 - Impute “missing” counts
 - Normalize / transform counts
- Nonnormal model
 - Discrete count models
 - e.g. Poisson, negative binomial
 - SCARlink (Mitra 2024 Nat Genet)
 - SCENT (Sakaue 2024 Nat Genet)



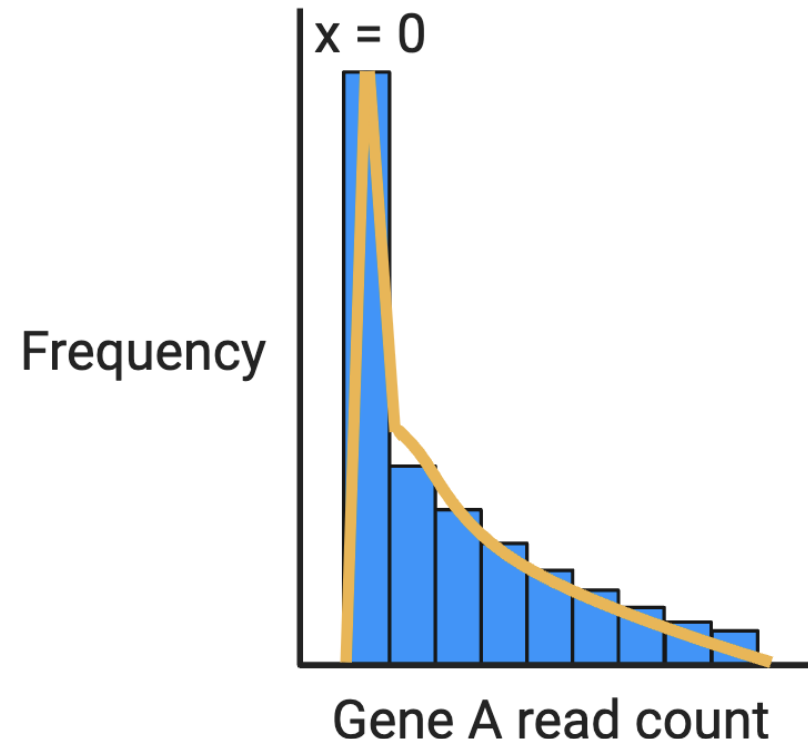
Sparsity and noise

Challenge:

- Single-cell data is sparse (many zeros) and noisy
 - Due to biological variation and incomplete sampling (“drop out”)
- Many regression-based models assume normality

Solutions:

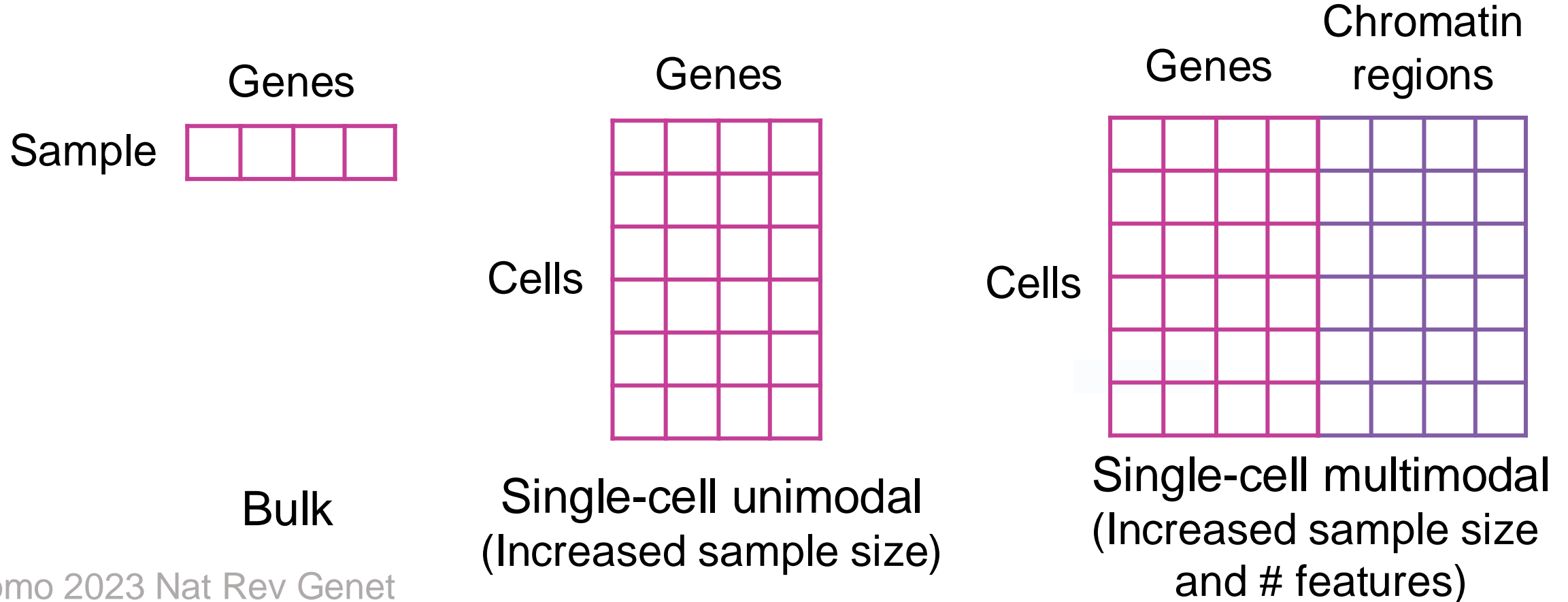
- Normal model
 - Aggregate single cells into metacells
 - Impute “missing” counts
 - Normalize / transform counts
- Nonnormal model
 - Discrete count models
 - e.g. Poisson, negative binomial
 - Zero inflated models
 - e.g. ZIP, ZINB



High dimensionality

Challenge:

- Single-cell data is highly dimensional (many data points)
- May increase computational complexity and hinder interpretability



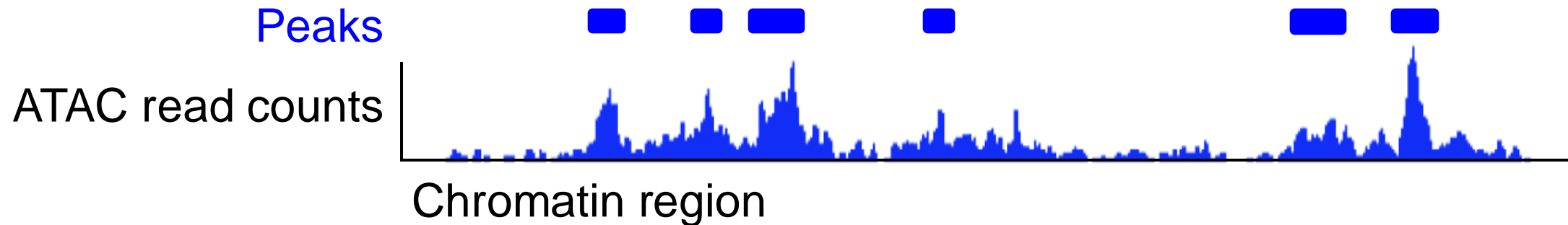
High dimensionality

Challenge:

- Single-cell data is highly dimensional (many data points)
- May increase computational complexity and hinder interpretability

Solutions:

- Reducing number of features
 - Feature selection
 - RNA: variable genes
 - ATAC: peaks, tiles
 - macs (Zhang 2008 Genome Biol)



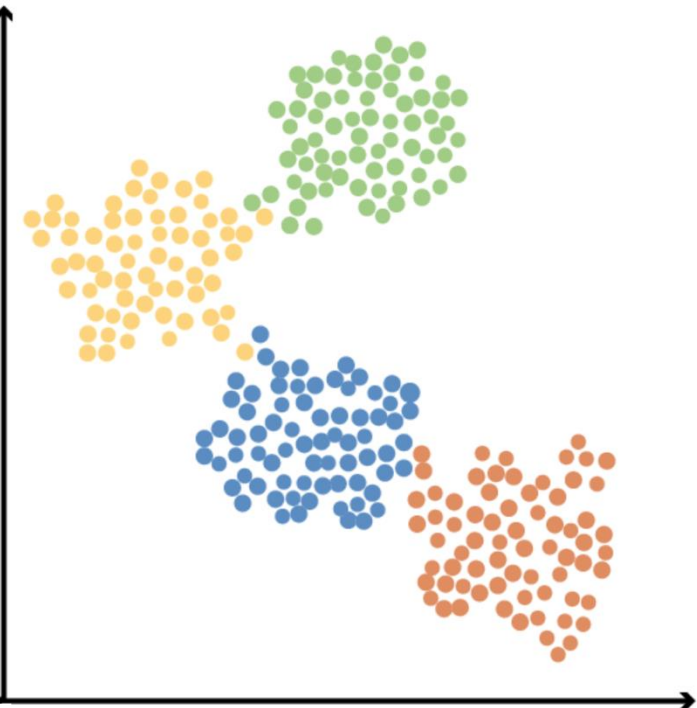
High dimensionality

Challenge:

- Single-cell data is highly dimensional (many data points)
- May increase computational complexity and hinder interpretability

Solutions:

- Reducing number of features
 - Feature selection
 - RNA: variable genes
 - ATAC: peaks, tiles
 - Dimension reduction
 - PCA
 - NMF



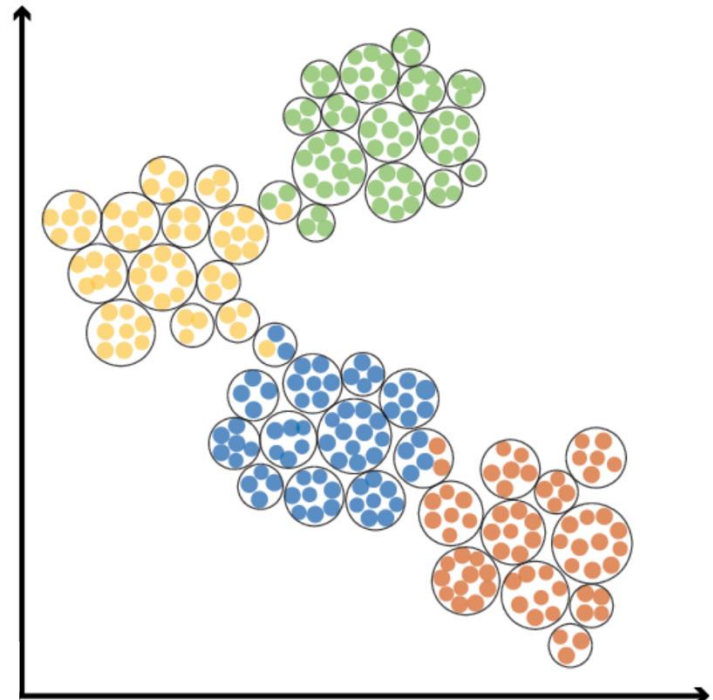
High dimensionality

Challenge:

- Single-cell data is highly dimensional (many data points)
- May increase computational complexity and hinder interpretability

Solutions:

- Reducing number of features
 - Feature selection
 - RNA: variable genes
 - ATAC: peaks, tiles
 - Dimension reduction
 - PCA
 - NMF
- Reducing sample size
 - Aggregate single cells into metacells
 - See previous citations



Outline

Technology

What is single-cell technology?

What is single-cell multiome technology?

Analysis methods

What questions can single-cell multiome data answer?

Technical challenges in single-cell multiome data analysis

Example: linking enhancers to genes

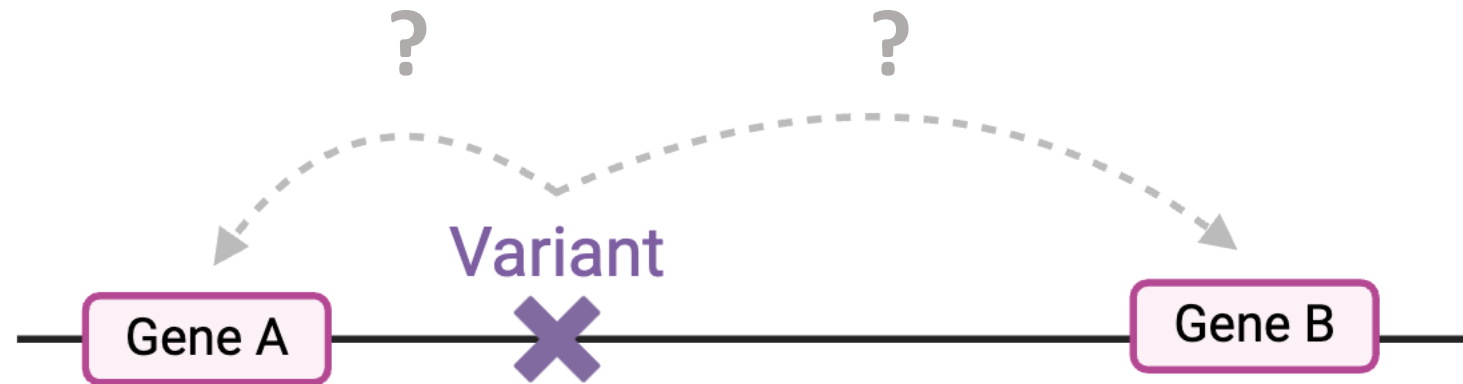
Mapping gene regulatory architectures is critical to understanding disease

Most disease-associated variants are noncoding.

(Hindorff 2009 *PNAS*, Nasser 2021 *Nature*, Boix 2021 *Nature*)

Noncoding variants often do not regulate the nearest gene.

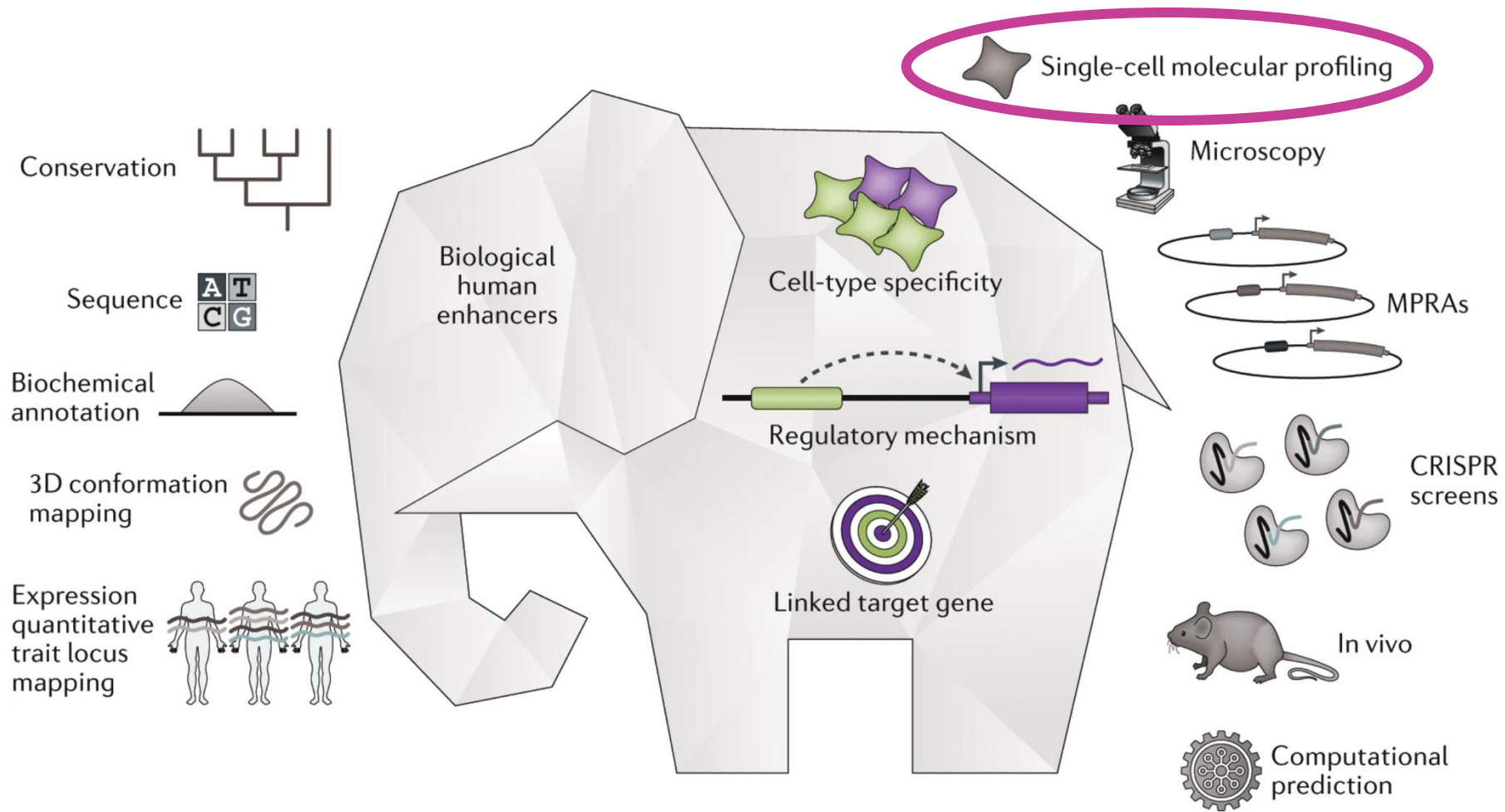
(Gusev 2016 *N.G.*, Gamazon 2018 *N.G.*, Weiner 2022 *AJHG*)



Linking regulatory variants to target genes is a critical step in: (Dey 2021 *Cell Genomics*)

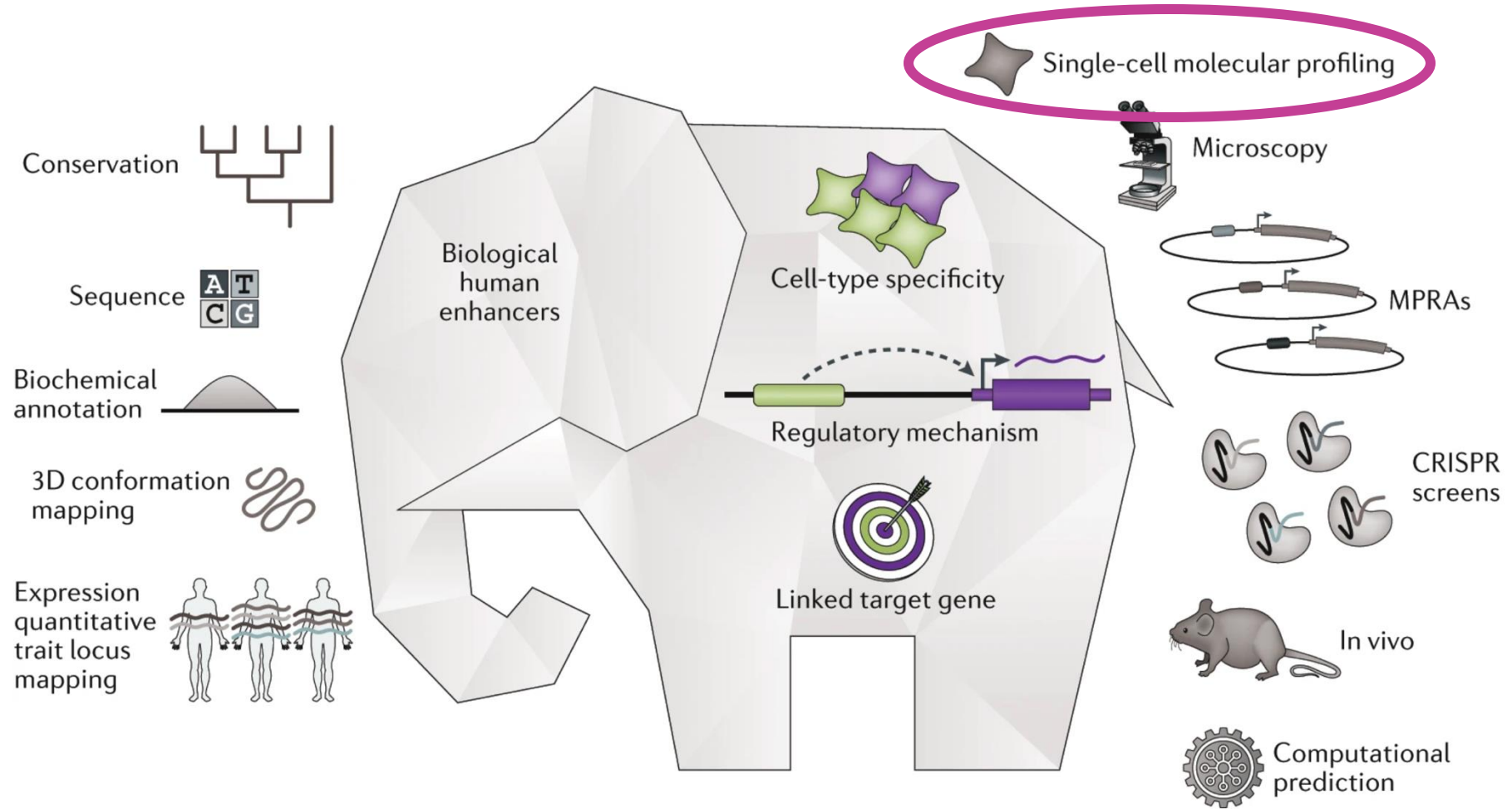
1. Interpreting GWAS discoveries
2. Nominating potential drug targets

Regulatory activity is interrogated using various lines of biological evidence



Regulatory activity is interrogated using various lines of biological evidence

Single-cell multiome data is...



Regulatory activity is interrogated using various lines of biological evidence

Single-cell multiome data is...

1. Context-specific

Conservation



Sequence



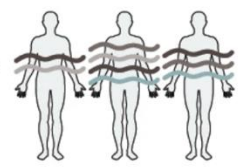
Biochemical annotation



3D conformation mapping



Expression quantitative trait locus mapping



Biological human enhancers

Cell-type specificity

Regulatory mechanism

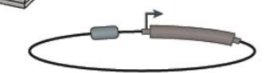
Linked target gene



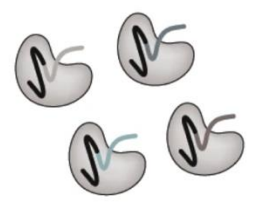
Single-cell molecular profiling



Microscopy



MPRAs



CRISPR screens



In vivo

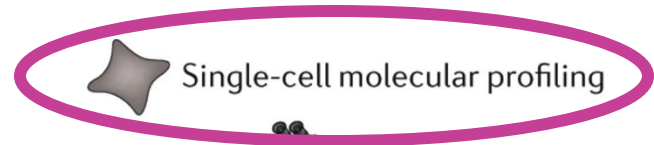
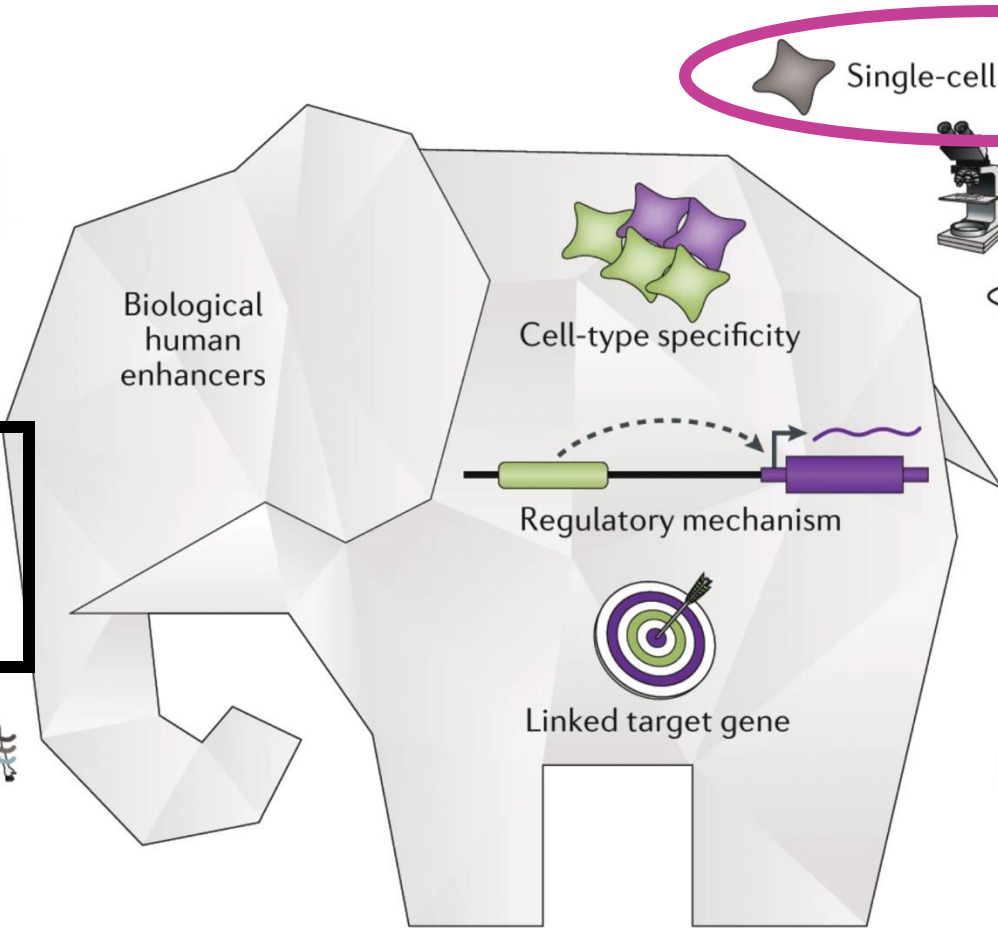
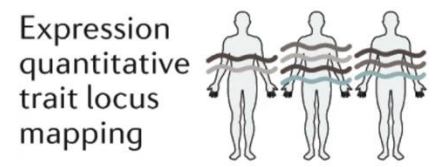
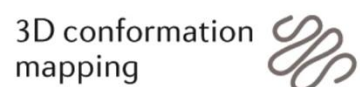
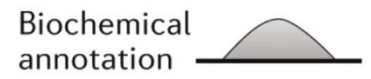
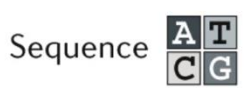


Computational prediction

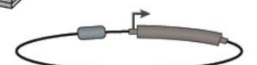
Regulatory activity is interrogated using various lines of biological evidence

Single-cell multiome data is...

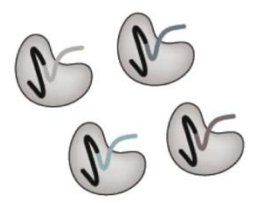
1. Context-specific



Microscopy



MPRAs



CRISPR screens



In vivo



Computational prediction

2. Closer to causality

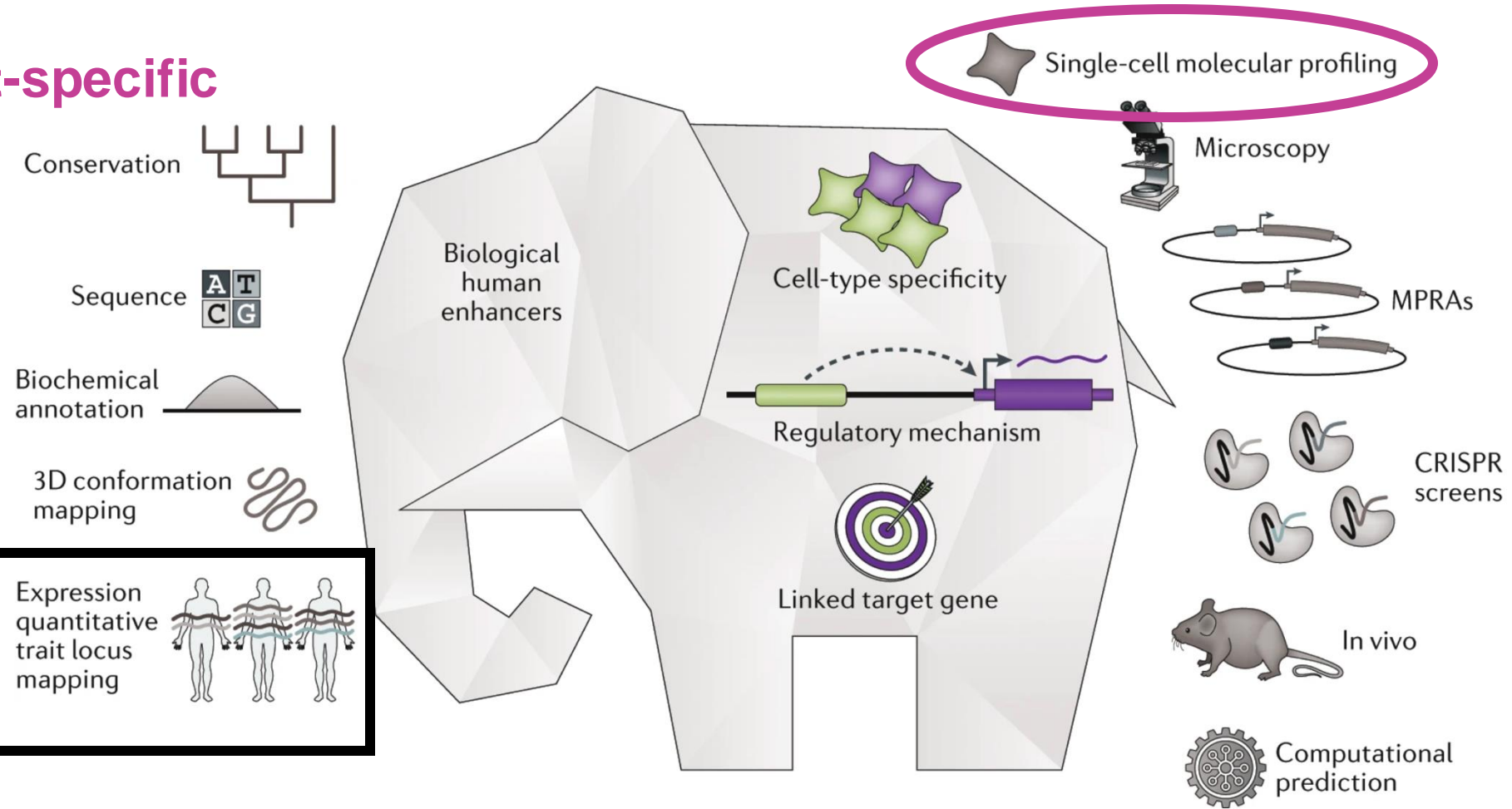
Regulatory activity is interrogated using various lines of biological evidence

Single-cell multiome data is...

1. Context-specific

2. Closer to causality

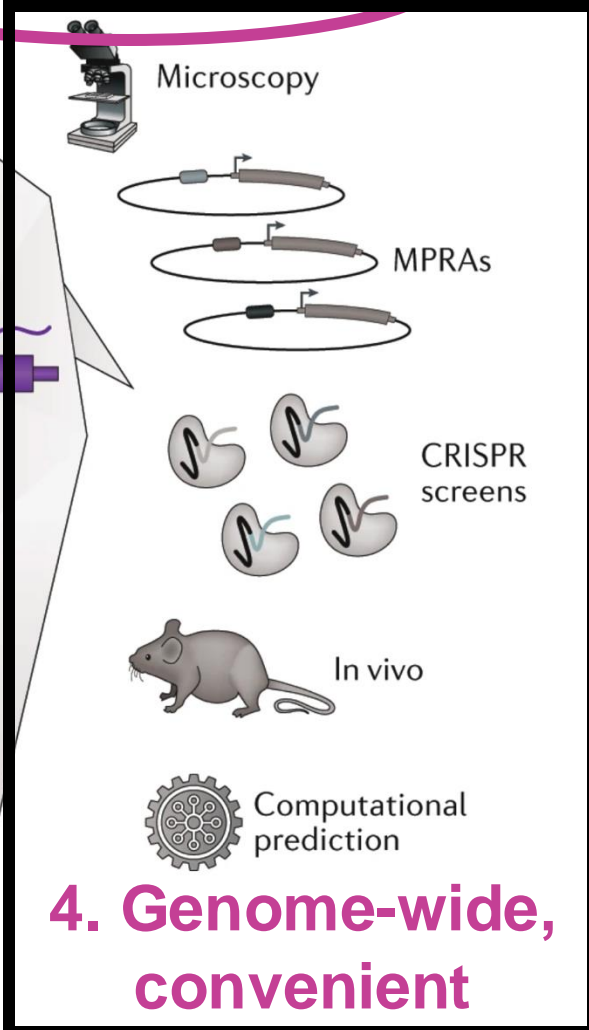
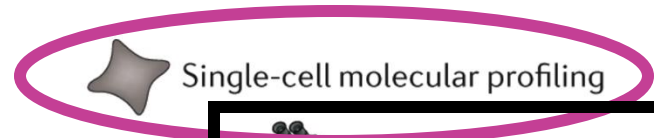
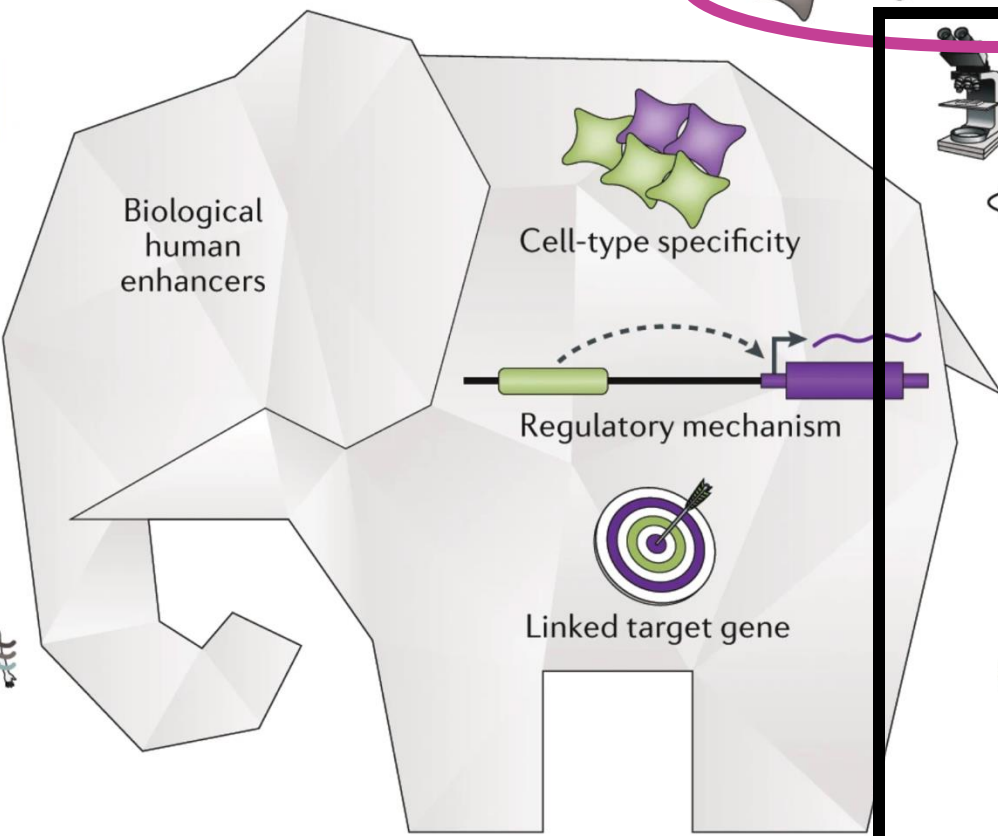
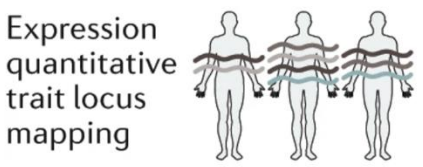
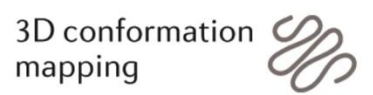
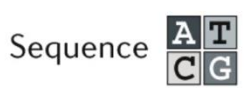
3. Easier to collect



Regulatory activity is interrogated using various lines of biological evidence

Single-cell multiome data is...

1. Context-specific

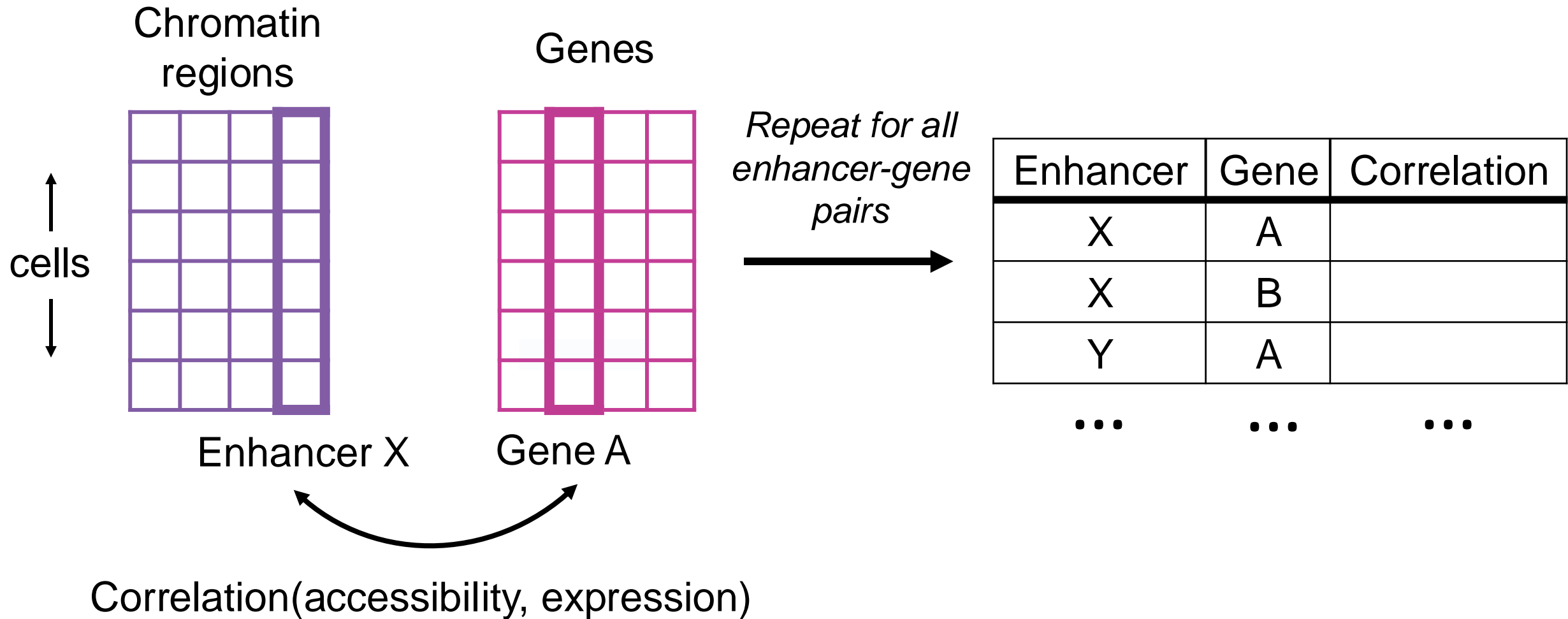


2. Closer to causality

3. Easier to collect

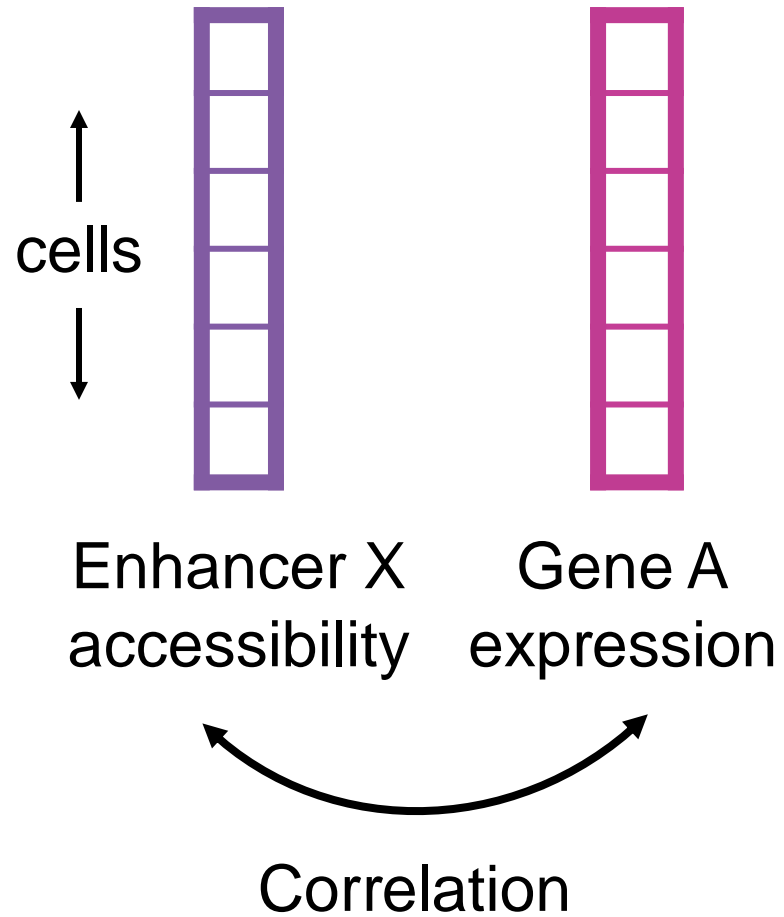
4. Genome-wide, convenient

Linking regulatory elements to their target genes



Examples: SCENT (Sakaue 2024 Nat Genet), SCARlink (Mitra 2024 Nat Genet), pgBoost (Dorans *In Press* Nat Genet), Signac (Stuart 2021 Nat Methods), ArchR (Granja 2021 Nat Genet)

Linking regulatory elements to their target genes



Method	Analysis unit	Model
SCENT ¹	Single cells	Poisson regression
SCARlink ²	Single cells	Poisson regression
Signac ³	Single cells	Linear correlation (after normalization)
ArchR ⁴	Metacells	Linear correlation

Conclusions

Single-cell multiome technologies jointly profile multiple biological modalities while retaining cellular heterogeneity.

Single-cell multiome analyses can target diverse biological questions.

Single-cell multiome methods use various strategies to handle data sparsity/noise and high dimensionality.

Several computational methods use single-cell RNA/ATAC multiome data to link regulatory elements to their target genes.

Forecast: multiome technologies are rapidly developing!

e.g. CITE-Seq: gene expression + cell surface protein

Emerging triple-modality technologies:

- scNMT-Seq and scChaRM-seq (Clark 2018 Nat Comm, Yan 2021 STAR Protoc)
 - Gene expression + chromatin accessibility + DNA methylation
- NEAT-Seq (Chen 2022 Nat Methods)
 - Gene expression + chromatin accessibility + intranuclear proteins

Reviewed in Vandereyken et al. 2023 Nat Rev Genet!

Thank you!