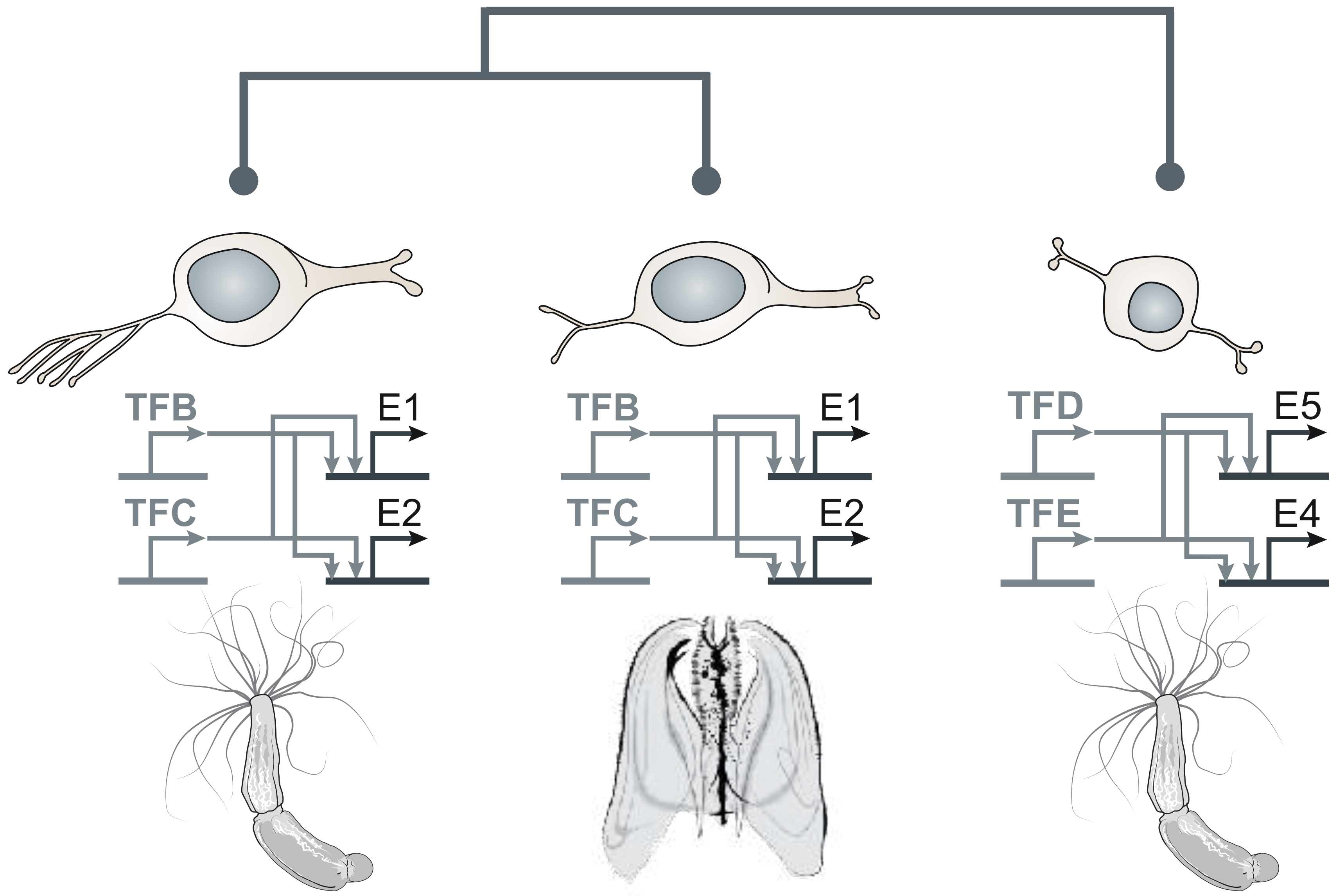
Introduction to single-cell functional genomics



Arnau Sebé-Pedrós

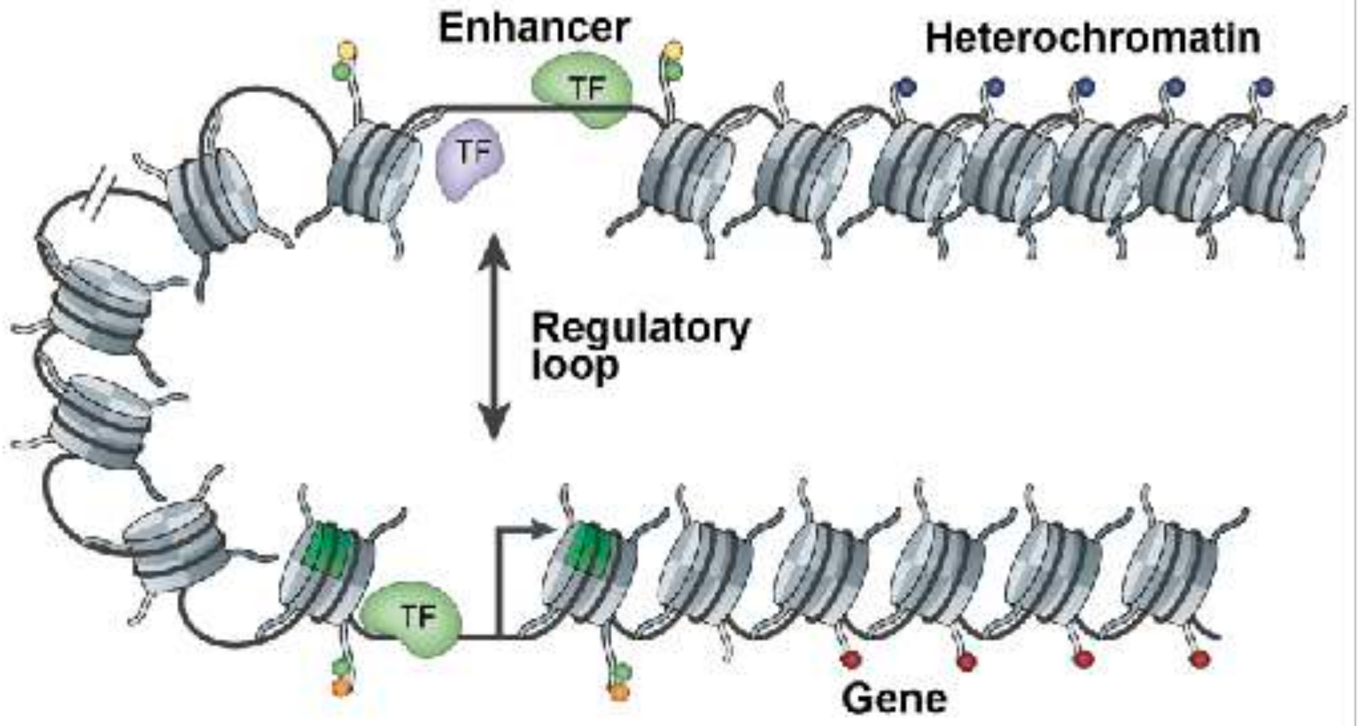
Centre for Genomic Regulation

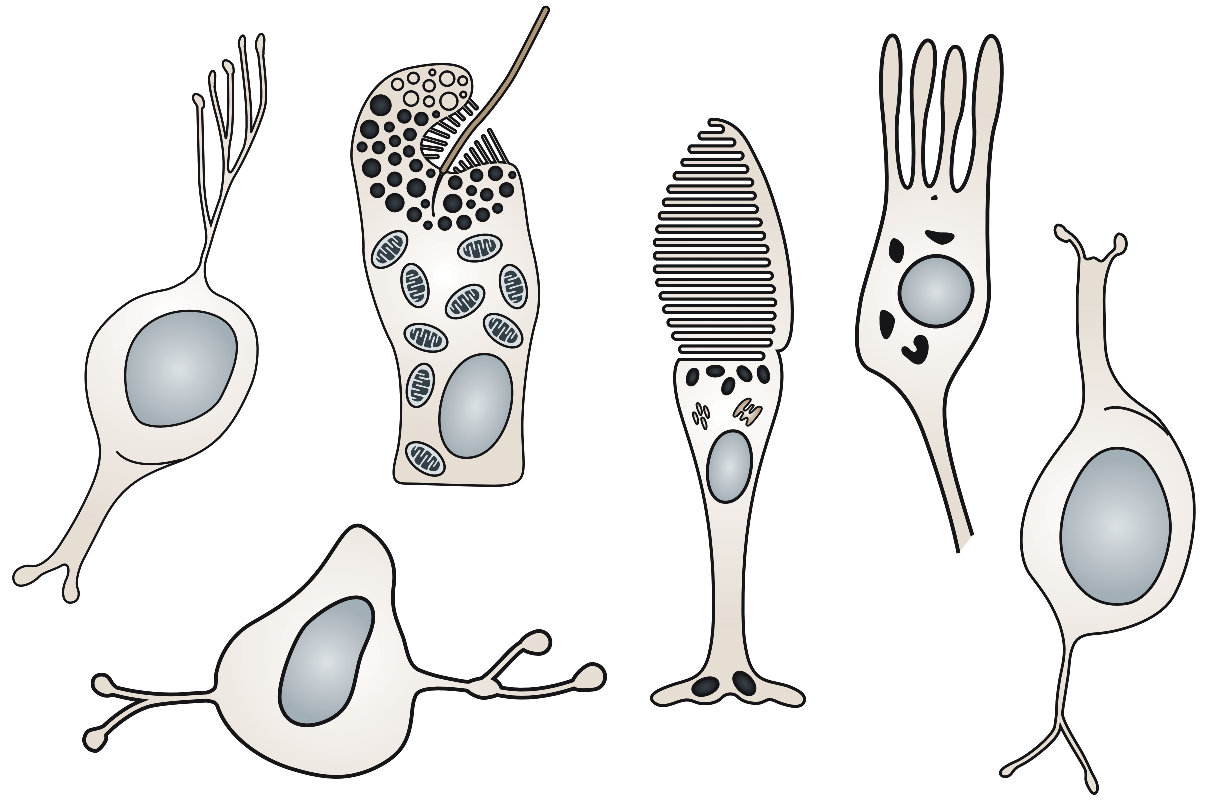
wellcome sanger institute

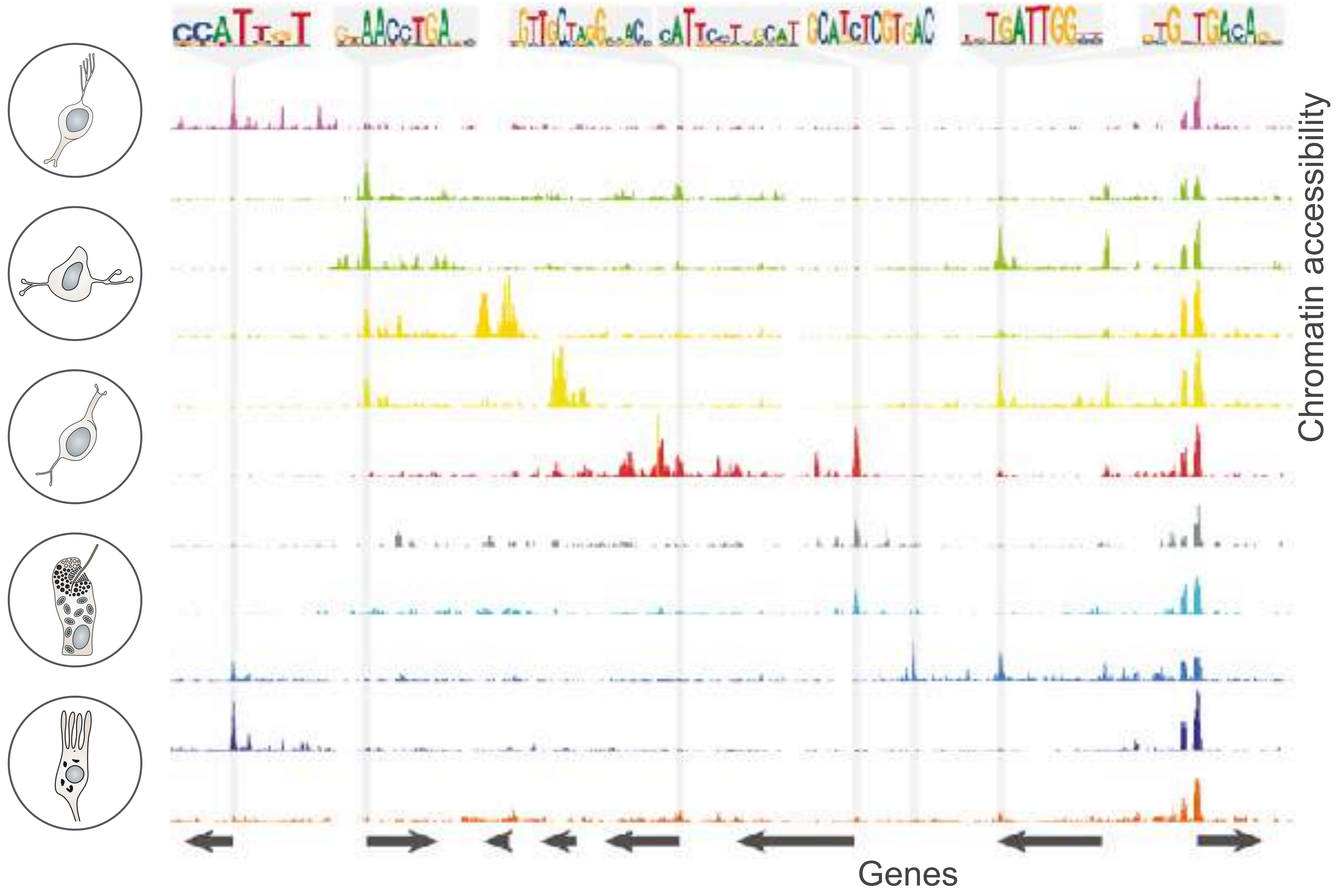
www.sebepedroslab.org

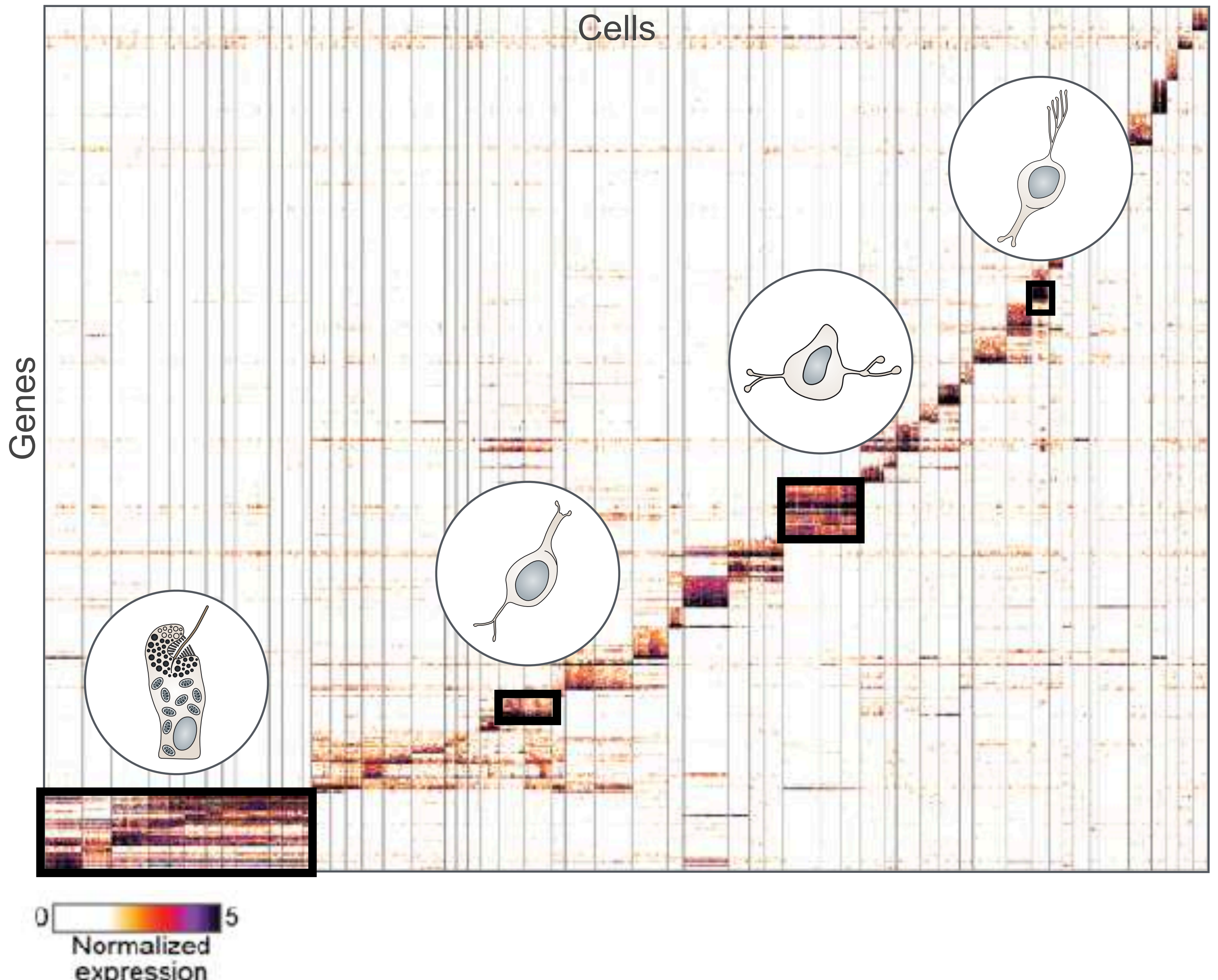
The single genomics revolution: cell type molecular profiling across the tree of life

Genome sequence → Genome regulation → Cell types





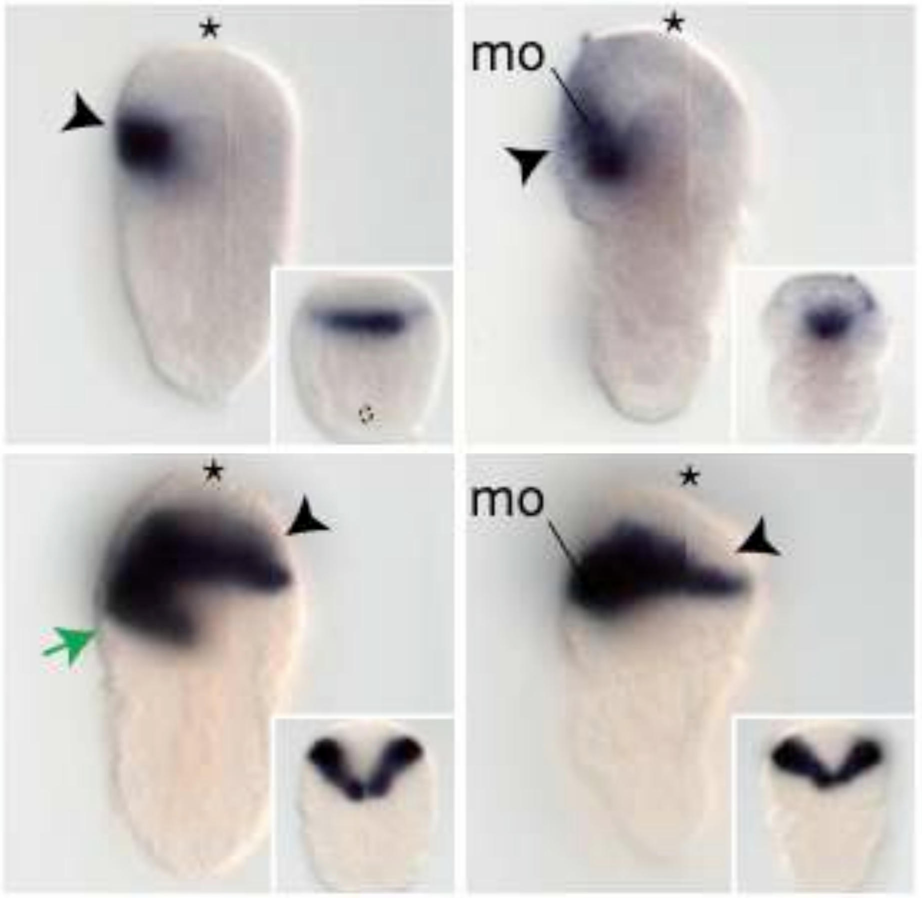




Cell type molecular fingerprinting: *in situs*, transcriptomics and single-cell transcriptomics

Gene expression pattern comparison (classical evo-devo)

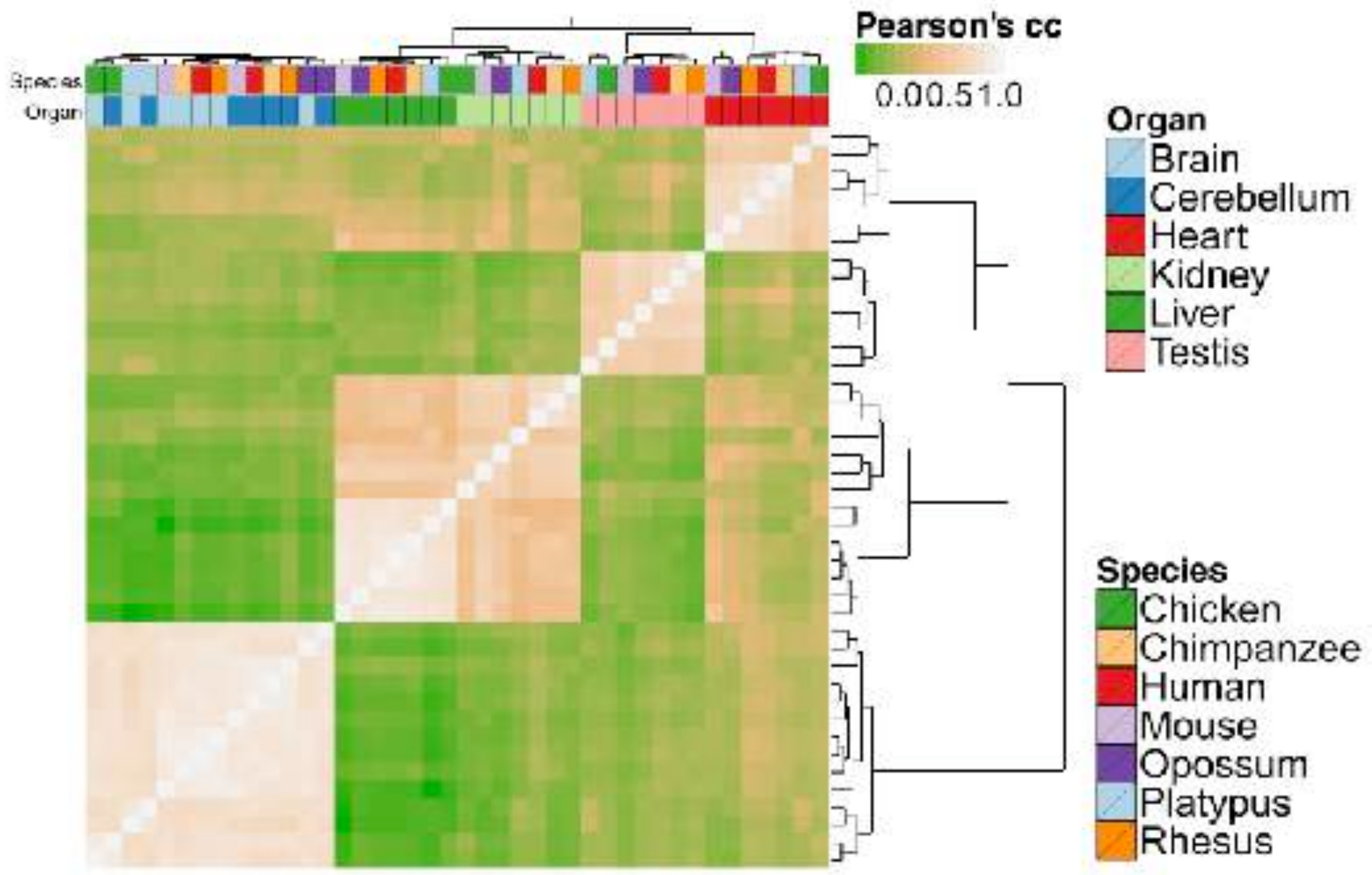
Problems:
- Need to define markers a priori.
- Low throughput (one or a few genes at the time)



Martin-Duran et al. 2016

Bulk tissue transcriptome comparisons

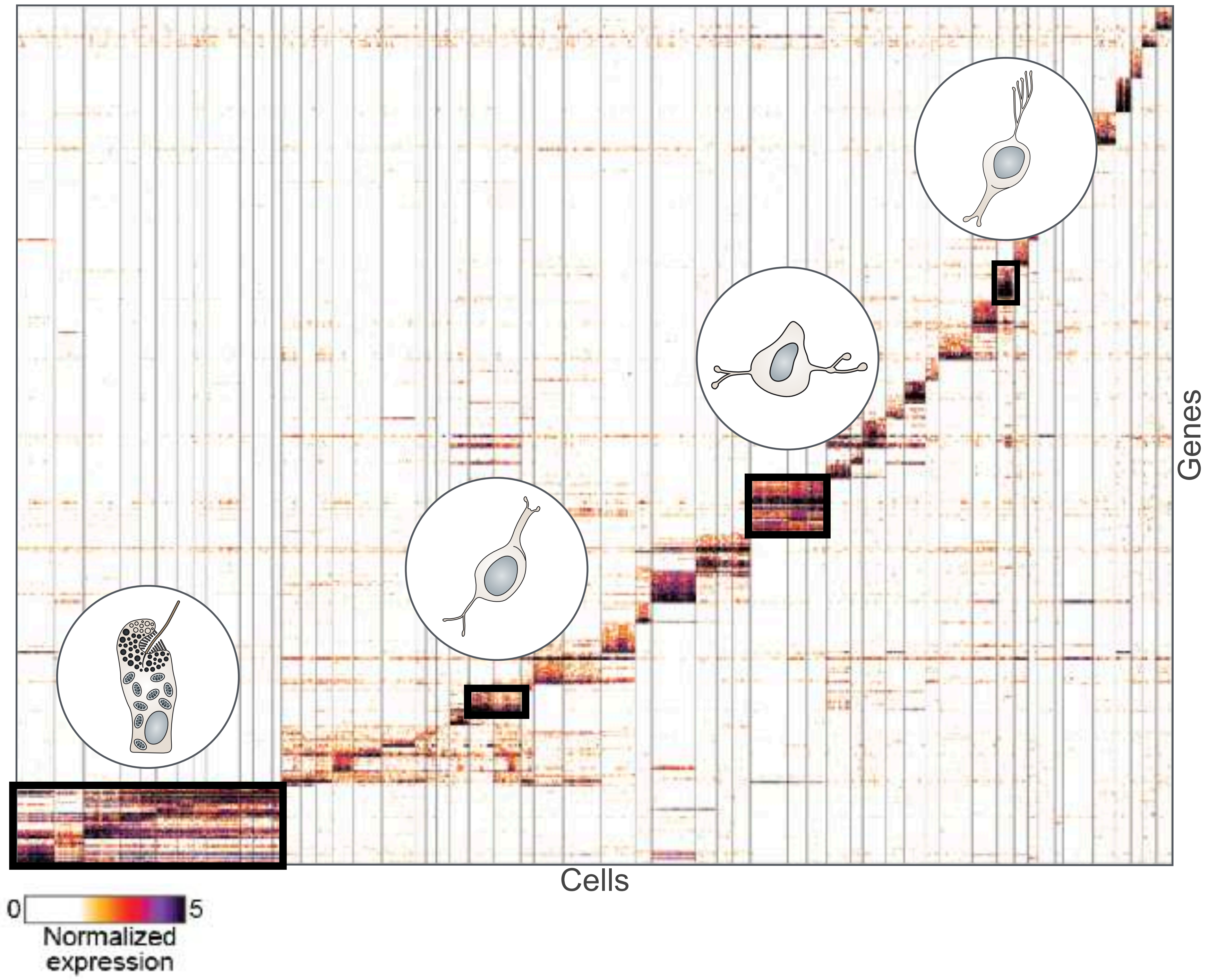
Problems:
- Need to dissect tissues/organs.
- Cellular heterogeneity within tissues.



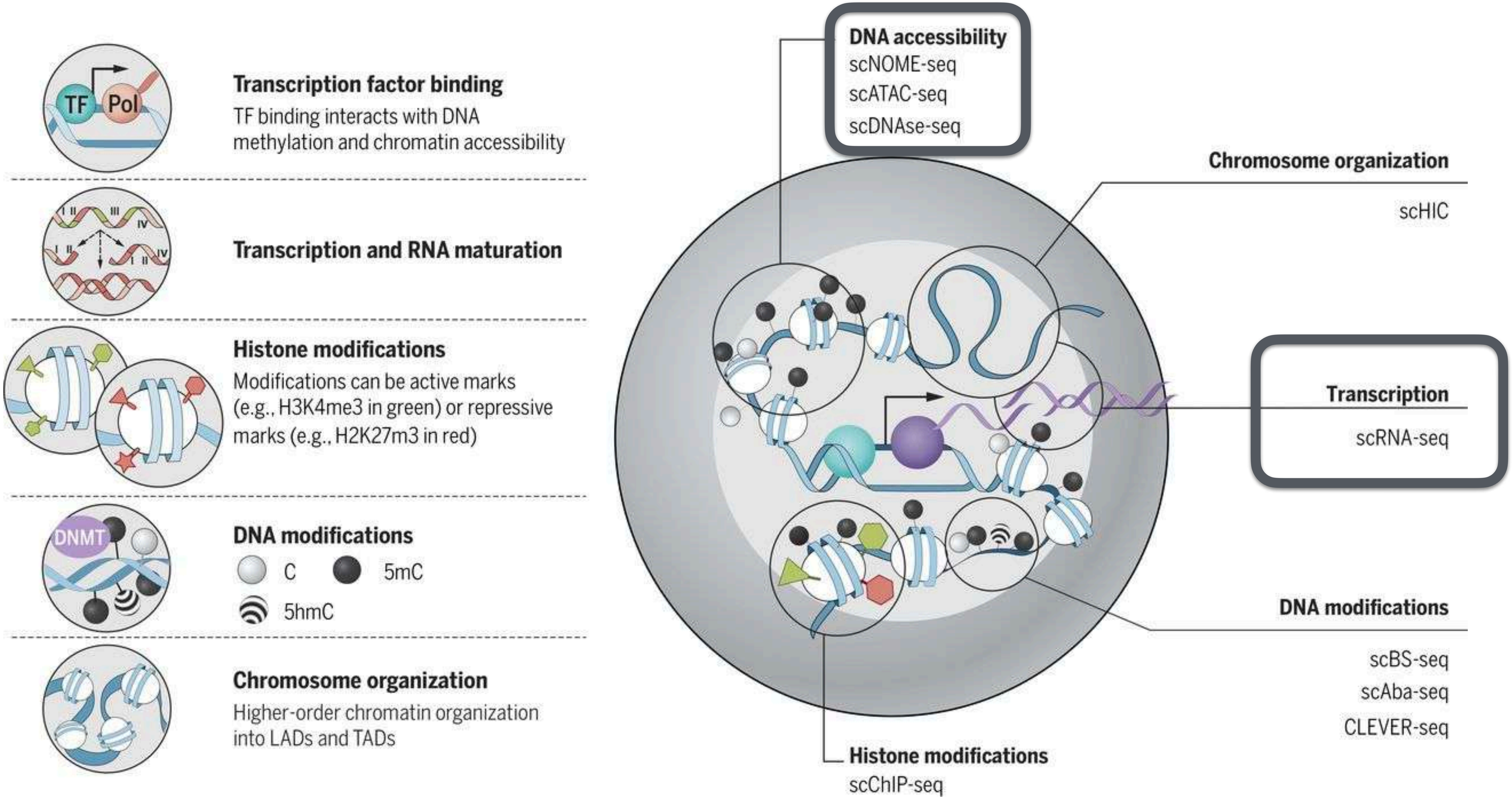
Breschi et al. 2016

Single-cell transcriptomics

1. No need to define marker genes *a priori.*
2. No need for tissue dissection -> Cellular resolution.
3. Low input material (non-culturable species).

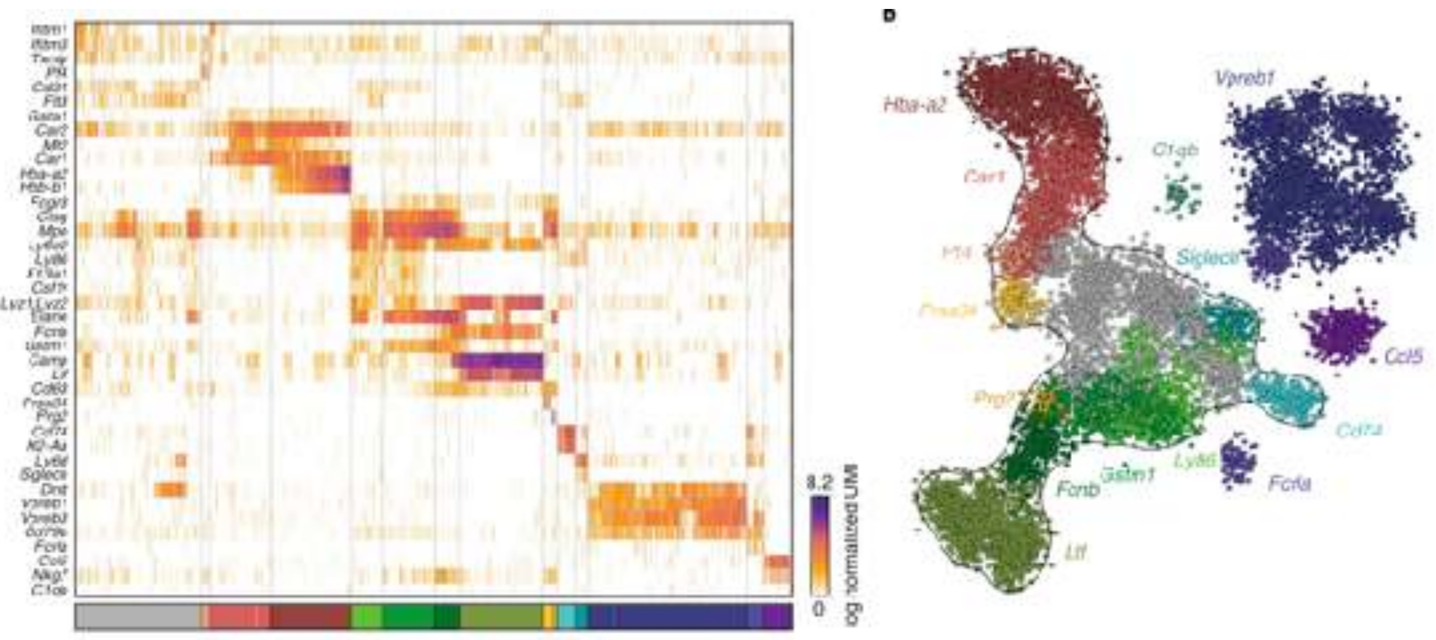


What can we (try) to measure in a single-cell

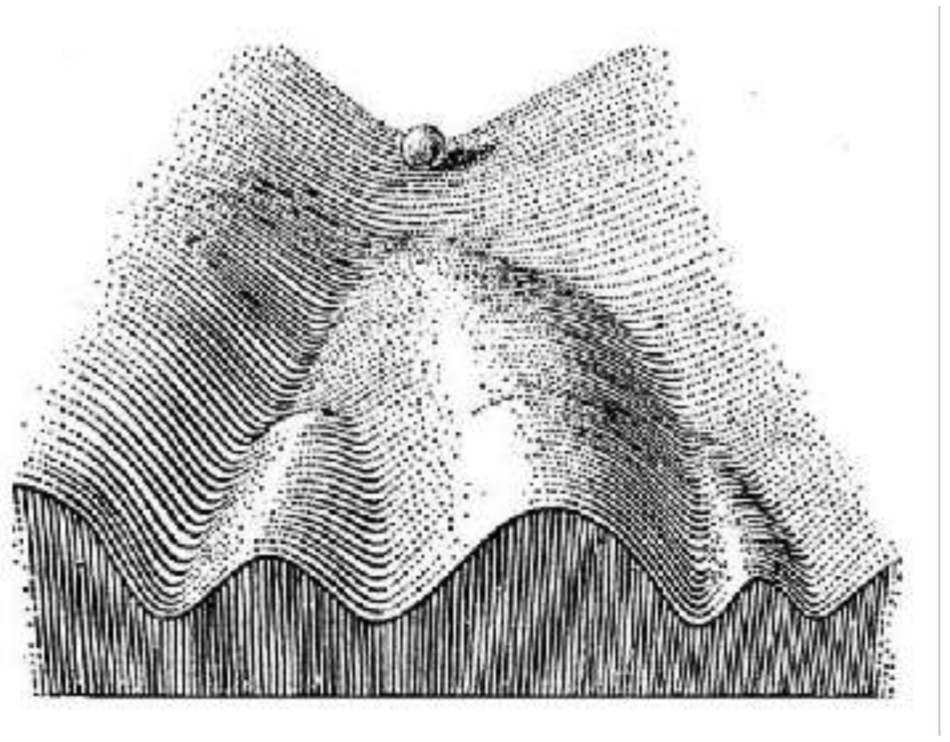


Applications of single-cell transcriptomics

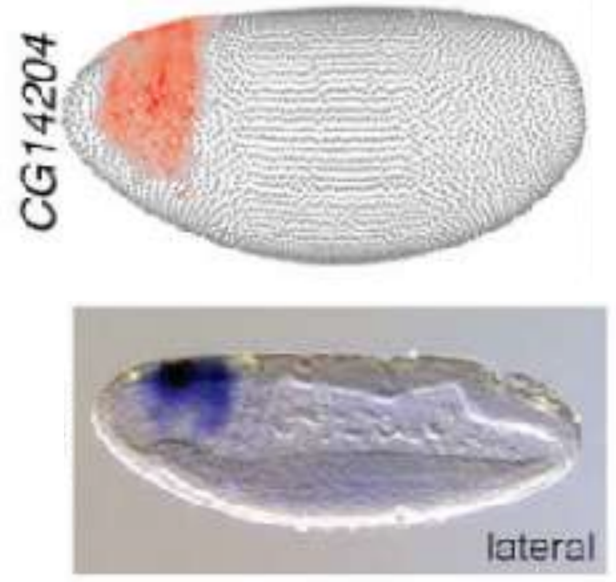
Cell type phenomenology (& variations)



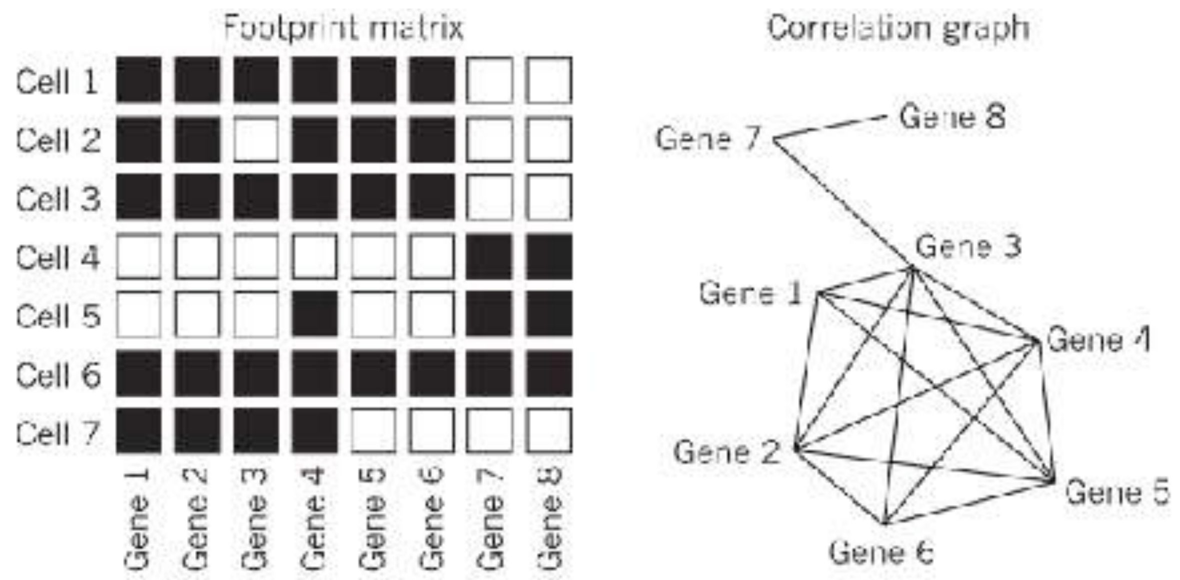
Temporal axis: cells in time



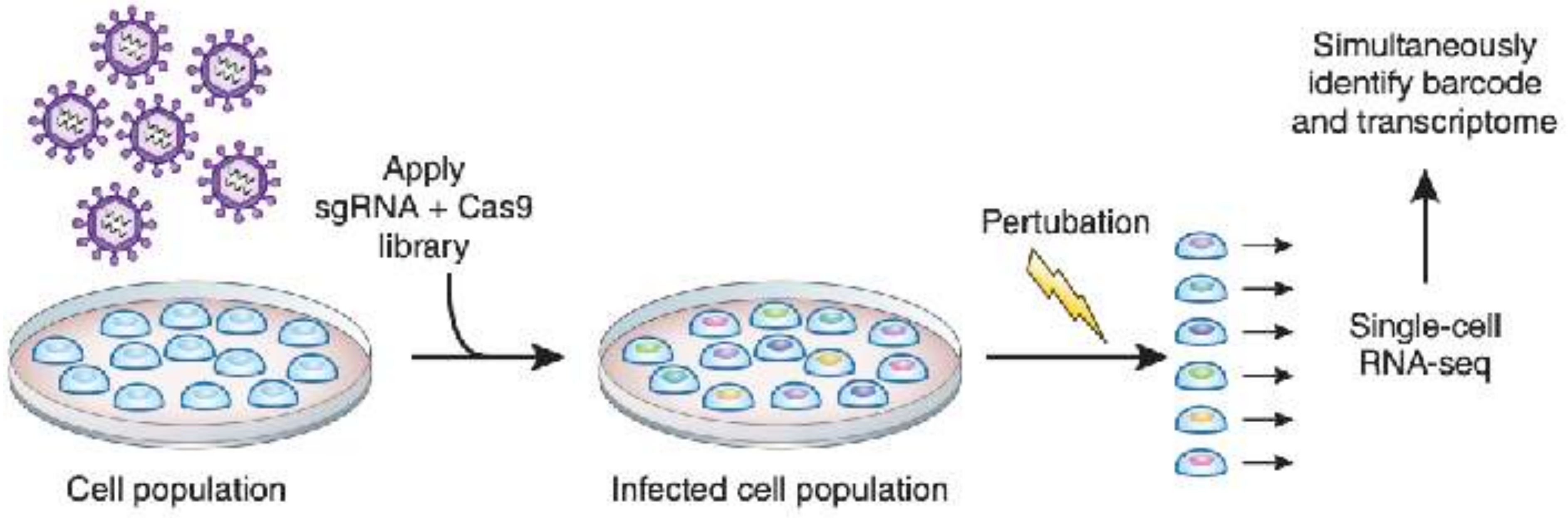
Spatial axis: cells in space



Co-regulated gene programs

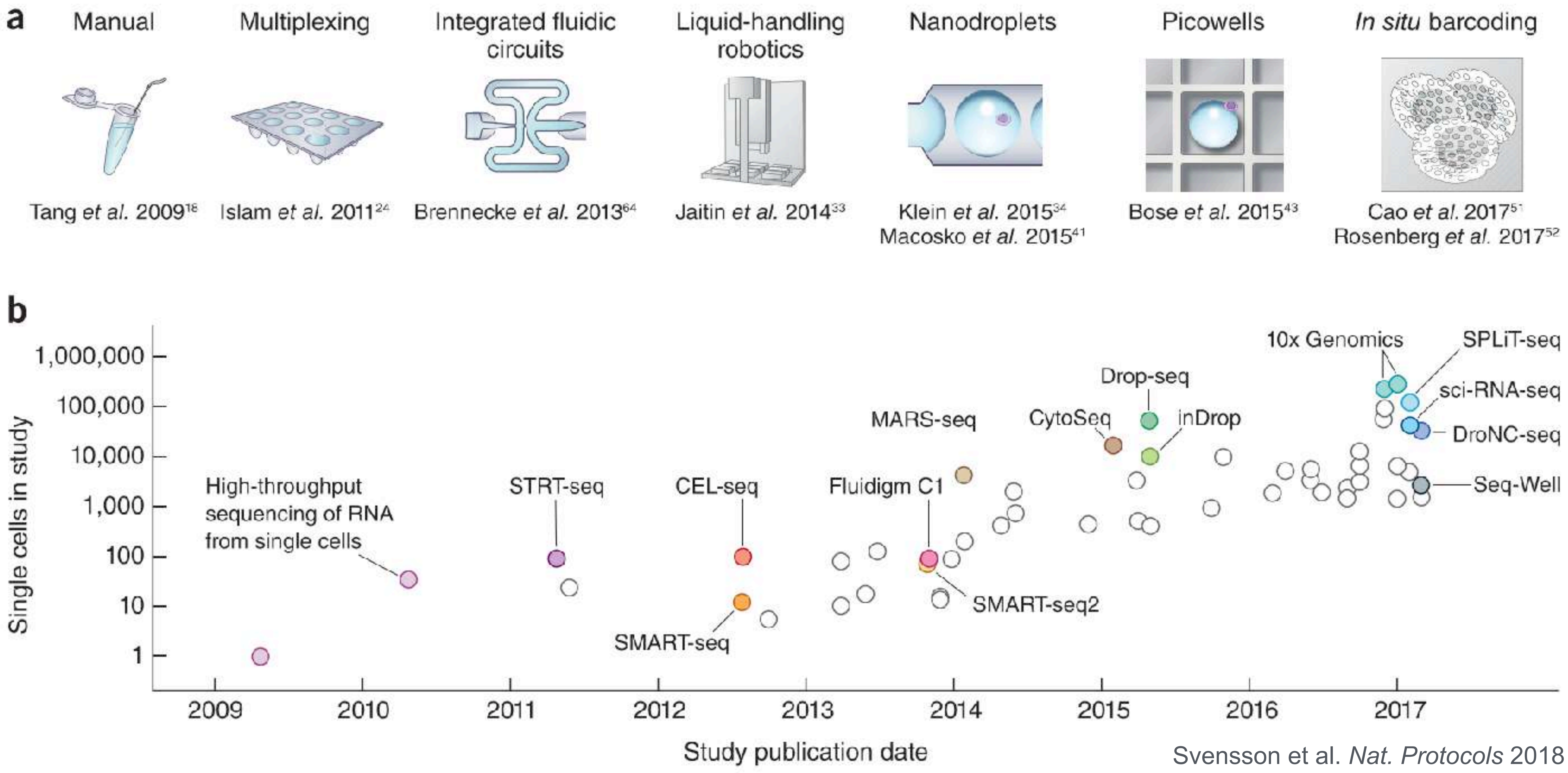


Experimental profiling: CRISPR screens, lineage tracing, etc.



Part 1 - Single-cell transcriptomics technologies

Exponential scaling of single-cell transcriptomics methods

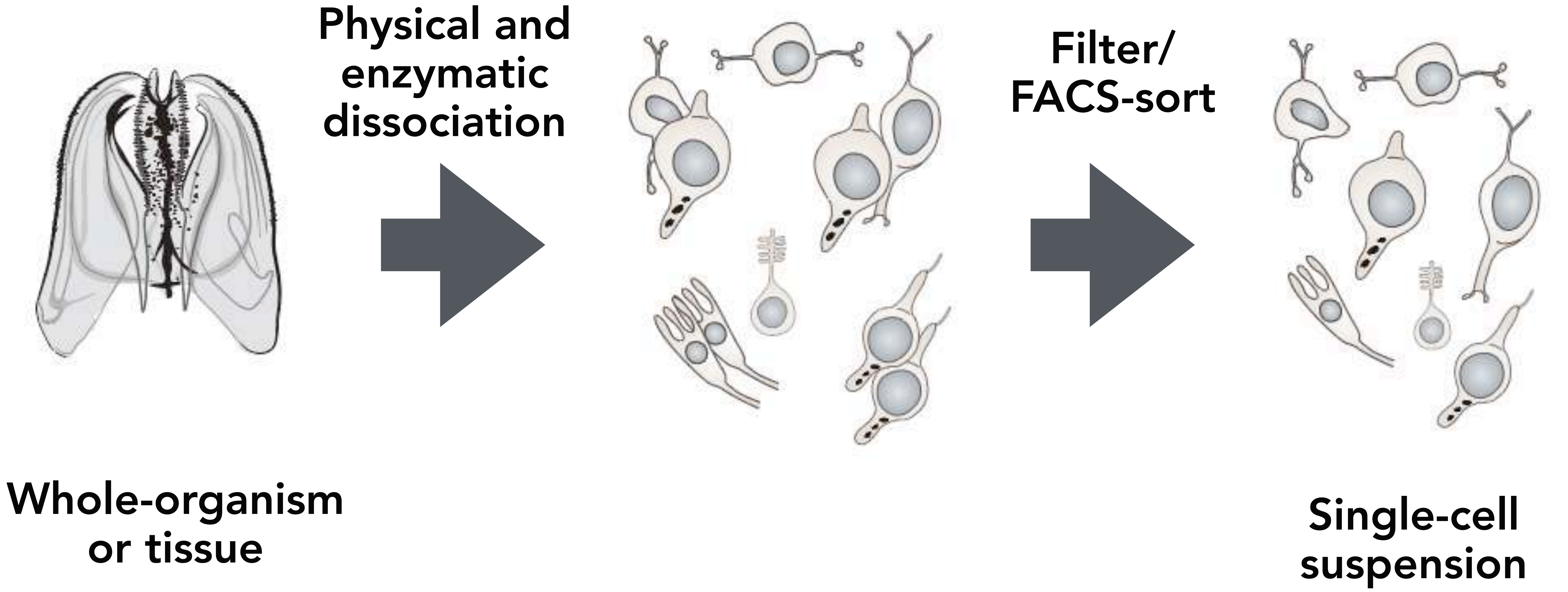


Svensson et al. *Nat. Protocols* 2018

The ideal single-cell method (and the reality)

- ✔ ***Universal*** in terms of cell size, type and state (and species)
- ☛ ***In situ*** measurements.
- ✘ No ***minimum input*** of number of cells to be assayed.
- ✘ Every cell is assayed, i.e.100% ***capture rate***.
- ✘ Every transcript in every cell is detected, i.e.100% ***sensitivity***.
- ☛ Every transcript is identified by its ***full-length sequence***.
- ✔ Transcripts are assigned correctly to cells, e.g. no ***doublets***.
- ☛ Additional ***multimodal*** *measurements.*
- ✔ ***Cost*** effective per cell.

Basic steps in single-cell transcriptomics: **from whole-organisms to cells**



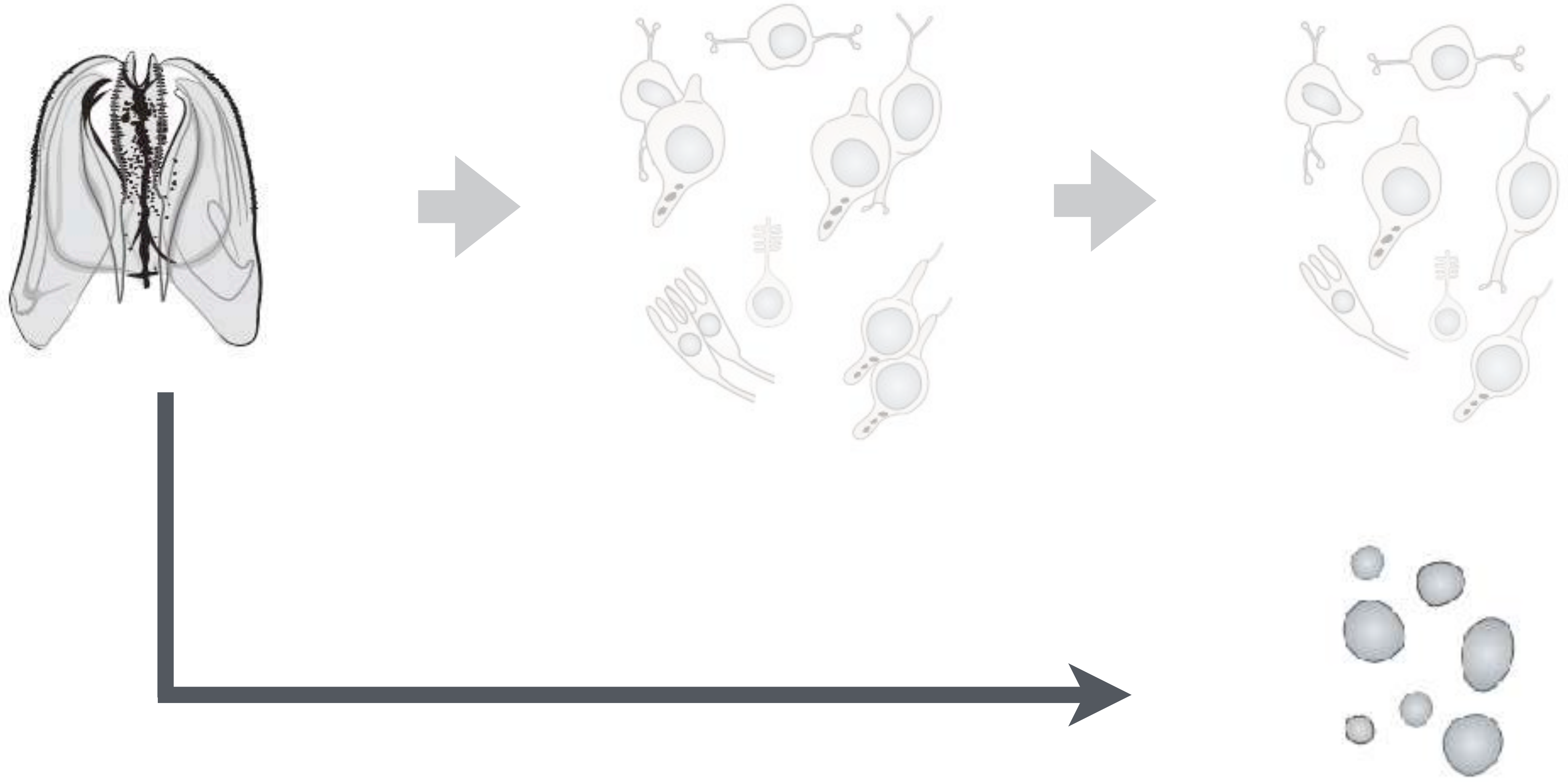
<u>Cell fixation/cryopreservation</u>: decoupling tissue processing from single-cell sequencing.

Methanol-based fixatives

Formaldehyde/Glyoxal/DSP and other cross-linkers

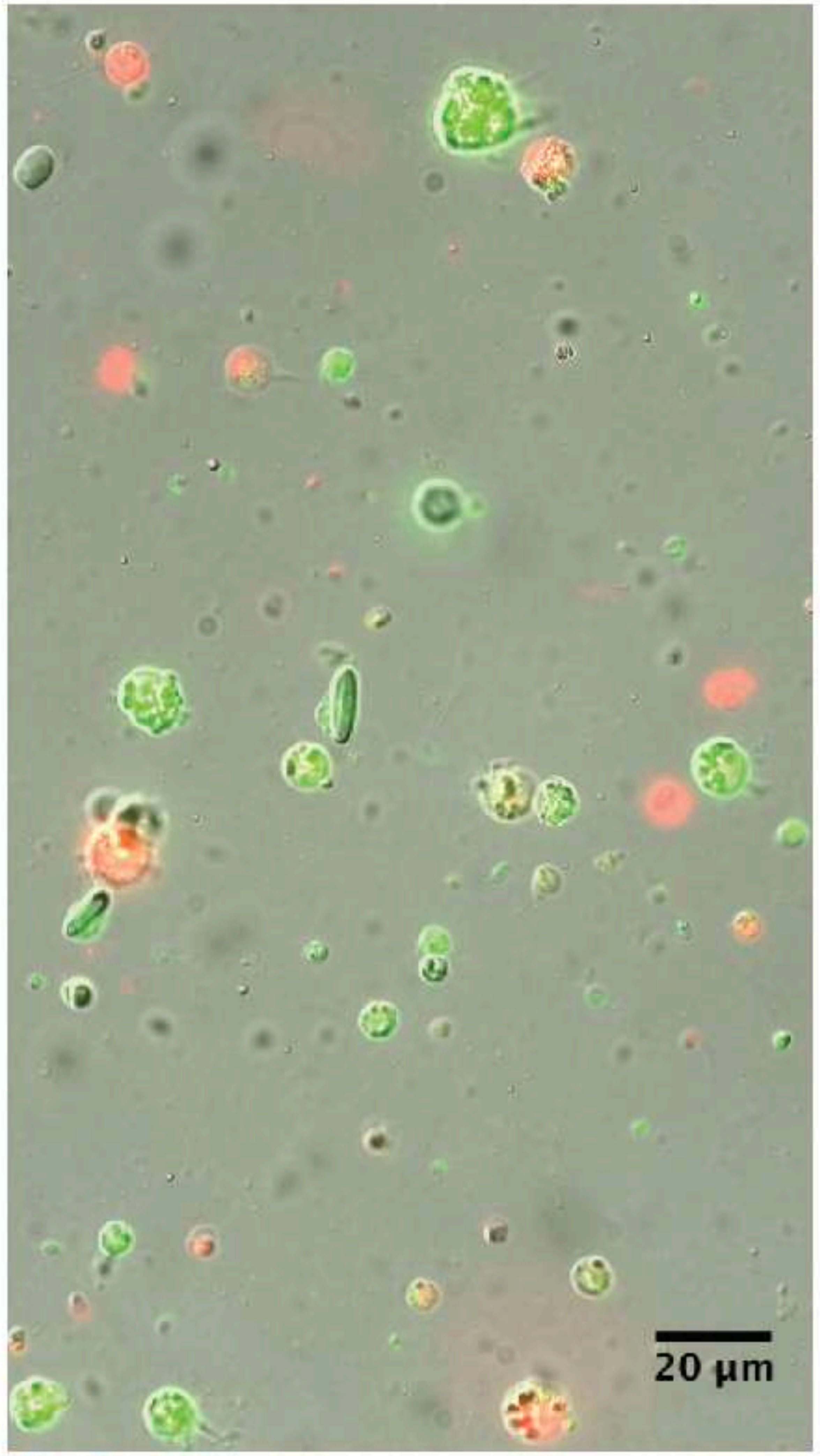
Cryopreservation

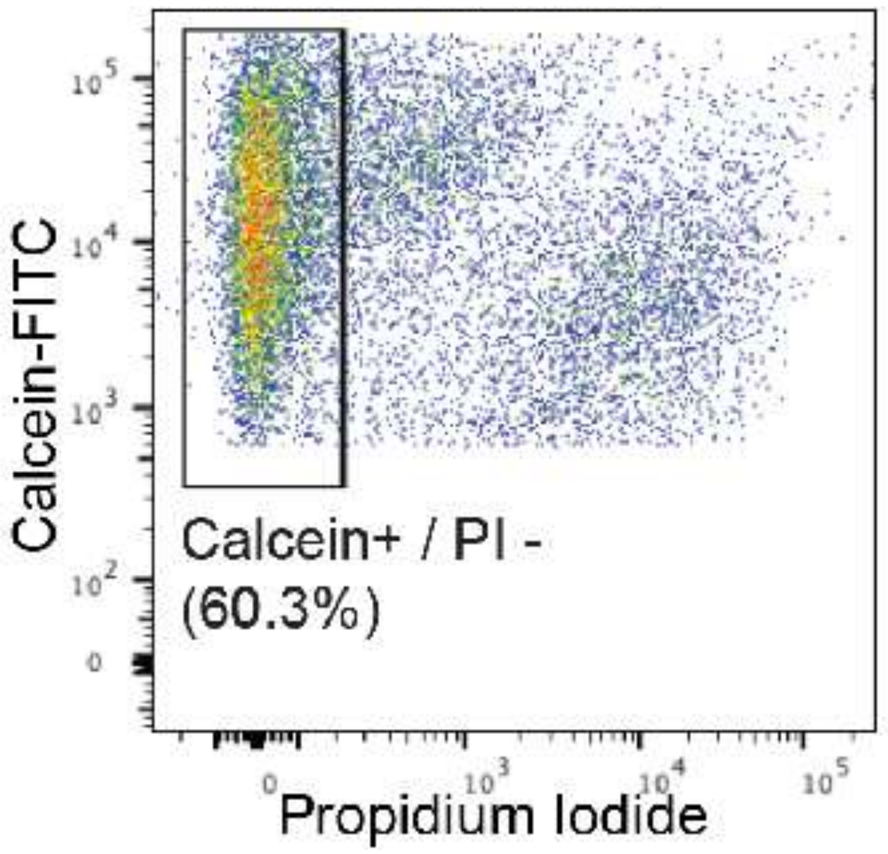
<u>Nuclei sequencing</u>: direct extraction from complex tissues (e.g. brain)



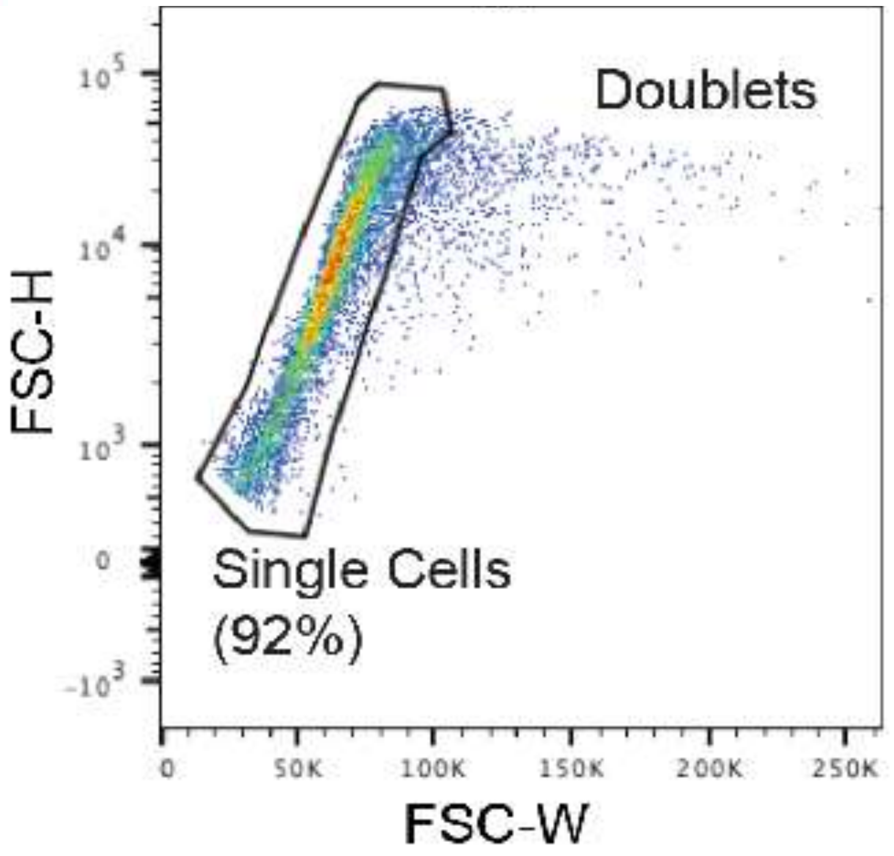
Why sample prep is the most important step in single-cell transcriptomics?

Cell death, debris and multiples





Dead cells and non-cellular particles

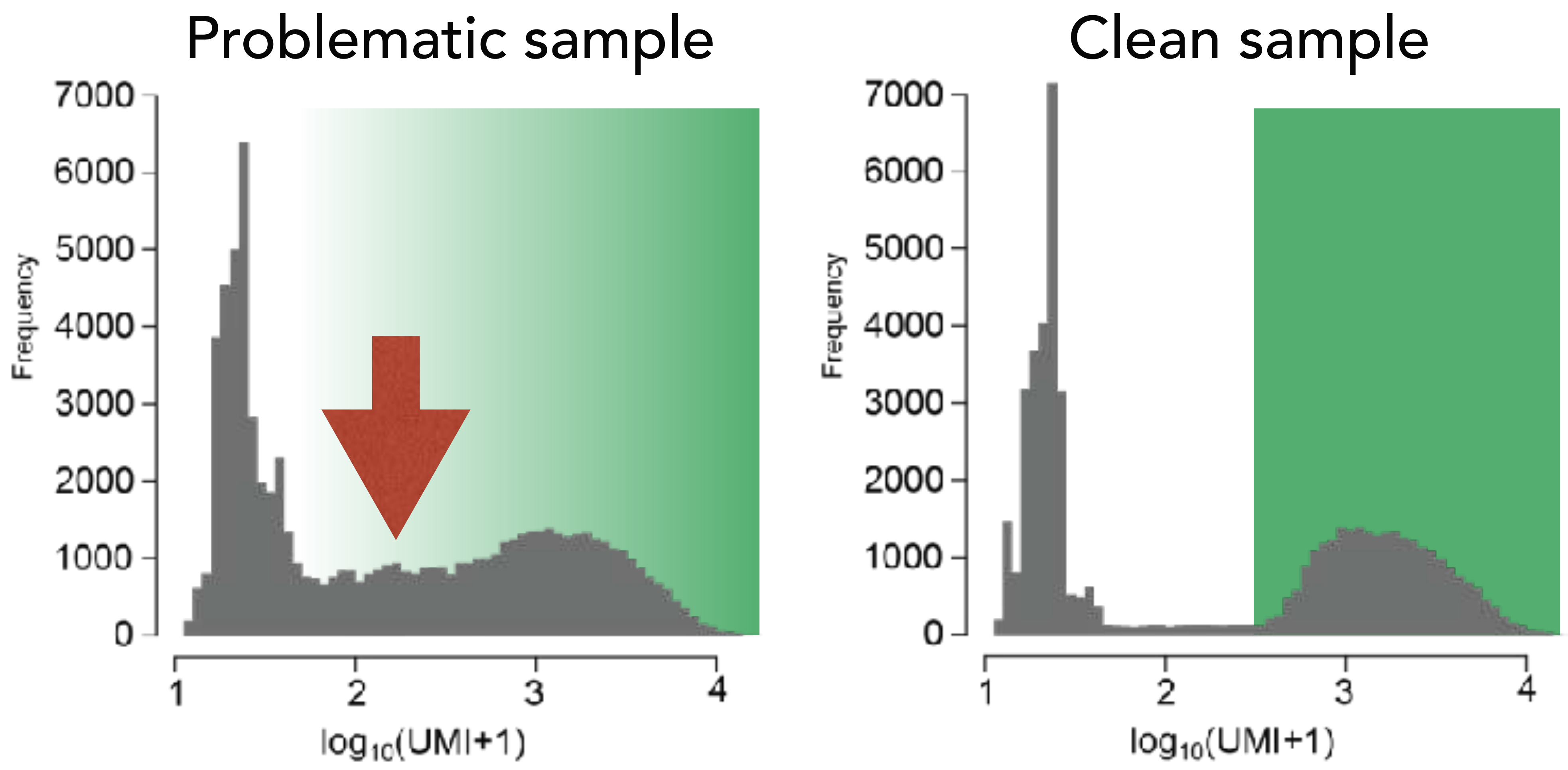


Physical doublets/ multiplets

Live cells (Calcein)
Dead cells (PI)

Sebe-Pedros et al. *Cell* 2018

Ambient RNA

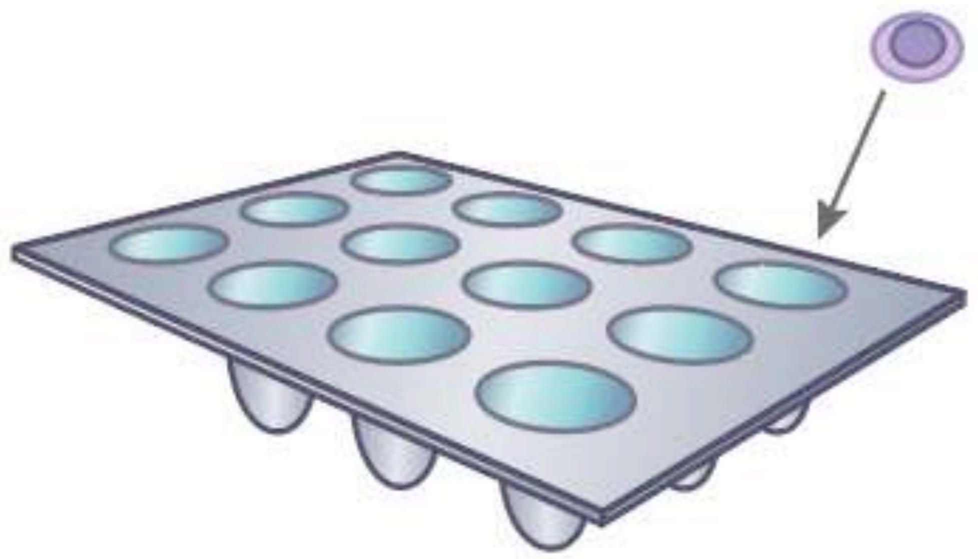


Problems:

- Difficult to determine cells from non-cells (empty barcodes)
- Transcritpionally quiescent cells (low UMIs/cell) are “swallowed” by background RNA signal
- Major factor explaining batch effects.

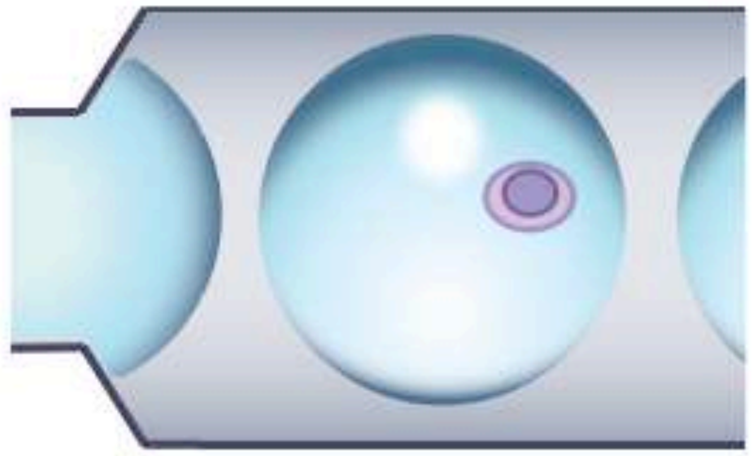
Cell encapsulation and lysis

Multi-well plates



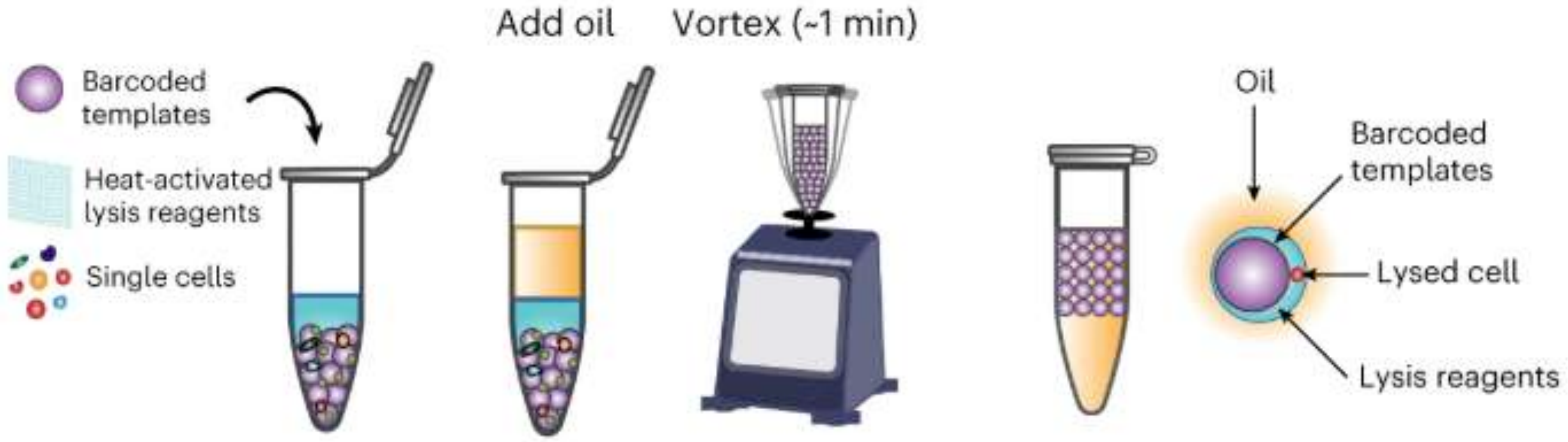
SMART-seq2
MARS-seq
mcSCRB-seq
CELseq2
Quartz-seq2

Droplets

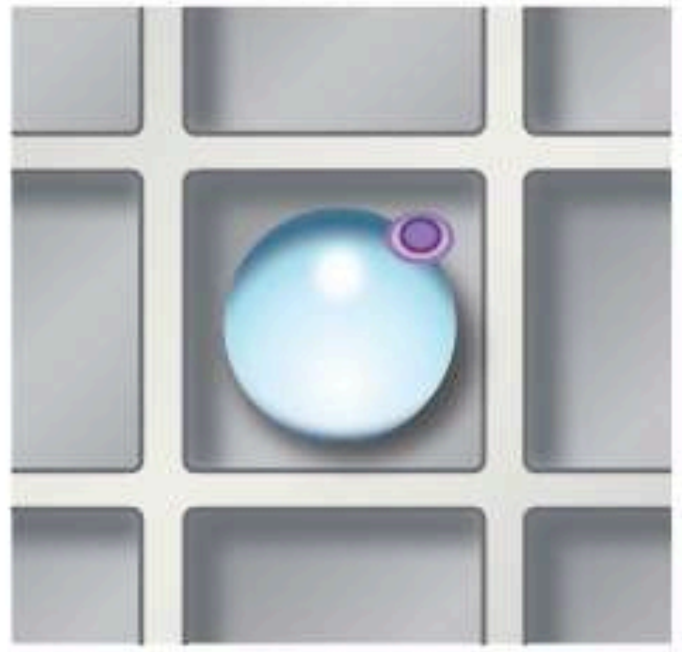


inDrops
Drop-seq
10X Chromium*

*PIP-seq/FluentBio**

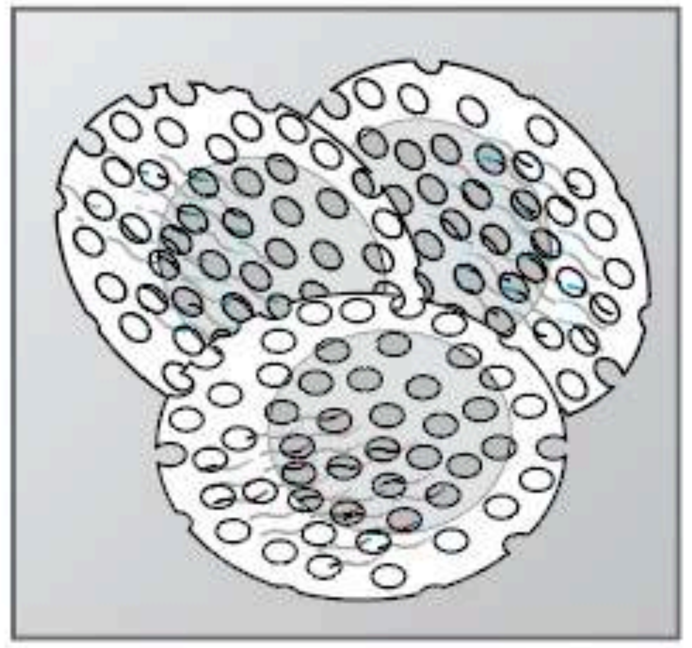


Nanowells



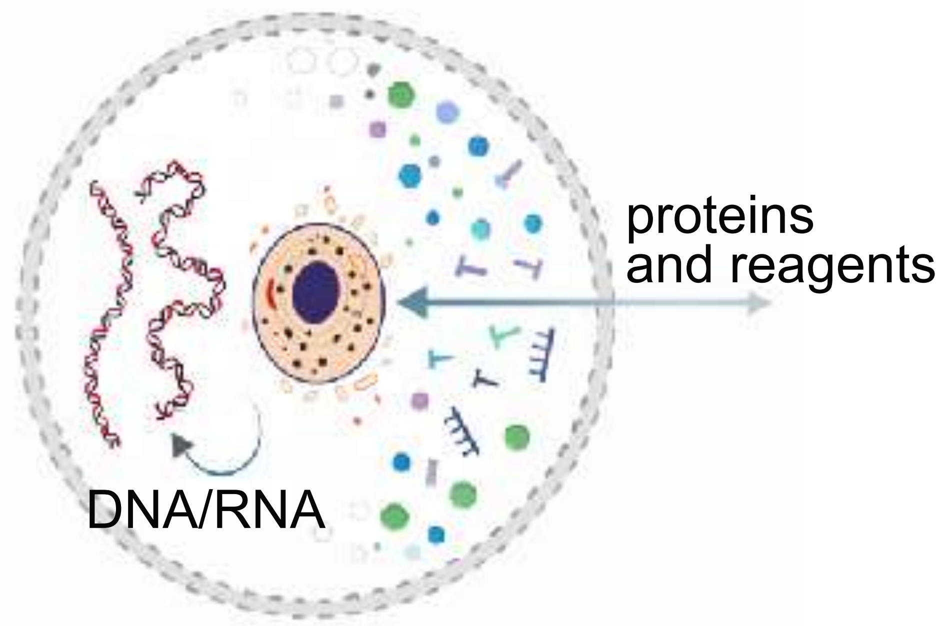
Microwell-seq
Seq-well
Tanaka ICell8*
BD Rhapsody*

in-cell barcoding



SPLIT-seq
sciRNA-seq
ParseBio* (Qiagen)
ScaleBio* (10X)

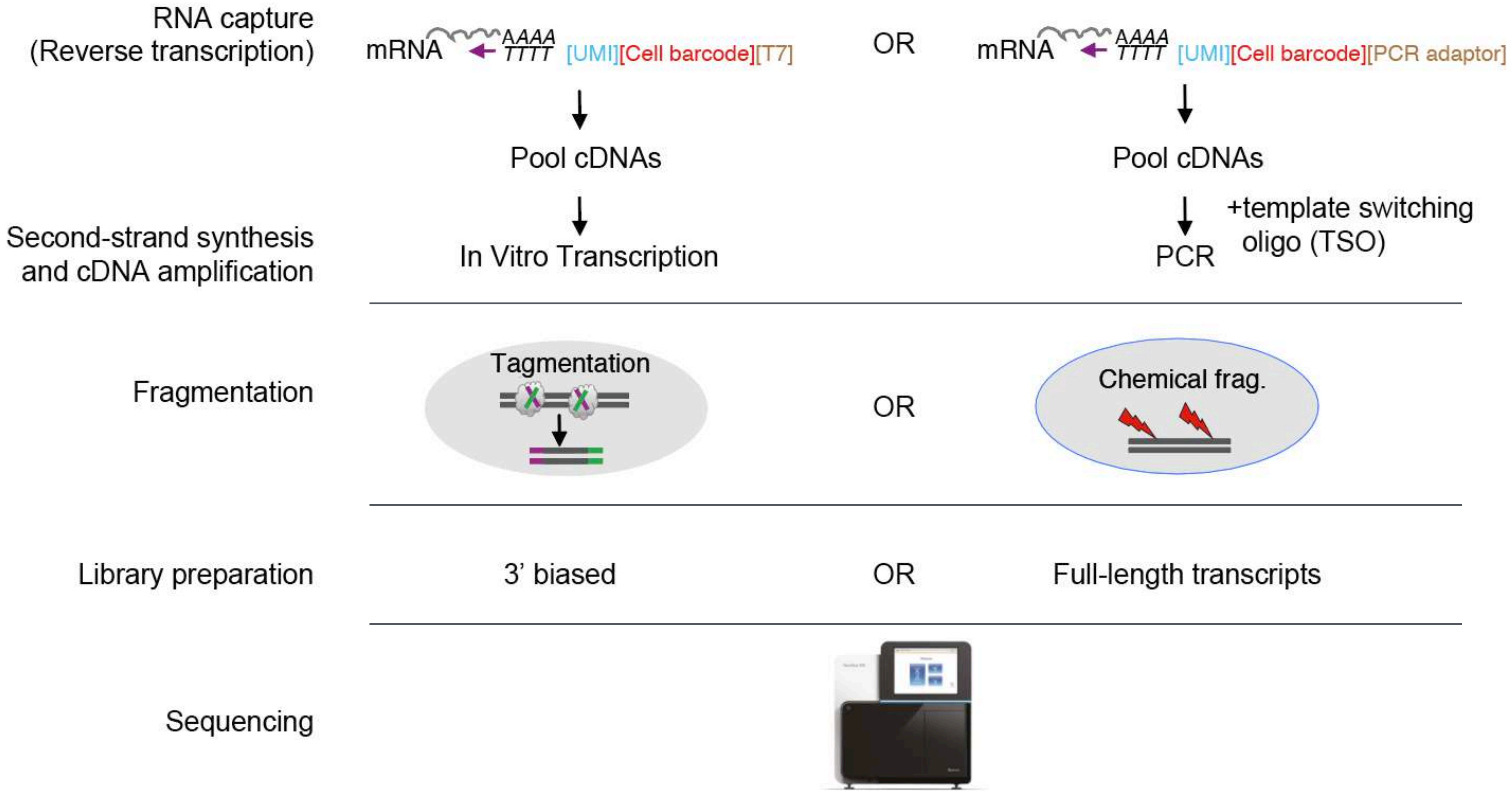
Semi-permeable capsules



AtrandiBio*

*commercial

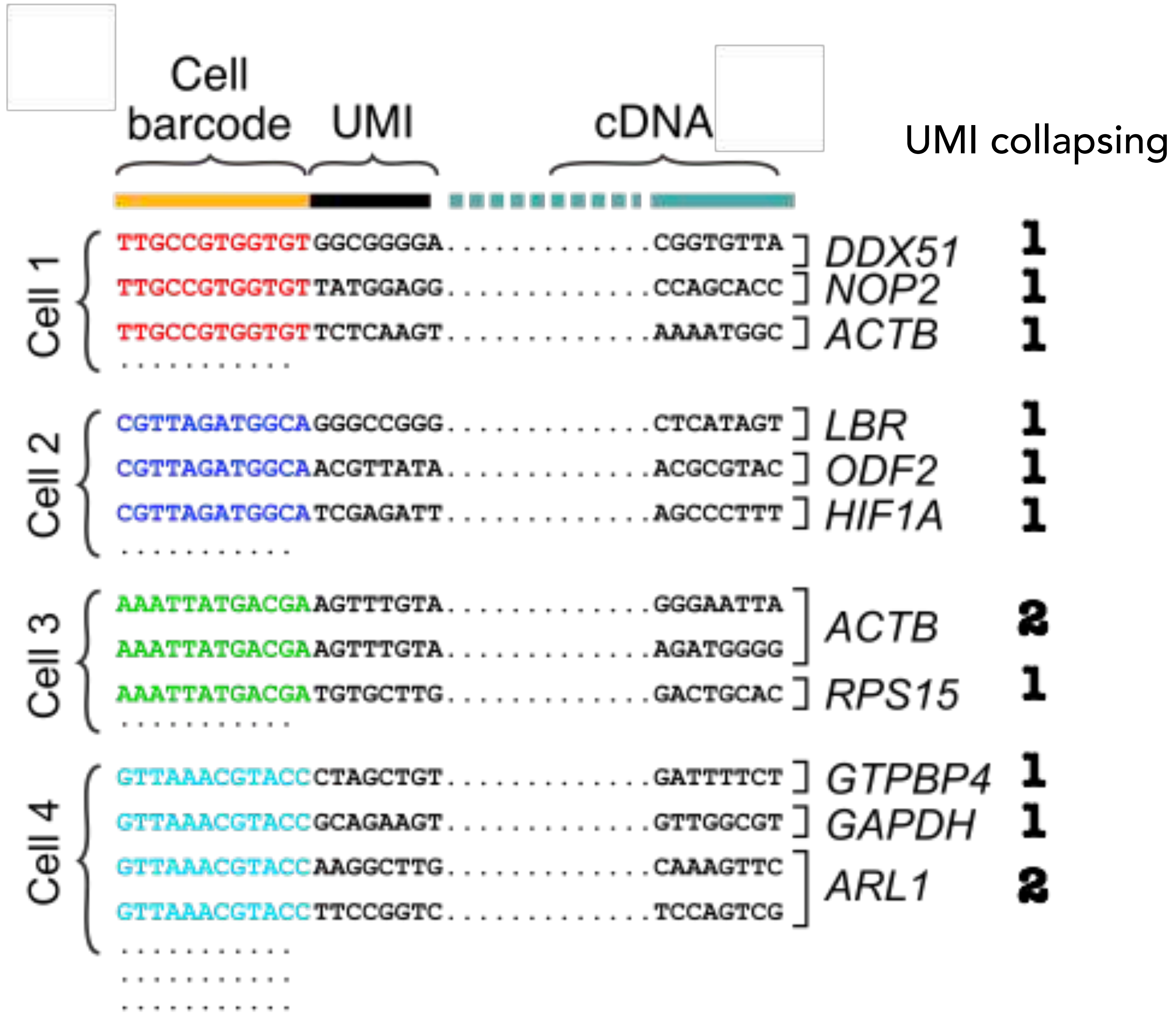
Basic steps in single-cell transcriptomics: **from RNA to cDNA libraries to sequences**



Unique Molecule Identifiers (UMIs) and ERCC spike-ins

Quantitative single-cell RNA-seq with unique molecular identifiers

Saiful Islam[1], Amit Zeisel[1], Simon Joost[2], Gioele La Manno[1], Pawel Zajac[1], Maria Kasper[2], Peter Lönnerberg[1] & Sten Linnarsson[1]



ERCC: External RNA Controls Consortium

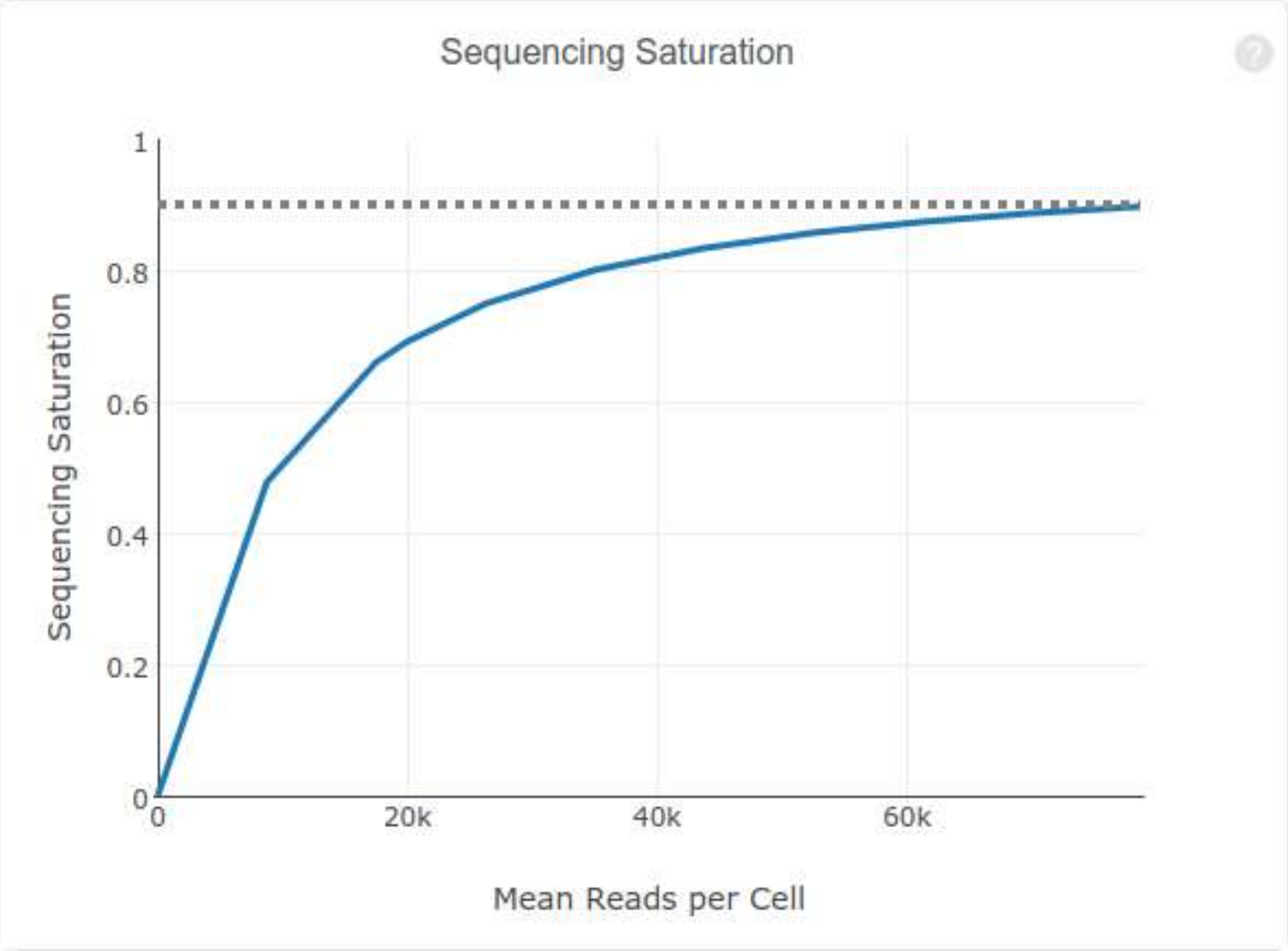
- Set of external RNA transcripts with known concentrations.
- Represent diverse lengths and sequence composition.
- Internal control used to measure method performance.
- Originally used for internal expression normalization.

How much should I sequence my cells?

- In most situations: ±30-50K reads per cell (e.g. 5 billion reads for 100K cells).
- Library saturation can be measured: reads/UMI (±4-5 is enough, 0.7-0.8 saturation)
- *De novo* cell type atlas *versus* resampling (can be shallower).
- Remember, for most applications: **More cells, better than more reads!**

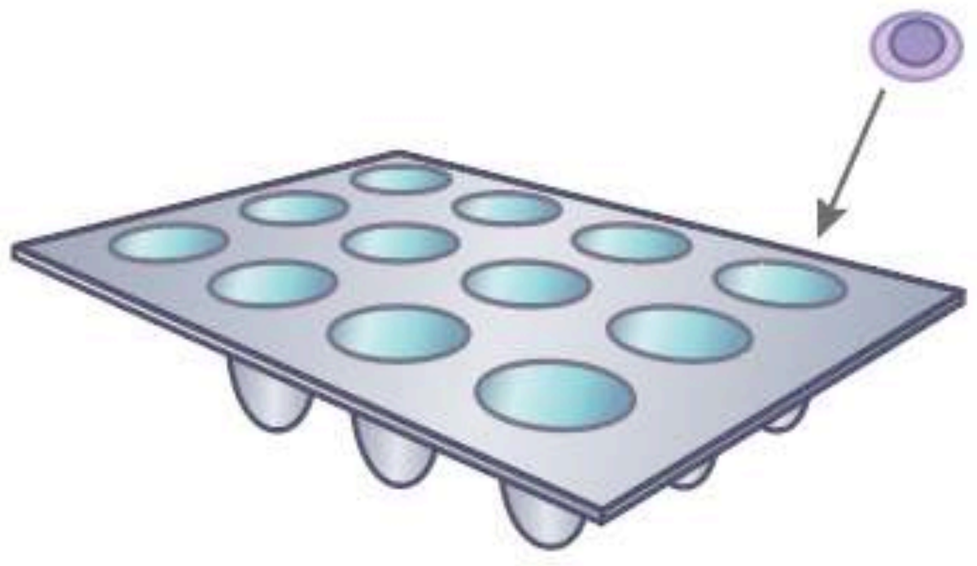
$$\text{Saturation} = 1 - \frac{n_{\text{dedup\_reads}}}{n_{\text{reads}}}$$

$$\text{Saturation} = \frac{n_{\text{duplicated\_reads}}}{n_{\text{reads}}}$$



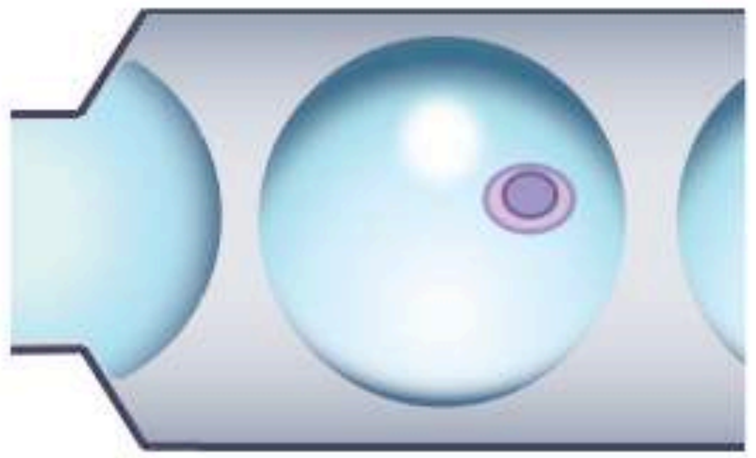
Three examples of scRNA-seq methods

Multi-well plates



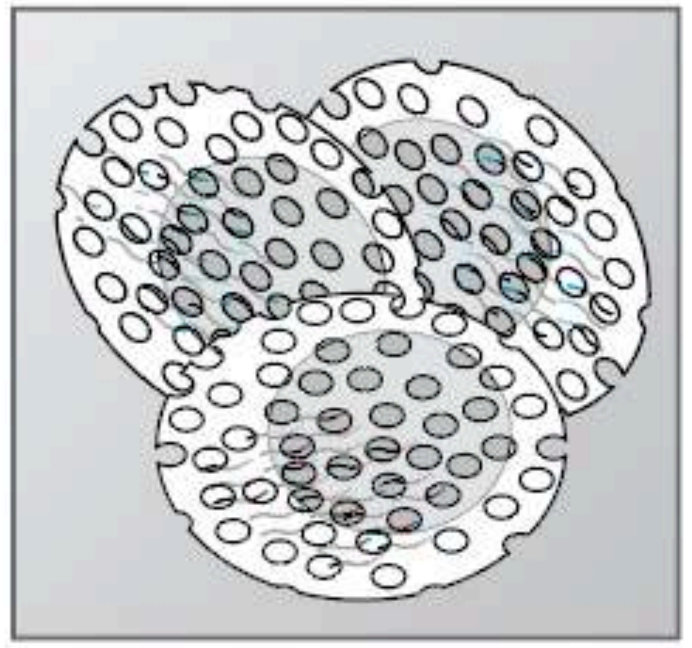
MARS-seq

Droplets



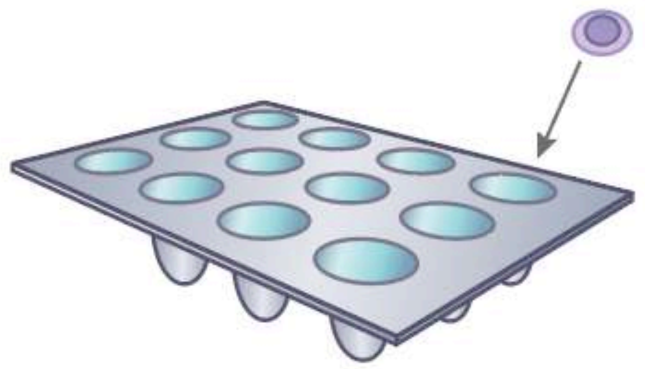
inDrops

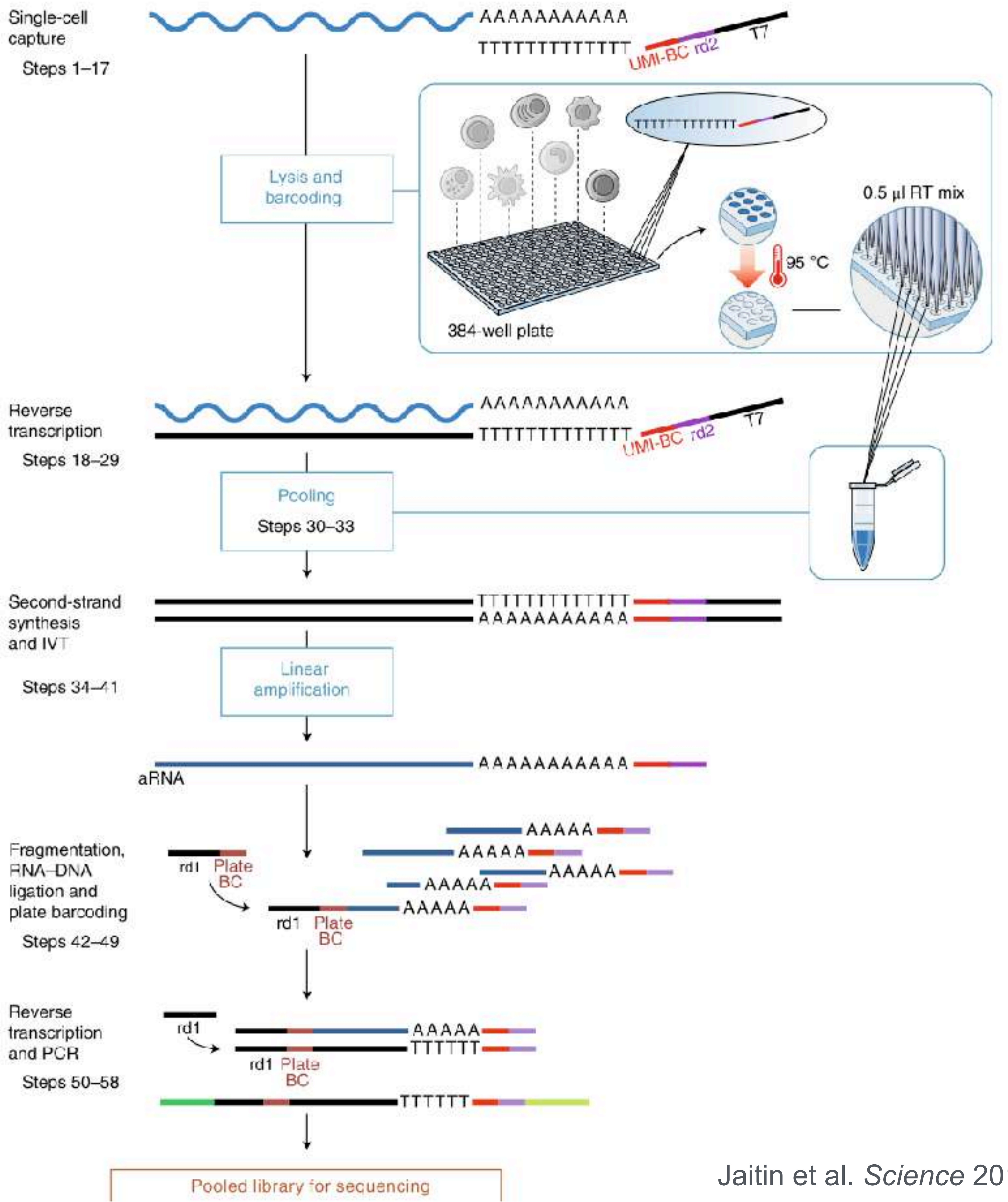
Combinatorial in-cell barcoding



sciRNA-seq

Example 1: MARS-seq **plate-based multi-tiered barcoding**





Jaitin et al. *Science* 2014

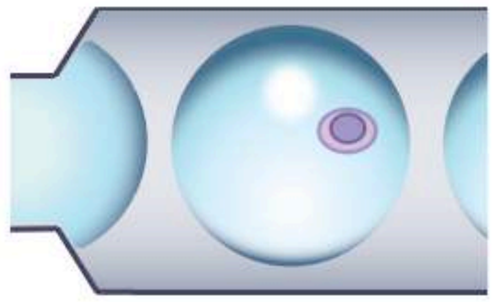
Pros:

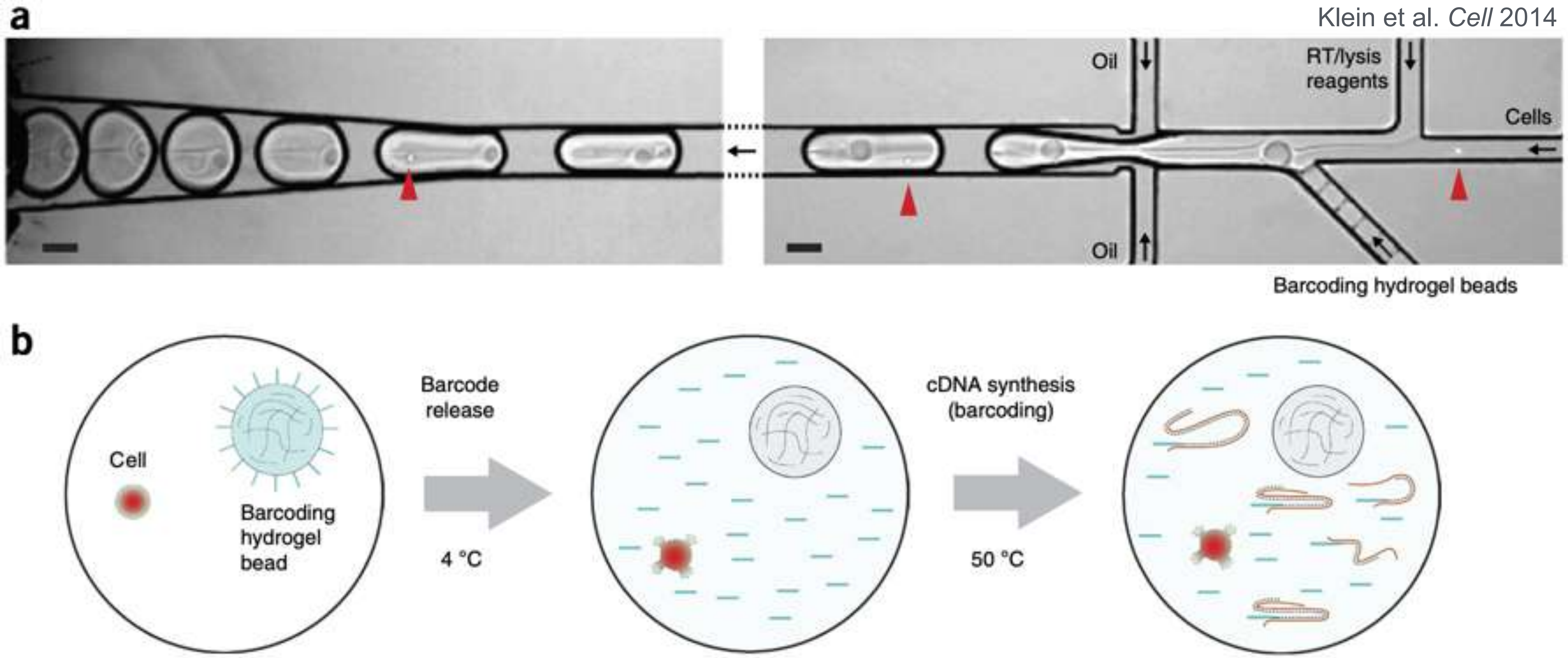
- Accurate selection of single-cell, possibility to target poulations.
- Transcriptome+FACS index data.
- Versatile (easy to modify)
- Harsh lysis conditions

Cons:

- Mid-throughput
- More expensive than (in-home) droplet methods.
- Needs FACS-sorting.
- Slow protocol

Example 2: inDrops **microfluidics droplet encapsulation and barcoding**

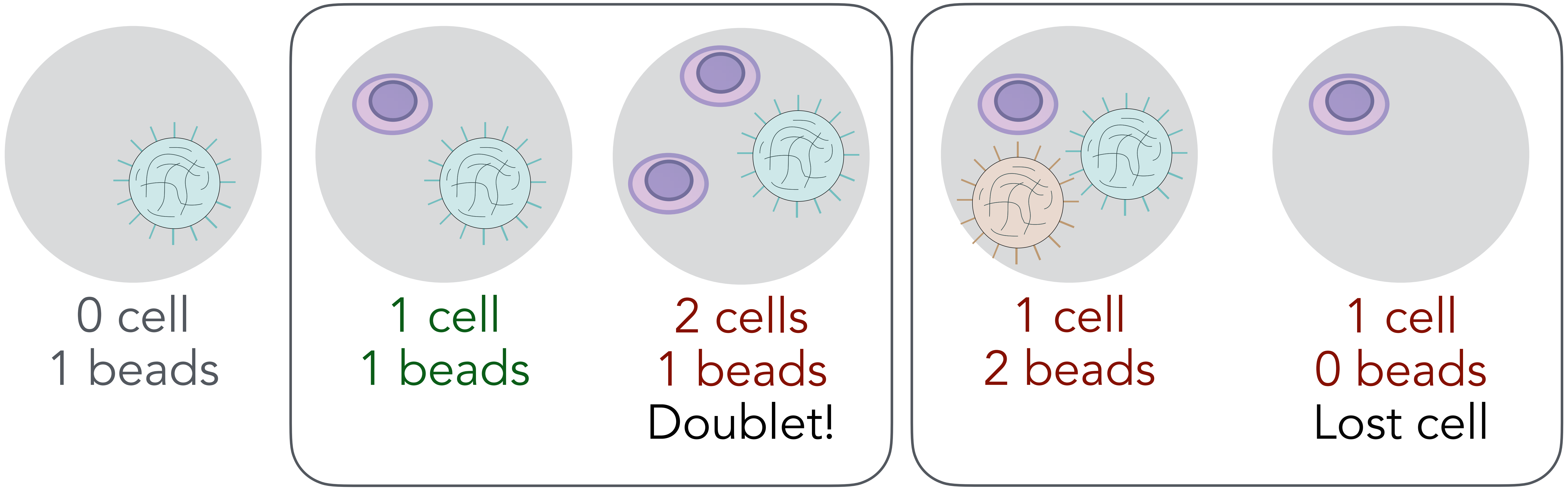




Pros:
- High-capture efficiency (±80%)
- Good sensitivity.
- Fast encapsulation

Cons:
- Doublet rates
- Mild cell lysis.
- Cell size limits (~30 um)

Poisson loading and capture rates



Cell encapsulation is explained by a Poisson distribution

$$\mathbb{P}(\text{droplet has } k \text{ cells}) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

λ is the average number of cells per

$$\lambda = \frac{N_{\text{cells loaded}}}{N_{\text{droplets}}}$$

UNLIKELY - bead encapsulation can be forced into a sub-Poisson distribution

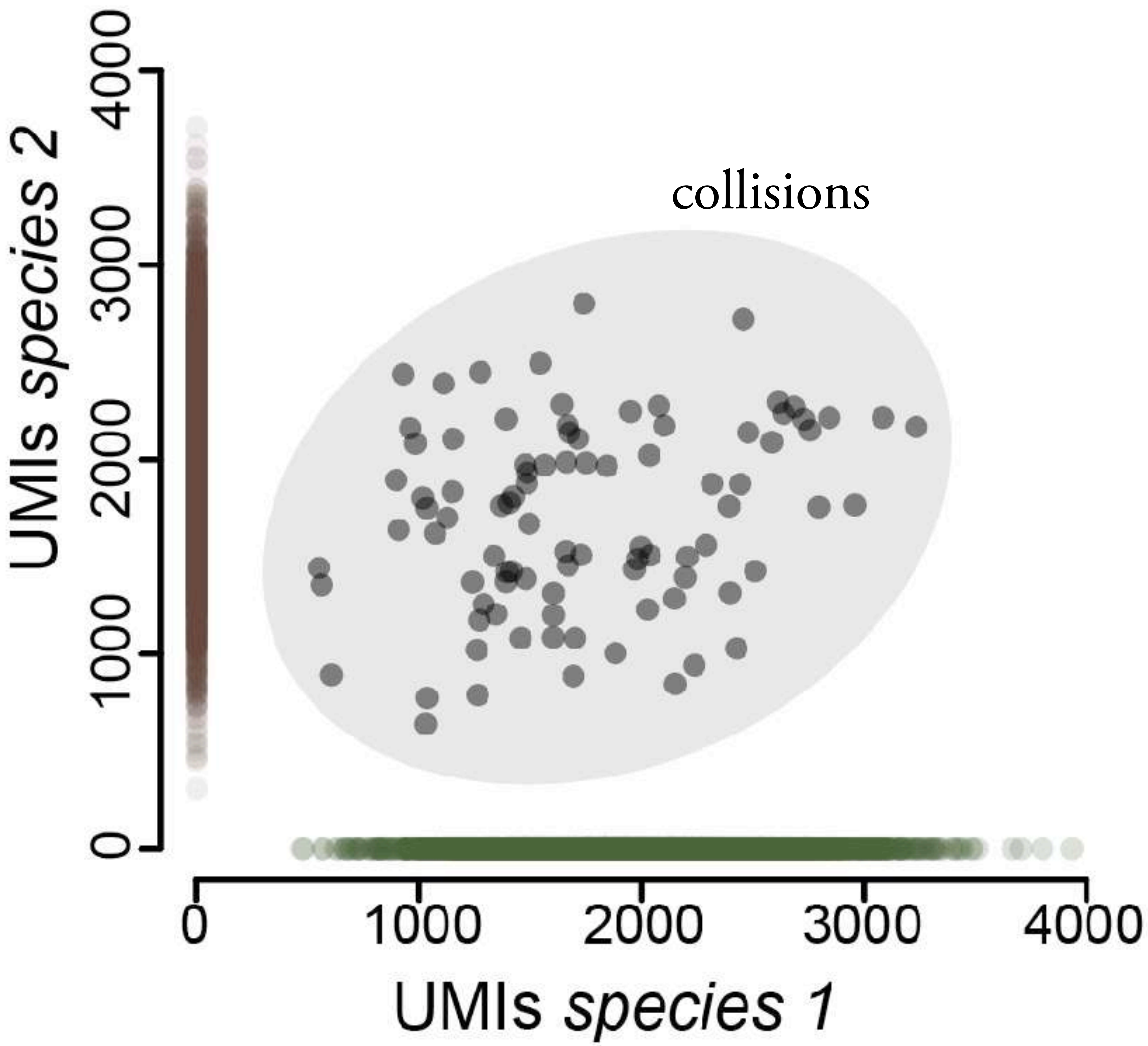
Using tighly packed hydrogel beads (10x chromium, Indrop) instead of polystirene beads (Drop-seq) massively reduce variance, resulting in practice in 1 bead per droplet.

Estimating technical multiplet rates

Multiplet Rate (%)	# of Cells Loaded	# of Cells Recovered
~0.4%	~800	~500
~0.8%	~1,600	~1,000
~1.6%	~3,200	~2,000
~2.3%	~4,800	~3,000
~3.1%	~6,400	~4,000
~3.9%	~8,000	~5,000
~4.6%	~9,600	~6,000
~5.4%	~11,200	~7,000
~6.1%	~12,800	~8,000
~6.9%	~14,400	~9,000
~7.6%	~16,000	~10,000

10X GENOMICS

Estimating technical multiplet rates



number of cells species 1

number of cells species 2

$$N = \frac{N_1 N_2}{N_{1,2}}$$

number of droplets

Observed collisions

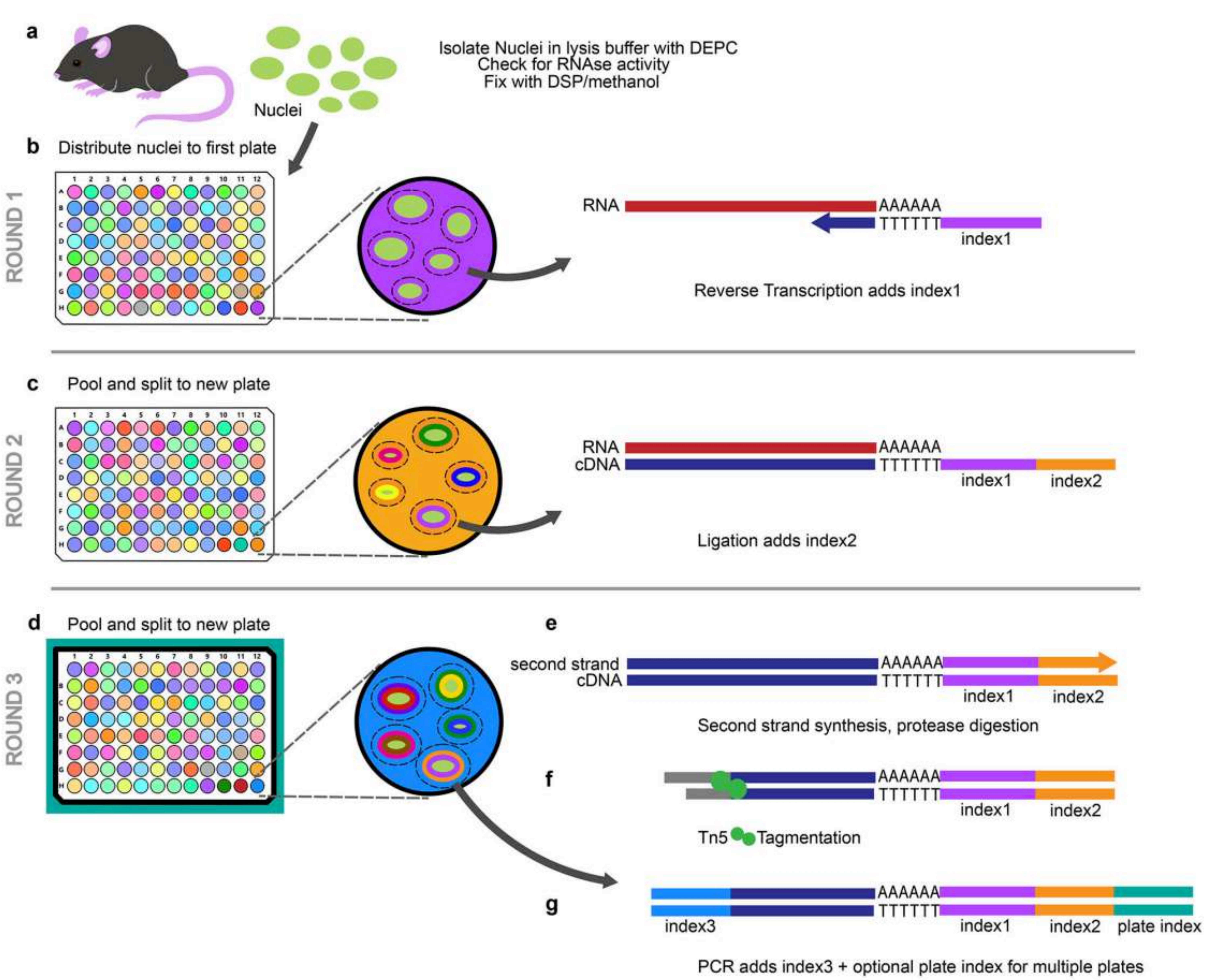
$$\mu_1 = -\ln\left(\frac{N - N_1}{N}\right) \quad \mu_2 = -\ln\left(\frac{N - N_2}{N}\right)$$

Average number of cells species 1 or 2

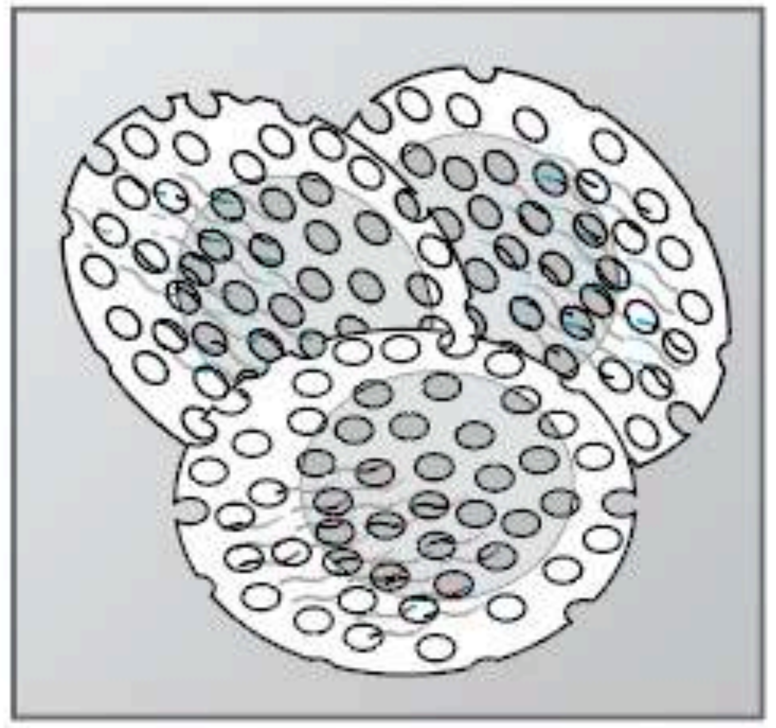
$$\boxed{M = 1 - \frac{(\mu_1 + \mu_2)e^{-\mu_1 - \mu_2}}{1 - e^{-\mu_1 - \mu_2}}}$$

Probability of a droplet with at least 1 cell containing multiple cells

Example 3: sci-RNA-seq3 **split&pool combinatorial barcoding**



Martin et al. *Nature Protocols* 2023



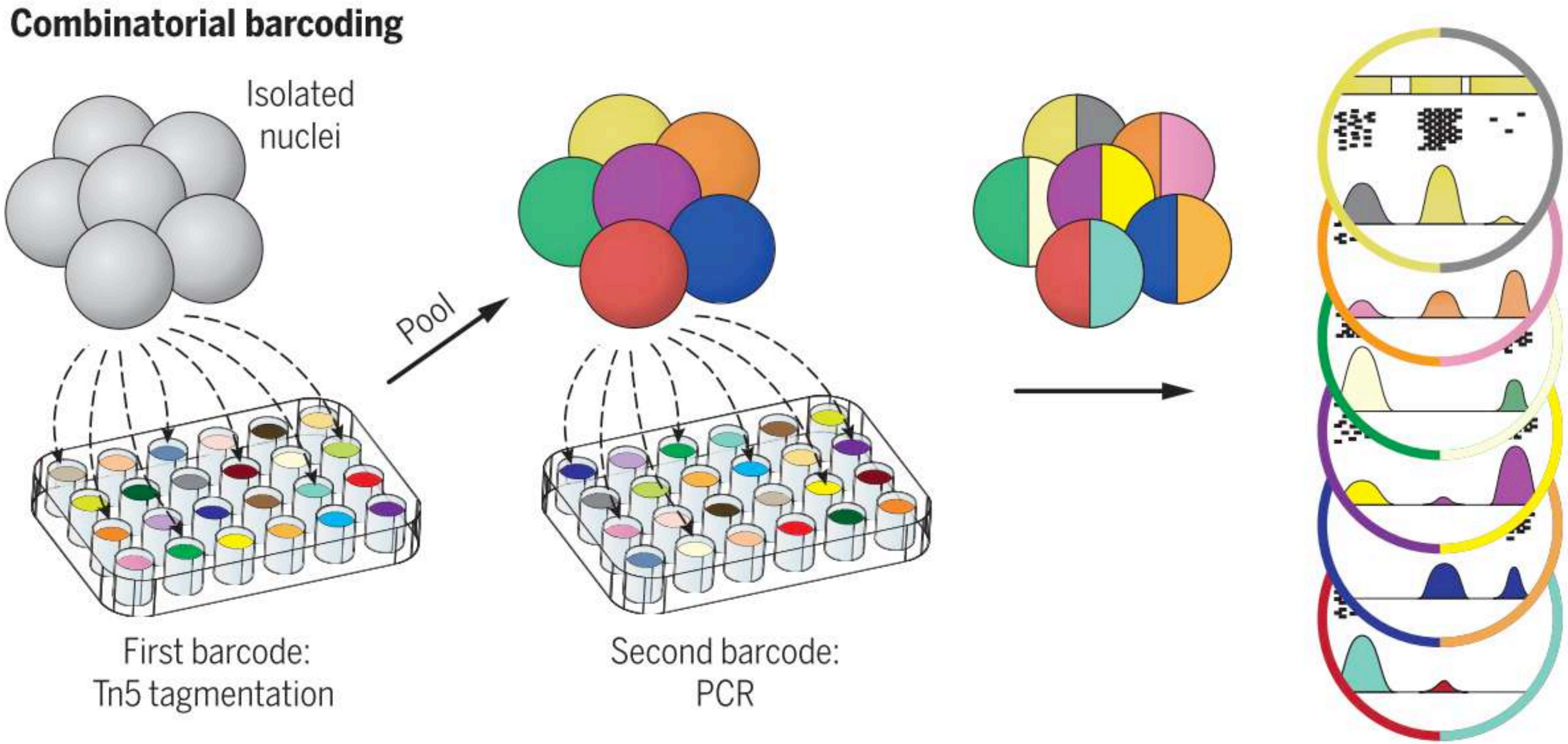
Pros:

- Extremely high throughput
- Very low per-cell costs, <0.1 USD)
- No equipment required*

Cons:

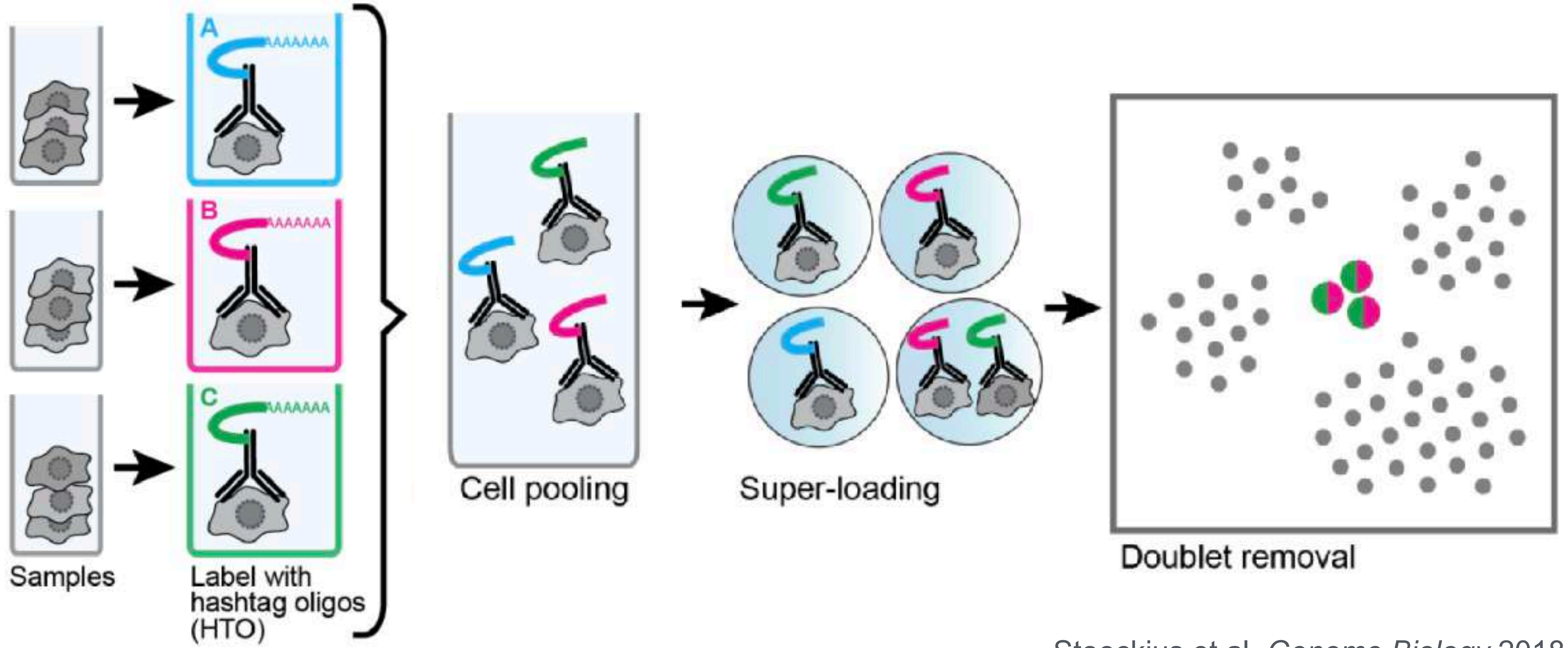
- Very low sensitivity
- Requires fixed cells/nuclei
- Expensive initial set-up (BCs)
- 3′-biased, no full-length.

Combinatorial barcoding is at the core of many single-cell genomics methods!



Cell hashing for sample overloading

Antibodies (against ubiquitous surface proteins) loaded with unique polyA barcodes



Stoeckius et al. *Genome Biology* 2018

Importantly, it also allows improved capture rates (for low input samples, combined)

Cell hashing without antibodies: ClickTag oligonucleotides

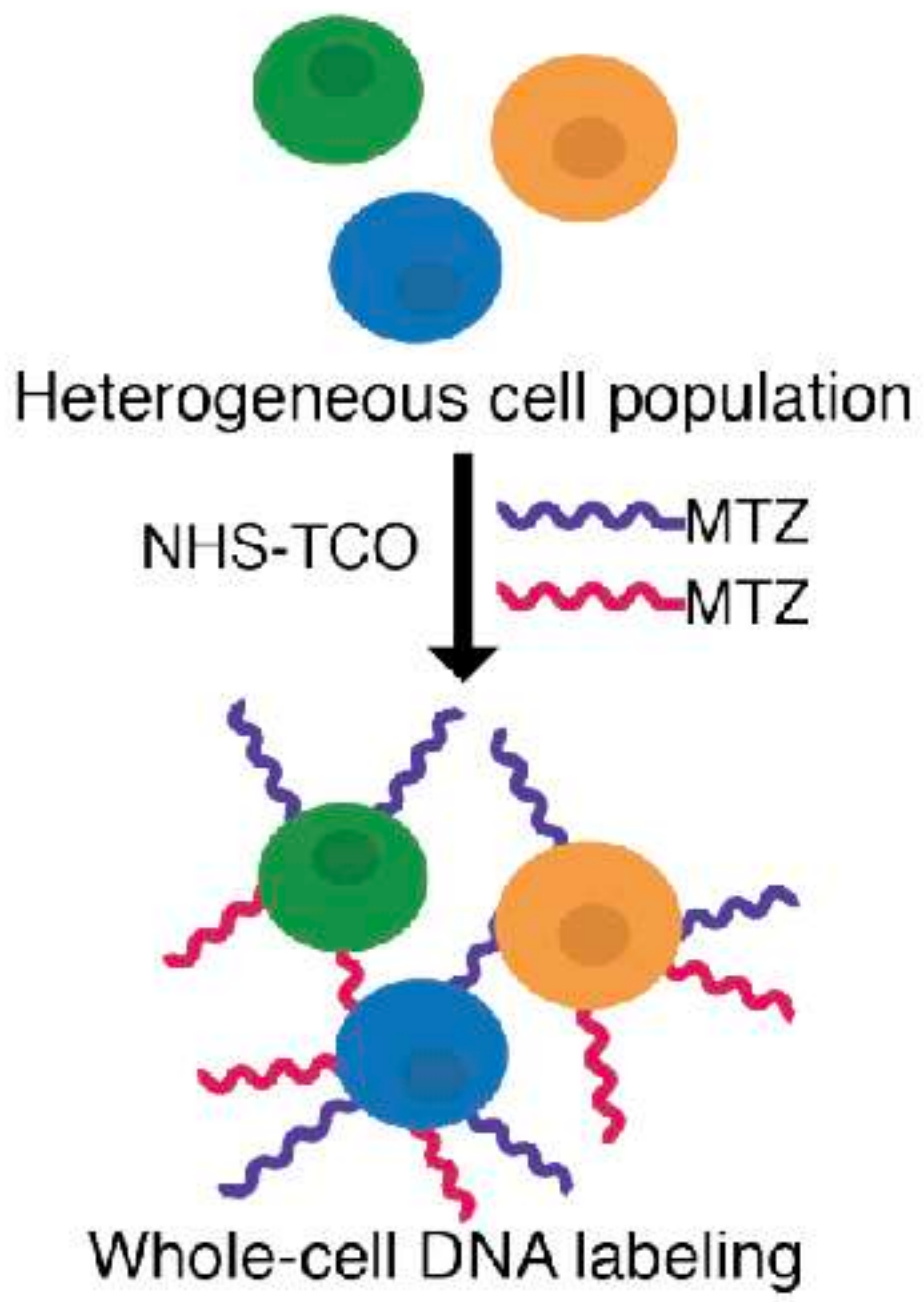
Universal sample multiplexing by chemically labelling cells.

Highly Multiplexed Single-Cell RNA-seq for Defining Cell Population and Transcriptional Spaces

Jase Gehring, Jong Hwee Park, Sisi Chen, Matthew Thomson, Lior Pachter

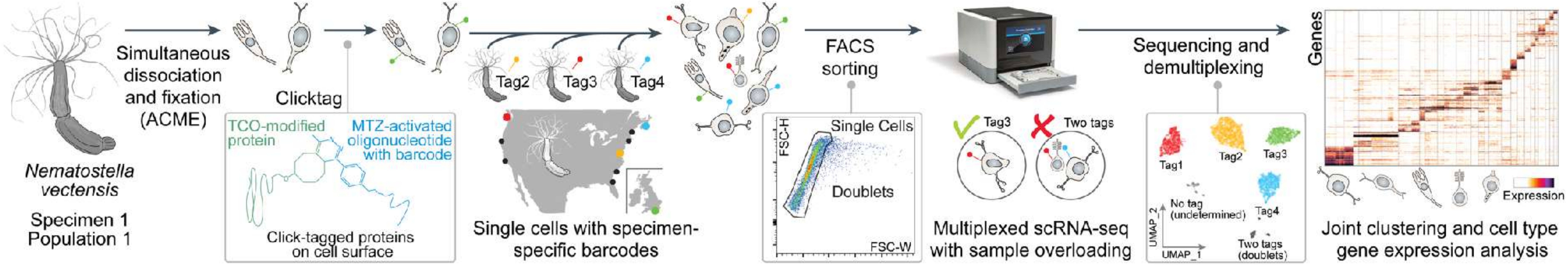
doi: https://doi.org/10.1101/315333

Methyltetrazine (MTZ)-activated barcoded oligonucleotides are attached to exposed NHS-reactive amines in a one-pot reaction.



Cell hashing without antibodies: ClickTag oligonucleotides

Example application: low-input, specimen-resolved scRNA-seq atlases



Open issues in scRNA-seq methods

- Sample prep (dissociation, nuclei extraction, etc.) is still the major bottleneck.
- Reaching very high capture efficiencies: studying small specimens (e.g. embryos) without pooling.
- Cell fixation/preservation: decoupling sampling from single-cell processing (e.g. field work).
- Trade-off between sensitivity & scalability/costs.
- Glass ceiling: sequencing costs… (new sequencing technologies, e.g. UltimaGenomics)

Part 2 - scRNA-seq analysis

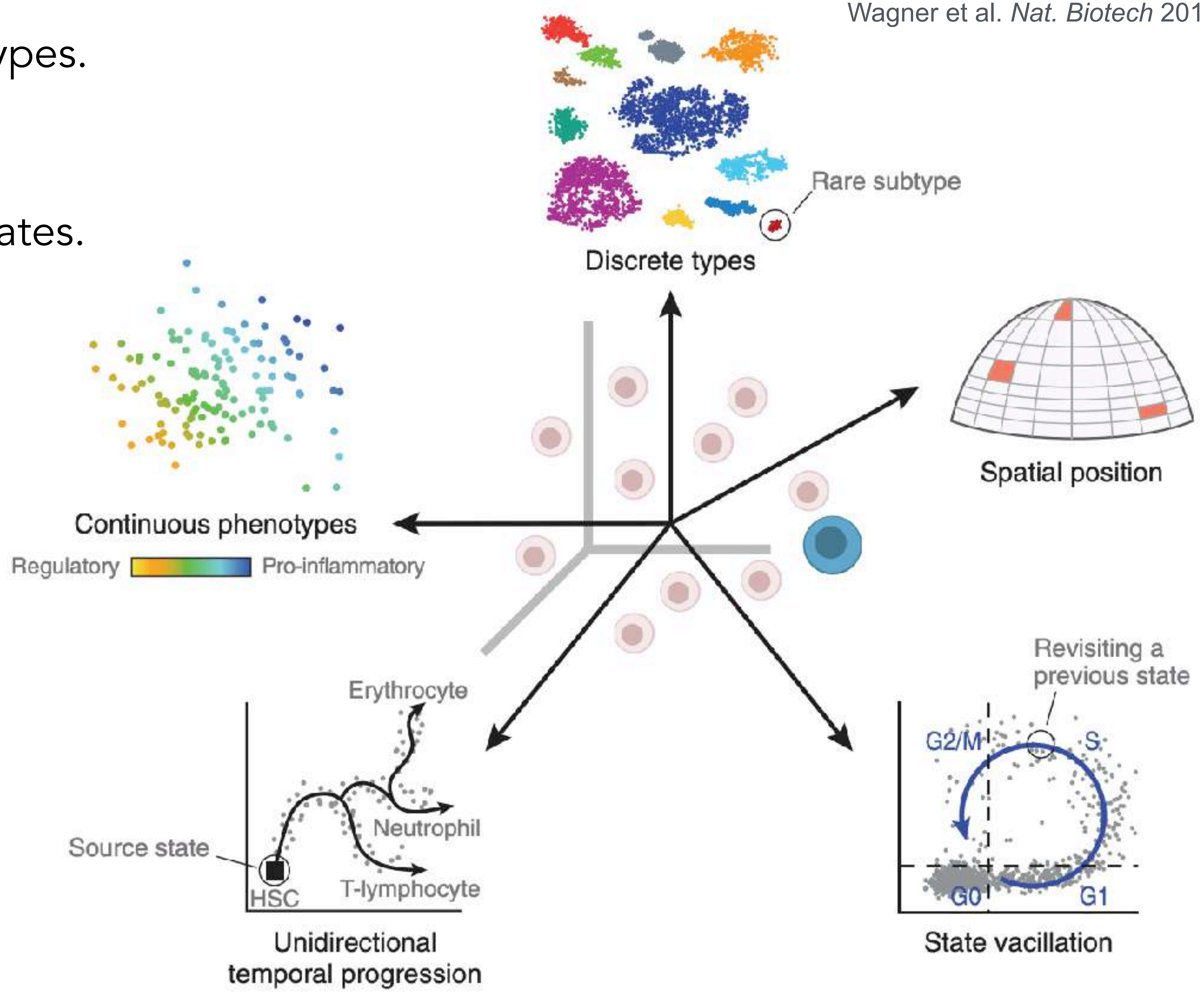
The vectors of cellular identity

Multiple factors define cell type identity:

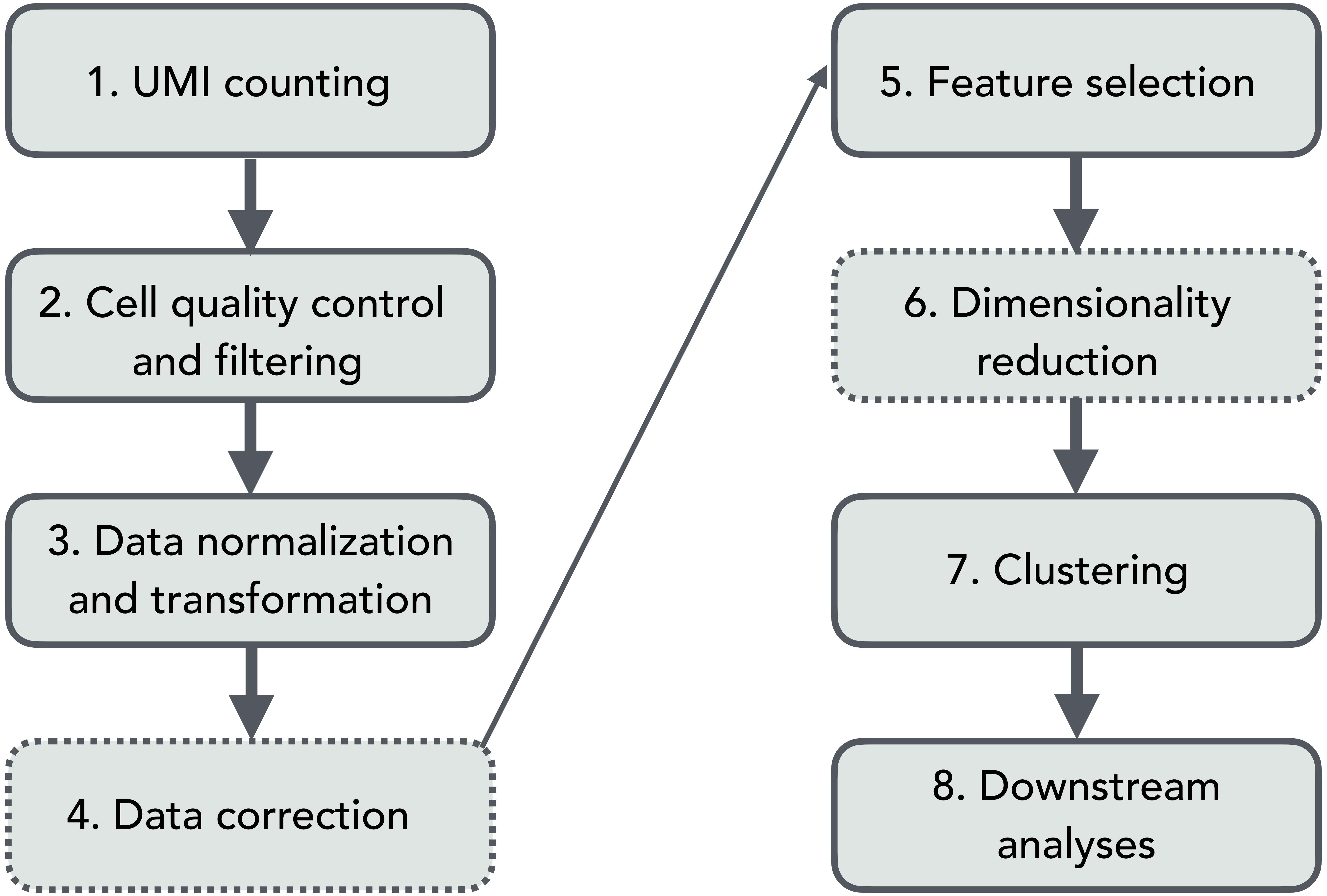
- Membership in a hierarchy/taxonomy of cell types.
- Time-dependent processes (e.g. cell cycle).
- Response to the environment/physiological states.
- Spatial position

In practice, in most situations the cell type identity signal dominates the transcriptional profile.

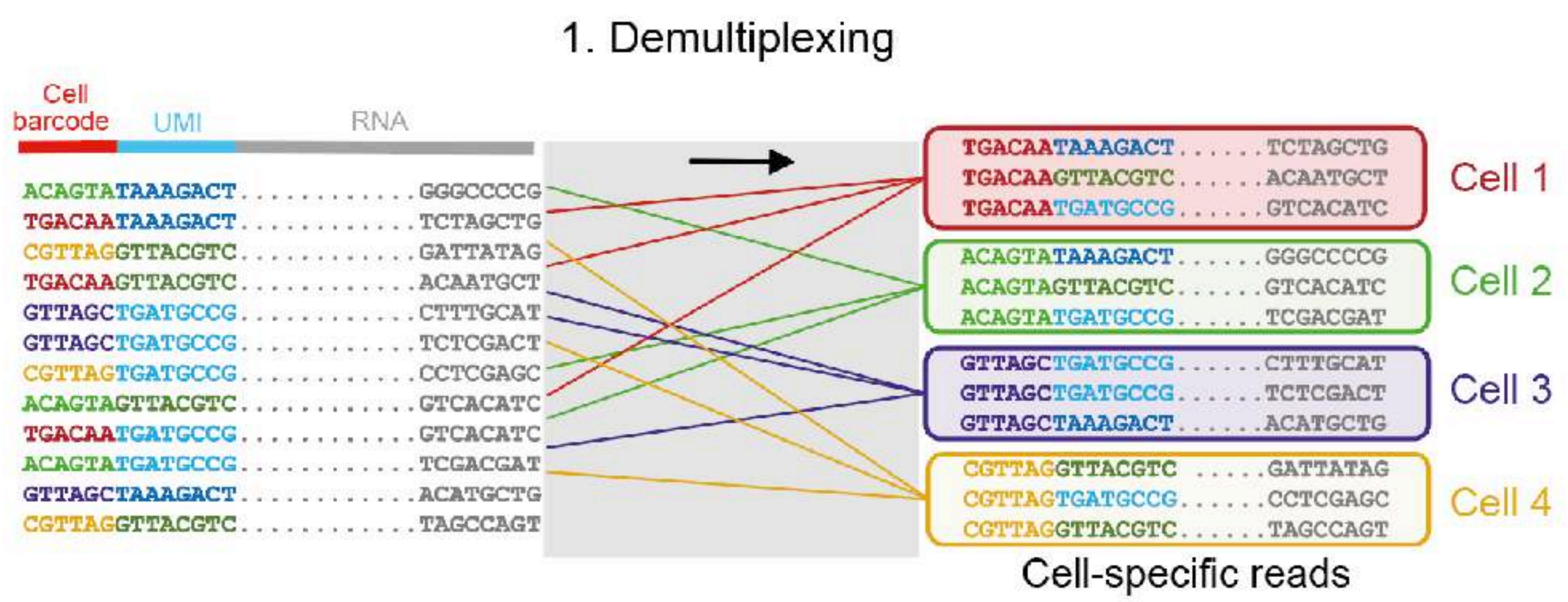
Wagner et al. *Nat. Biotech* 2017



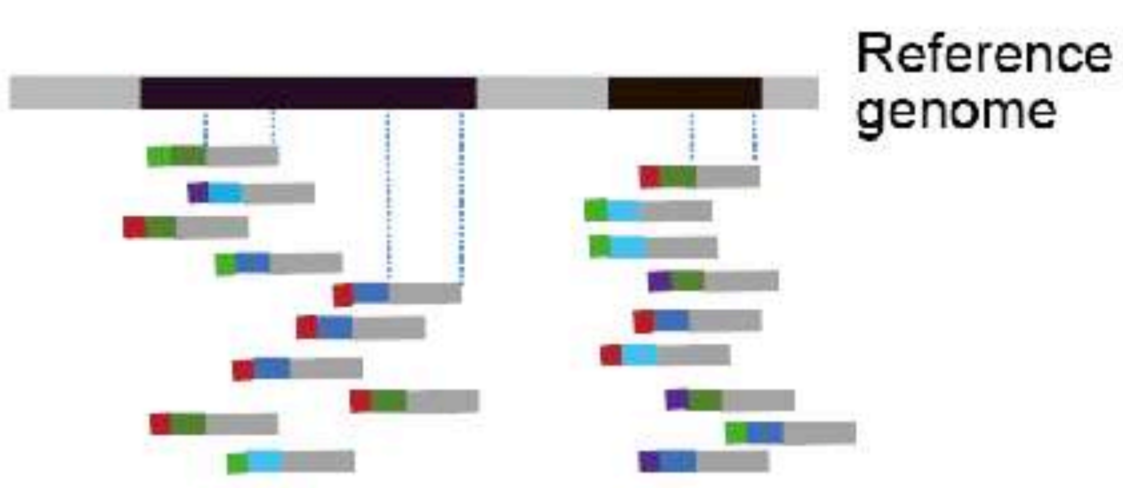
Standard scRNA-seq analysis pipeline



1. Demultiplexing and transcript counting (assigning reads to cells and to genes)



2. Mapping



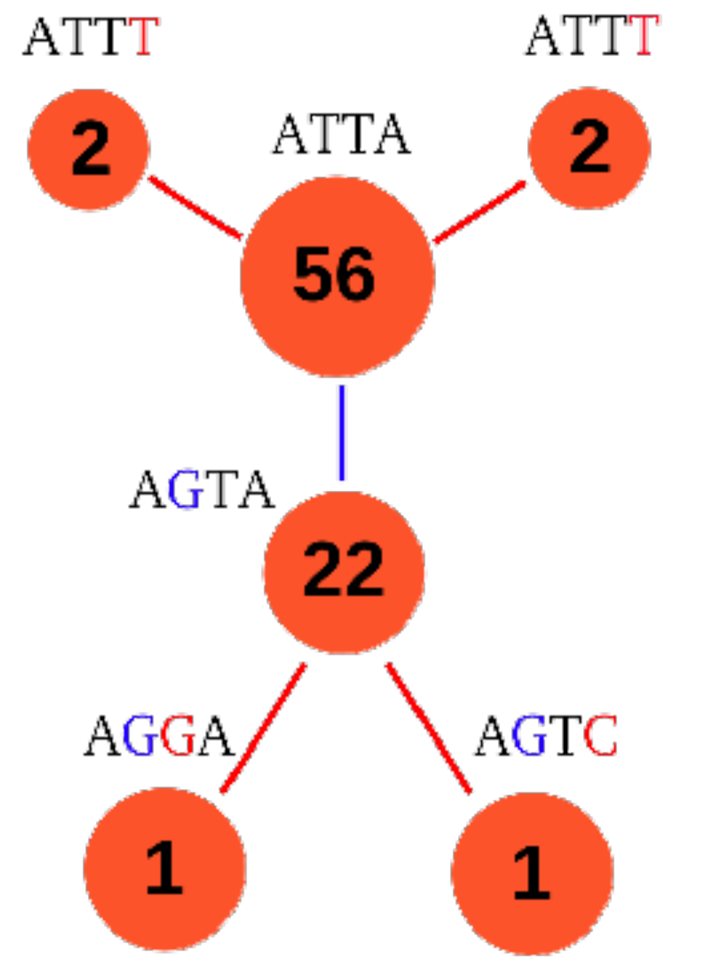
Reads to genes

3. Quantification

	Cell1	Cell2	...	CellN
Gene1	3	2	.	13
Gene2	2	3	.	1
Gene3	1	14	.	18
...	.	.	.	.
...	.	.	.	.
...	.	.	.	.
GeneM	25	0	.	0

Gene expression matrix

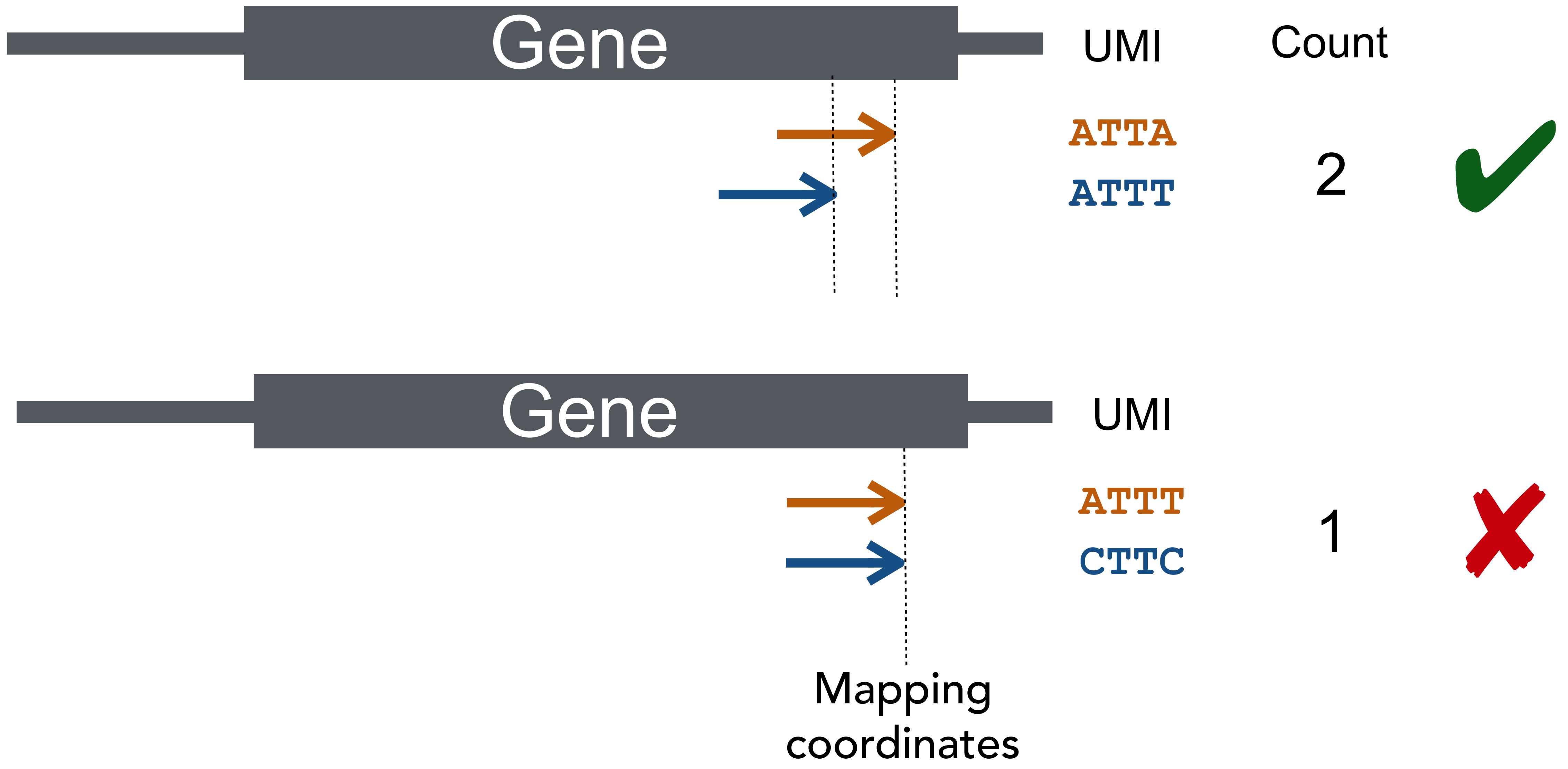
Modified from Lafzi et al. *Nat. Protocols* 2018



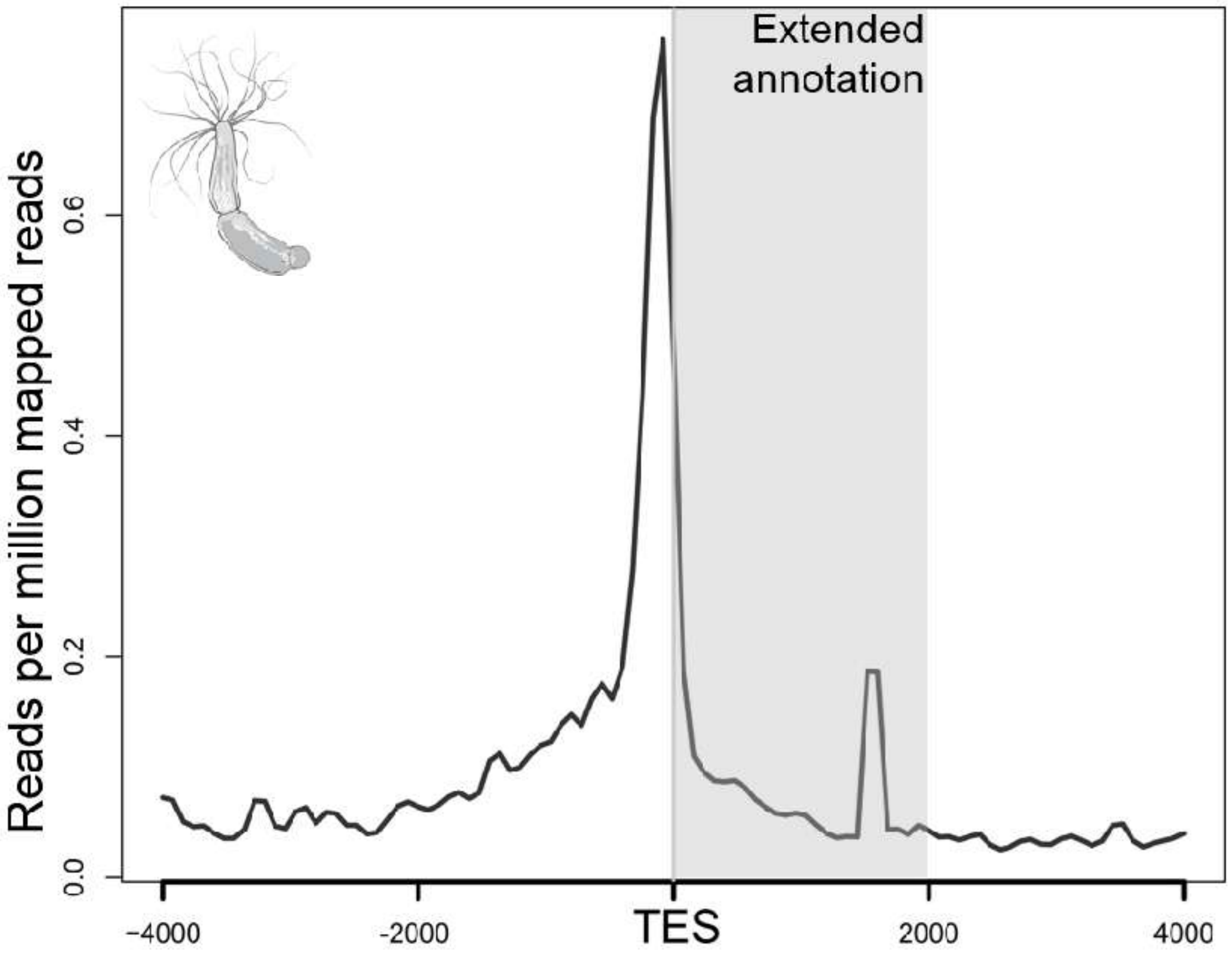
Sequencing errors
PCR errors

Informative features:

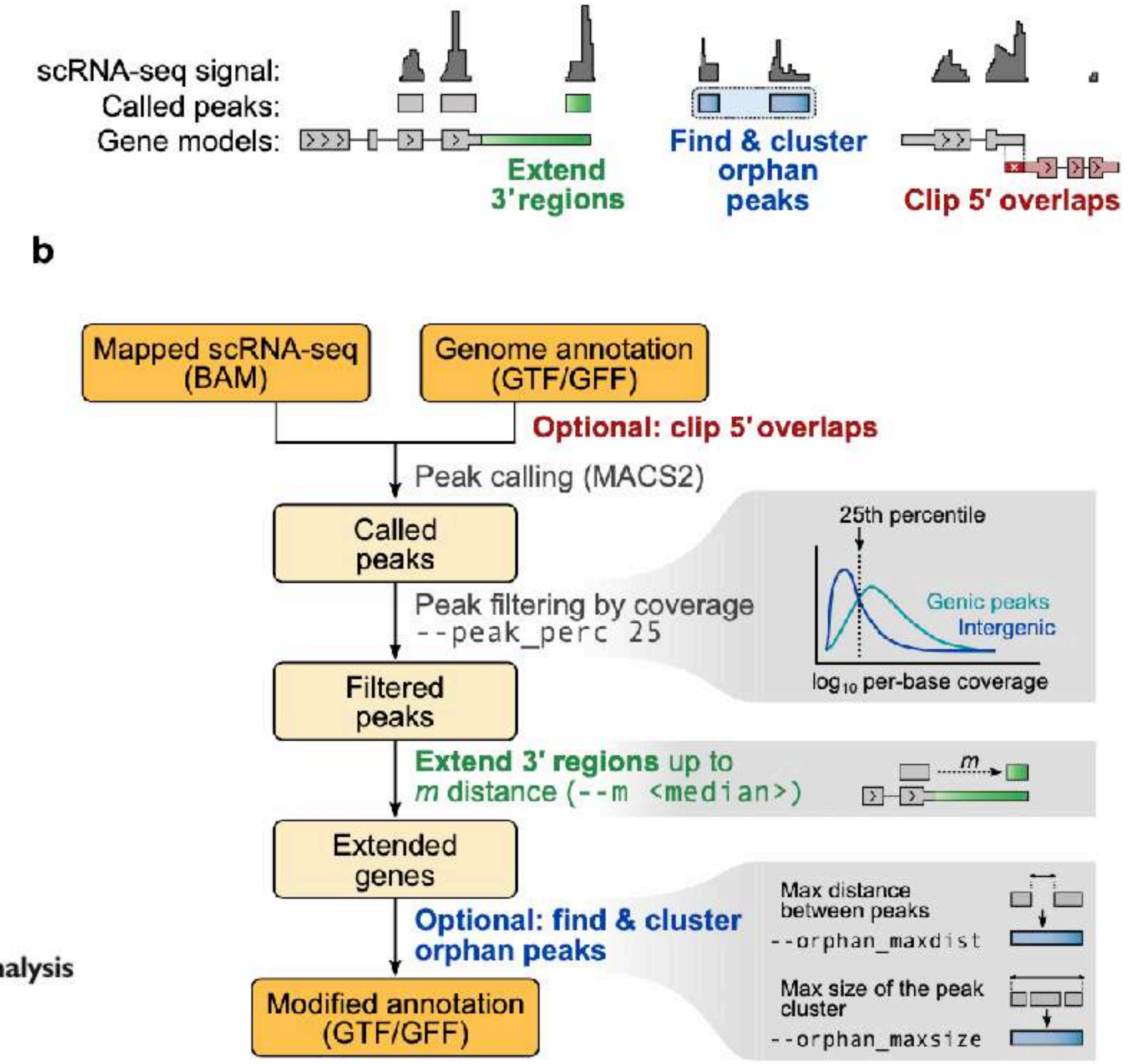
1. Base-call quality
2. Adjacency structure
3. Gene expression level
4. Mapping position



The impact of **incomplete gene models** in scRNA-seq data analysis



Recover unassigned reads by
3′ extension and intergenic bins



bioRχiv
THE PREPRINT SERVER FOR BIOLOGY

GeneExt: a gene model extension tool for enhanced single-cell RNA-seq analysis

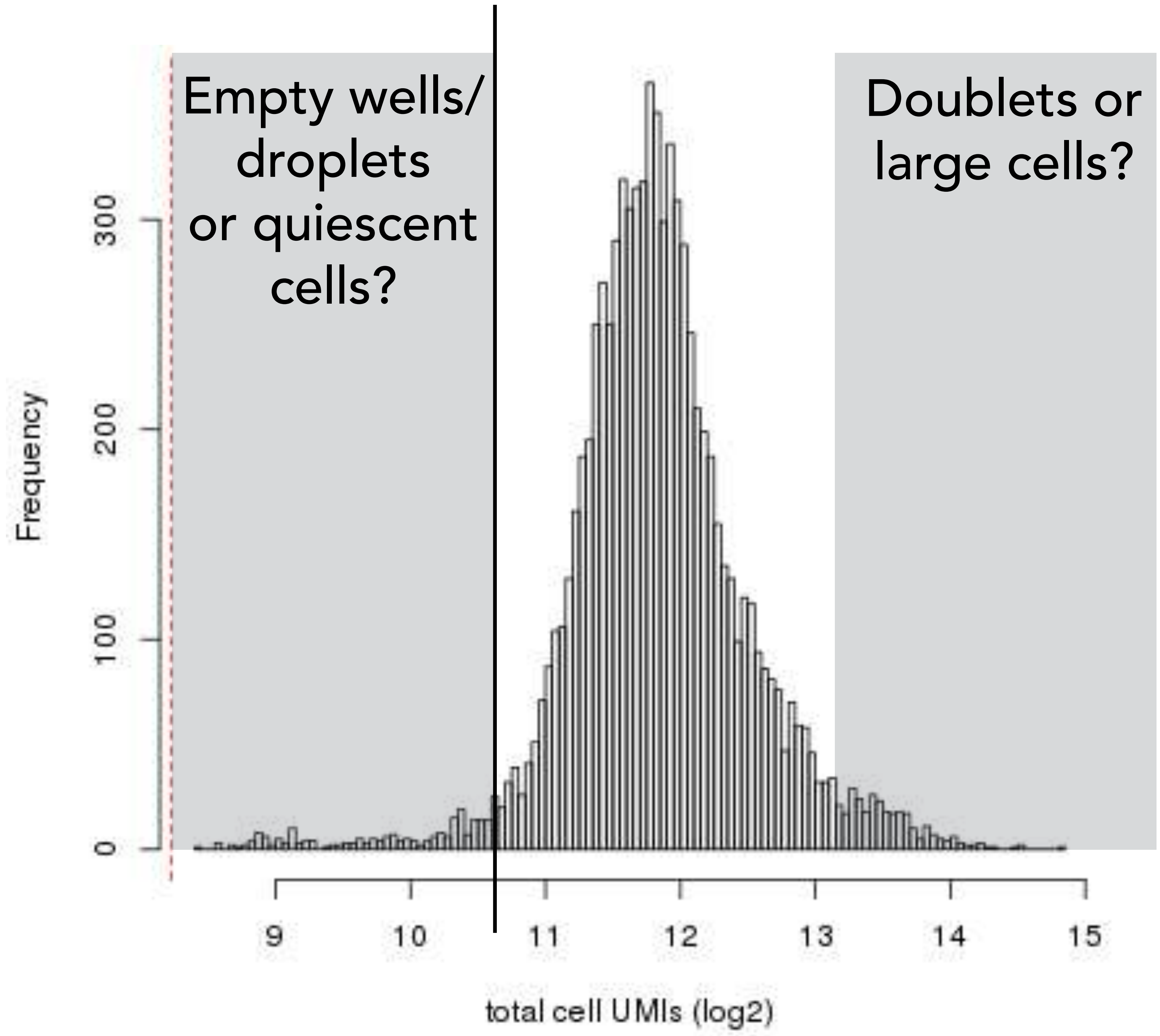
Grygoriy Zolotarov, Xavier Grau-Bové, Arnau Sebé-Pedrós

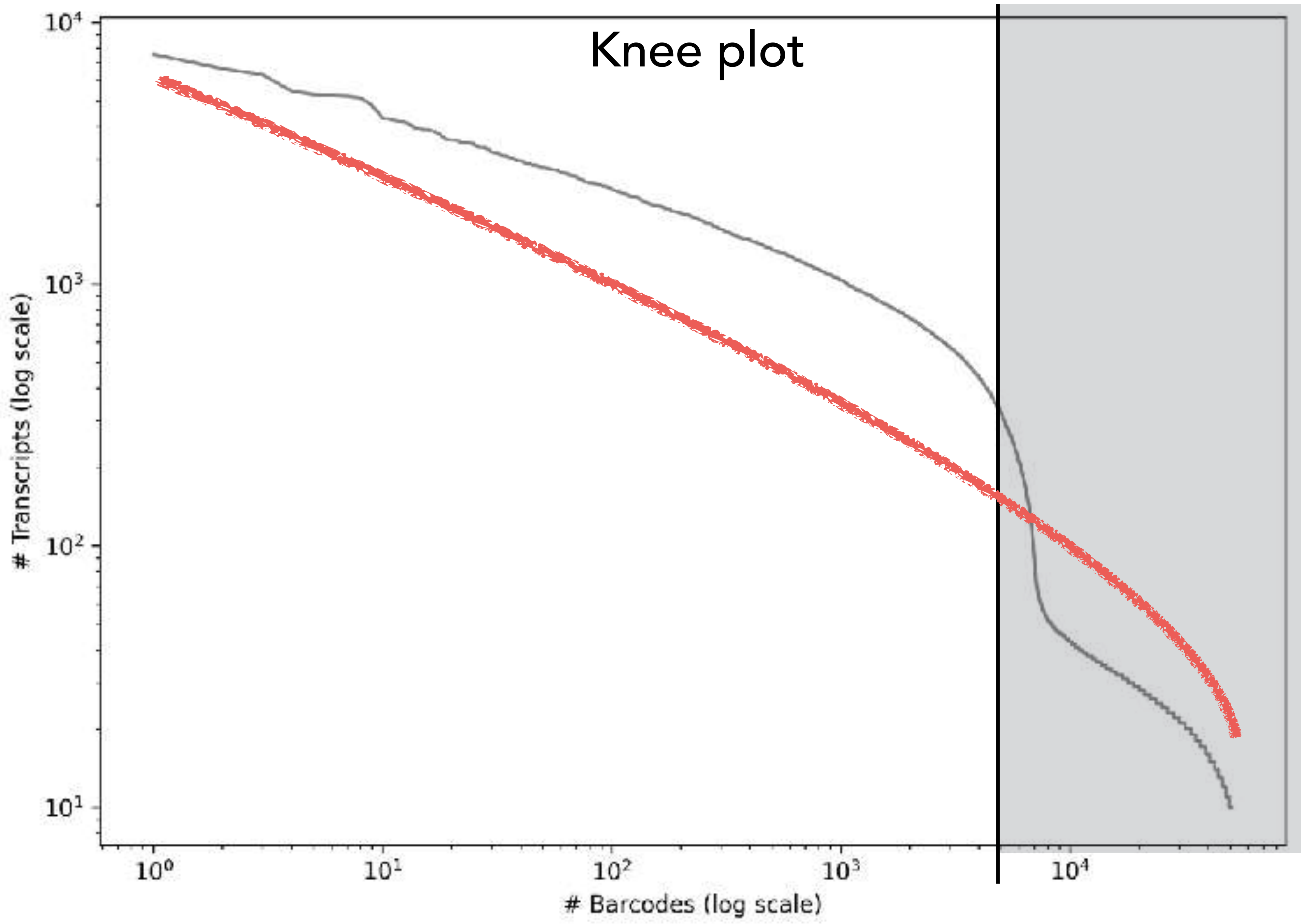
doi: https://doi.org/10.1101/2023.12.05.570120

2. Calling cells from non-cells and filtering bad cells

Informative features:

1. UMI counts per cell (cell size)
2. mitochondrial genes
3. ribosomal rRNAs
4. initial cell input (expected N of cells)





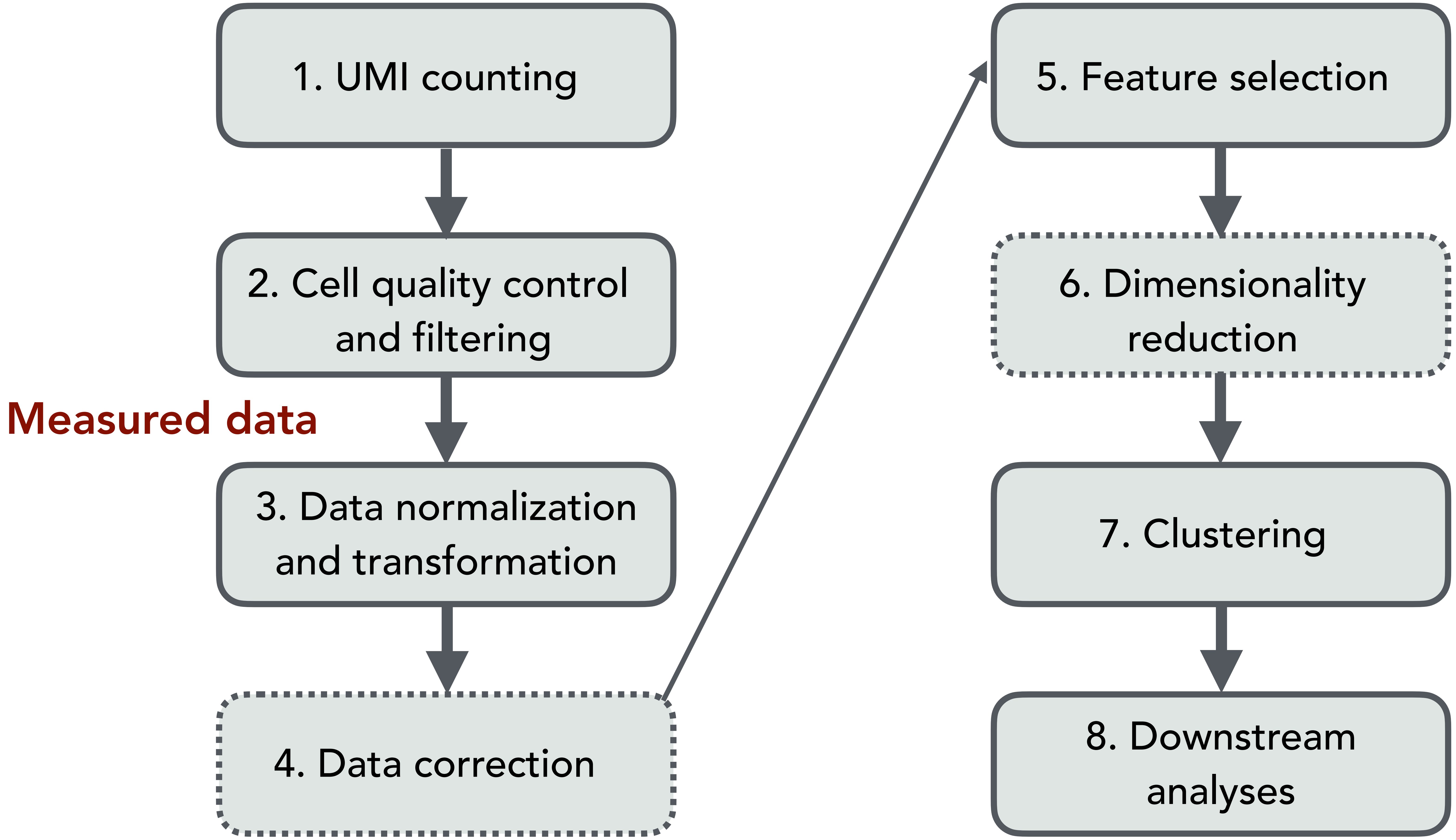
Example tools/strategies:

- *dropEst*: cell calling, based of cell UMI counts distribution. Used by CellRanger.
- *emptyDrops*: cell calling, based on deviations from background RNA distribution.
- *Scrublet*: doublet identification by simulation from observed expression.
- *DoubletFinder*: similar to Scrublet.

General QC tips:

- Be permissive
- Do not attempt to model what we don't understand
- Perform QC iteratively

Standard scRNA-seq analysis pipeline

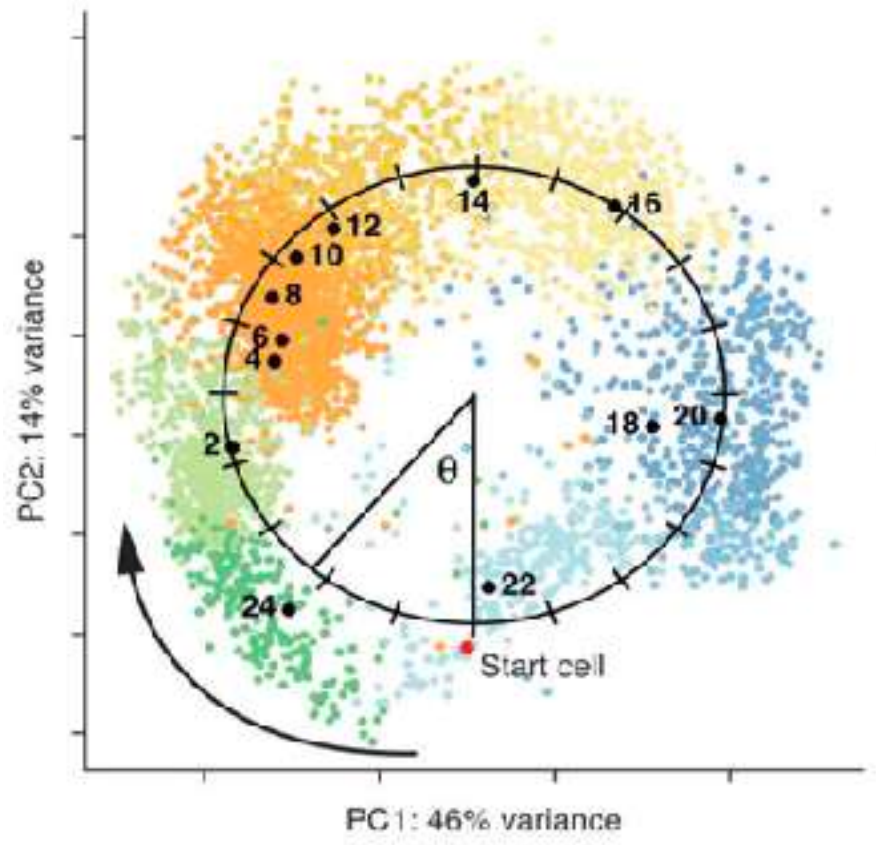


3. Data normalization and transformation

- Count depth scaling (scaling factor: 10,000 or 1,000,000).
- Random downsampling (only if small cell size variance, severe data loss)
- Size factor estimation (e.g. in *scran*), assumes most genes stable, diff. technical
- Parametric normalisation (e.g. neg binomial), principled variance stabilisation.
- No normalisation, if you use similarity metrics that are scale-invariant (e.g. correlation).
- Binary transformation
- Model-based latent representations (good for data integration/batch correction, e.g. *scVI*)
- Log-transformation: stabilise variance and reduce skewness (compress large values). Often used with count depth scaling.

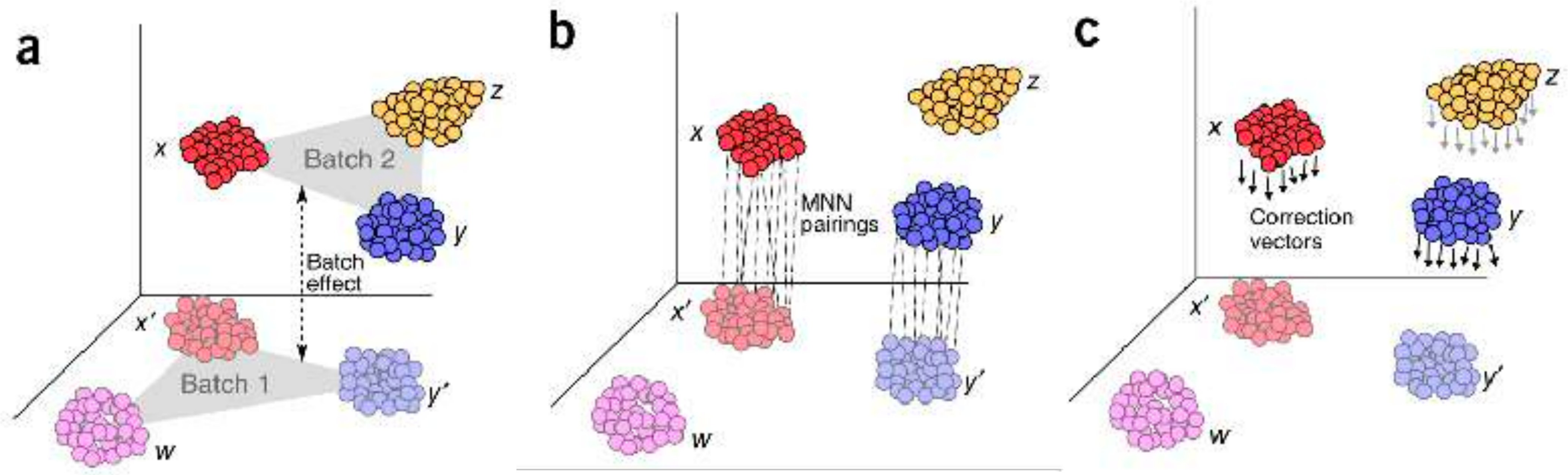
4. Data correction: regressing out unwanted covariates and imputing data

1. Biological effects. e.g. cell cycle.



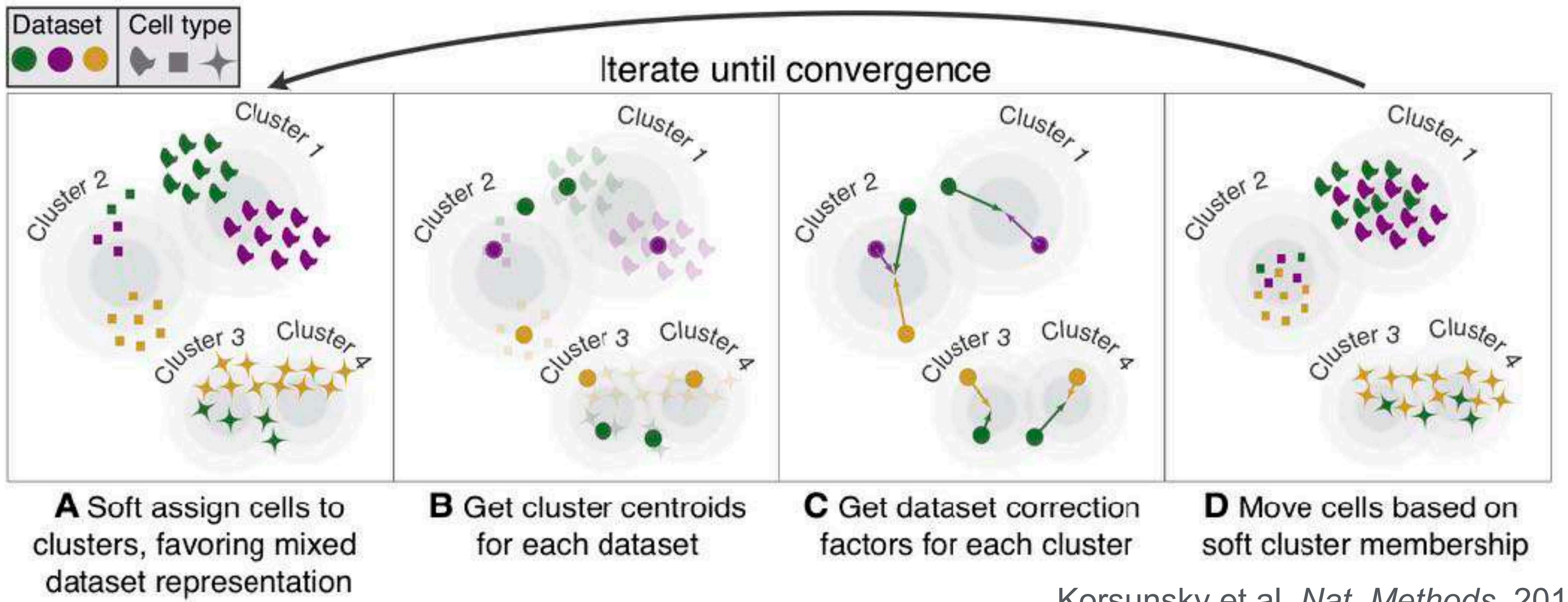
2. Batch effects. Methods:

- Identify and remove "batchy" genes.
- Mutual Nearest Neighbors (MNN): handles compositional differences between datasets.



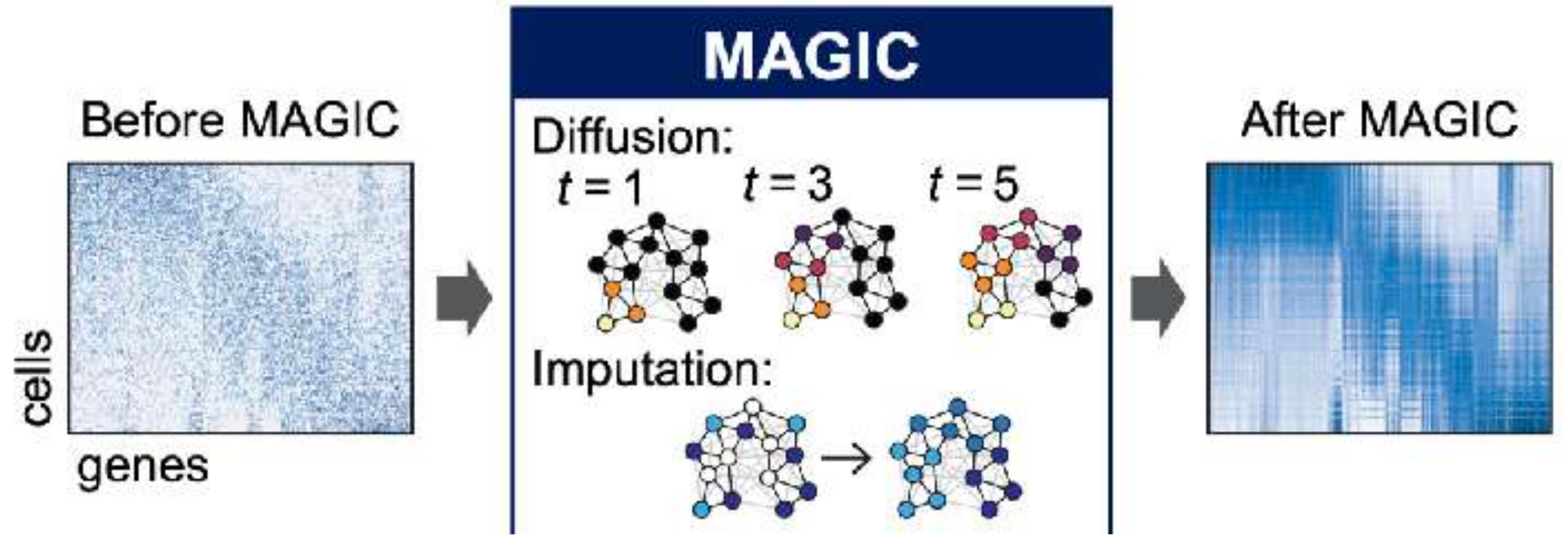
Haghverdi et al. *Nat. Biotech.* 2018

- Harmony



Korsunsky et al. *Nat. Methods.* 2019

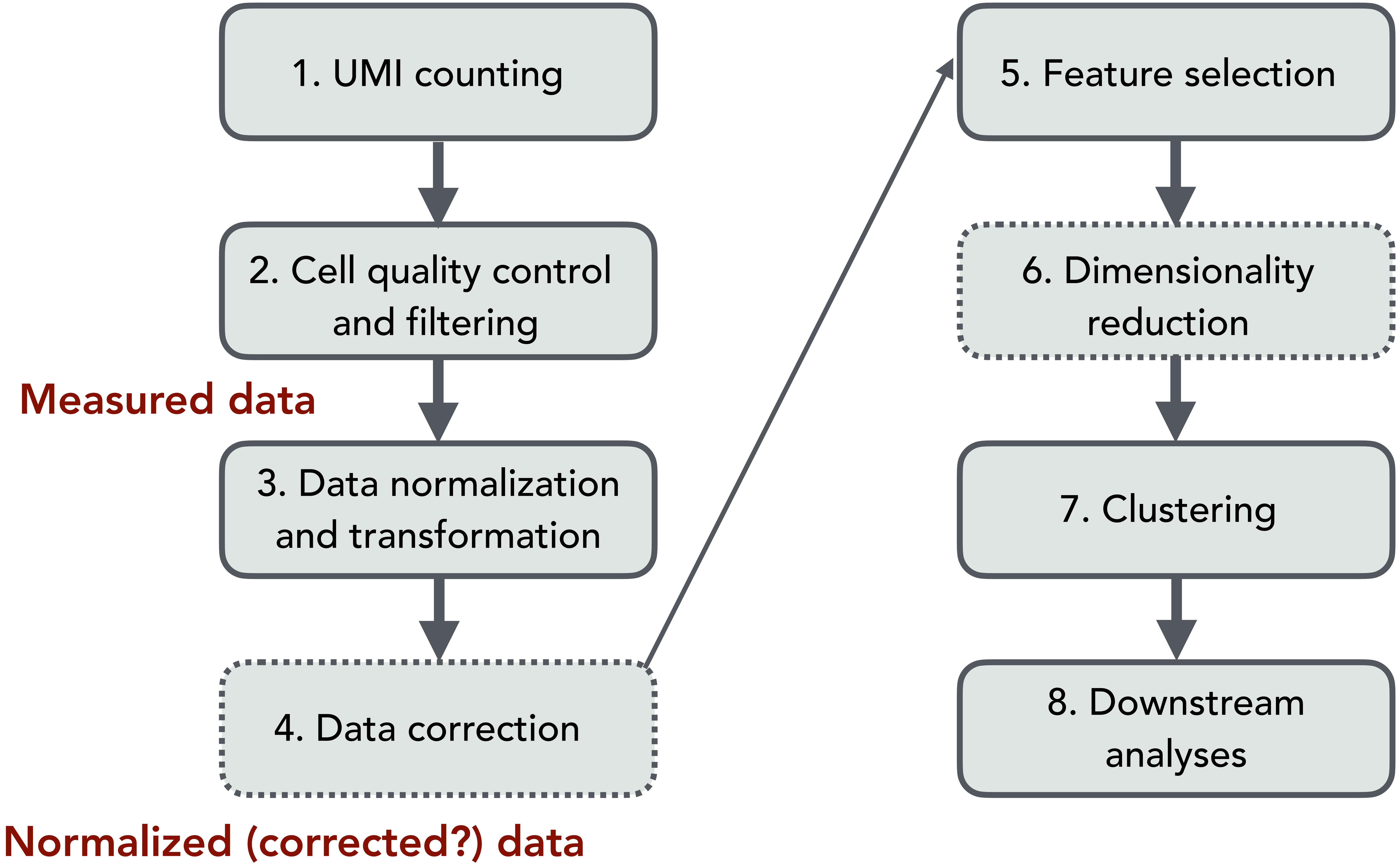
3. Data imputation to compensate for the sparsity of single-cell data



van Dijk et al. *Cell* 2018

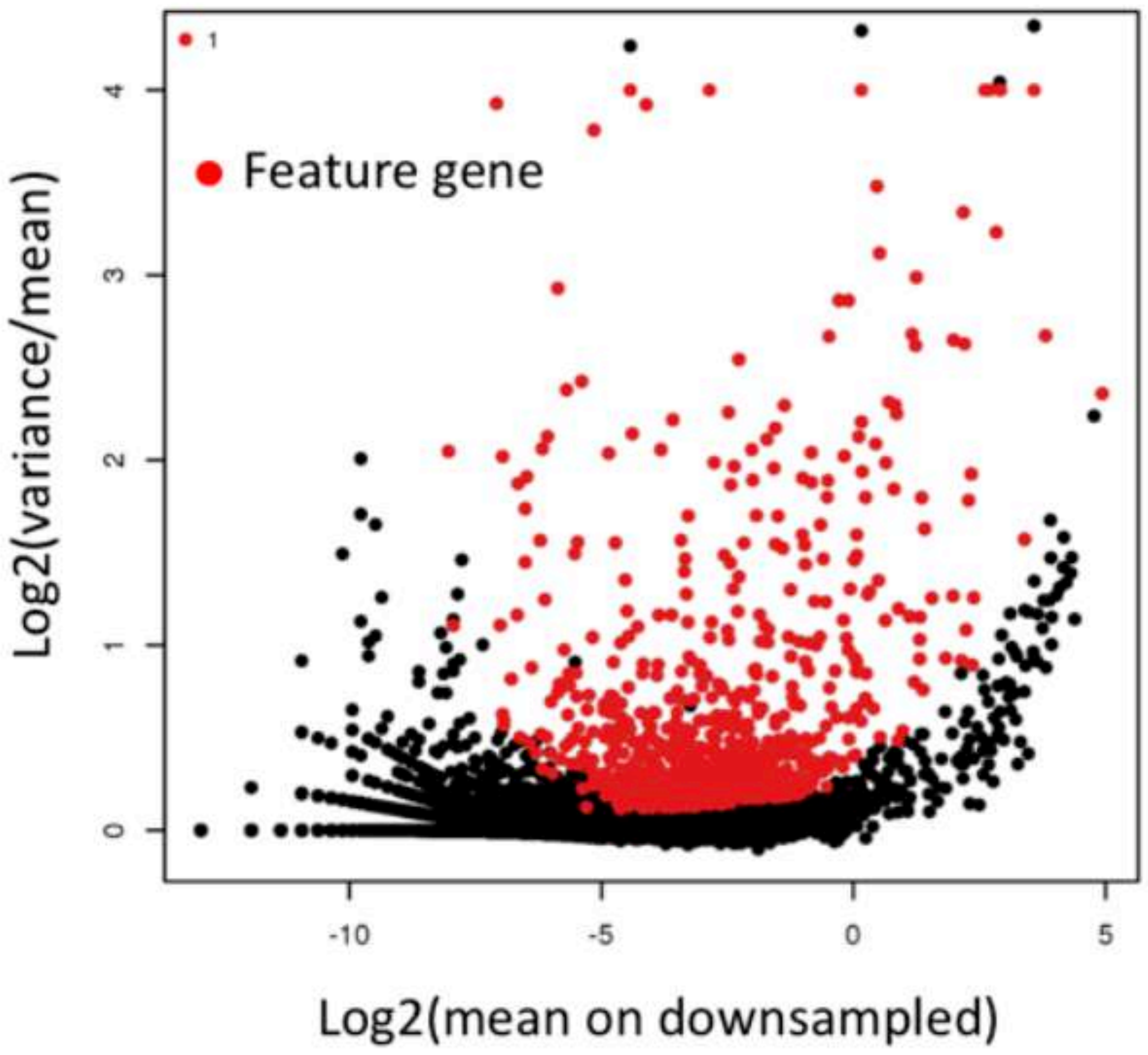
Generally a bad idea - instead, use metacells!

Standard scRNA-seq analysis pipeline



5. Feature selection: variable genes for downstream clustering

Select genes with high variance (normalized by the mean)
and a minimal total expression.

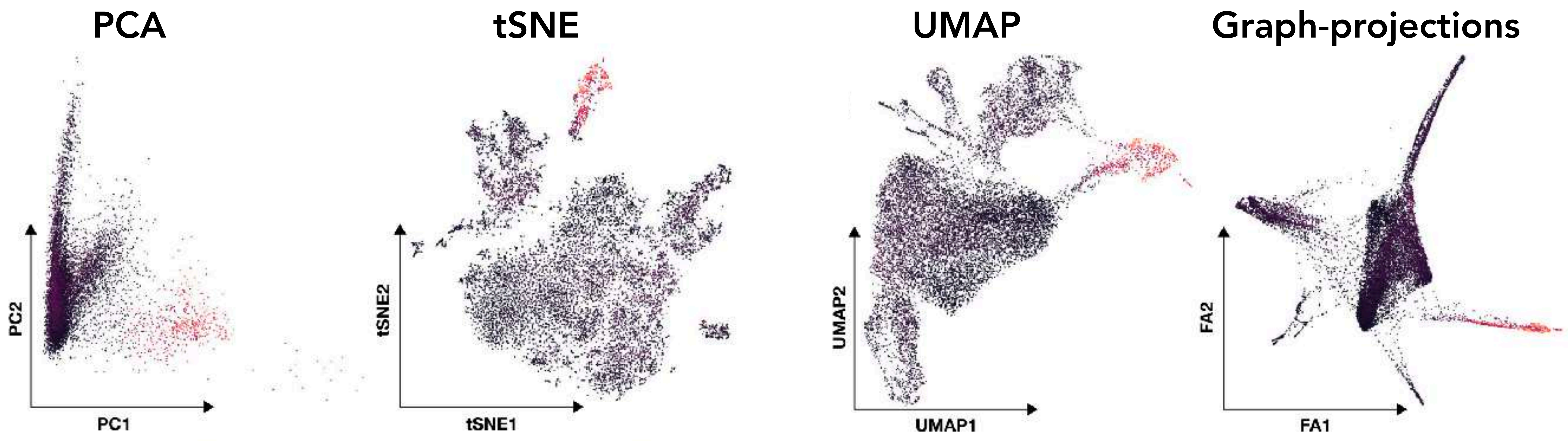


Not critical how many genes we select (usually 1,000s)

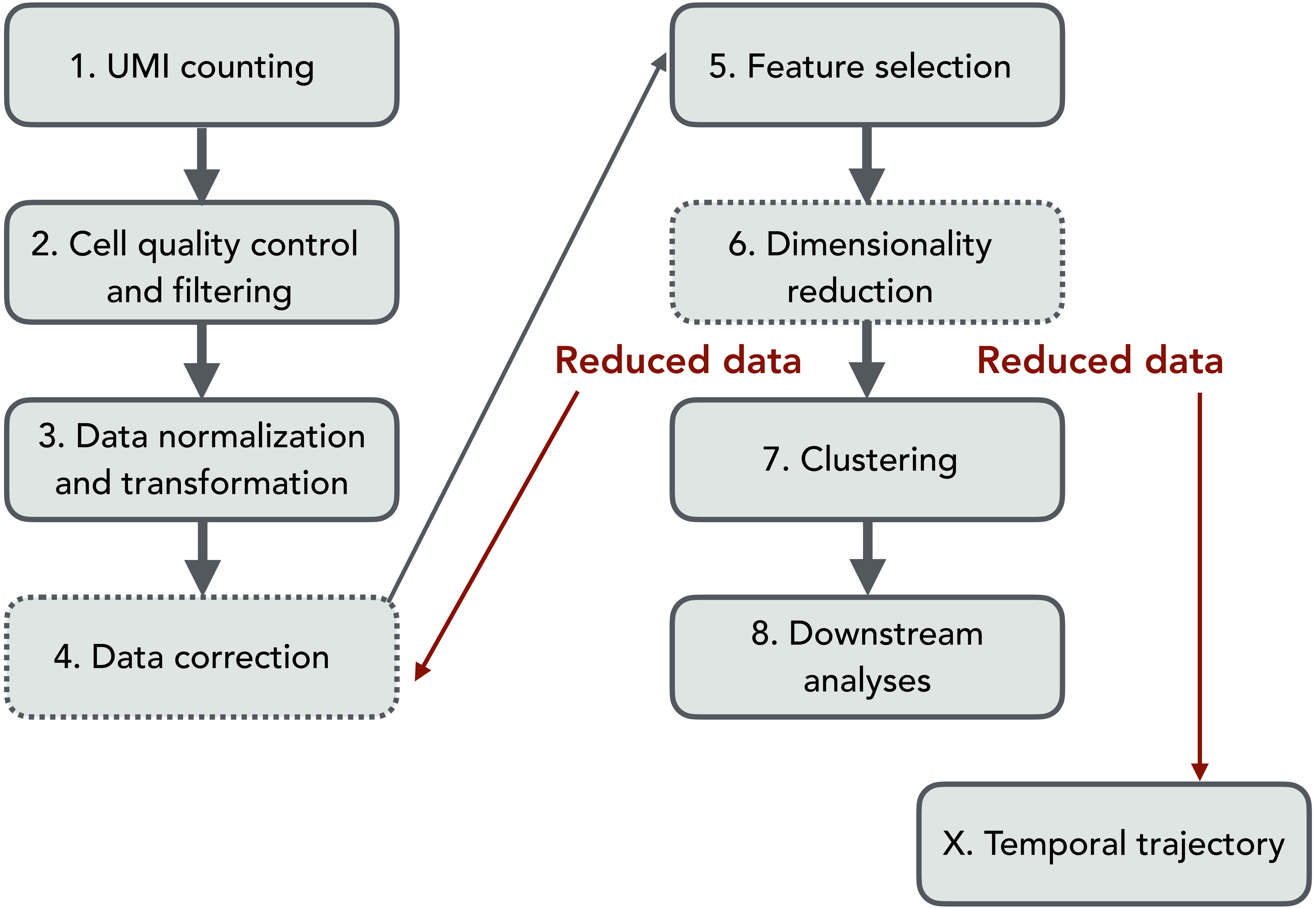
6. Dimensionality reduction

1. Summarization: reduces data to essential components for downstream analyses.
   E.g. PCA (clustering), Diffusion maps (trajectory).

2. Visualization: project dataset in two dimensions.



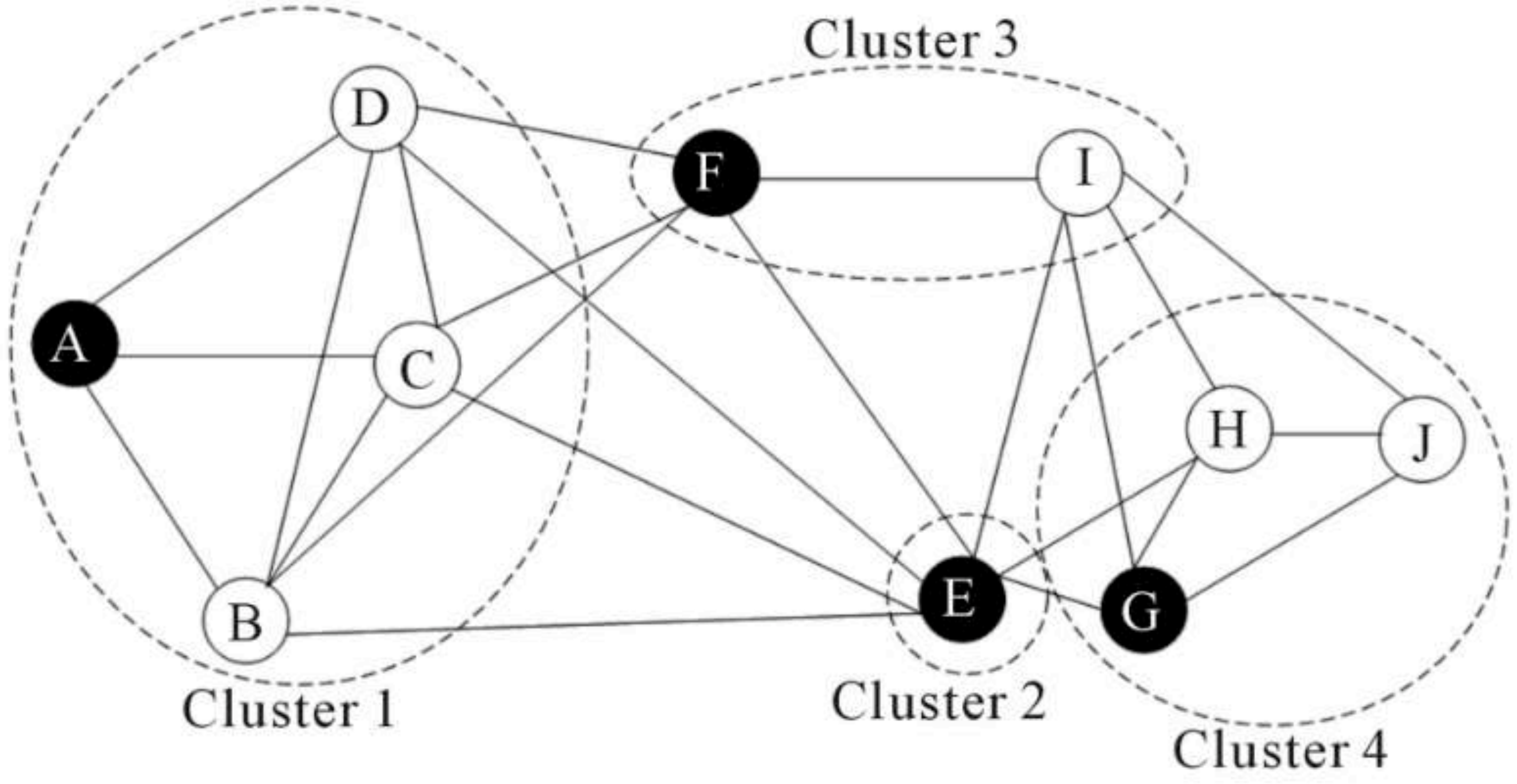
Standard scRNA-seq analysis pipeline



7. Cell clustering

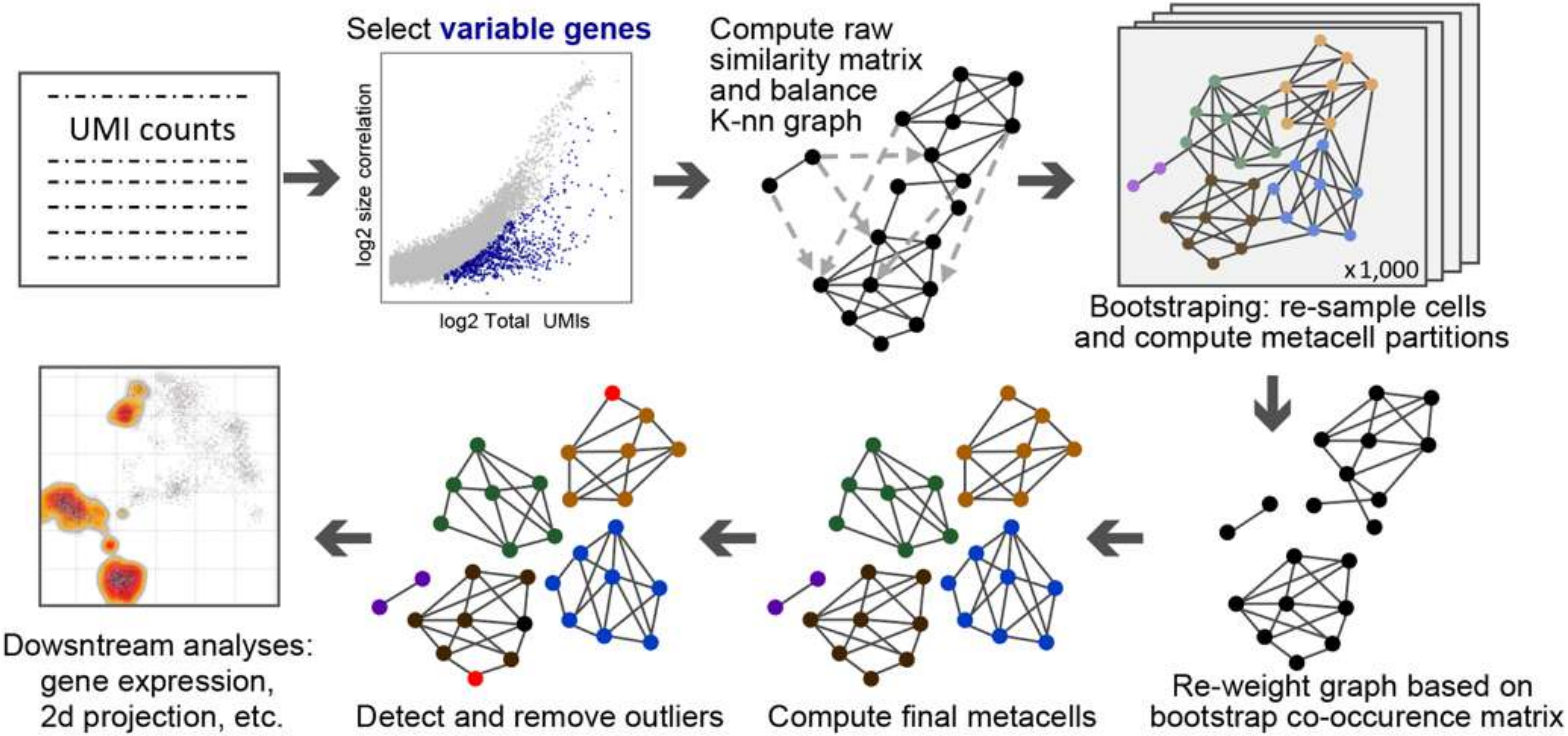
1. Cell-cell distance matrix. E.g. correlation-based, cosine similarity, Euclidean distance in PC-reduced space.

2. Cell clustering:
   i. Clustering algorithms. E.g. HC, k-means.
   ii. Graph-partitioning algorithm: k-NN graph construction followed by community detection (e.g. Louvain algorithm).



Cells (nodes) connected to *K* most similar cells.

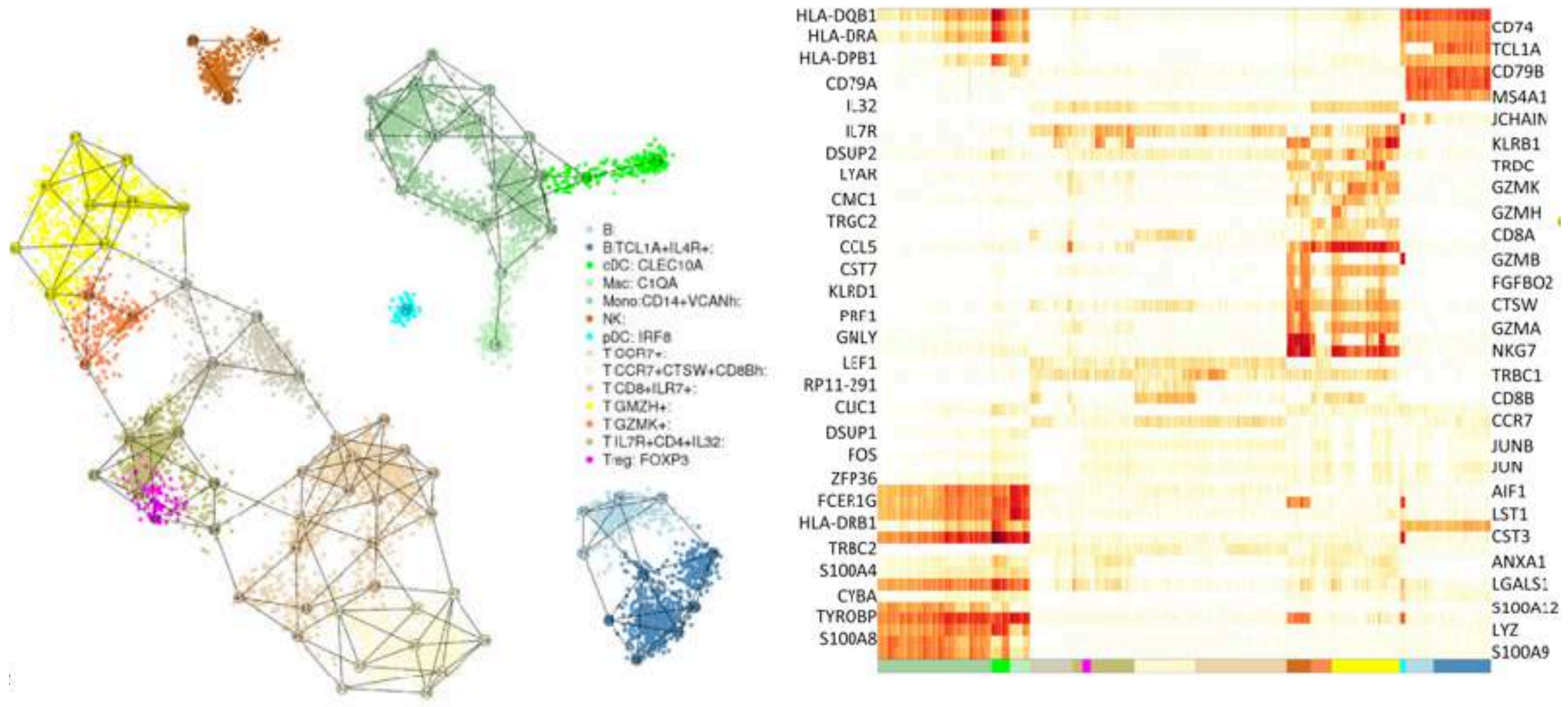
MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions



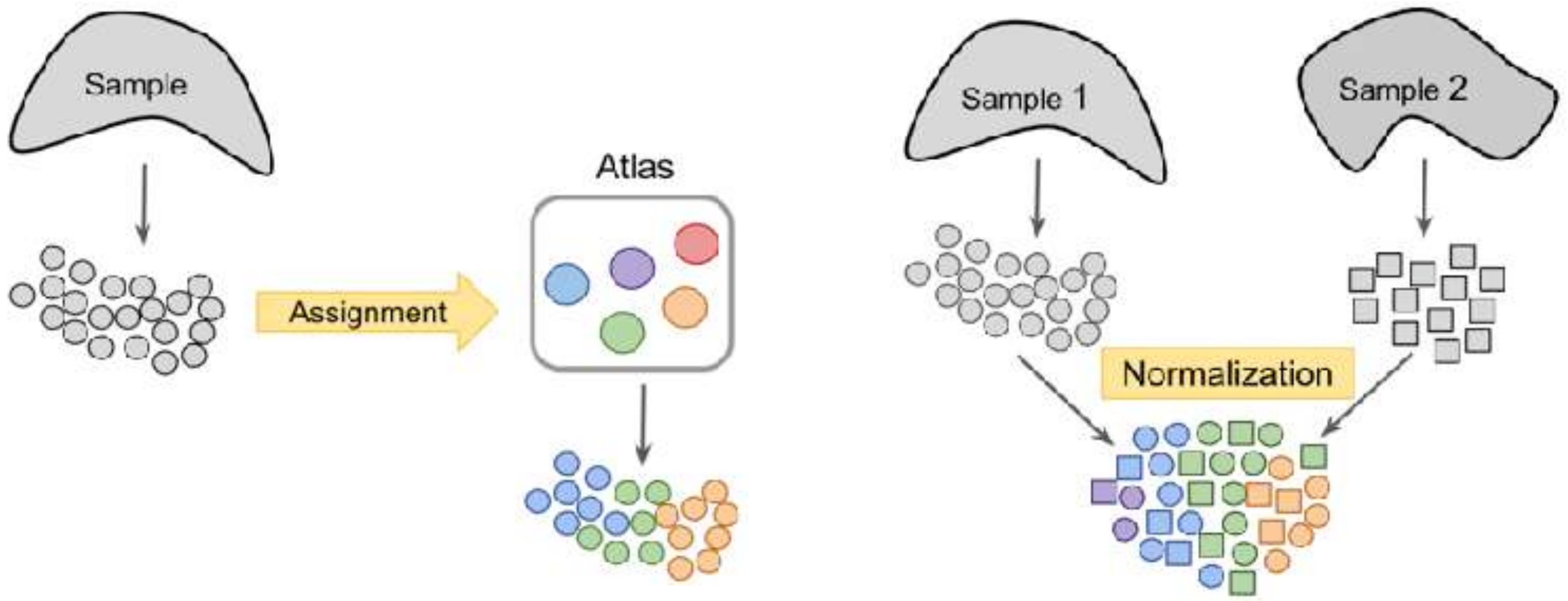
Baran, et al. *Genome Biology* 2019

8. Downstream data analysis: annotation, integration, gene module inference, etc.

Project expression of known *marker genes* and/or the top specific genes per each cluster.



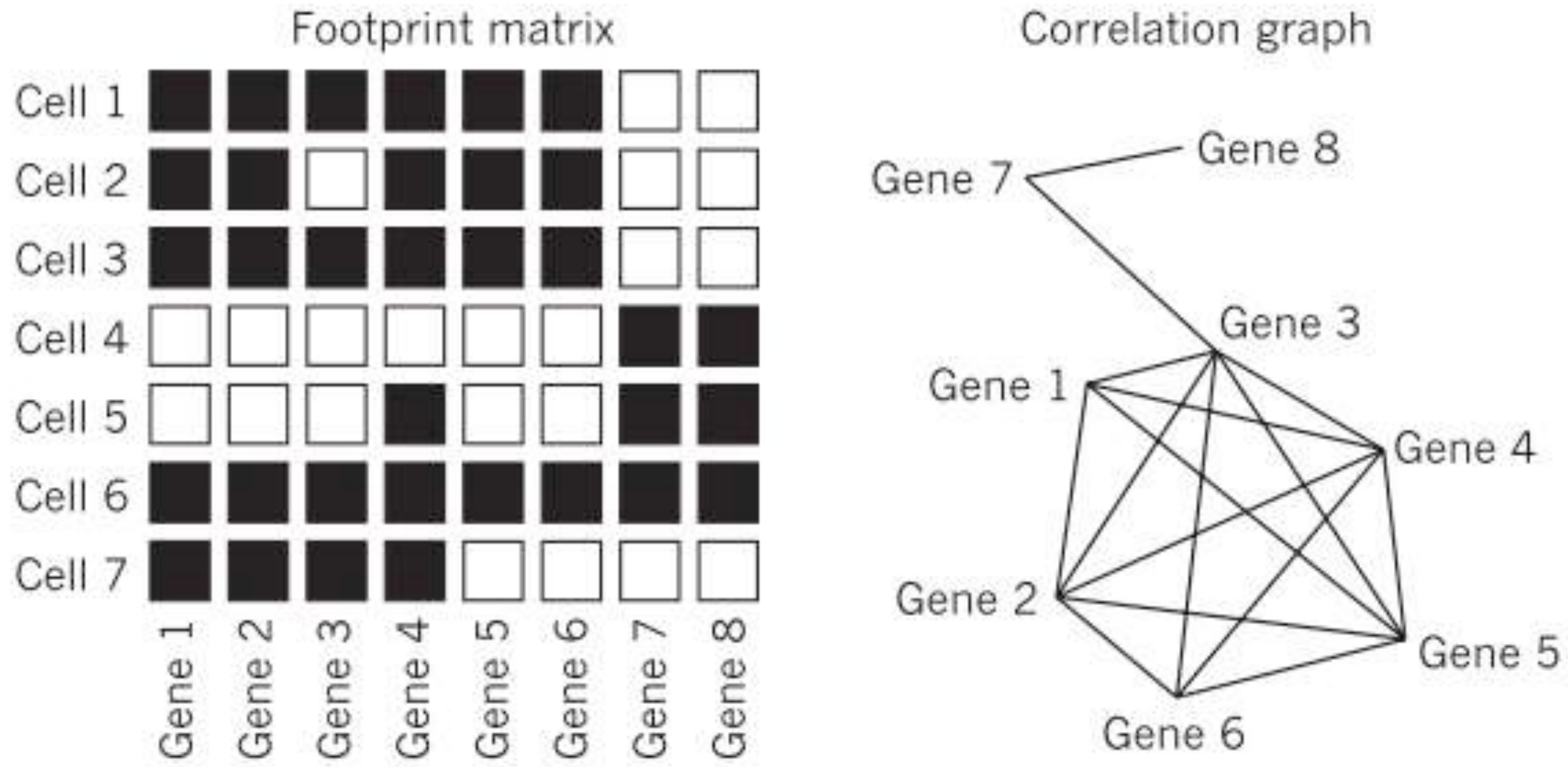
Comparative cell type annotation



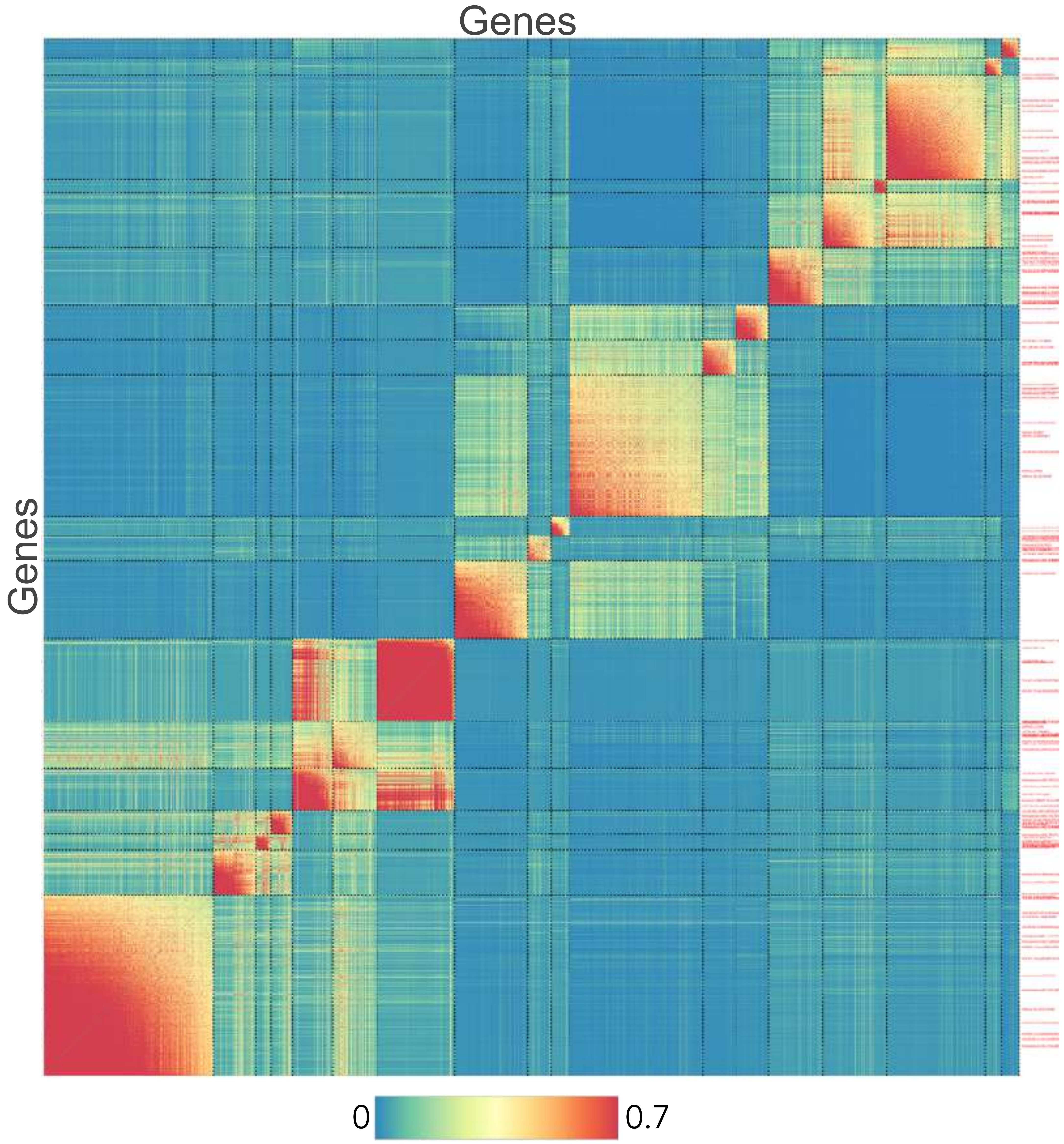
Projection to a reference atlas

Alternative: joint analysis

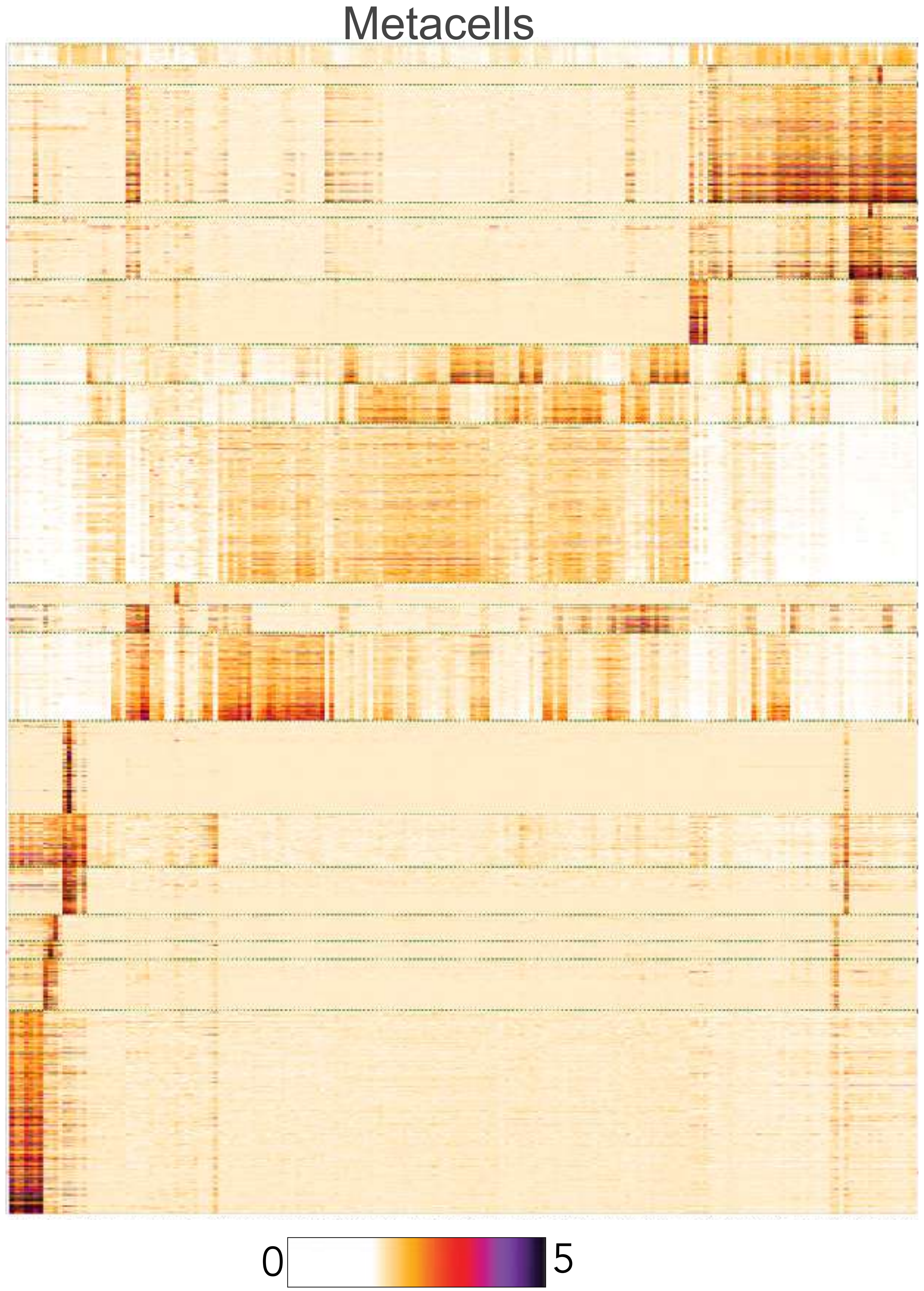
Gene-gene expression correlation to infer co-regulated gene modules



8. Downstream data analysis: gene modules



Gene-gene expression correlation (over metacells)



Normalized gene expression

8. Downstream data analysis: cross-species comparisons, overview of strategies

1. **Ortholog selection:**
- strict one-to-one orthologs
- homologs
- Protein Language Models

2. **Resolution**
- single cells
- cell clusters/cell types

3. **Comparison strategies:**
- gene expression correlation
- train classifiers
- sample integration
- DL universal cell embeddings

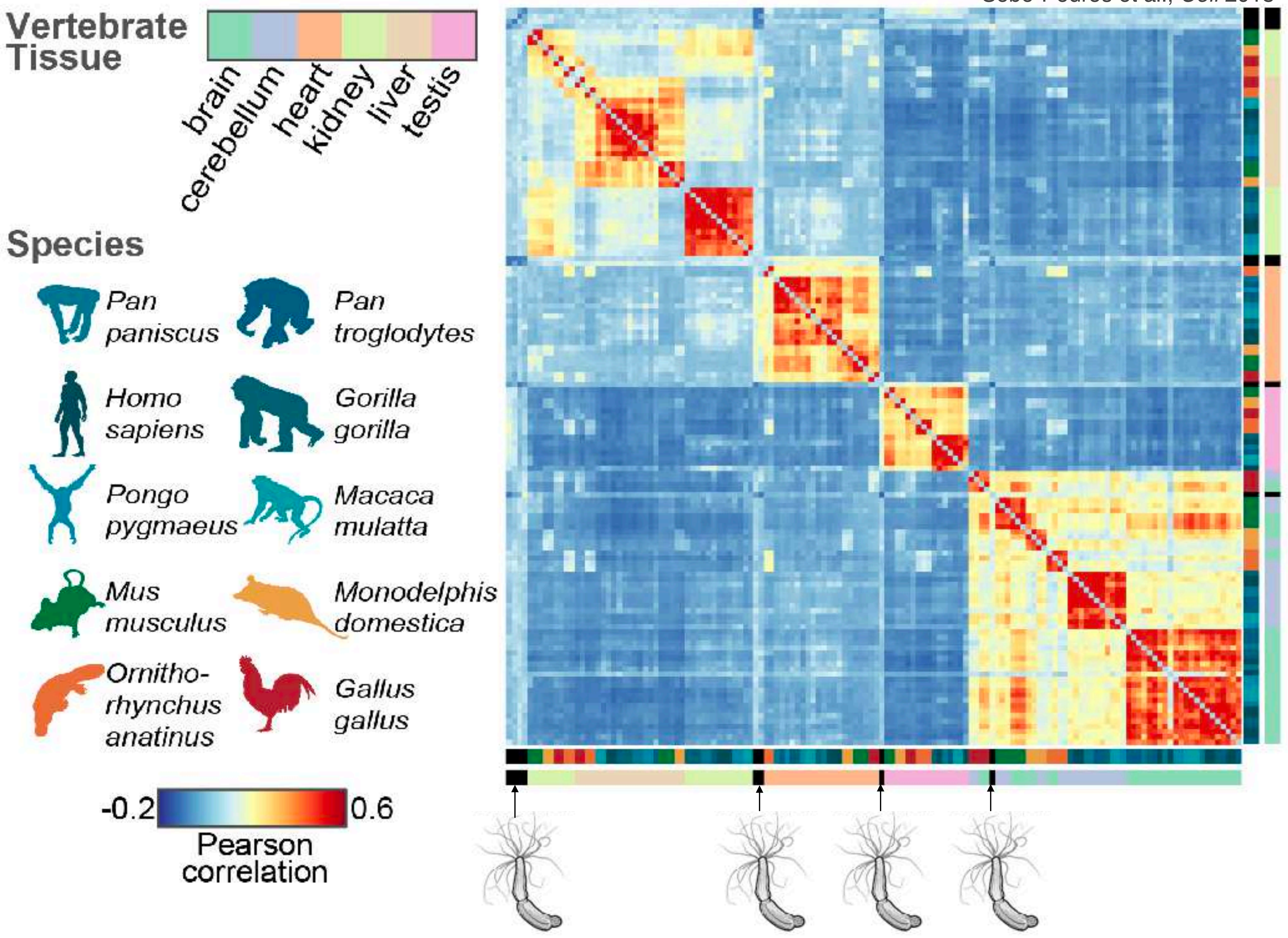
8. Downstream data analysis: cross-species comparisons, examples

1. Ortholog selection:
- strict one-to-one orthologs
- homologs
- Protein Language Models

2. Resolution
- single cells
- cell clusters/cell types

3. Comparison strategies:
- gene expression correlation
- train classifiers
- sample integration
- DL universal cell embeddings

Sebe-Pedros et al., *Cell* 2018



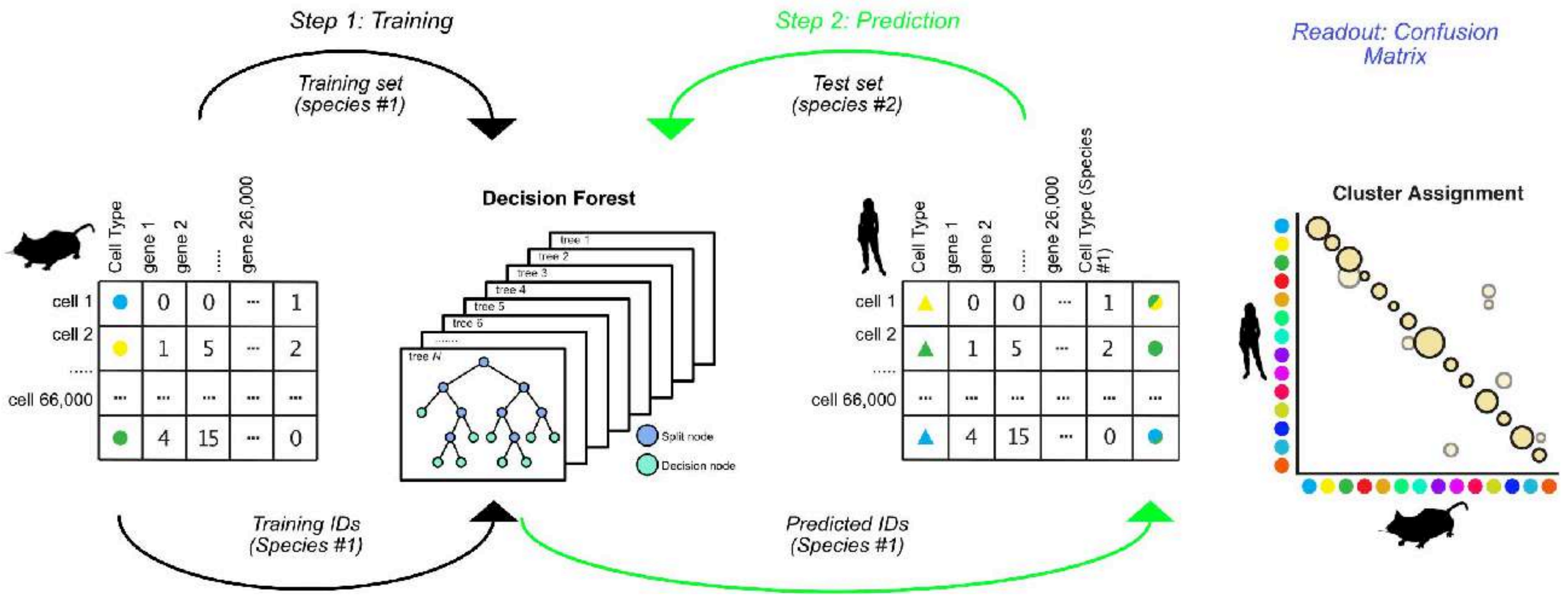
8. Downstream data analysis: cross-species comparisons, examples

1. Ortholog selection:
   - strict one-to-one orthologs
   - homologs
   - Protein Language Models

2. Resolution
   - single cells
   - cell clusters/cell types

3. Comparison strategies:
   - gene expression correlation
   - train classifiers
   - sample integration
   - DL universal cell embeddings

Random forest classifiers trained in one species and applied to another



Shafer *Front. Cell Dev. Biol.* 2019

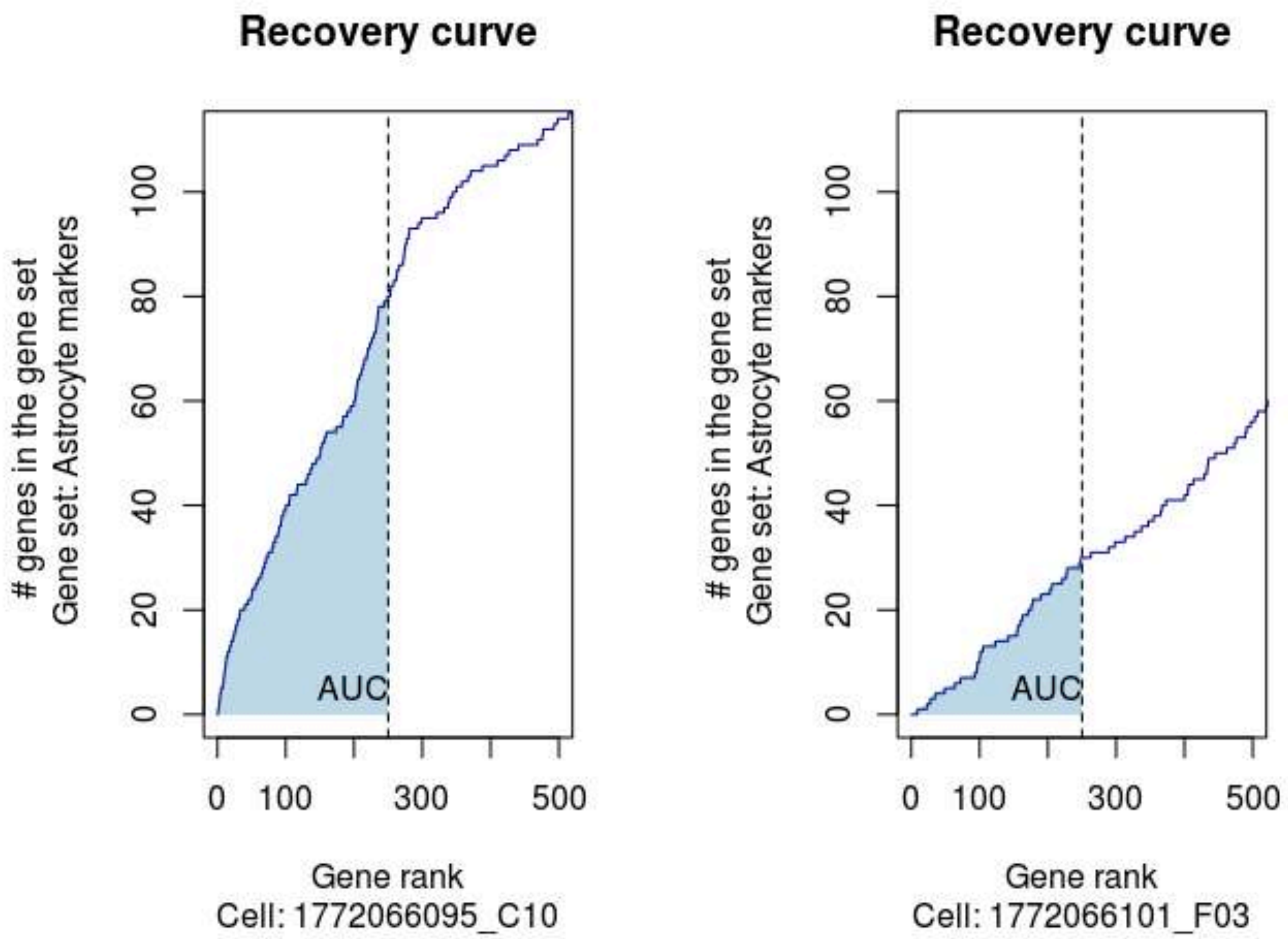
8. Downstream data analysis: cross-species comparisons, examples

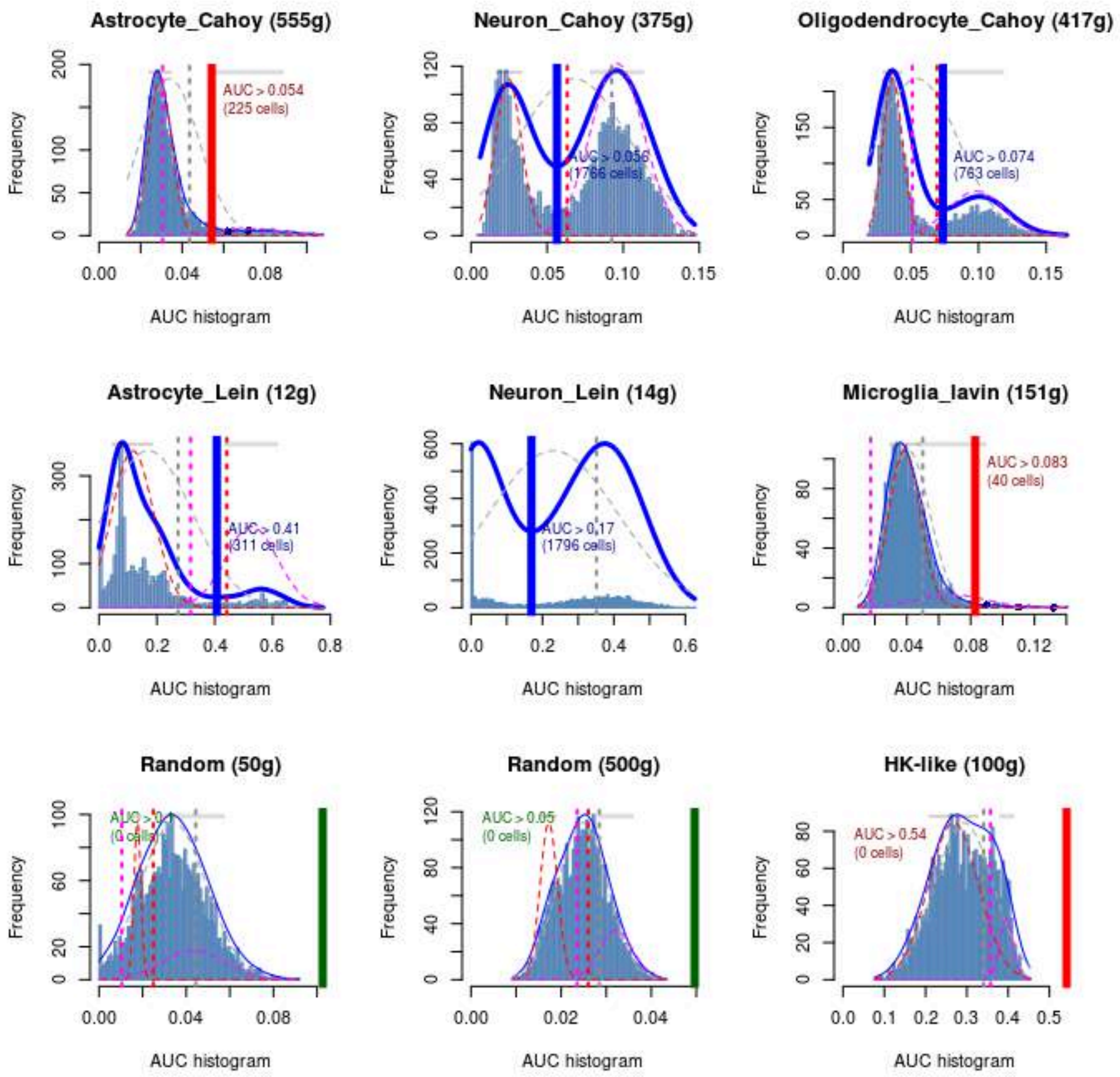
1. Ortholog selection:
- strict one-to-one orthologs
- homologs
- Protein Language Models

2. Resolution
- single cells
- cell clusters/cell types

3. Comparison strategies:
- gene expression correlation
- train classifiers
- sample integration
- DL universal cell embeddings

AUCell: Area Under the Curve for Gene Sets





https://www.bioconductor.org/packages/release/bioc/html/AUCell.html

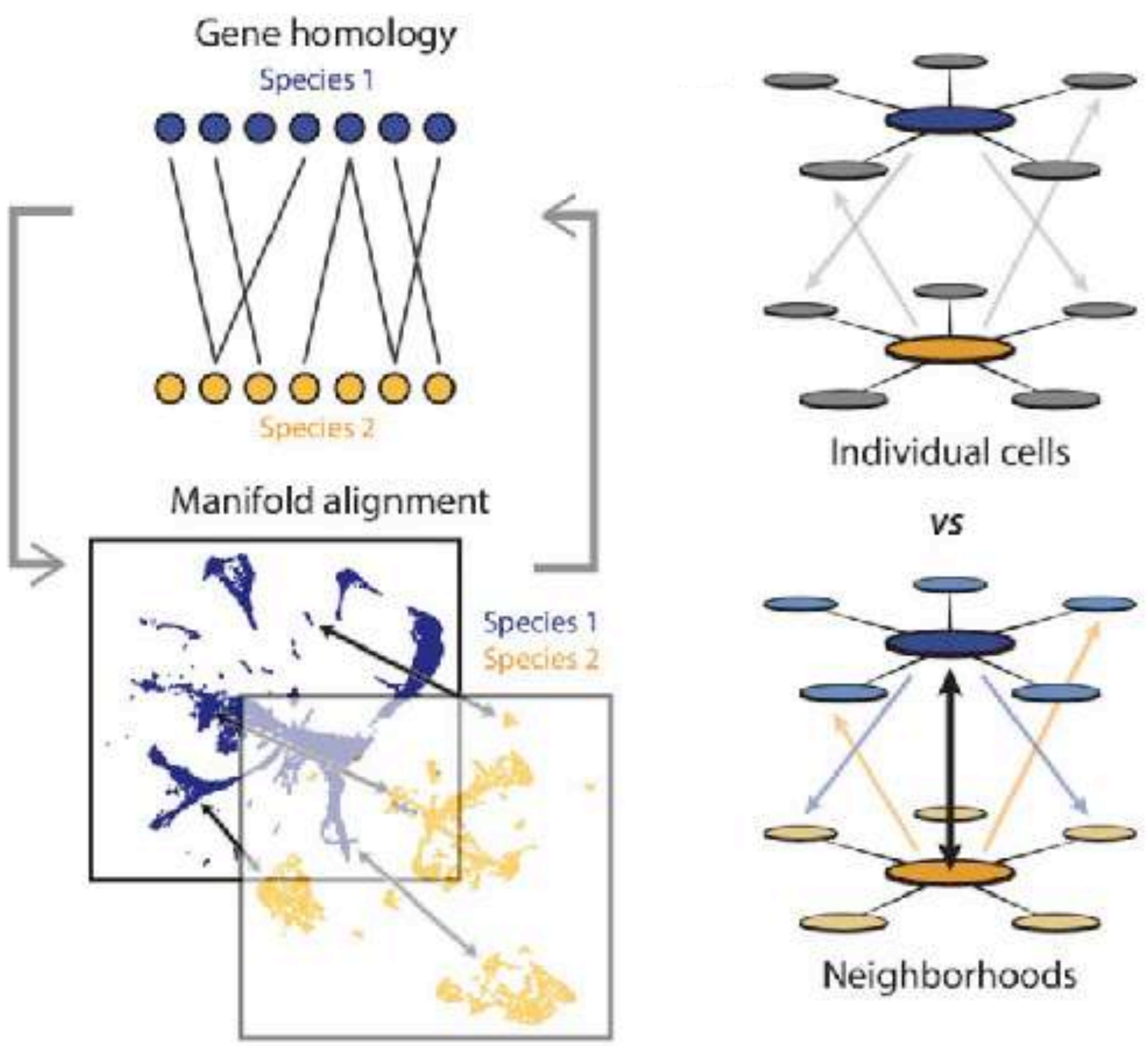
8. Downstream data analysis: cross-species comparisons, examples

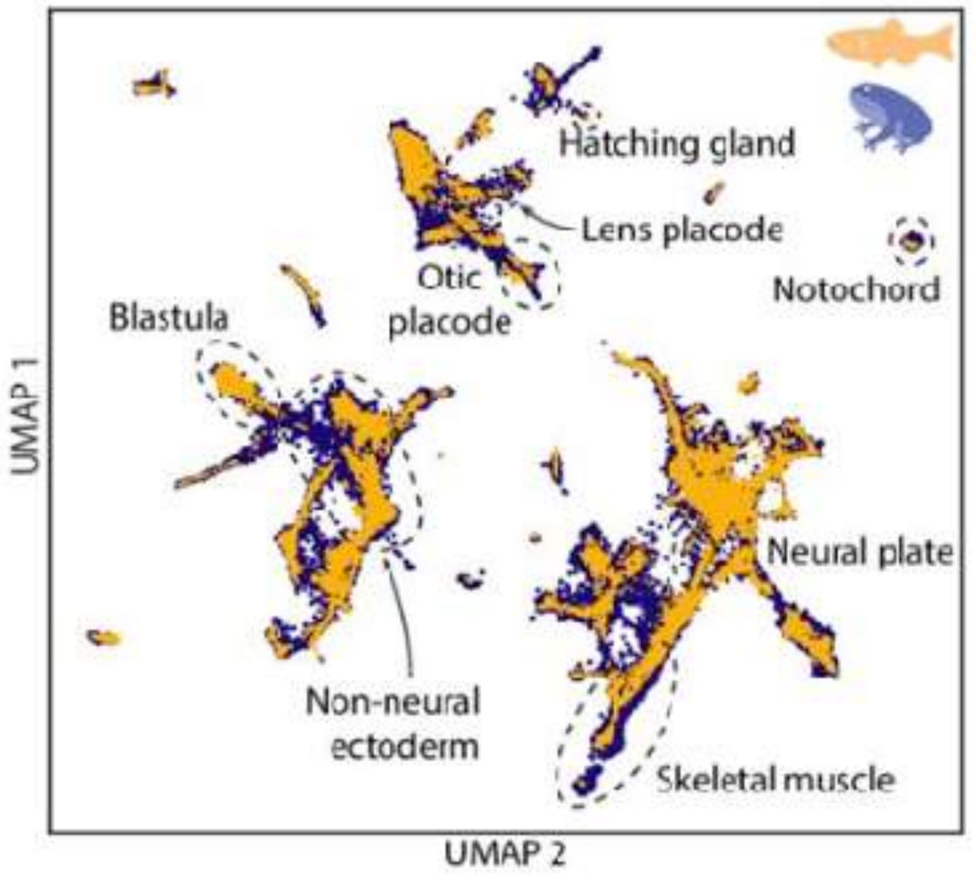
1. **Ortholog selection:**
- strict one-to-one orthologs
- homologs (many-to-many)
- Protein Language Models

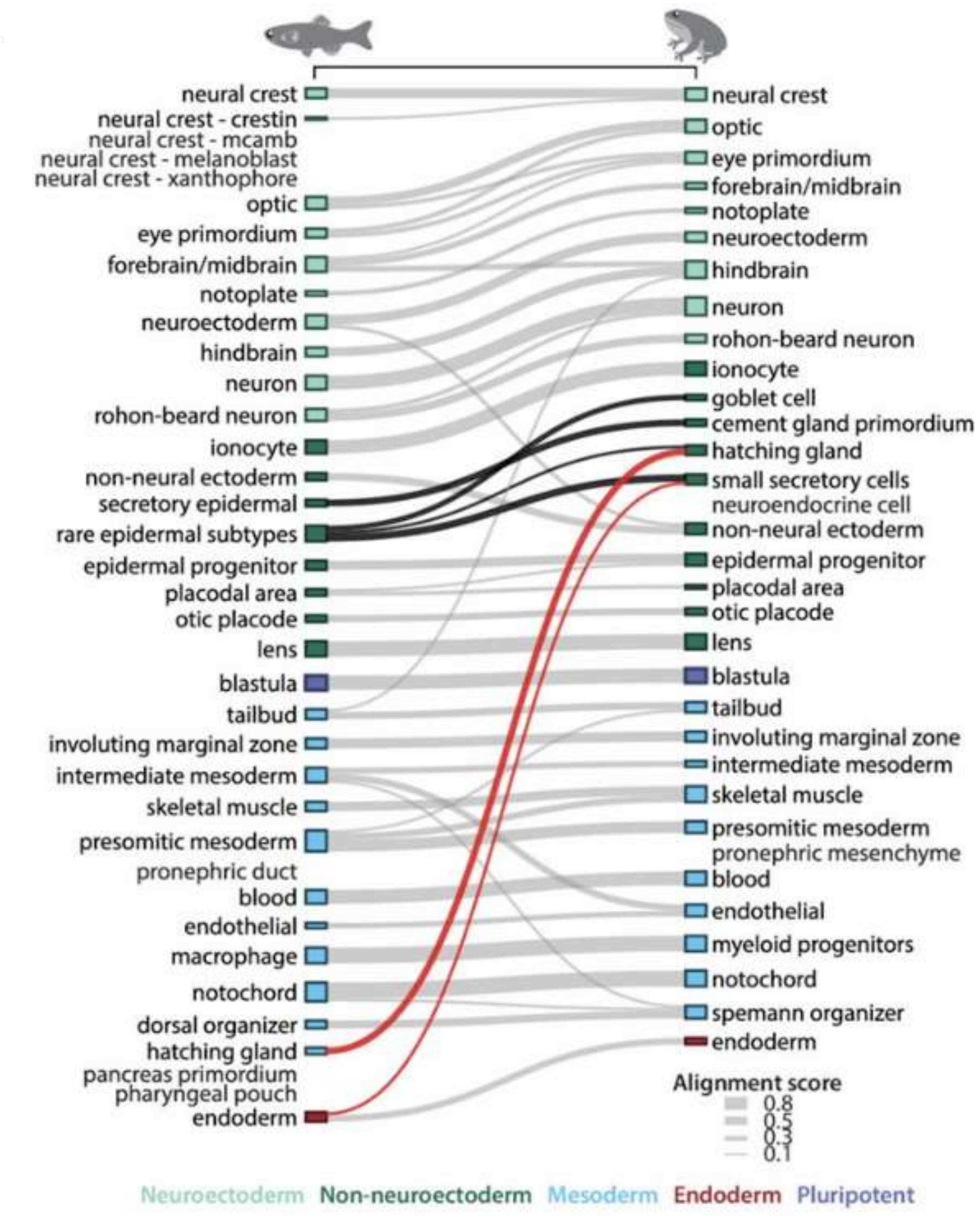
2. **Resolution**
- single cells
- cell clusters/cell types

3. **Comparison strategies:**
- gene expression correlation
- train classifiers
- sample integration
- DL universal cell embeddings

SAMap: cross-species self-assembling manifolds







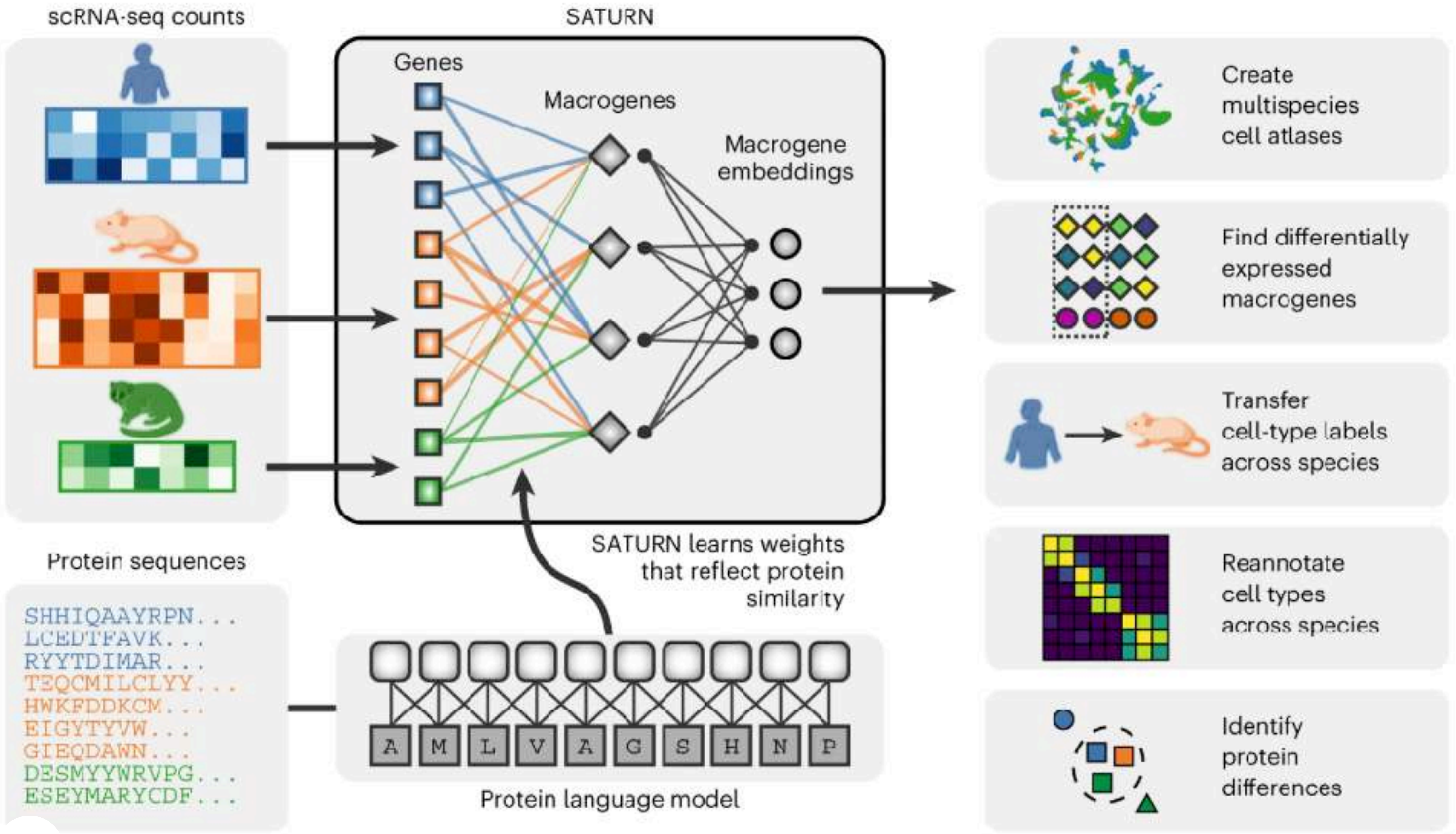
Tarashansky et al. *eLife* 2021

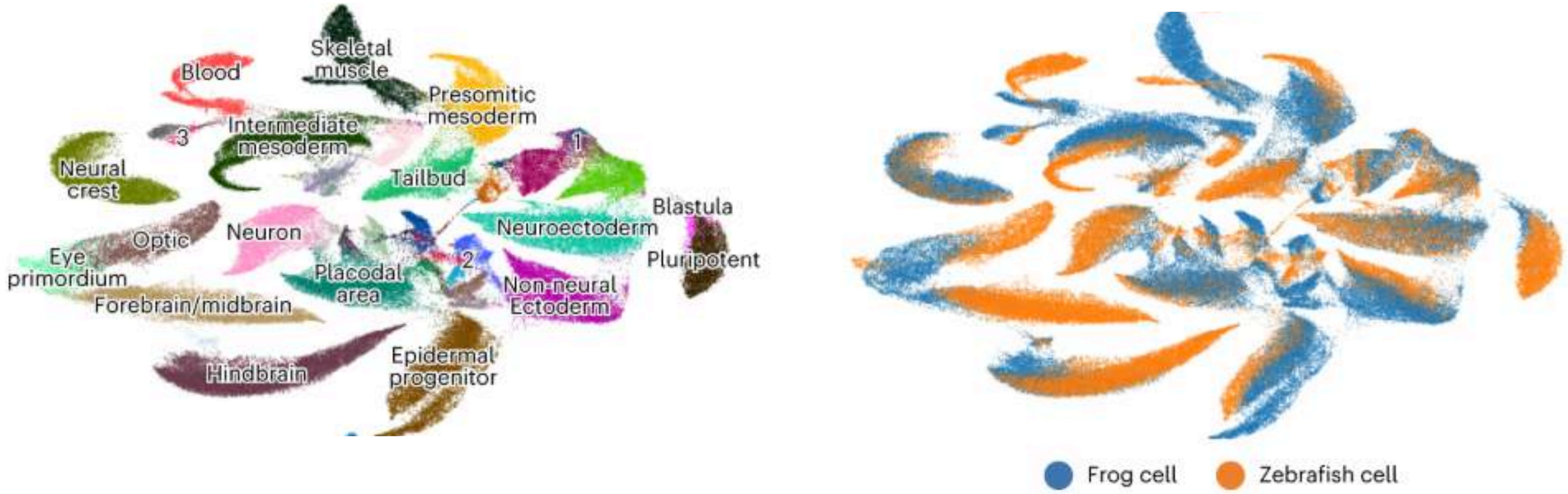
8. Downstream data analysis: cross-species comparisons, examples

1. Ortholog selection:
- strict one-to-one orthologs
- homologs
- Protein Language Models

2. Resolution
- single cells
- cell clusters/cell types

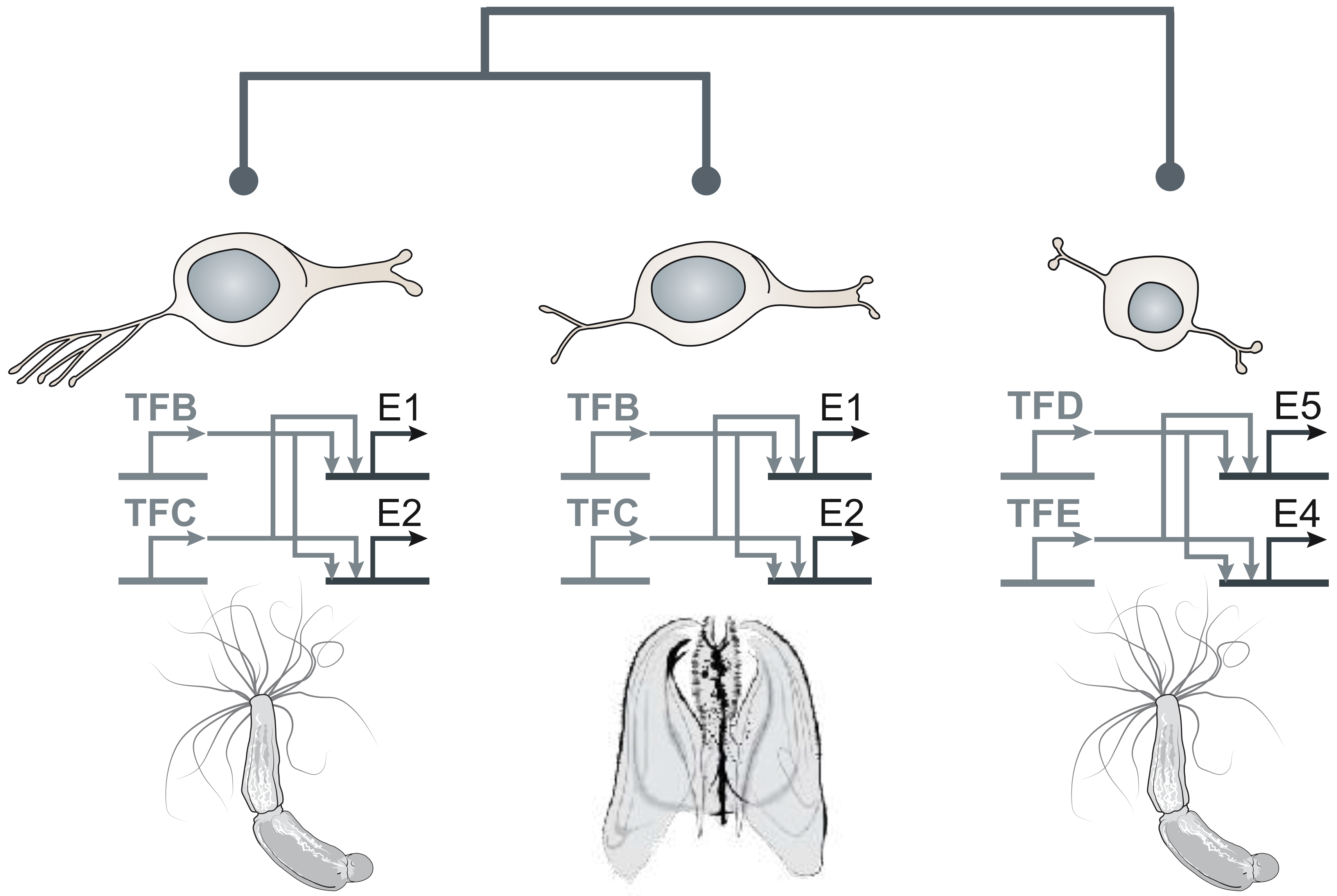
3. Comparison strategies:
- gene expression correlation
- train classifiers
- sample integration
- DL universal cell embeddings





Rosen et al. *Nature Methods* 2024

Early animal cell type
diversity, evolution and regulation



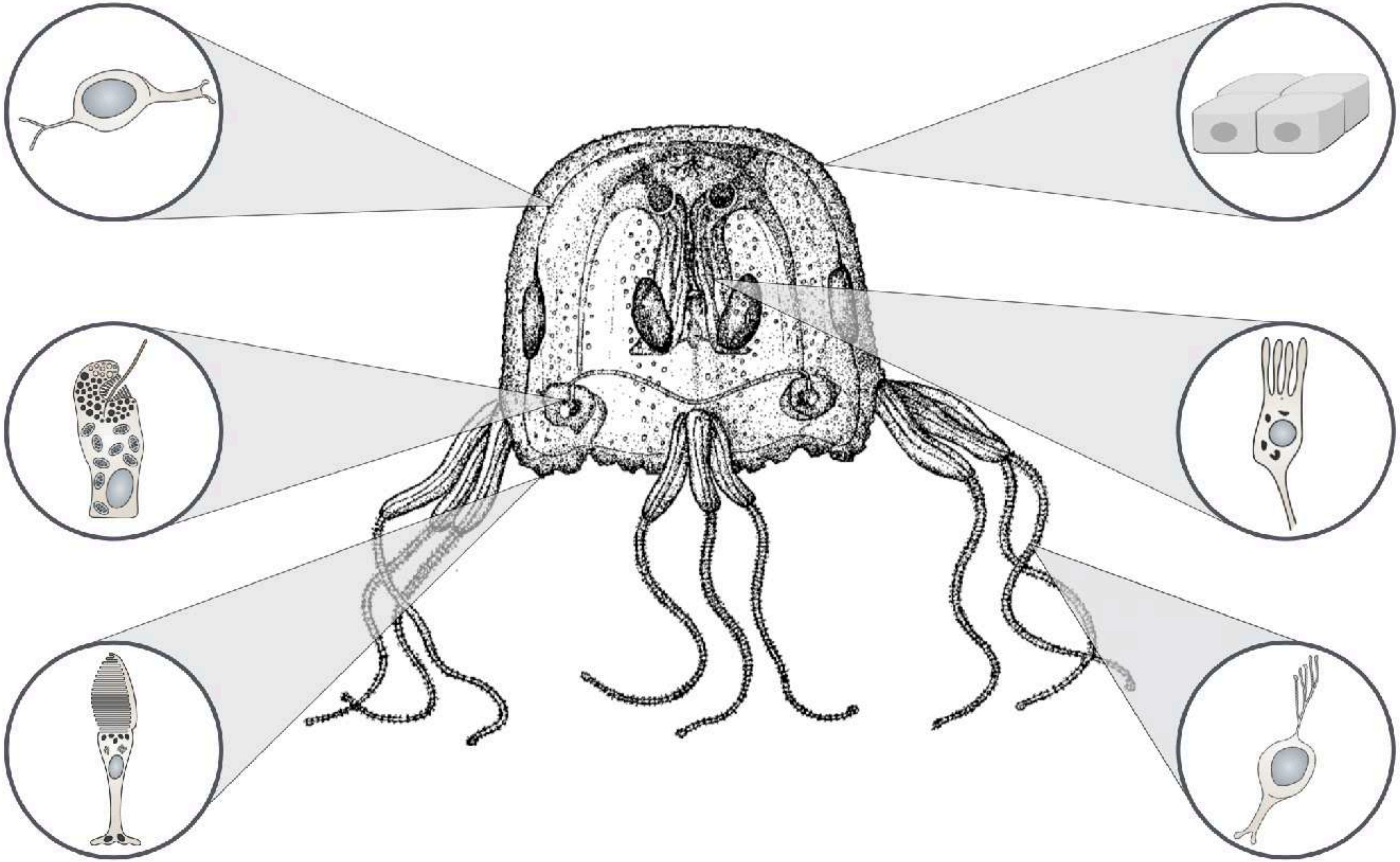
Arnau Sebé-Pedrós

CRG Centre for Genomic Regulation

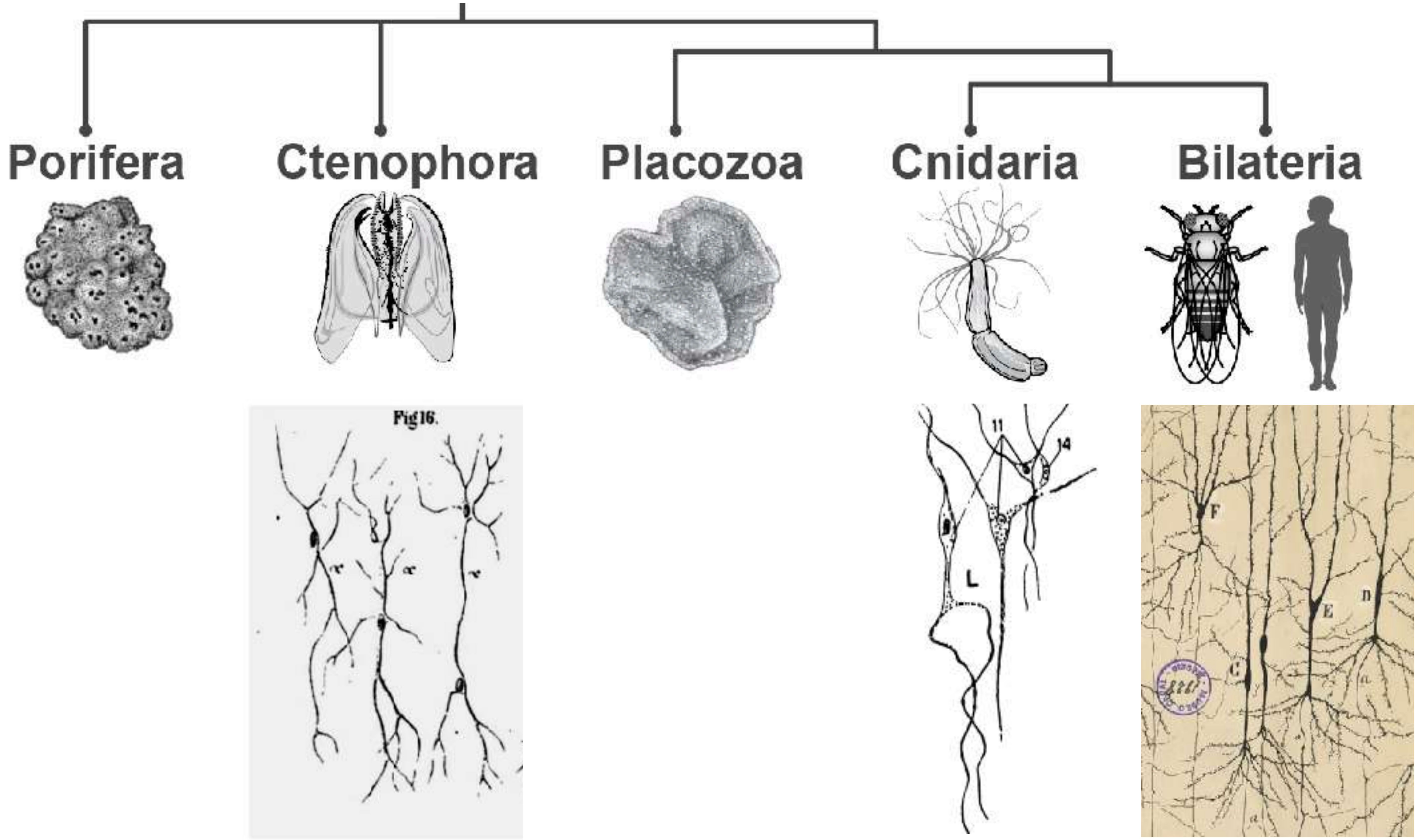
wellcome sanger institute

www.sebepedroslab.org

Cell types are the **functional and evolutionary units** of animal multicellularity



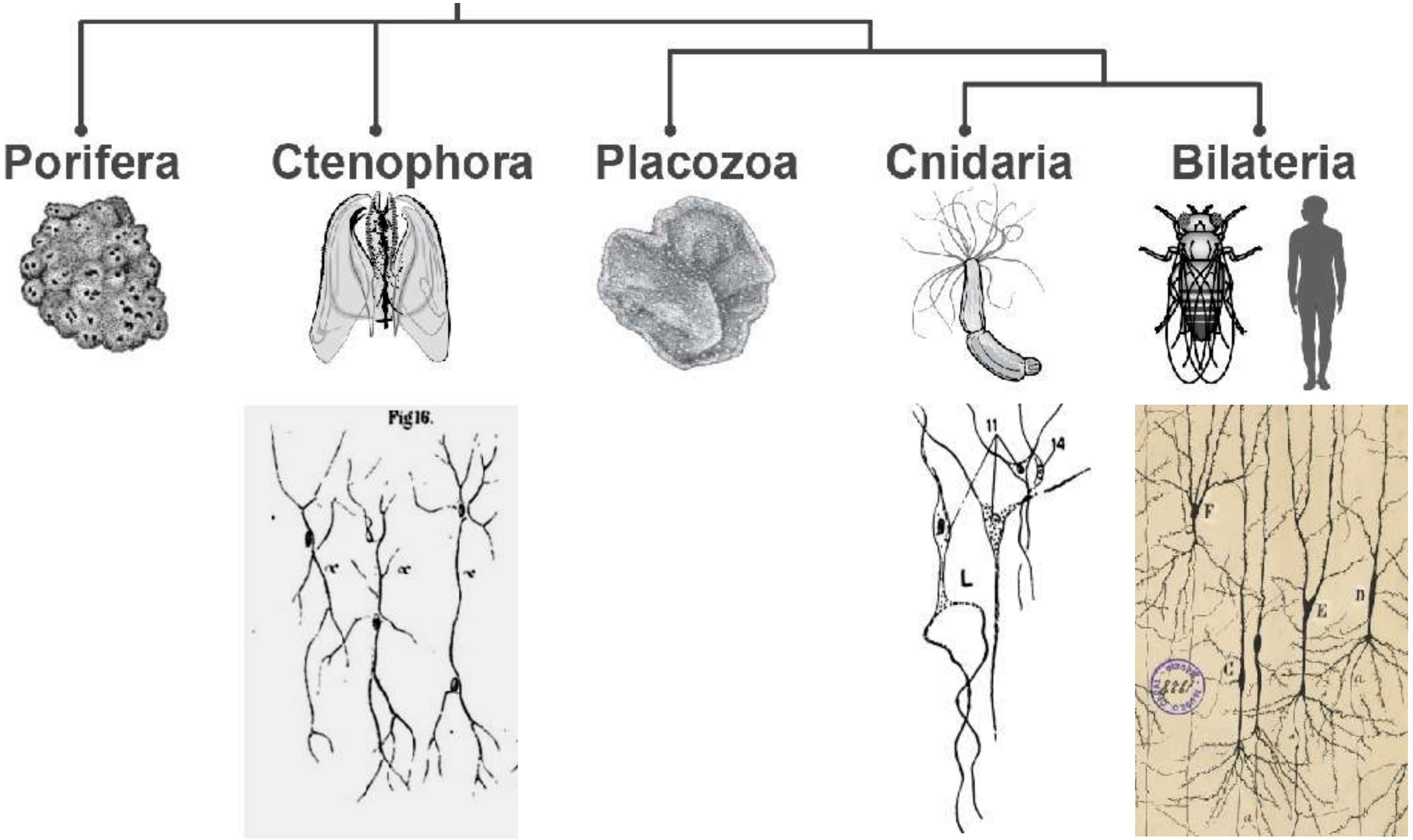
Cell types are the **functional and evolutionary units** of animal multicellularity



R. Hertwig, 1880 H. Hyman, 1949 Ramon y Cajal, 1890

Morphological similarities across animal phyla suggest **conserved cell types**

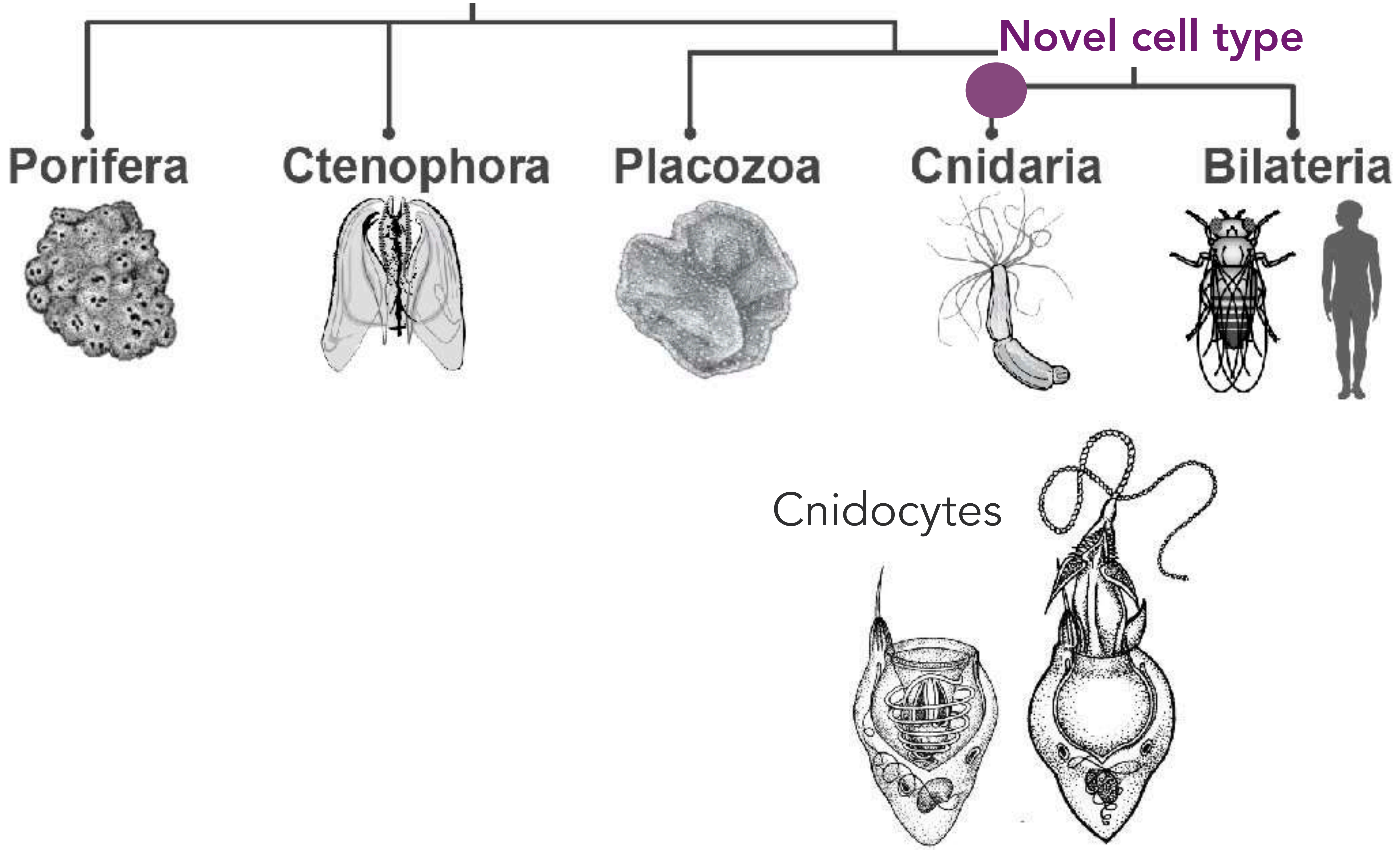
Cell types are the **functional and evolutionary units** of animal multicellularity



R. Hertwig, 1880 H. Hyman, 1949 Ramon y Cajal, 1890

A major question is **when cell types originated**

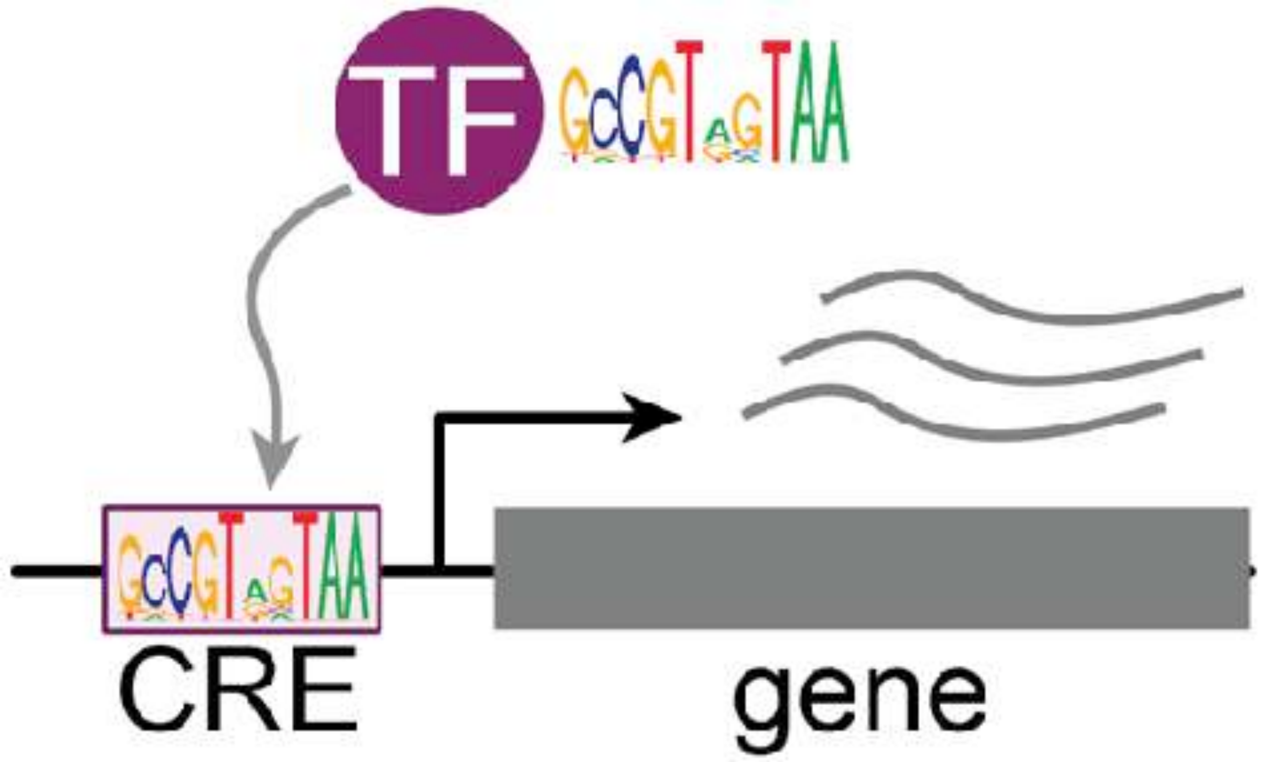
Cell types are the **functional and evolutionary units** of animal multicellularity

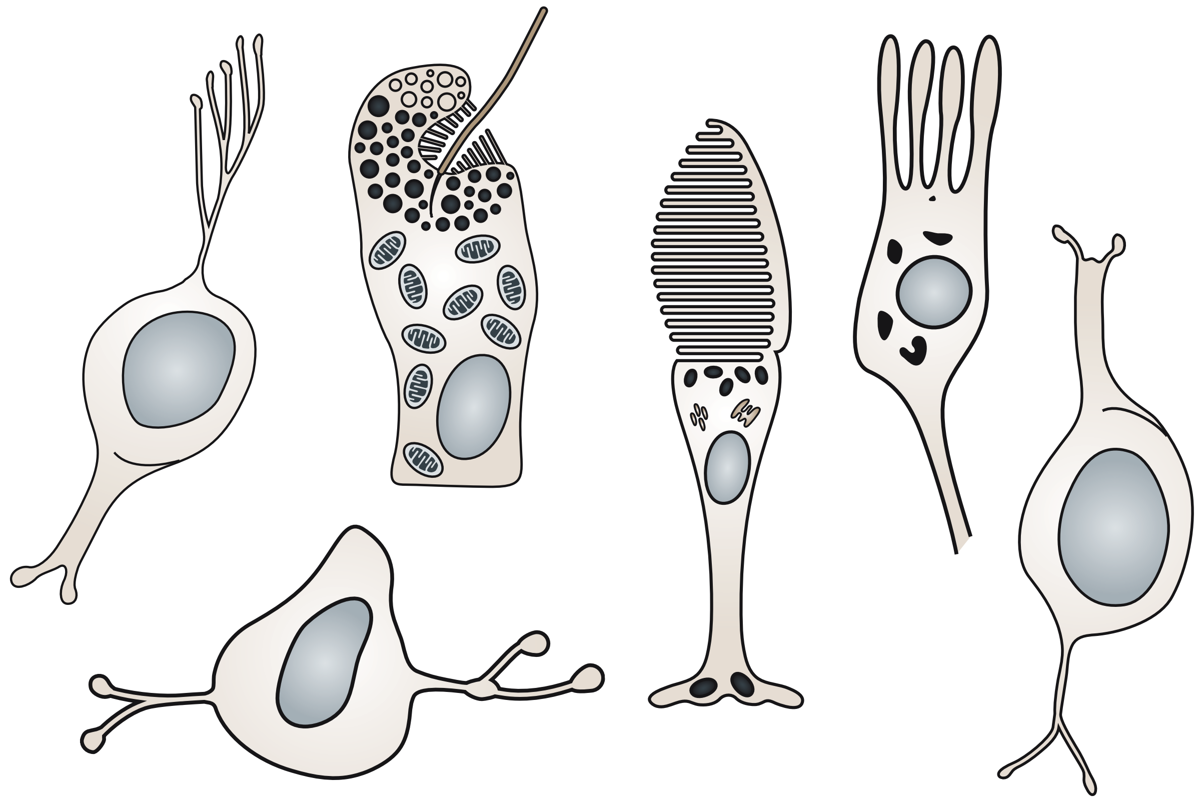


A major question is when cell types originated **and how novel cell types evolve**

Cell types are genetically defined by specific **regulatory programs**

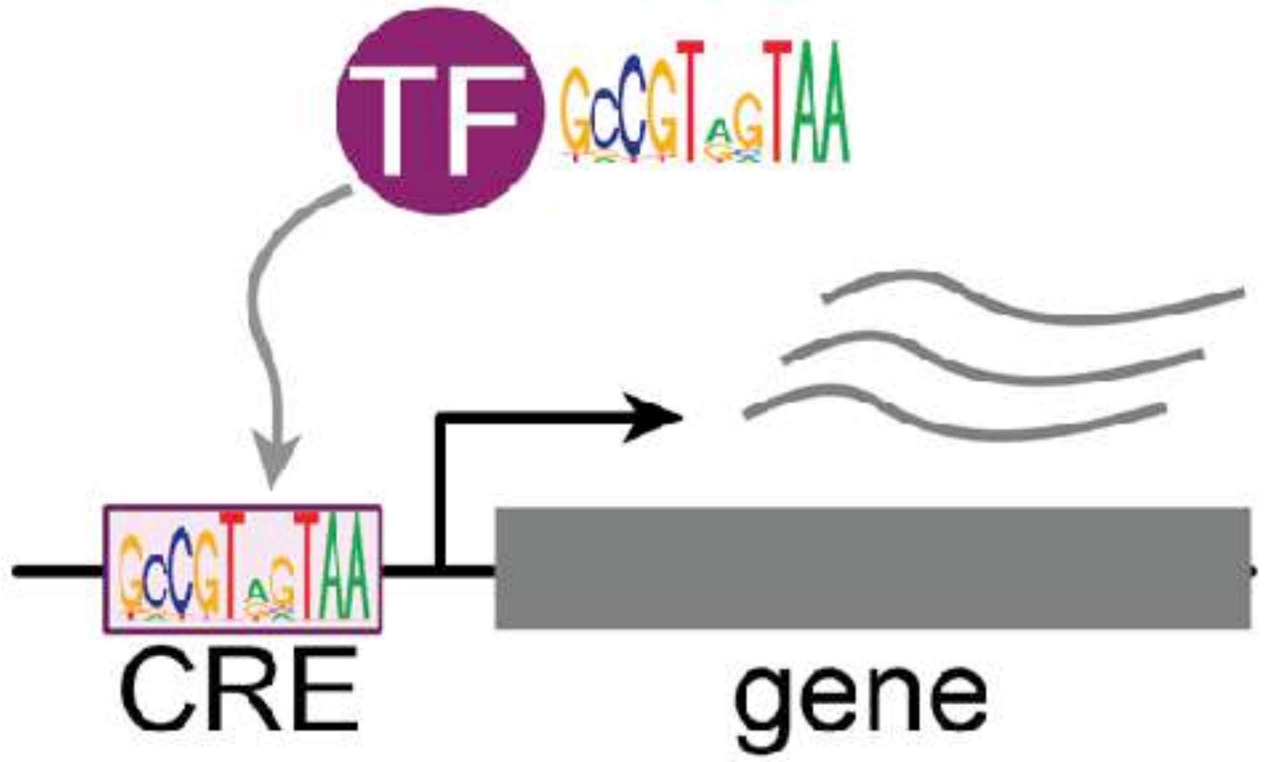
Genome sequence → Genome regulation → Cell types

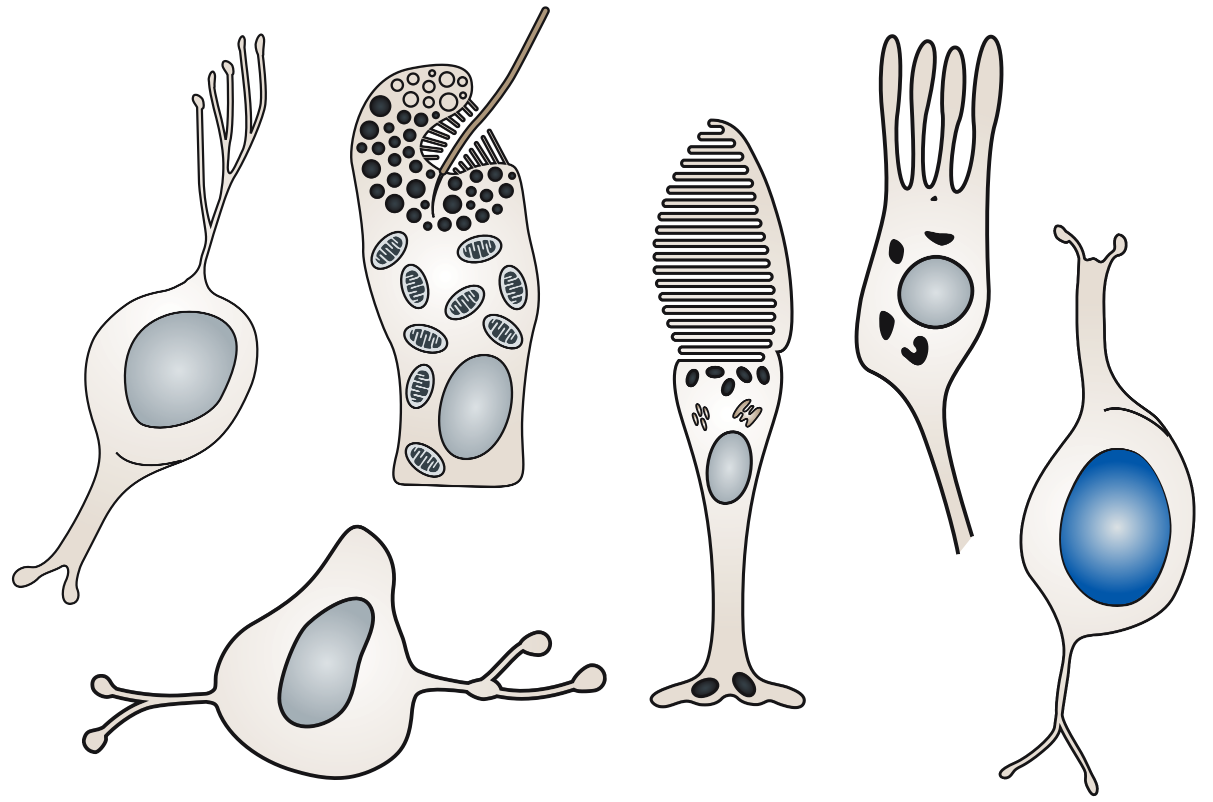




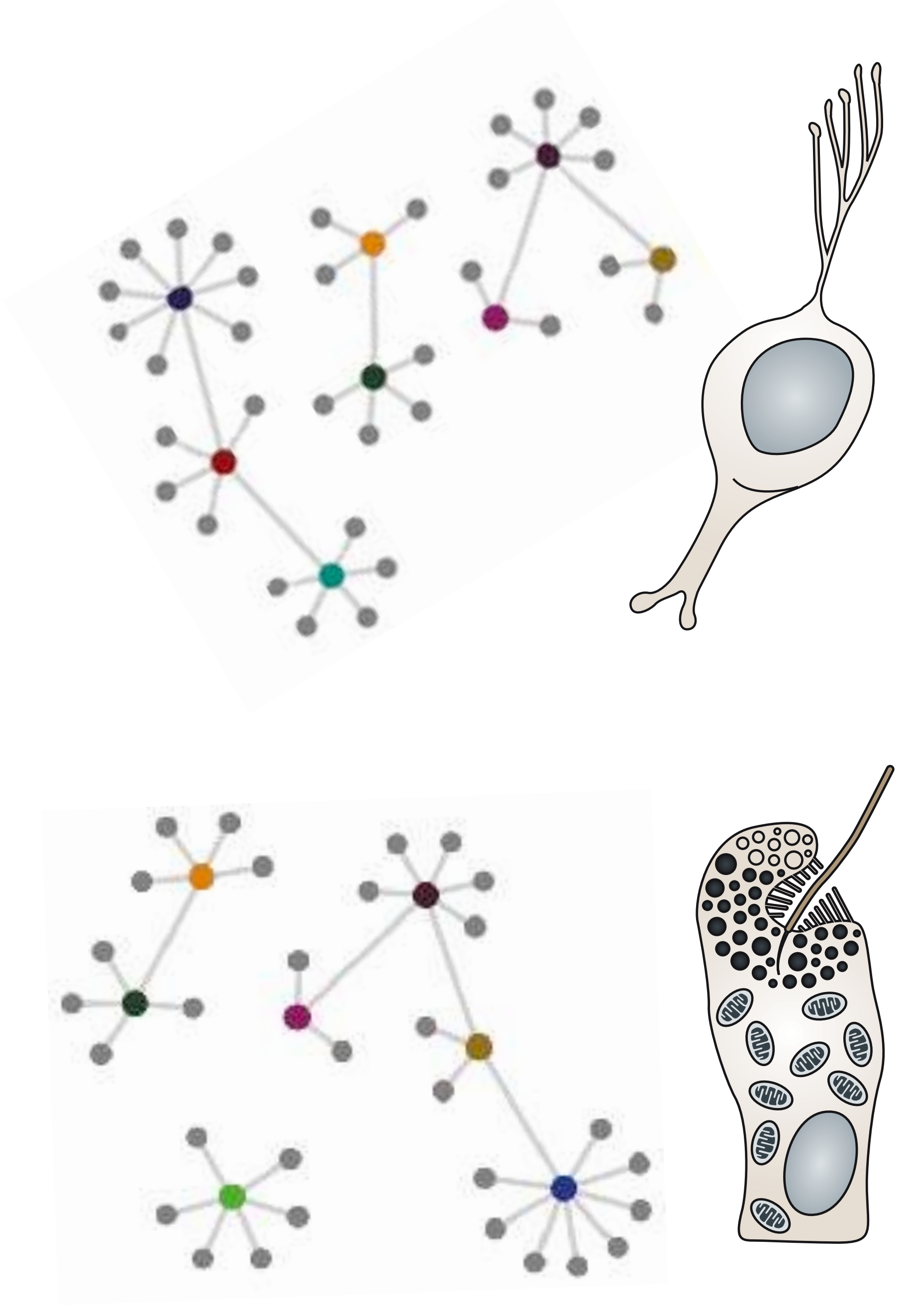
Cell types are genetically defined by specific **regulatory programs**

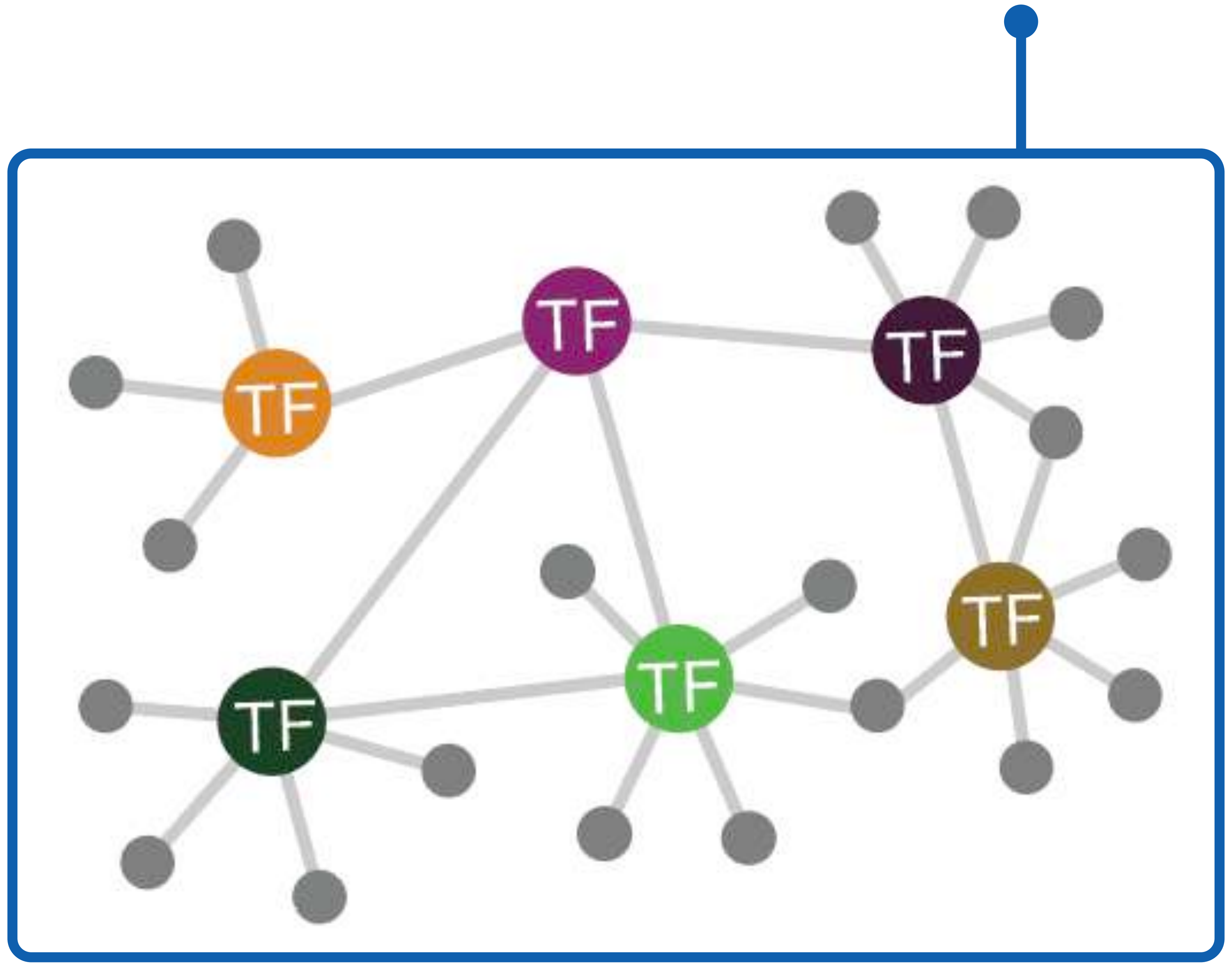
Genome sequence → Genome regulation → Cell types





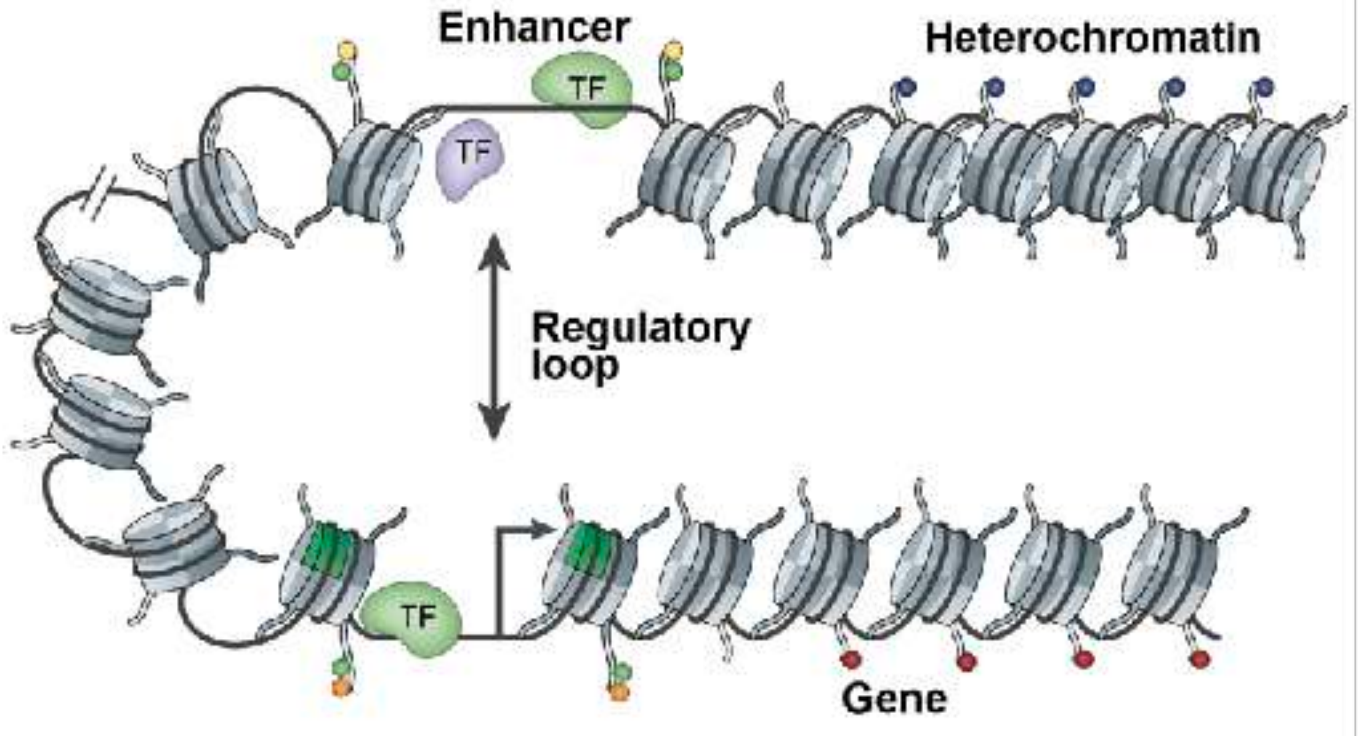
Study cell type evolution by defining and comparing cell identity programs

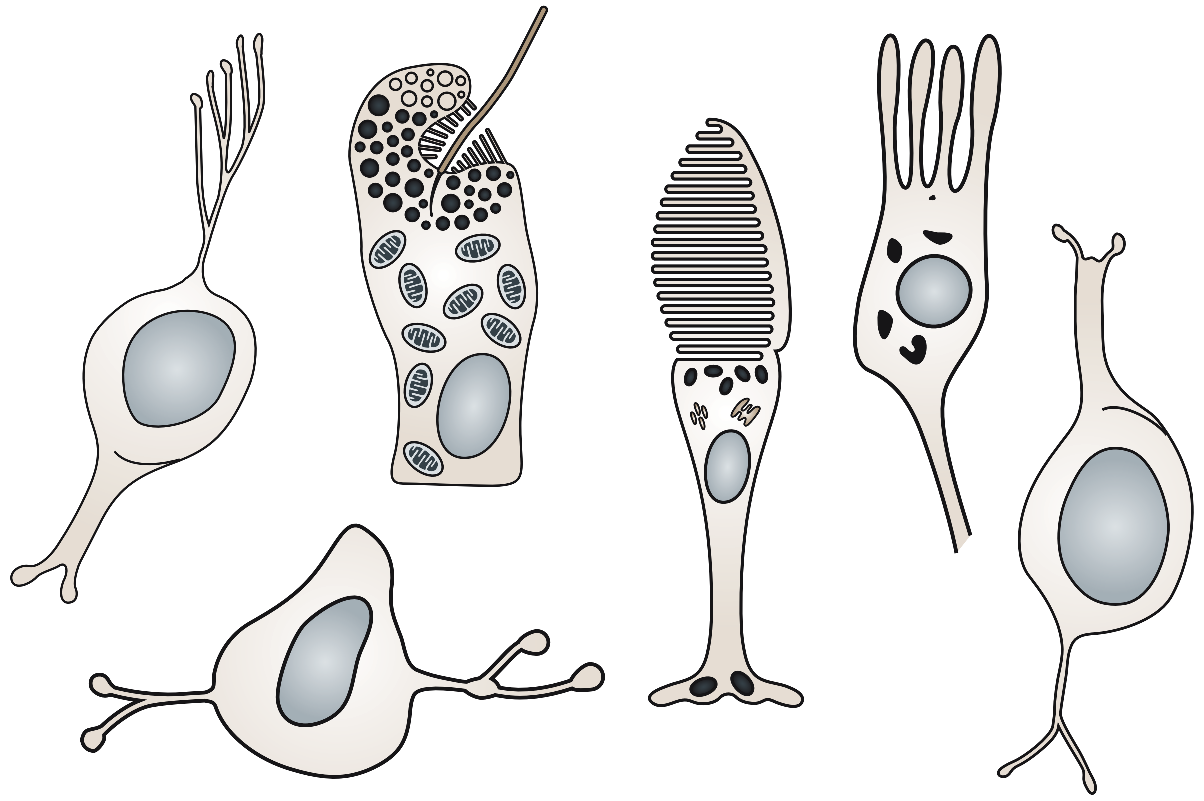


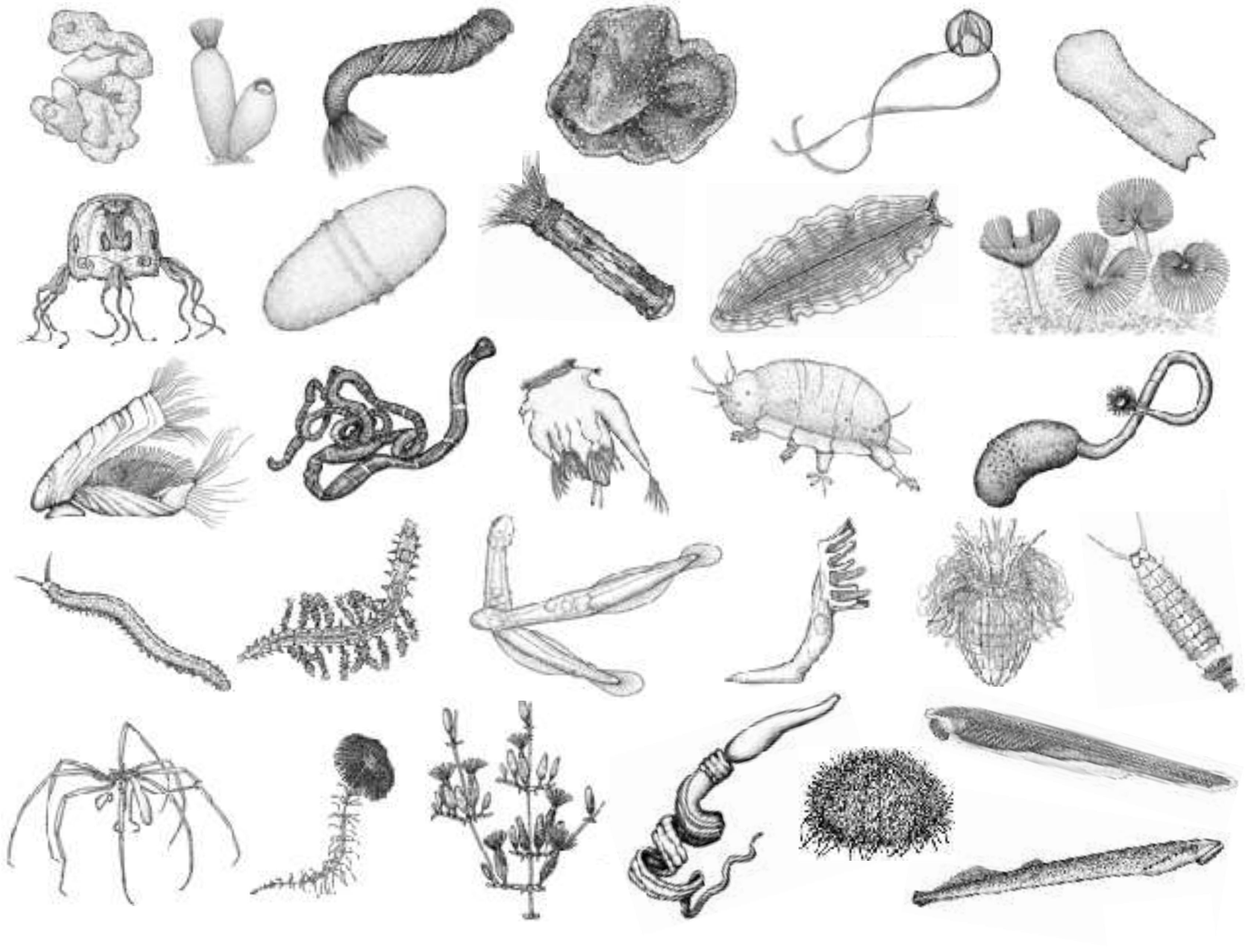


Phylogenetic sampling biases preclude the systematic comparative study of cell types

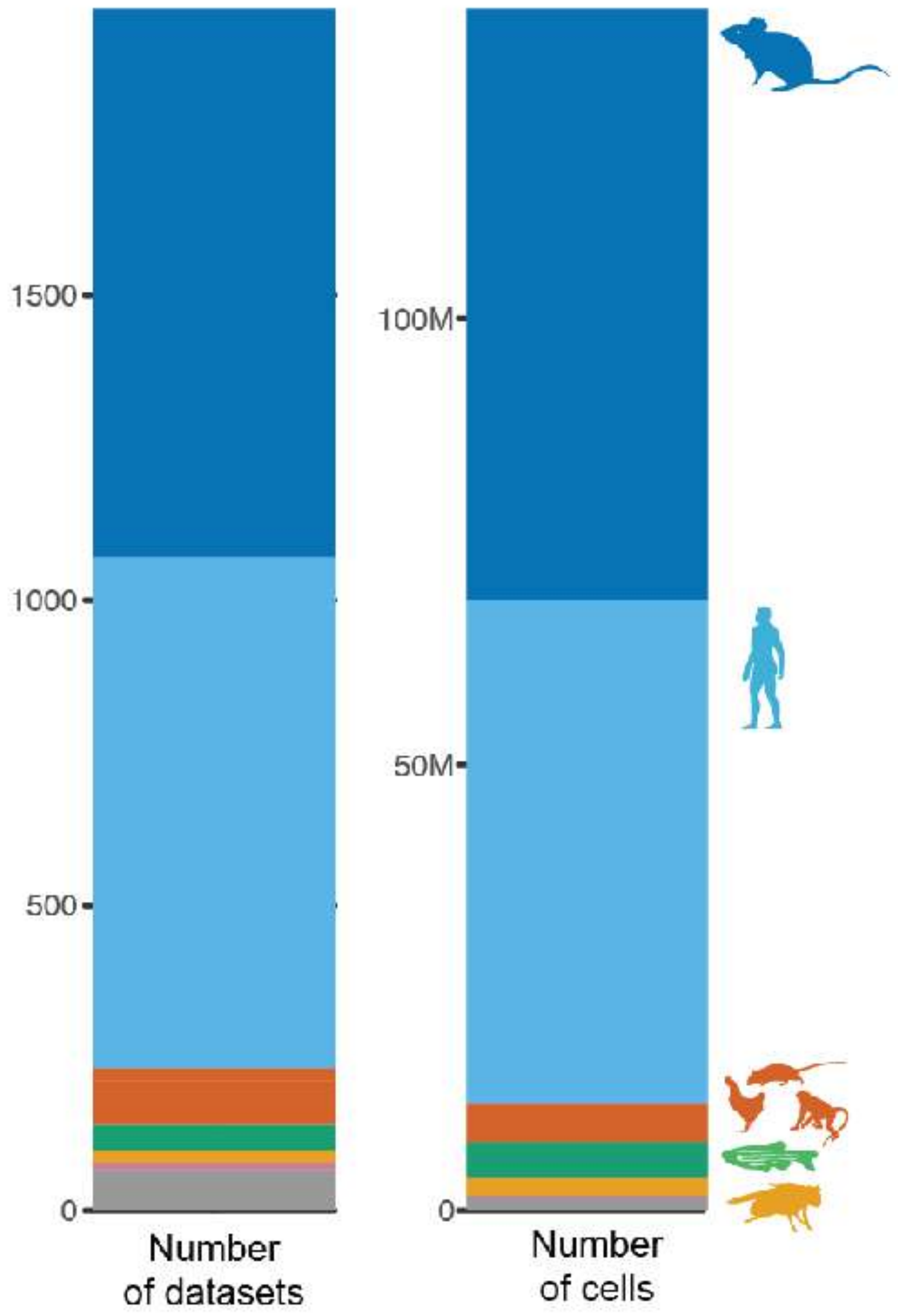
Genome sequence → Genome regulation → Cell types

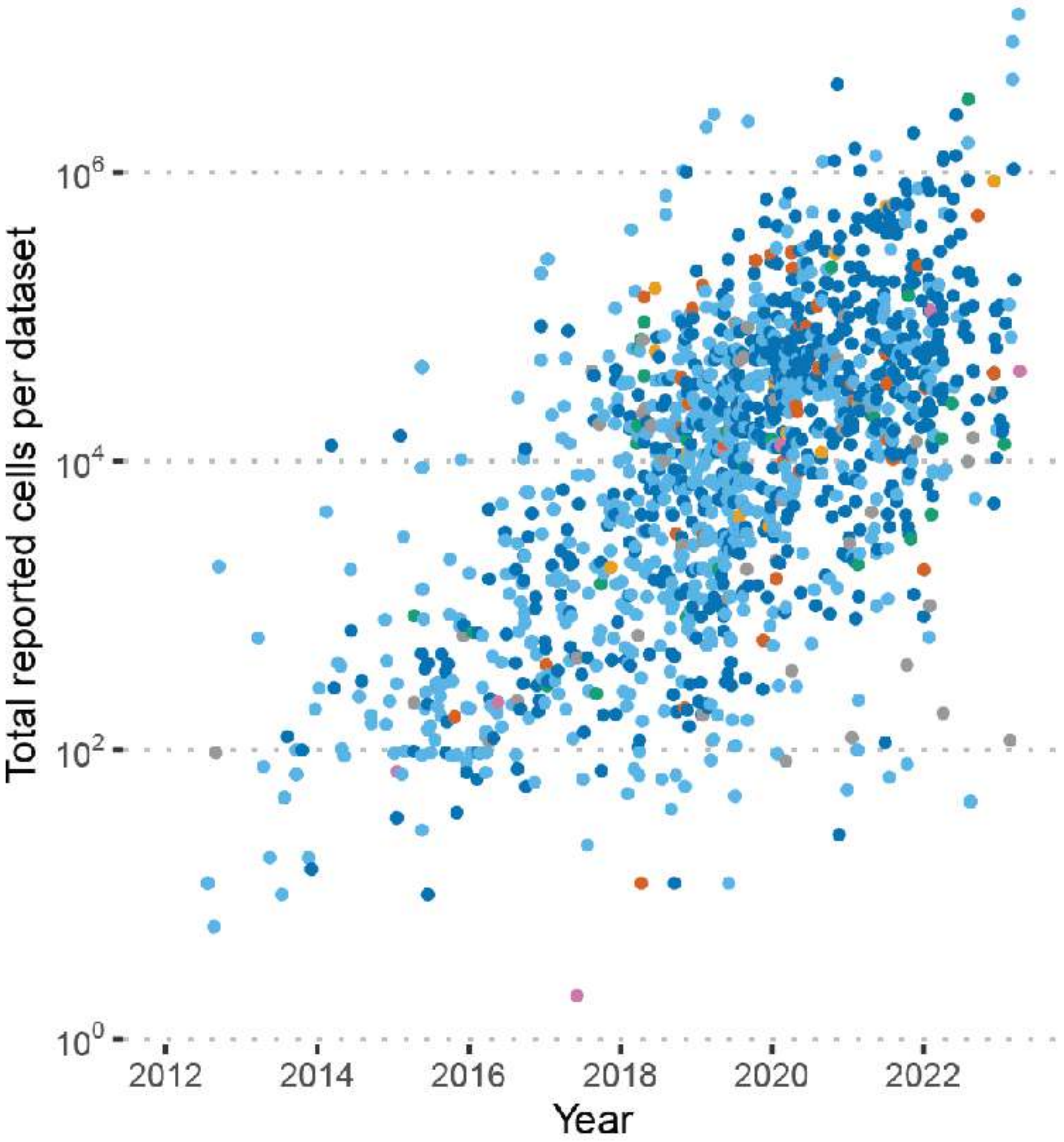






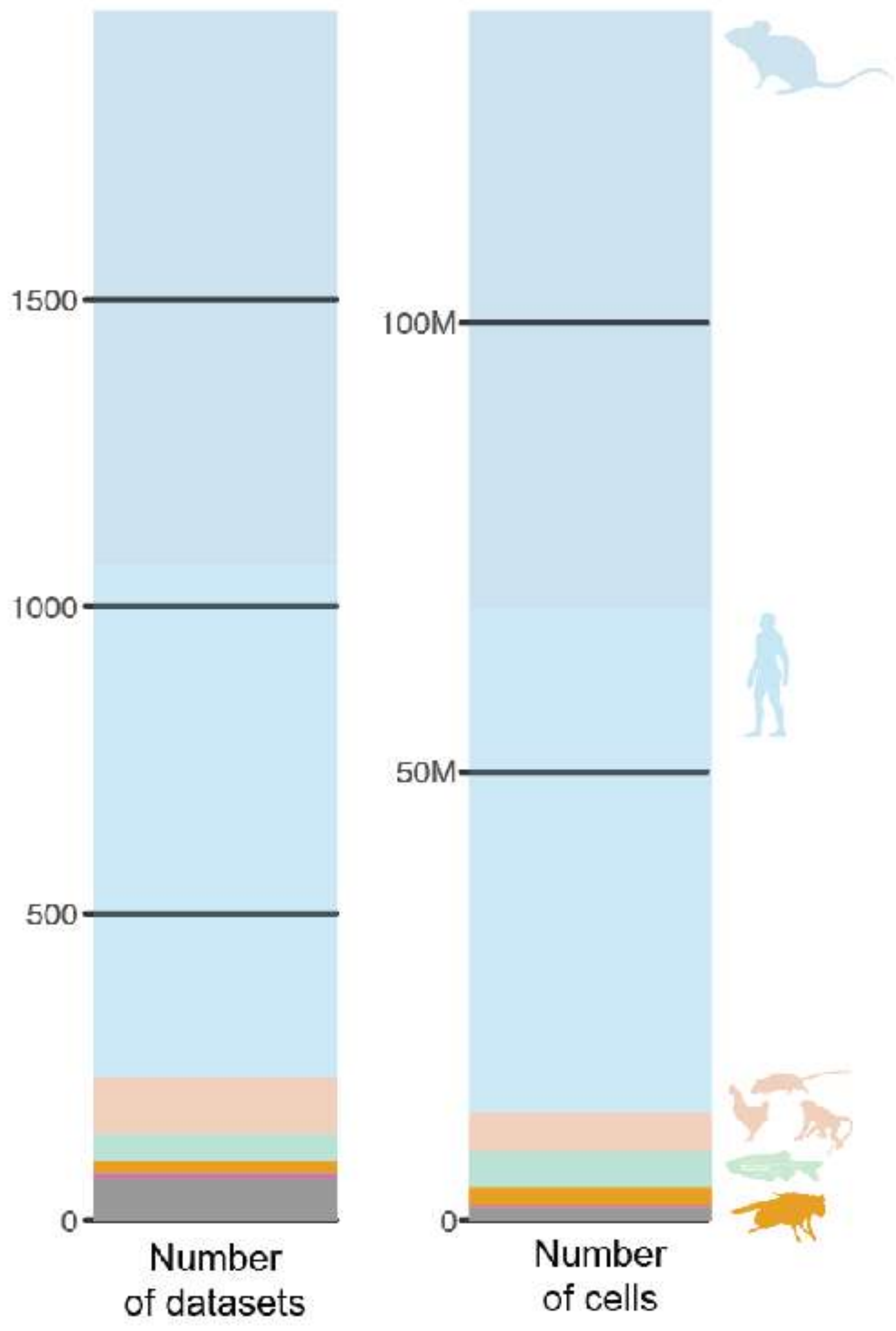
Single-cell transcriptomics: **phylogenetic state-of-the-art**

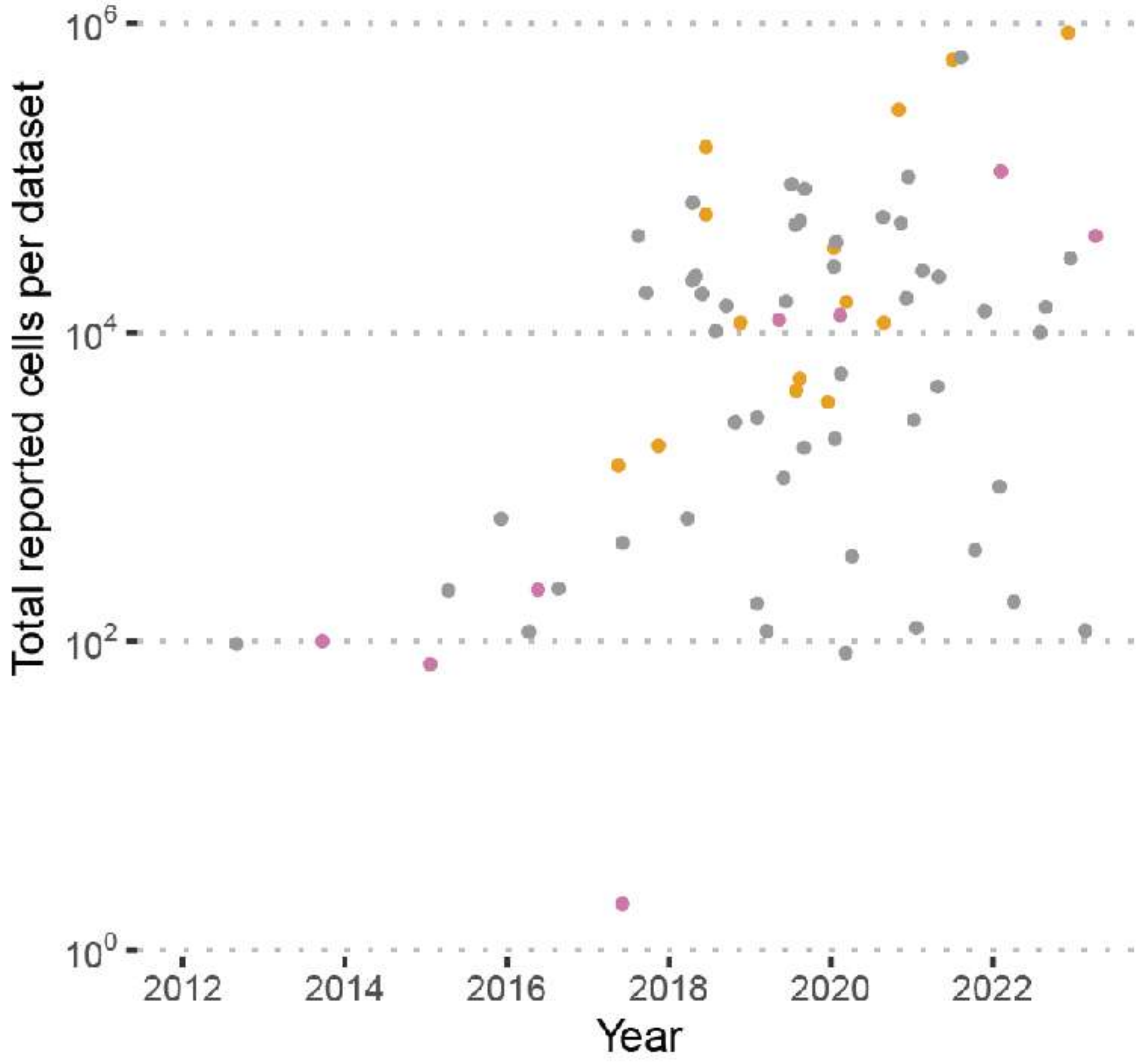




Modified from Svensson, Beltrame & Pachter, Database 2020

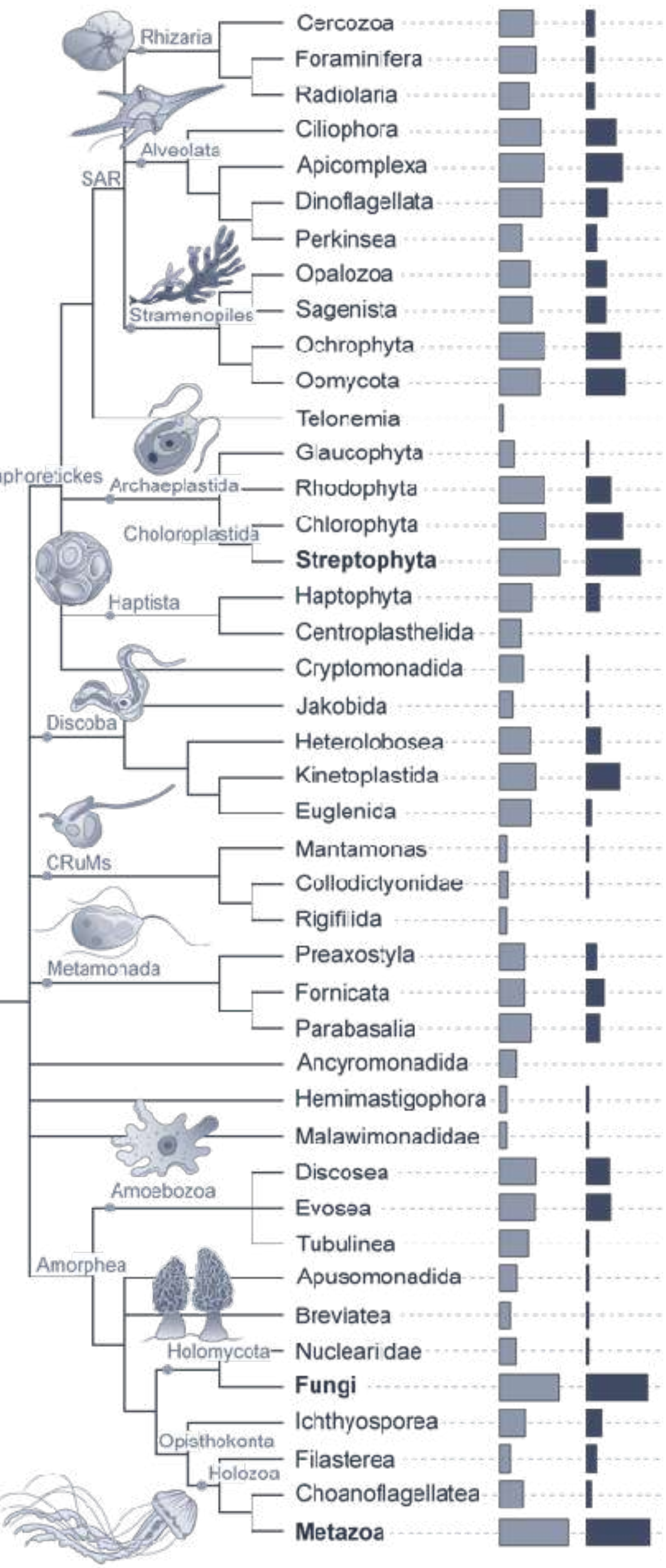
Single-cell transcriptomics: **phylogenetic state-of-the-art**

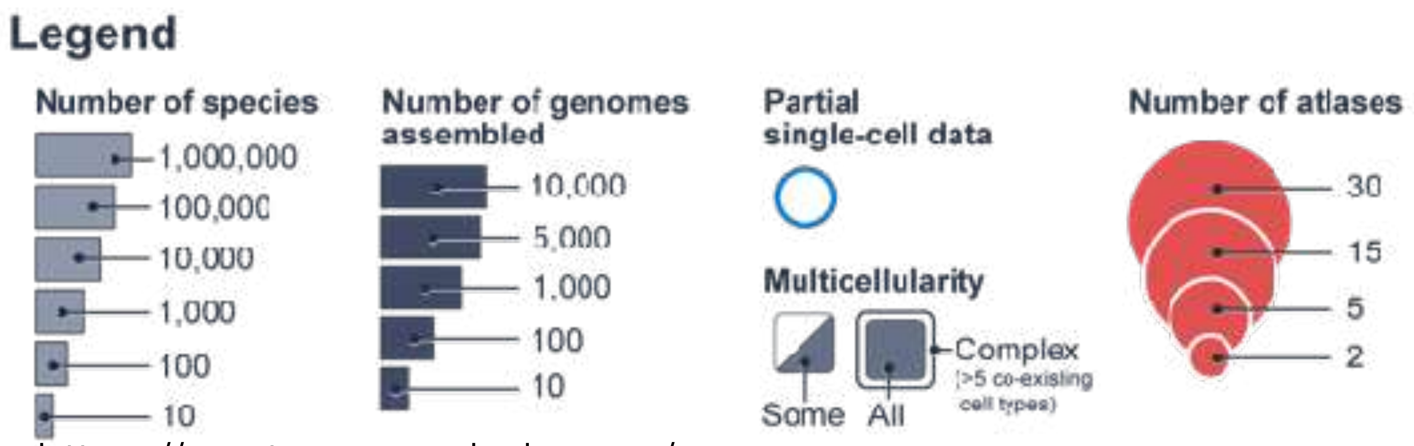




Modified from Svensson, Beltrame & Pachter, Database 2020

Single-cell transcriptomics: **phylogenetic state-of-the-art**





https://goat.genomehubs.org/

Biodiversity Cell Atlas

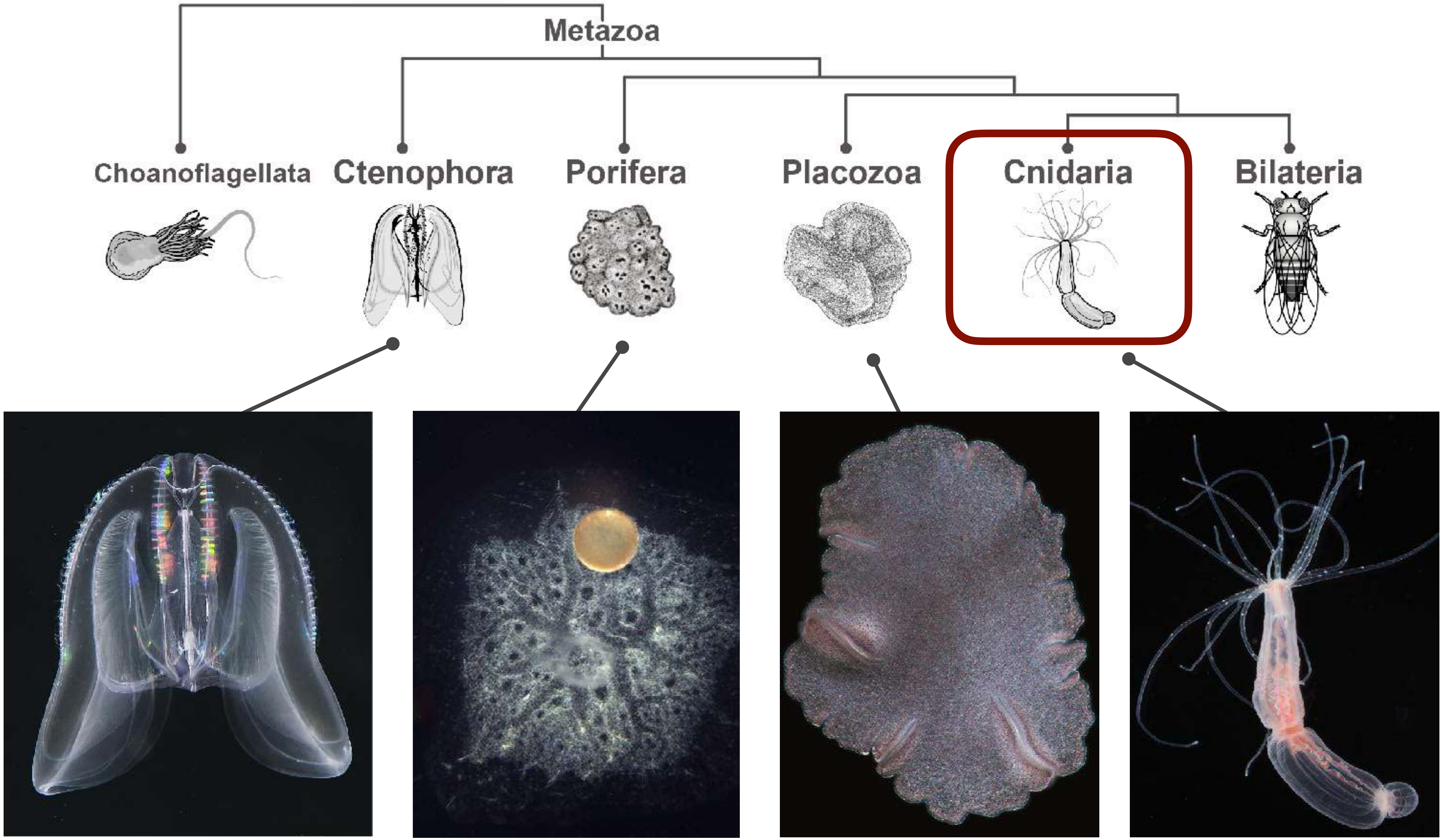
https://www.biodiversitycellatlas.org/

- Taxonomic prioritization and coordination
- Methods *decision tree* and validated protocols
- Shared atlas standards relevant across species
- Scale-up phylogenetic coverage

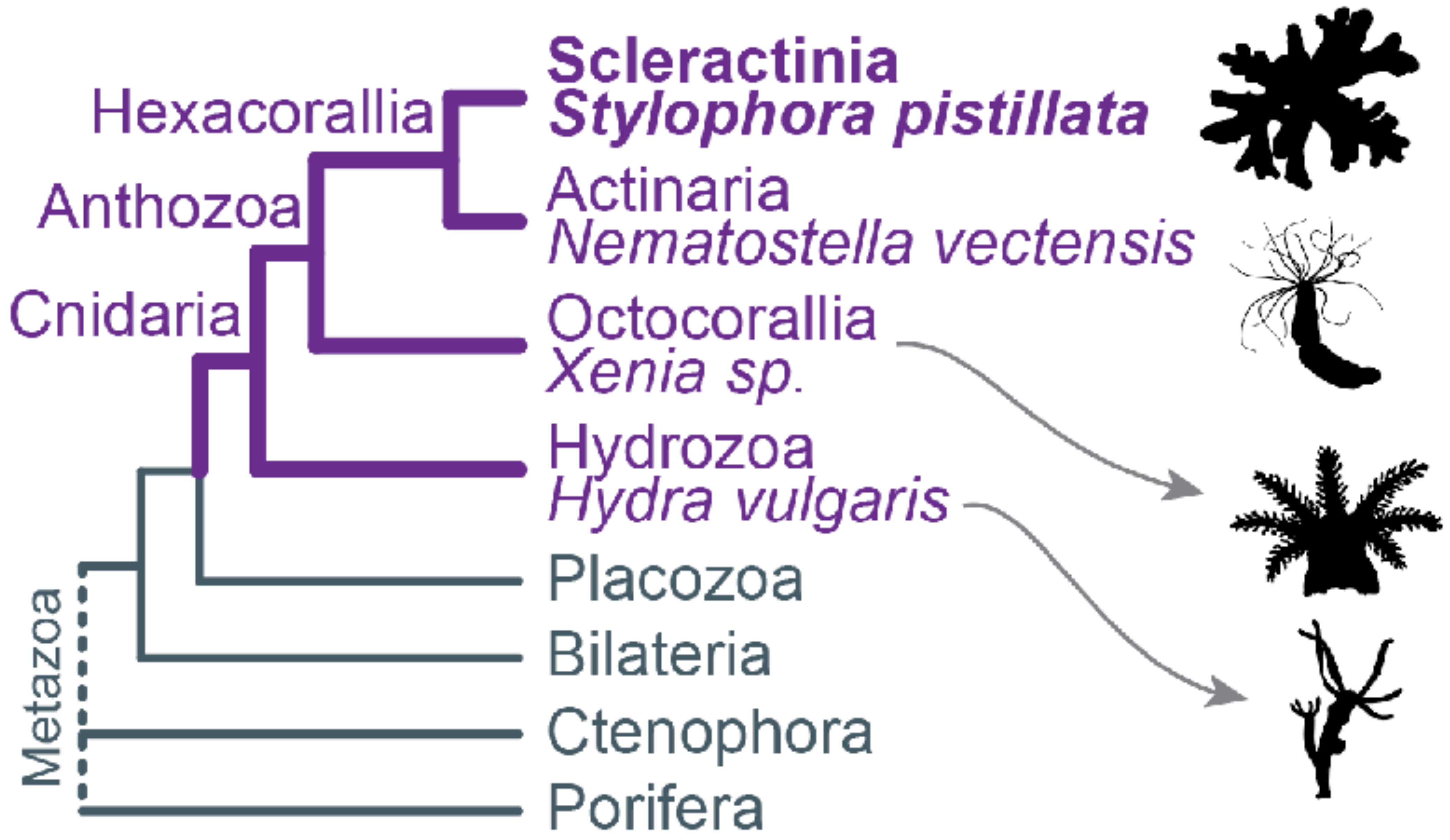
GORDON AND BETTY MOORE FOUNDATION

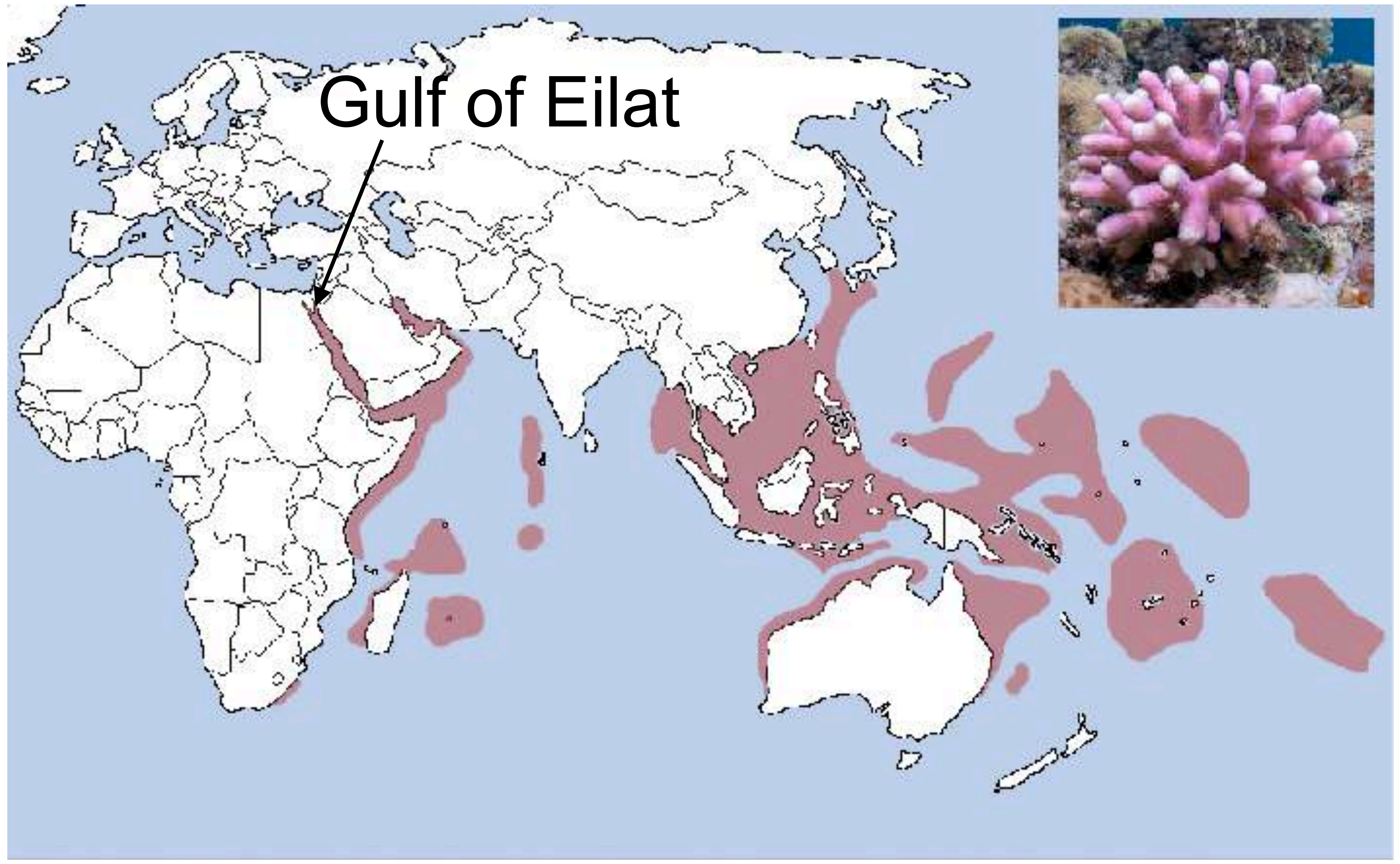
Sebé-Pedrós, Tanay, Lawniczak, Arendt et al, *Nature* 2025

Story 1: Coral cell type diversity and evolution



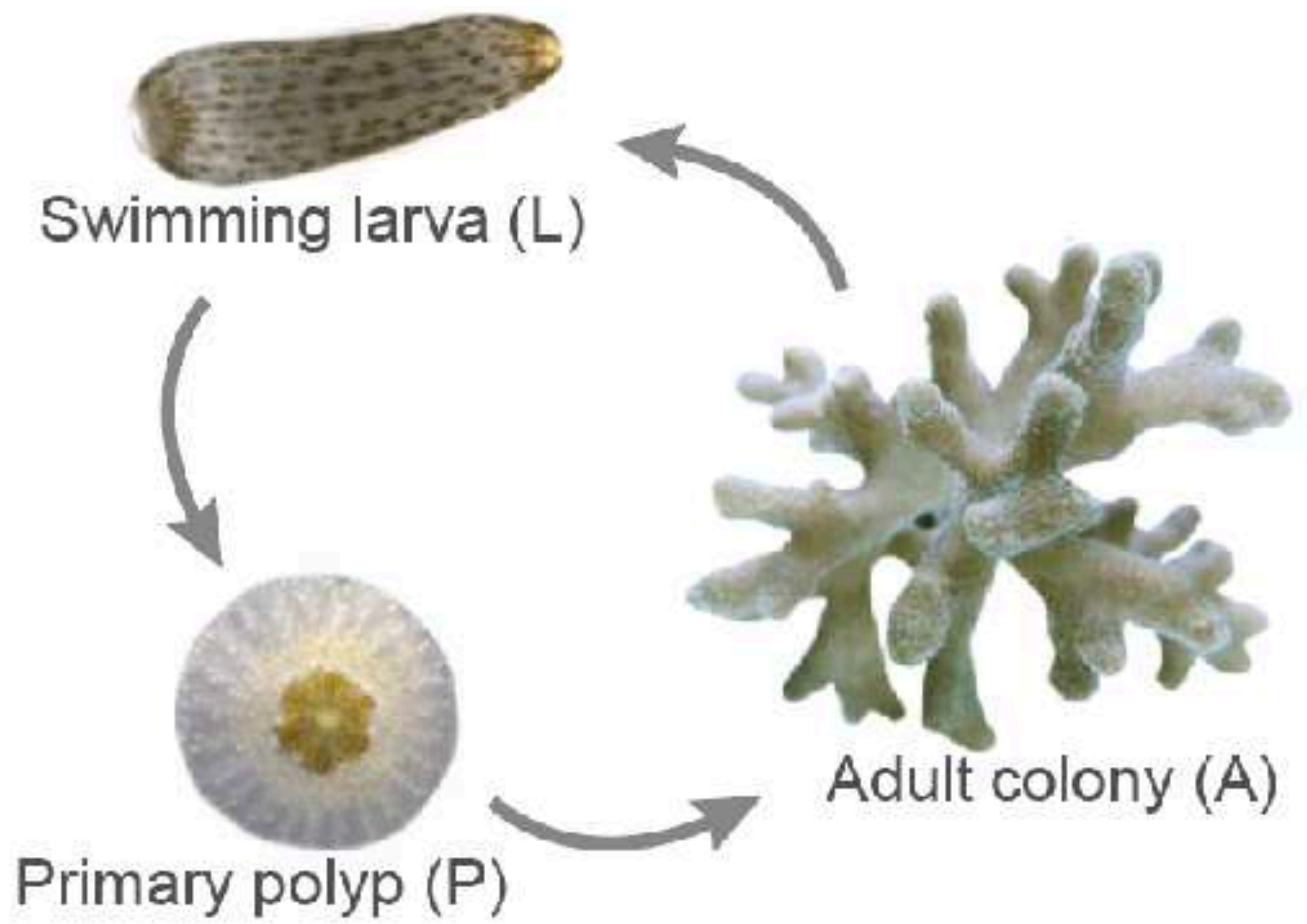
A multi-stage cell atlas reveals **stony coral cell type diversity and evolution**

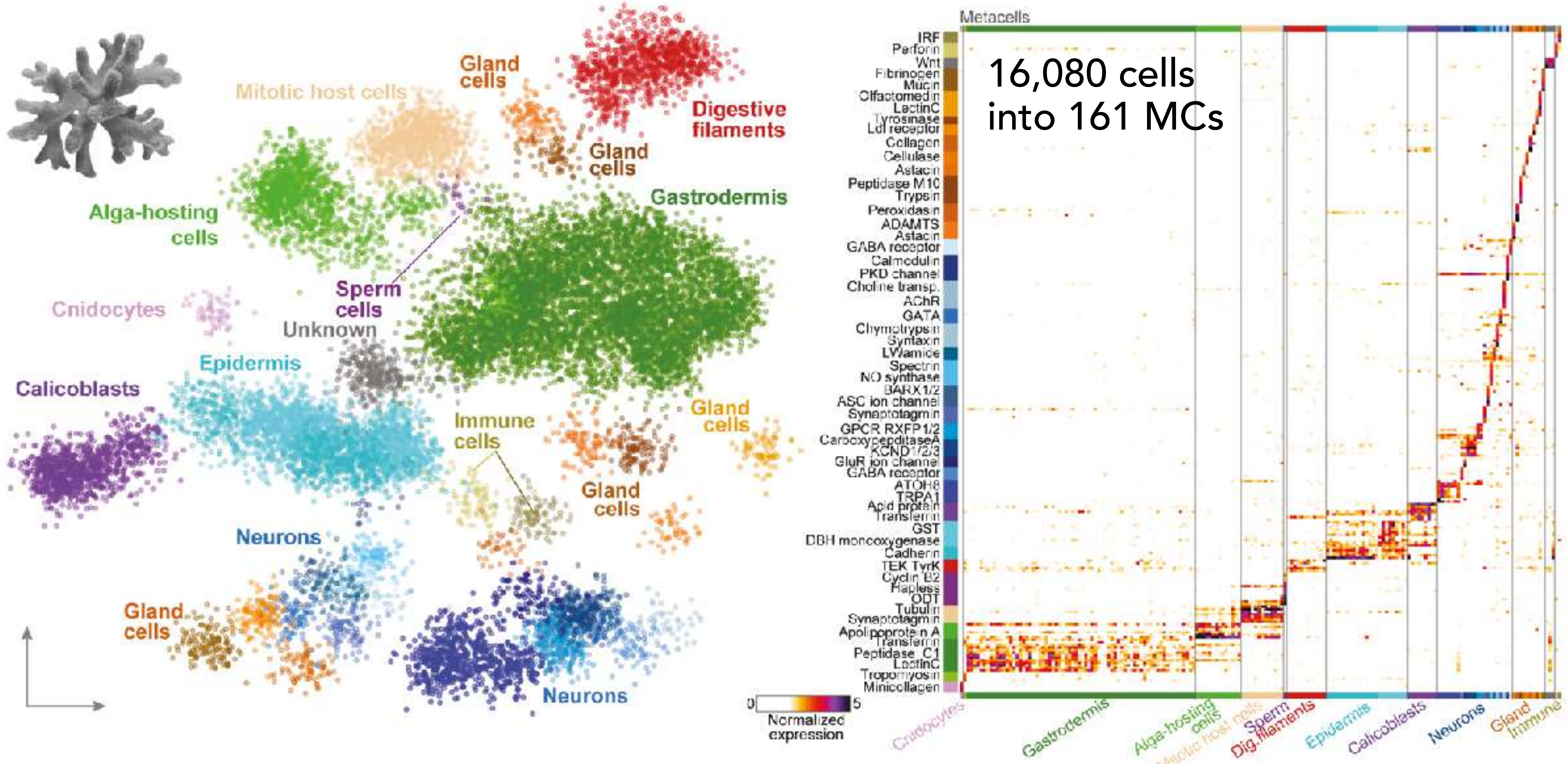


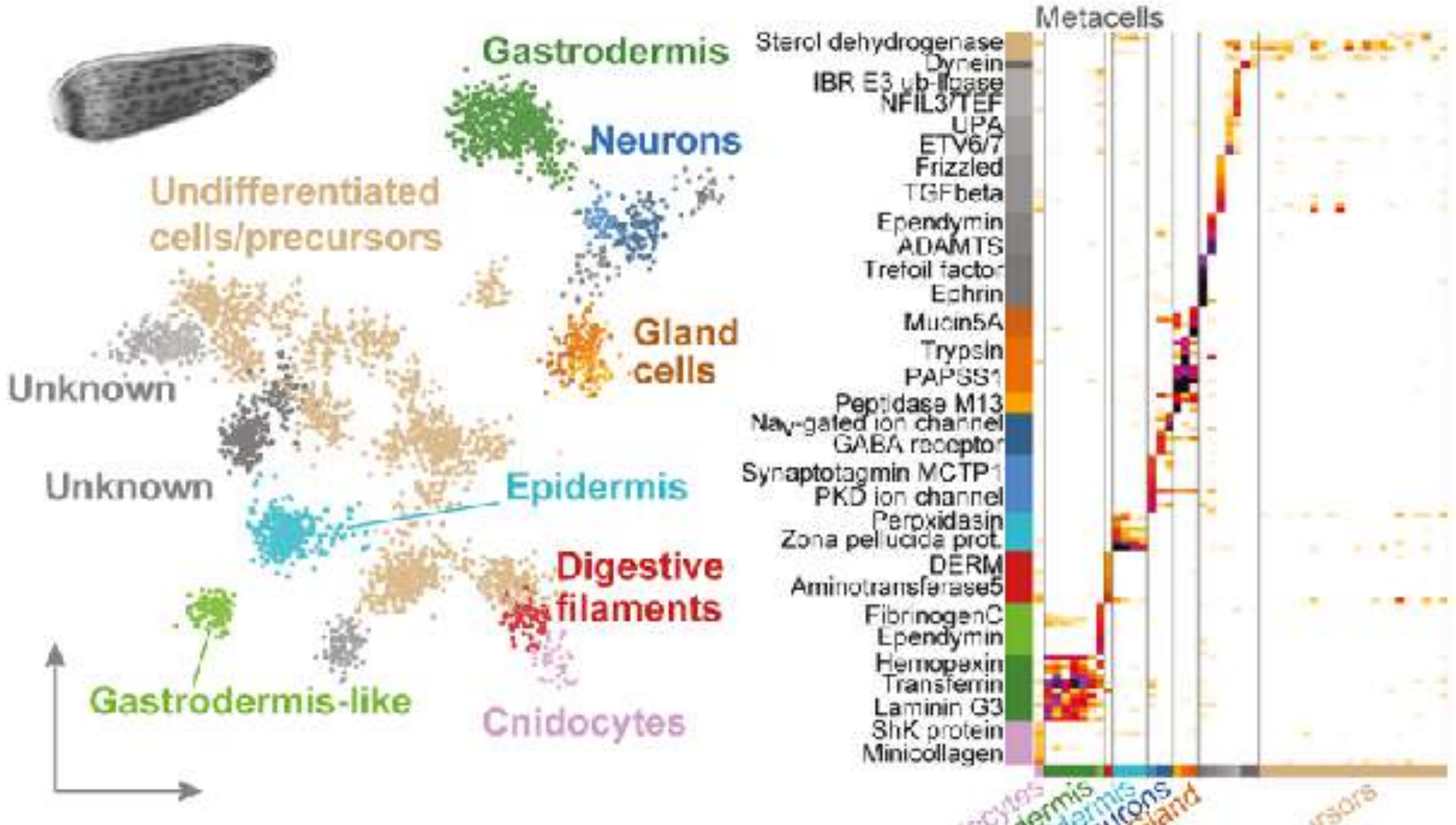


Stylophora pistillata cell type atlas

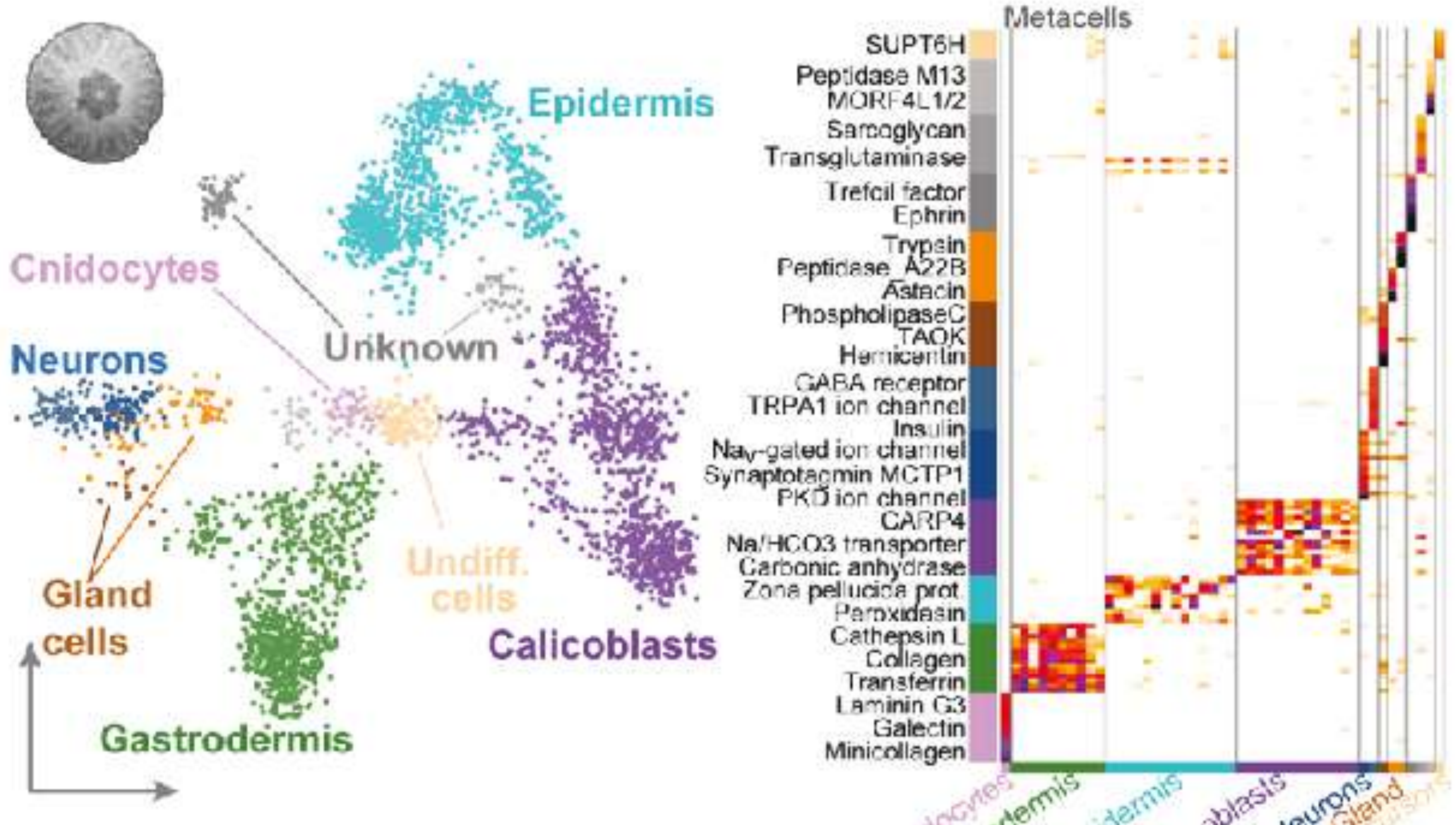
A multi-stage cell atlas reveals **stony coral cell type diversity and evolution**







3,571 cells



3,140 cells

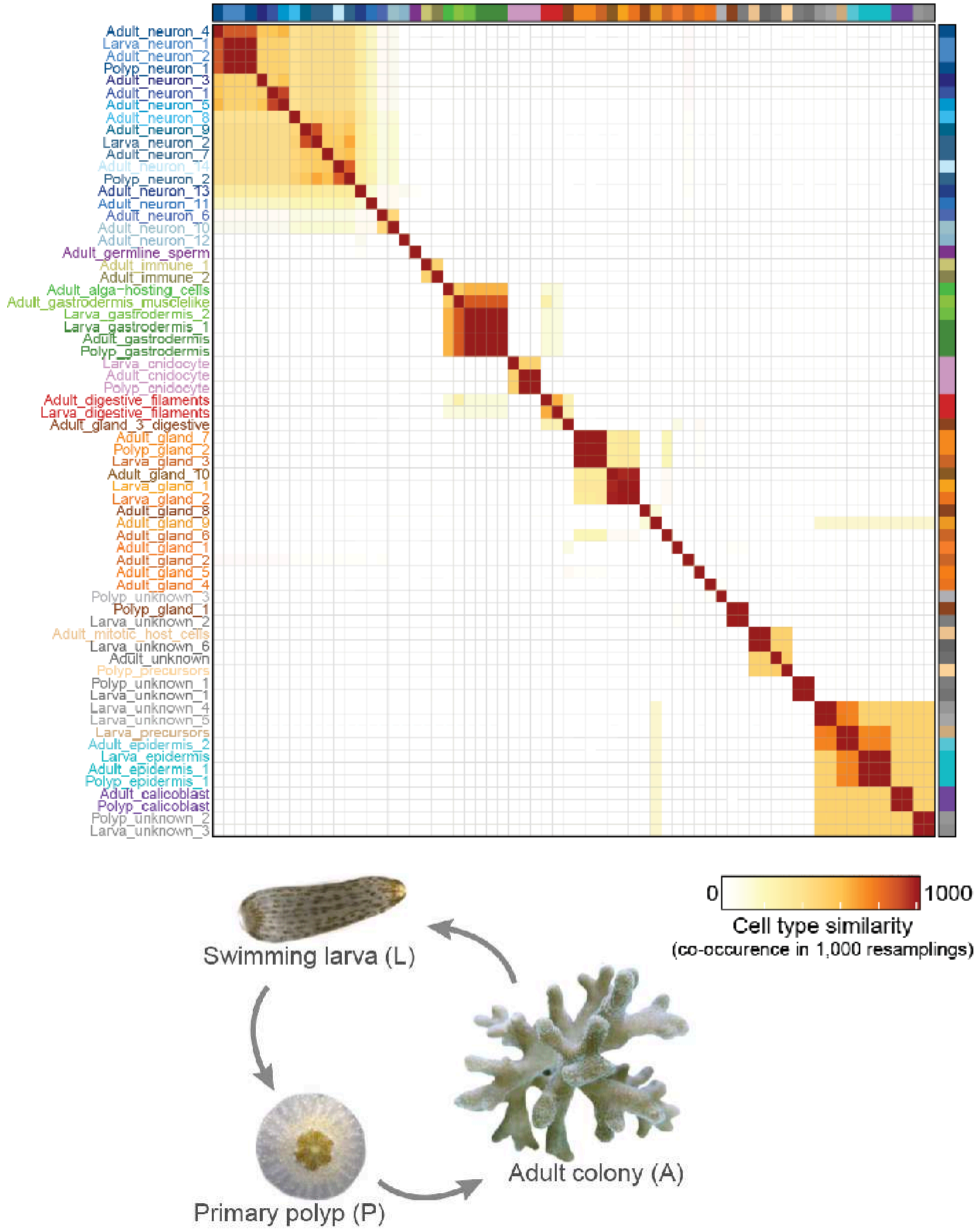
Shani Levy

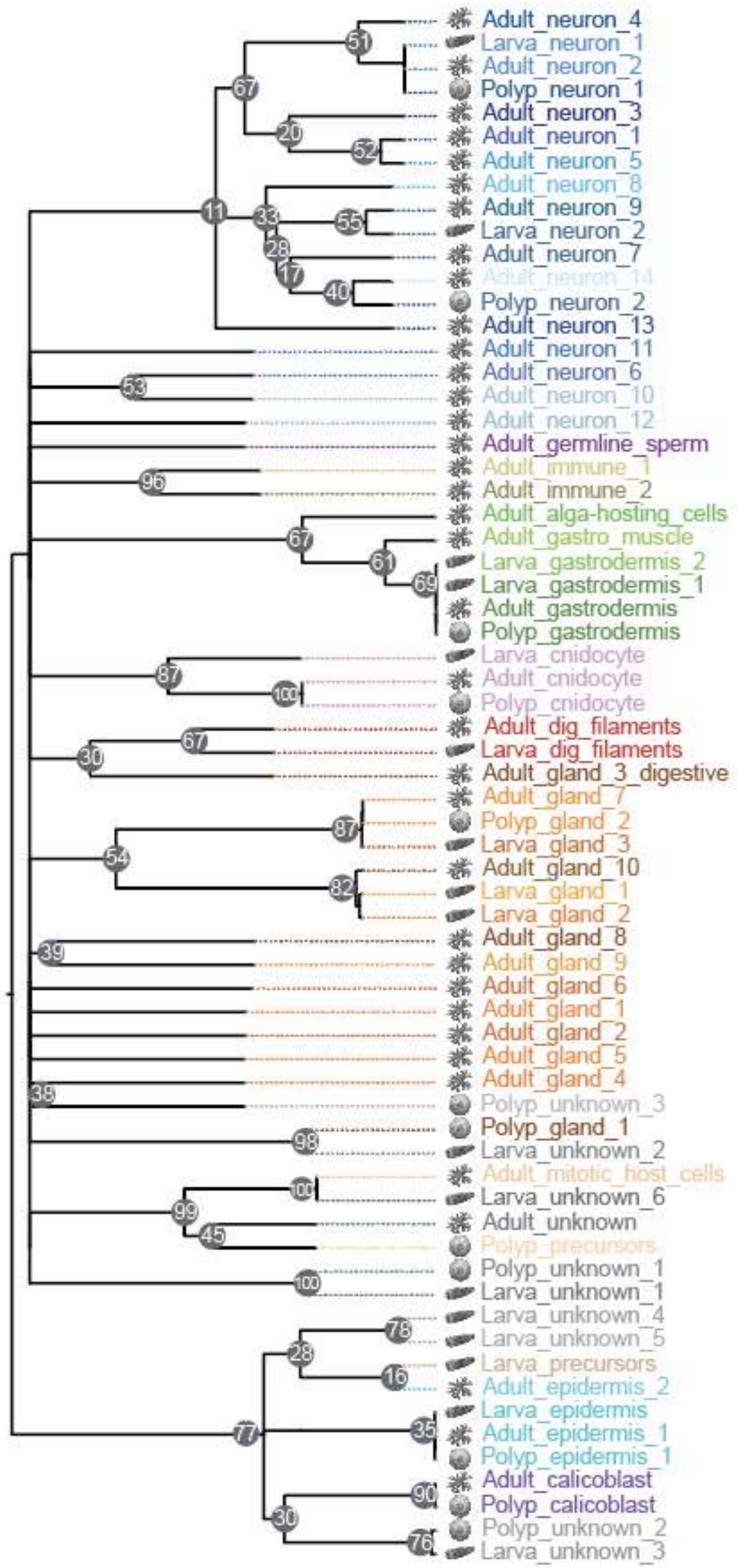
Anamaria Elek

Levy, Elek, et al. *Cell* 2021

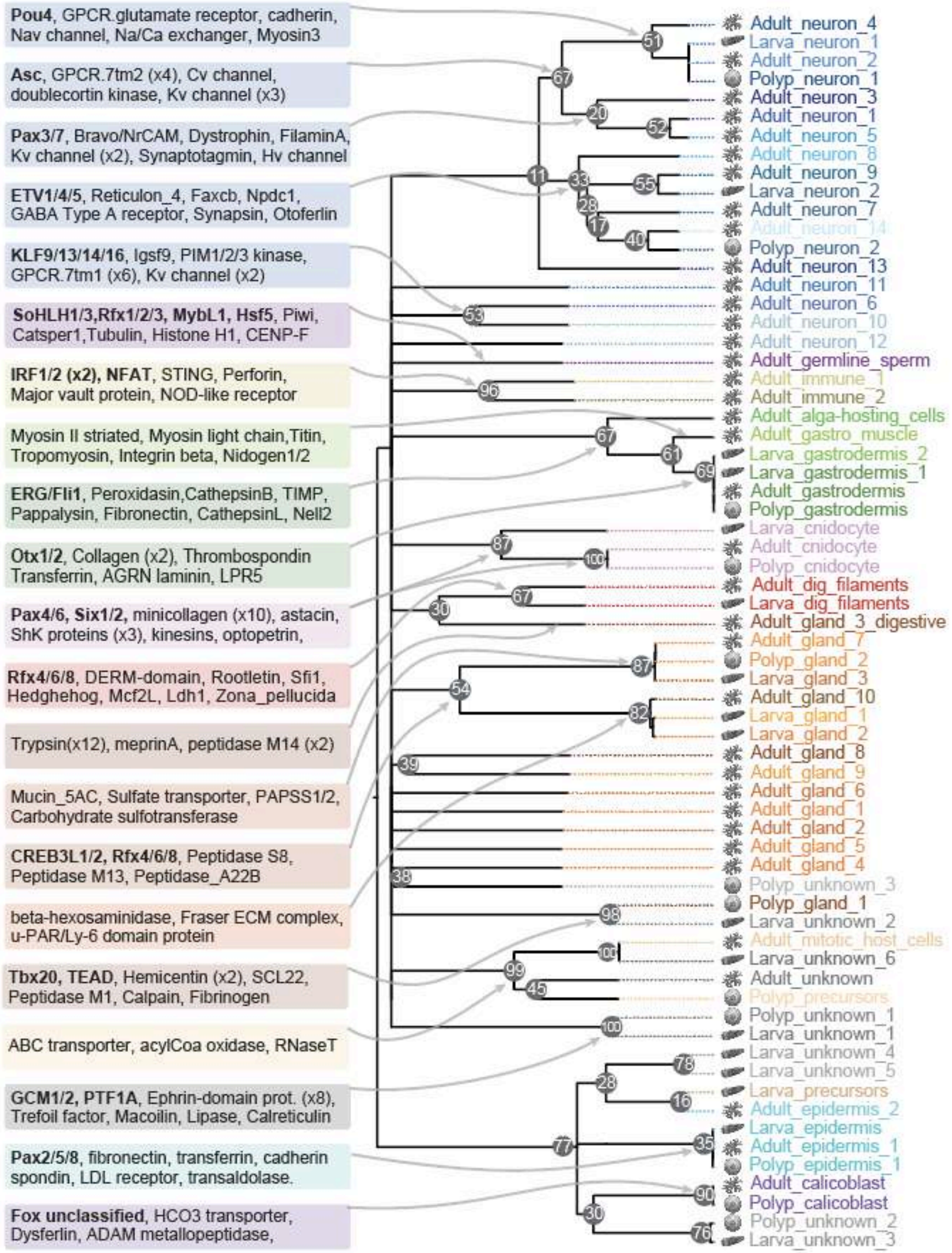
Stylophora cell atlas interpretation: ***in situ* hybridization validations**

Cross-stage comparisons

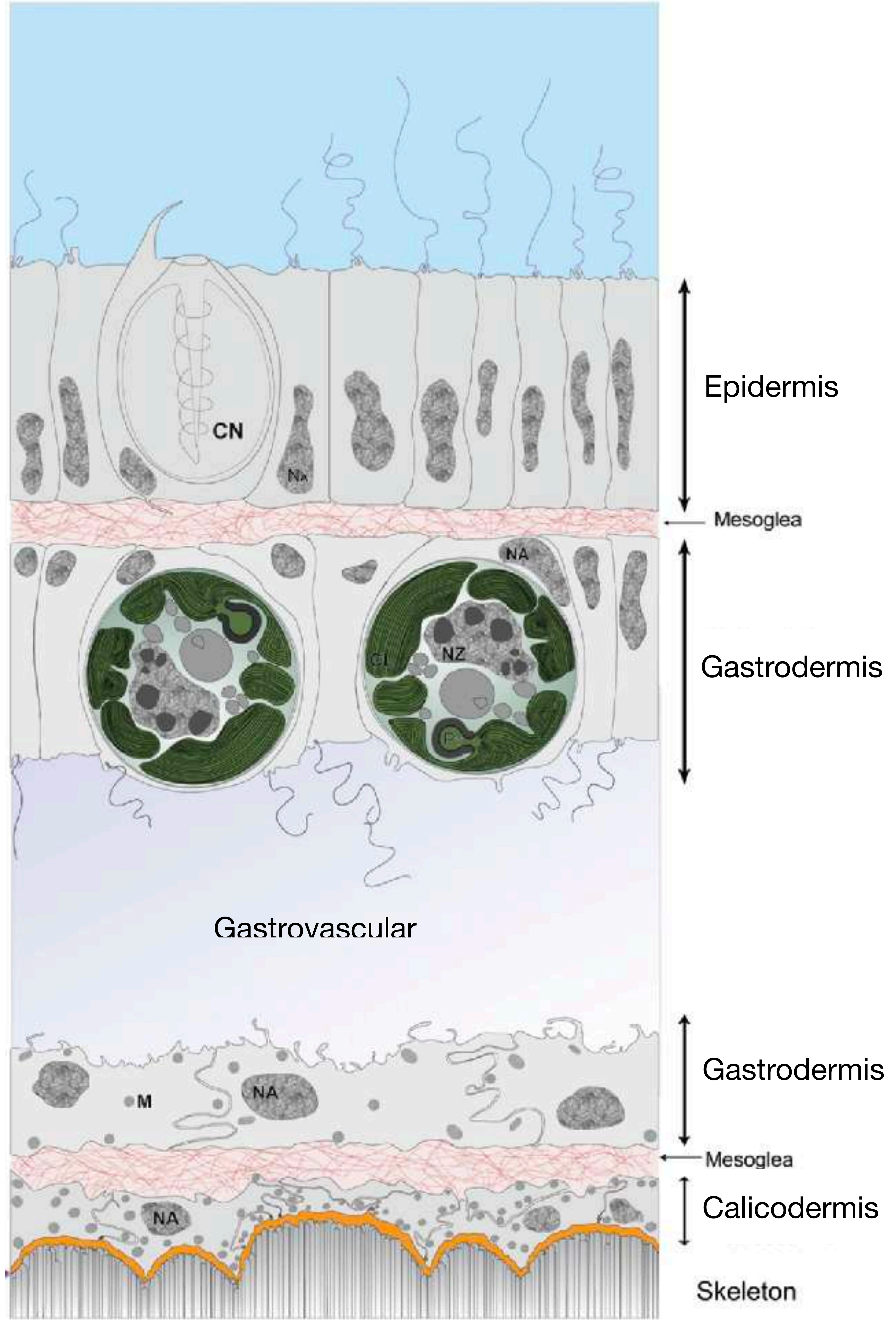




Shared and cell type-specific genes

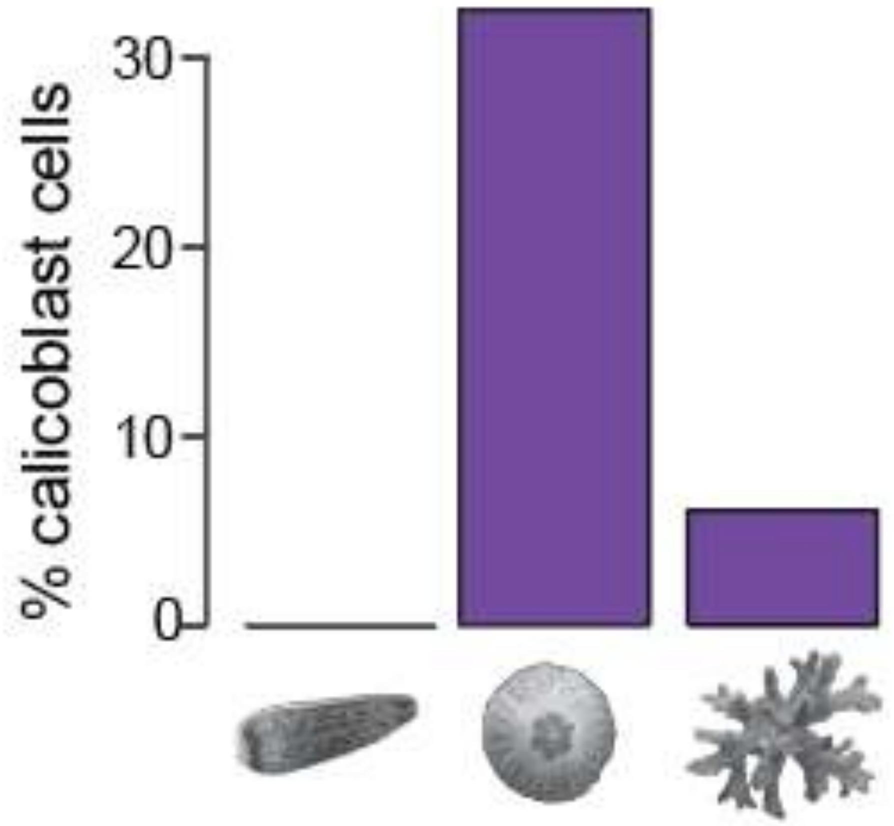


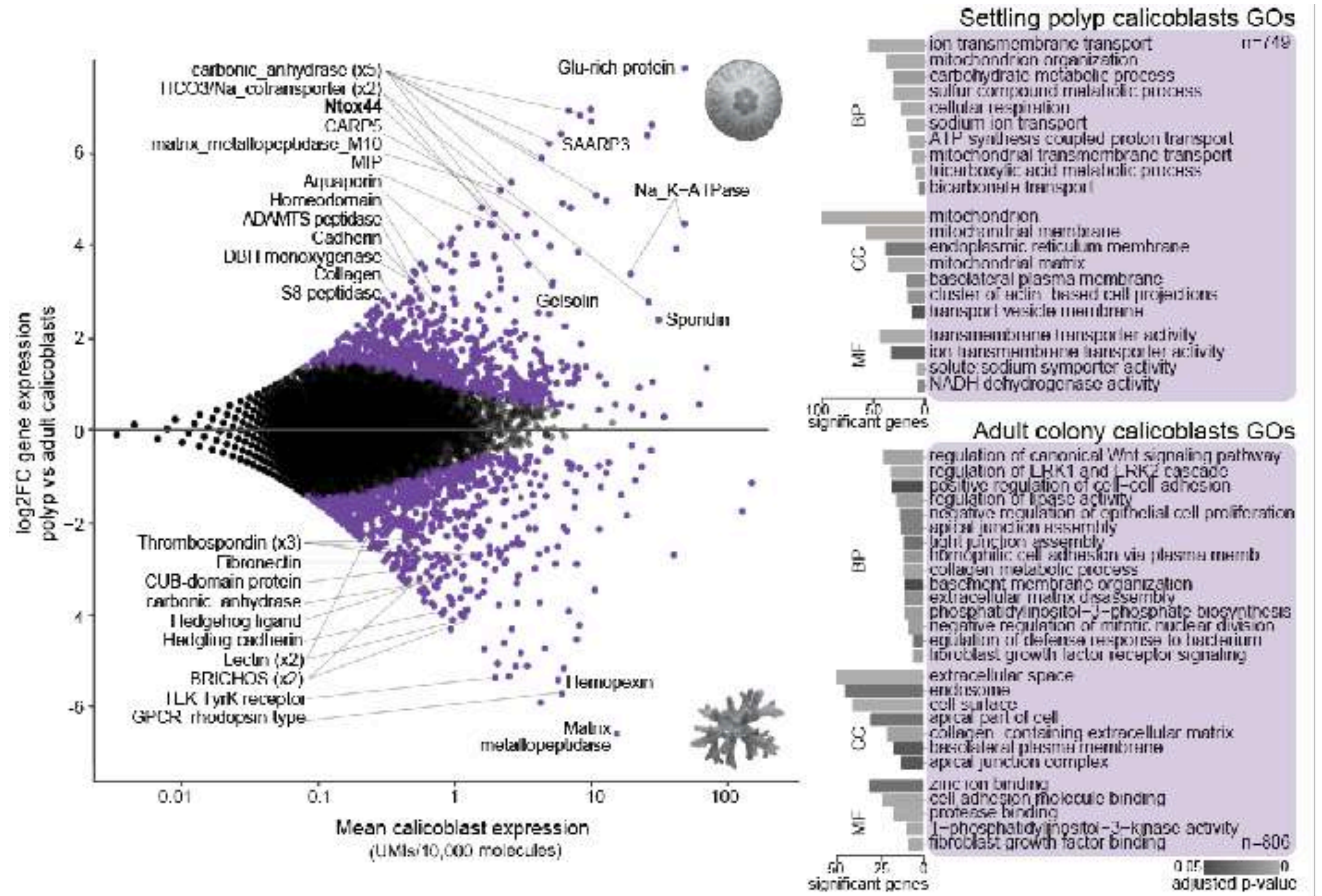
Transcriptional dynamics of skeleton formation



Calicoblasts are transcriptionally similar to epidermal cells

Calicobasts are abundant in settling polyps, absent in larva





Skeleton production metabolism

Fox XP_022788808_1

Homeobox XP_022801442_1

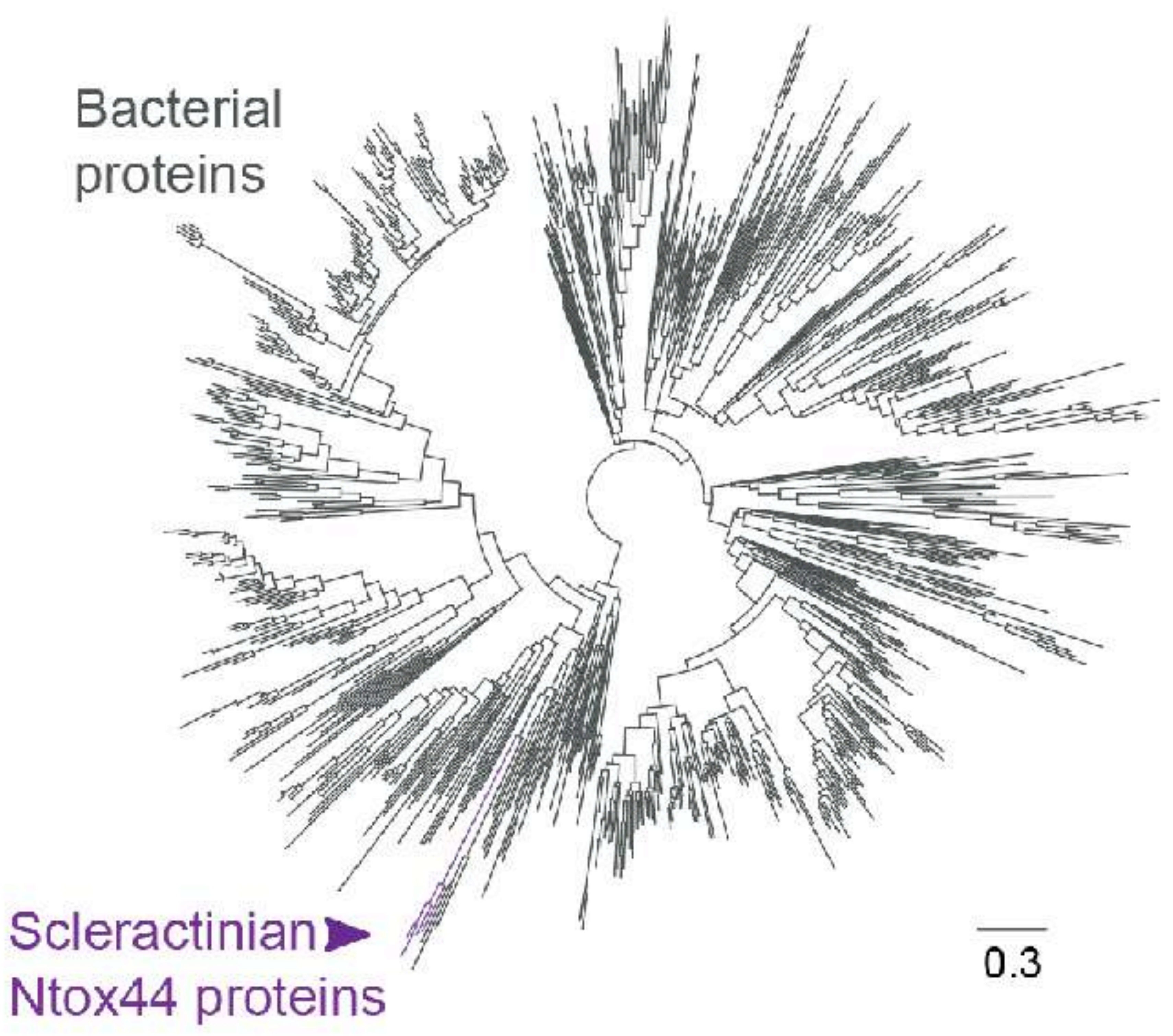
Epidermal-like identity

Fox XP_022788808_1

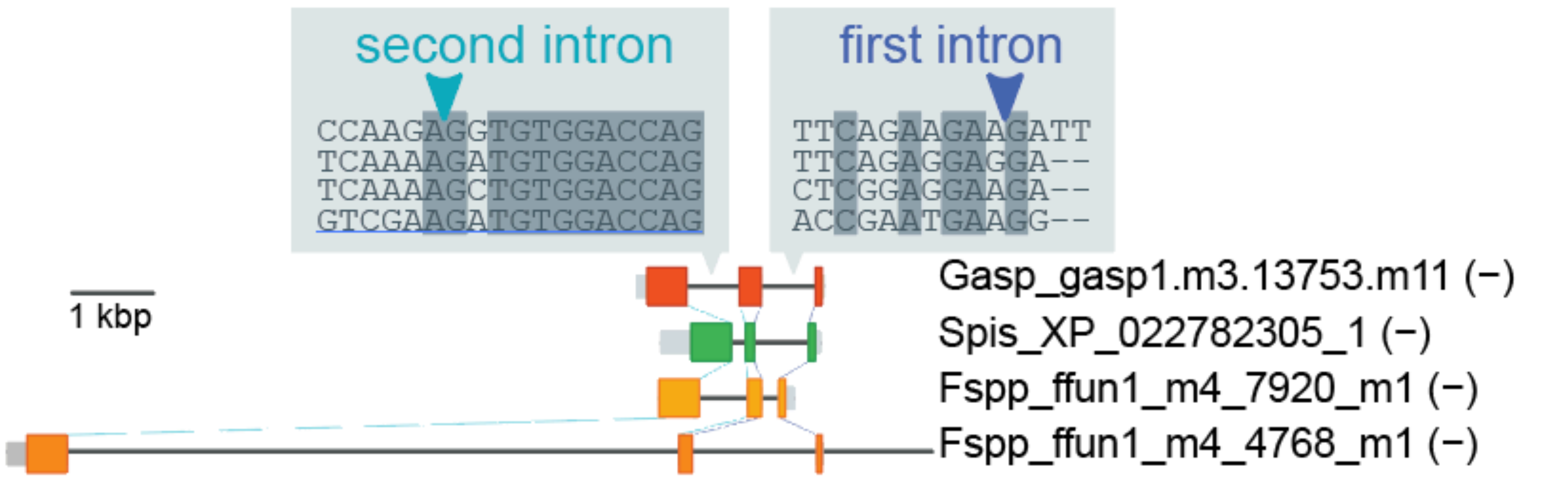
Bach/Nfe2

Pax2/5/8 (Epidermal TF)

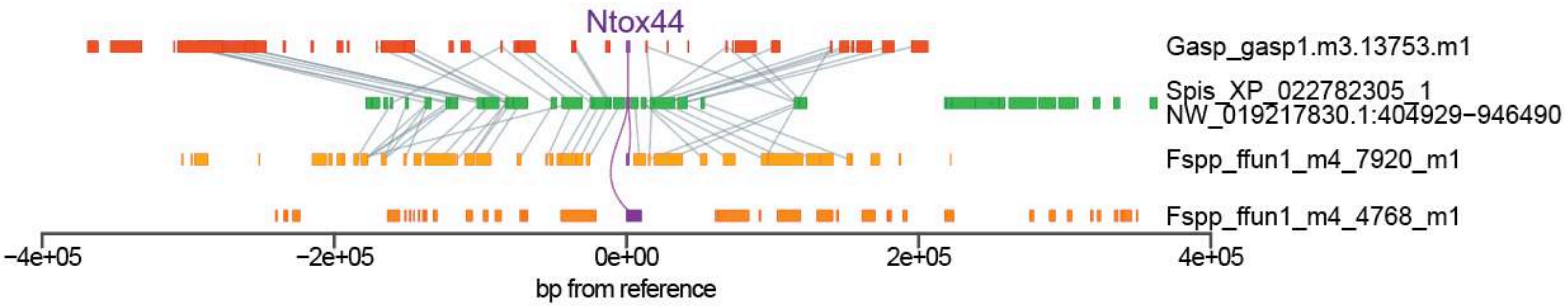
A bacteria-to-corals HGT toxin expressed in calicoblasts during skeleton formation



Secretion signal peptide and conserved intronic positions

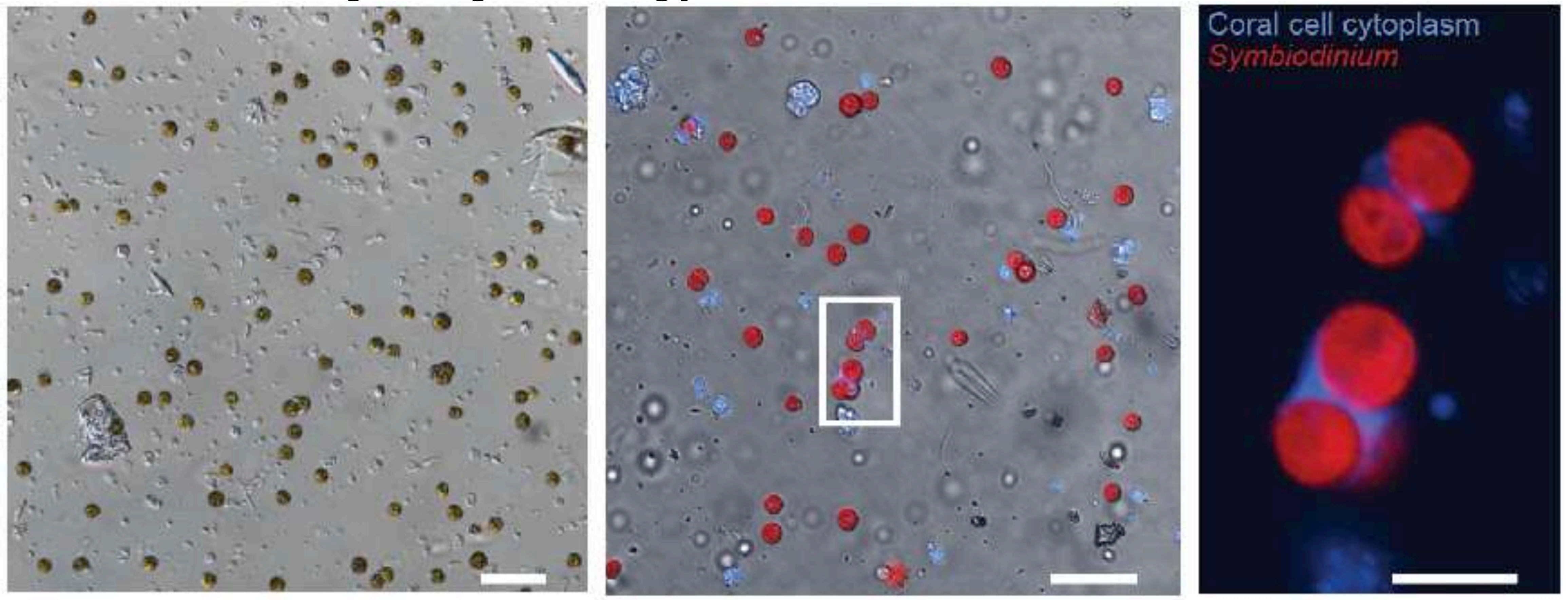


Syntenic gene neighborhoods



Host-symbiont gene expression at single-cell resolution

Host cells targeting strategy



USF1 (bHLH), Zic1 (zfC2H2)

Leloir pathway -> Galactose metabolism

Fatty acid metabolism (Elov, Pas2,…)

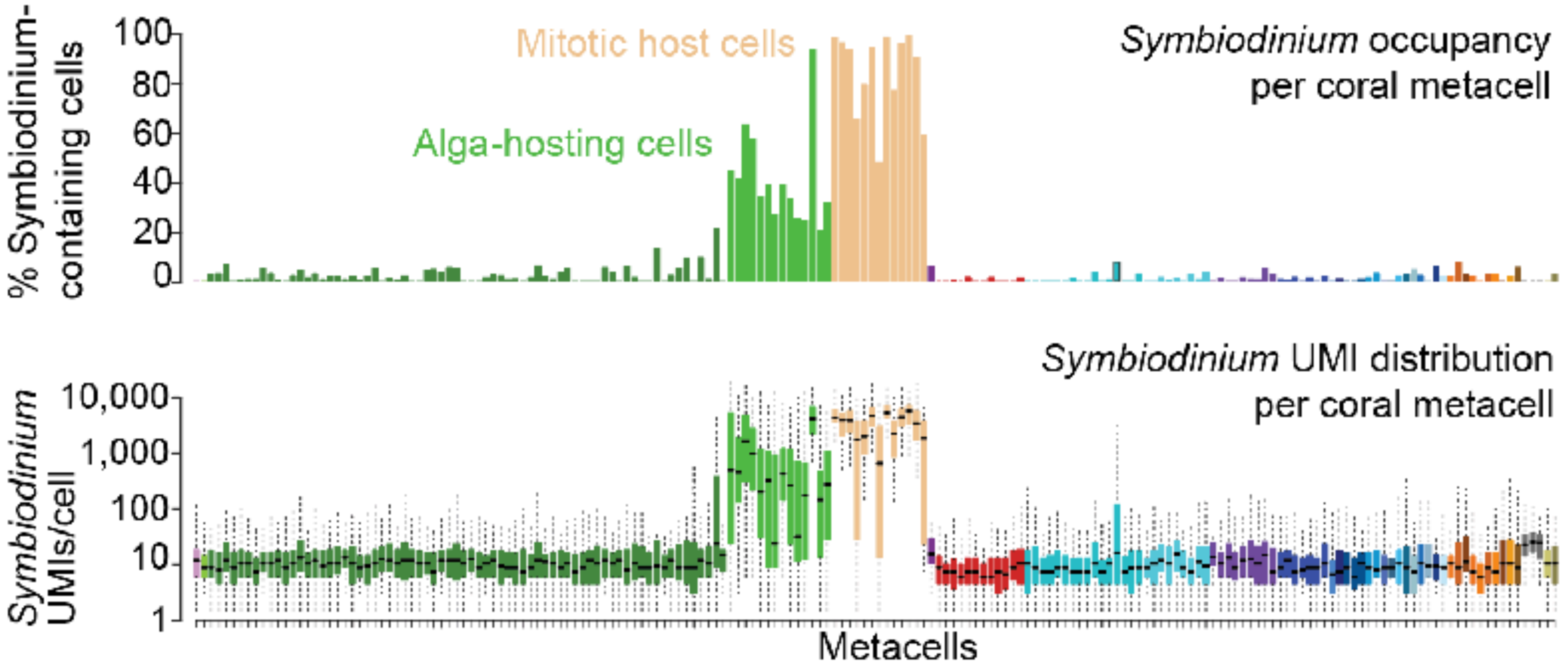
Lipid transporters (NPC1, ApoD)

Carbonic anhydrase -> CO_2 availability

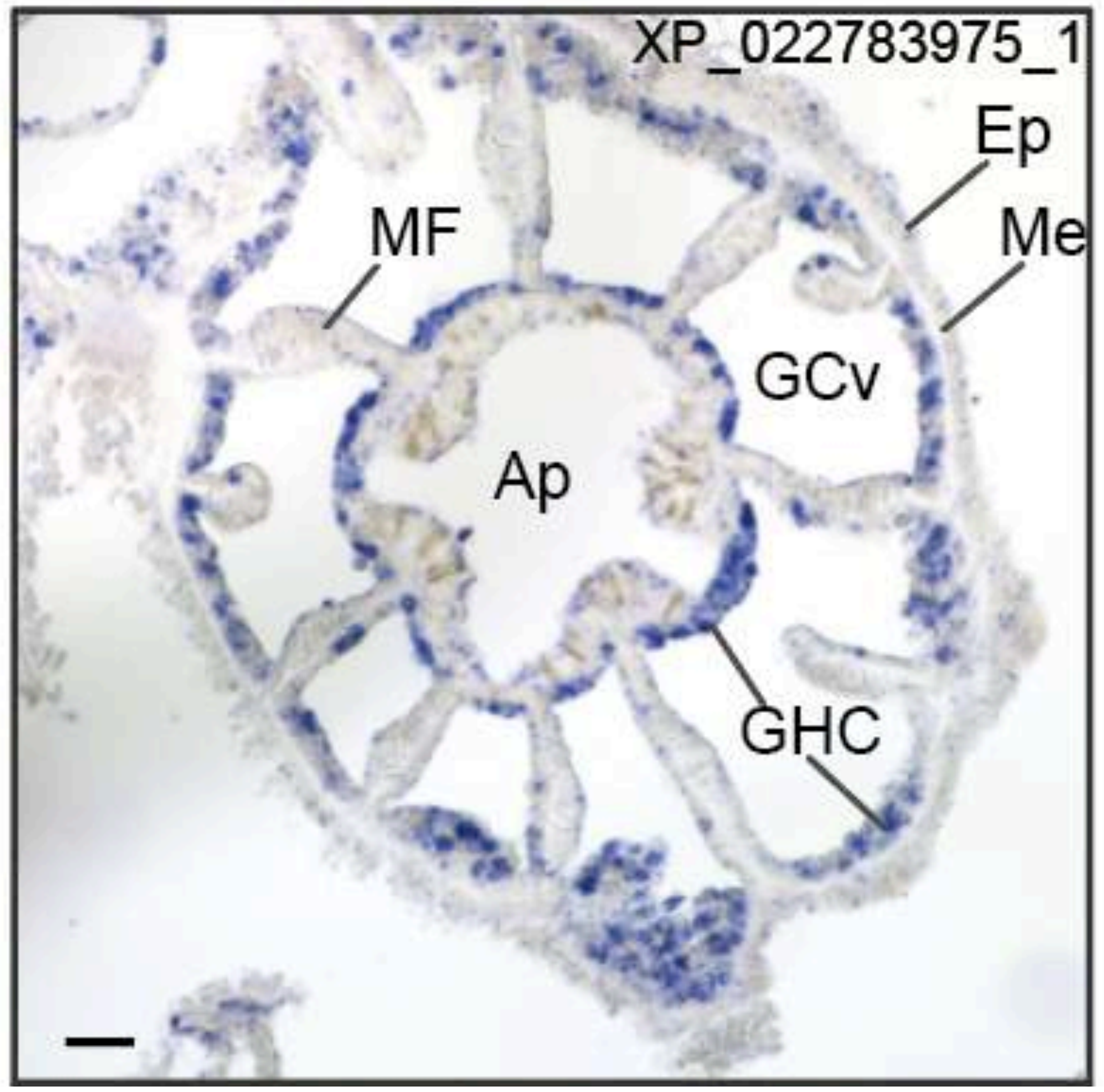
Glutathione pathway -> Oxidative stress

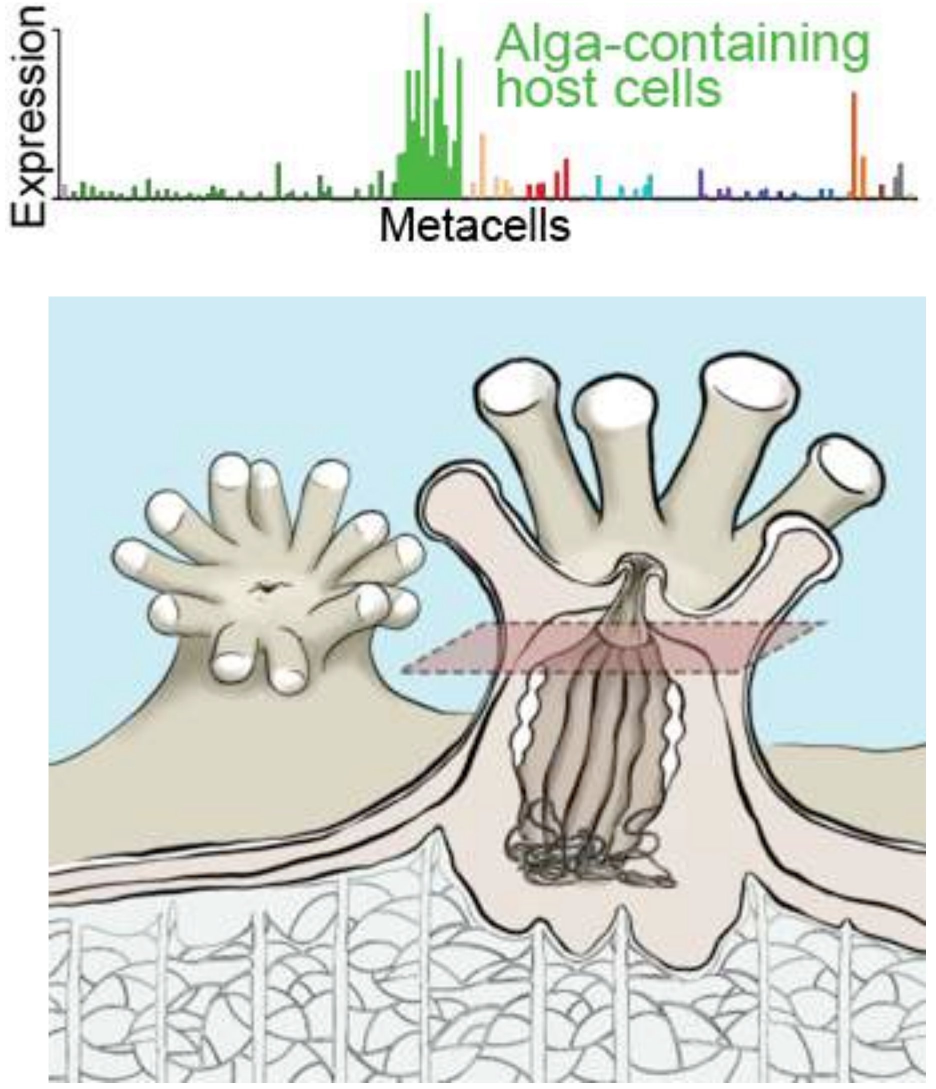
Ammonium transporters

Aminoacid transporters

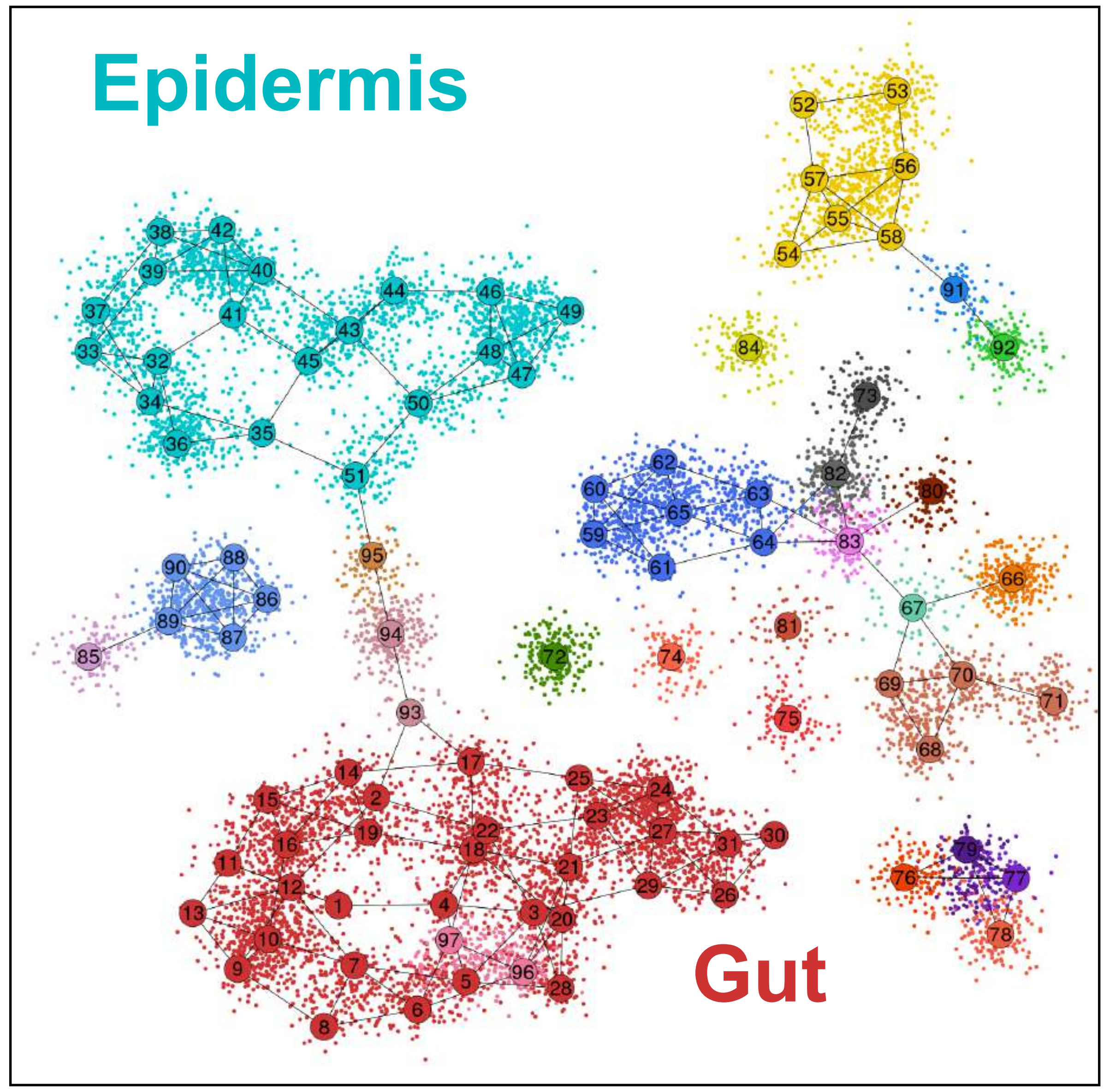


Symbiodium occupancy across coral cell types

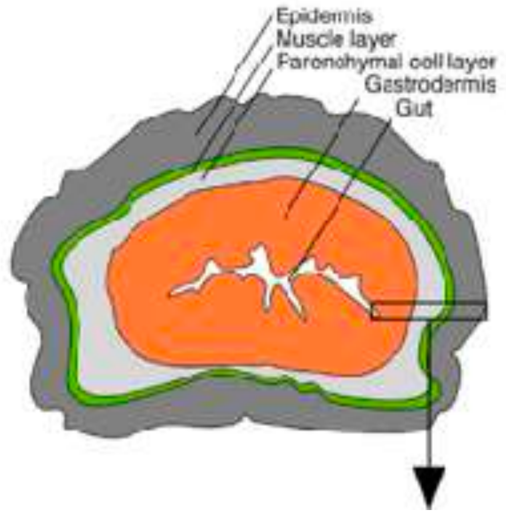




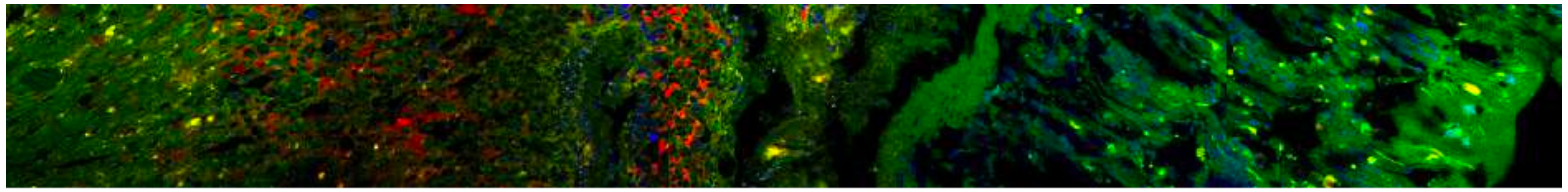
Mapping symbioses at single-cell resolution

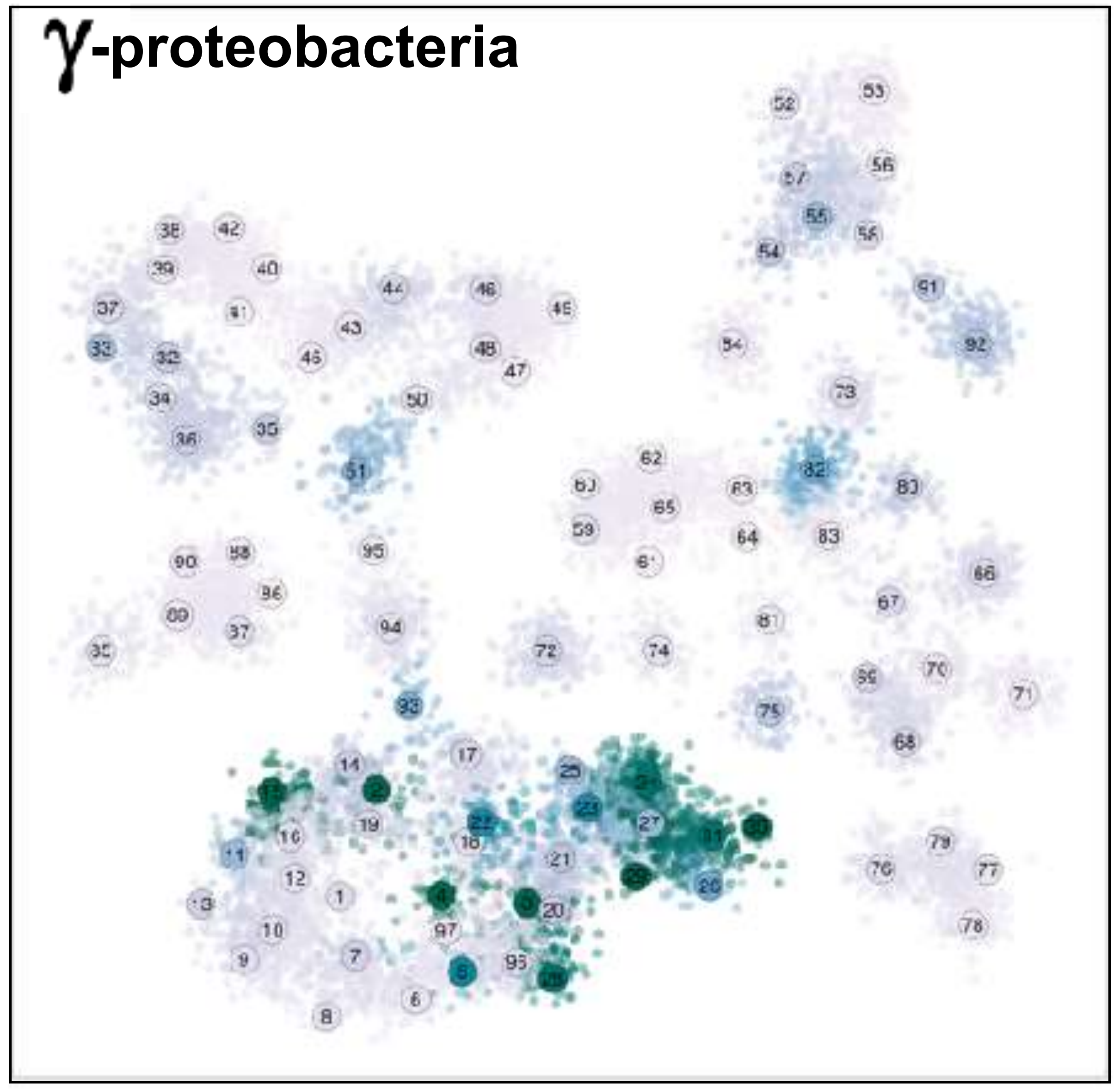


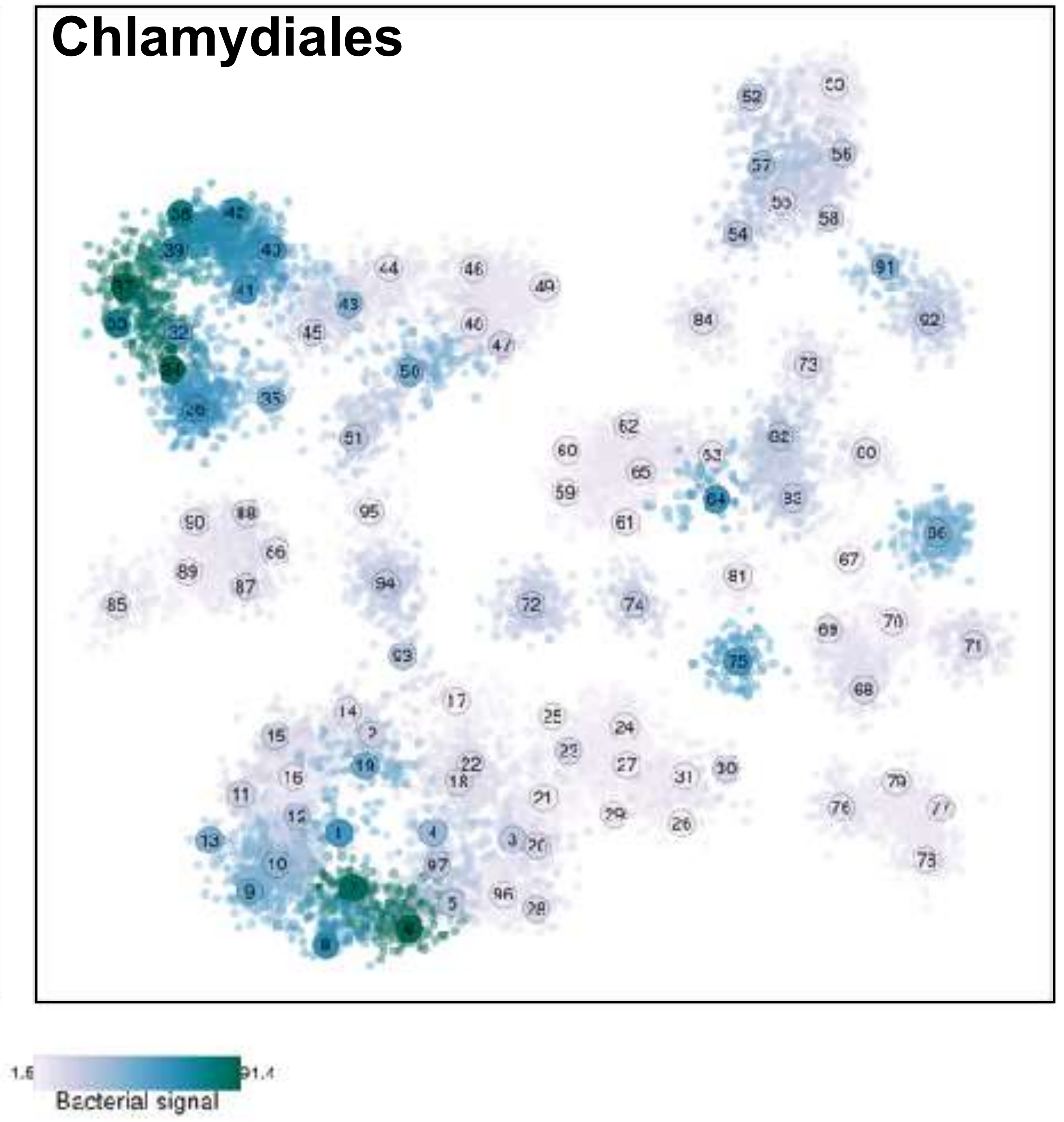
Xenoturbella cell atlas of endosymbiotic interactions

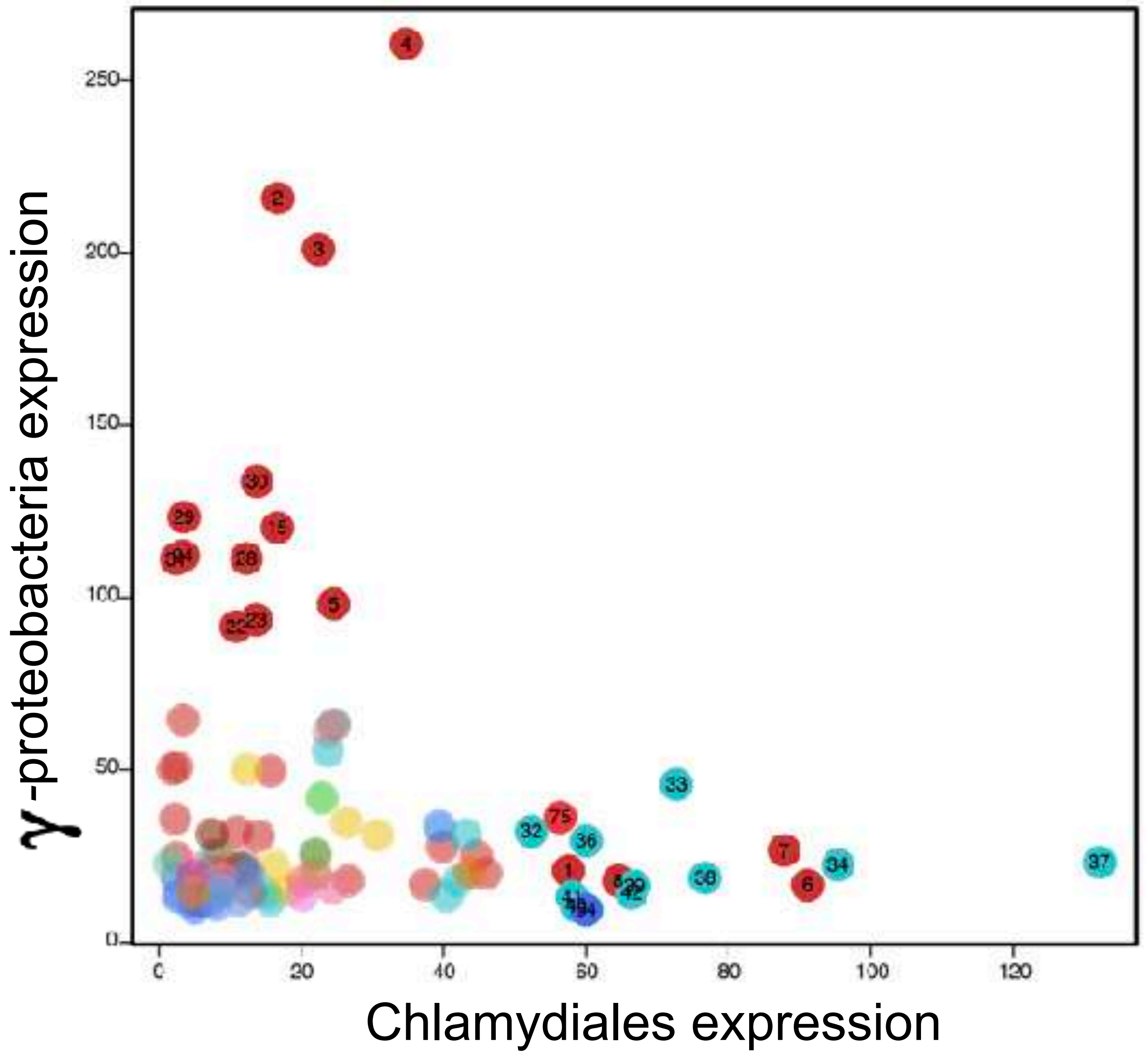


proteobacteria 16S FISH



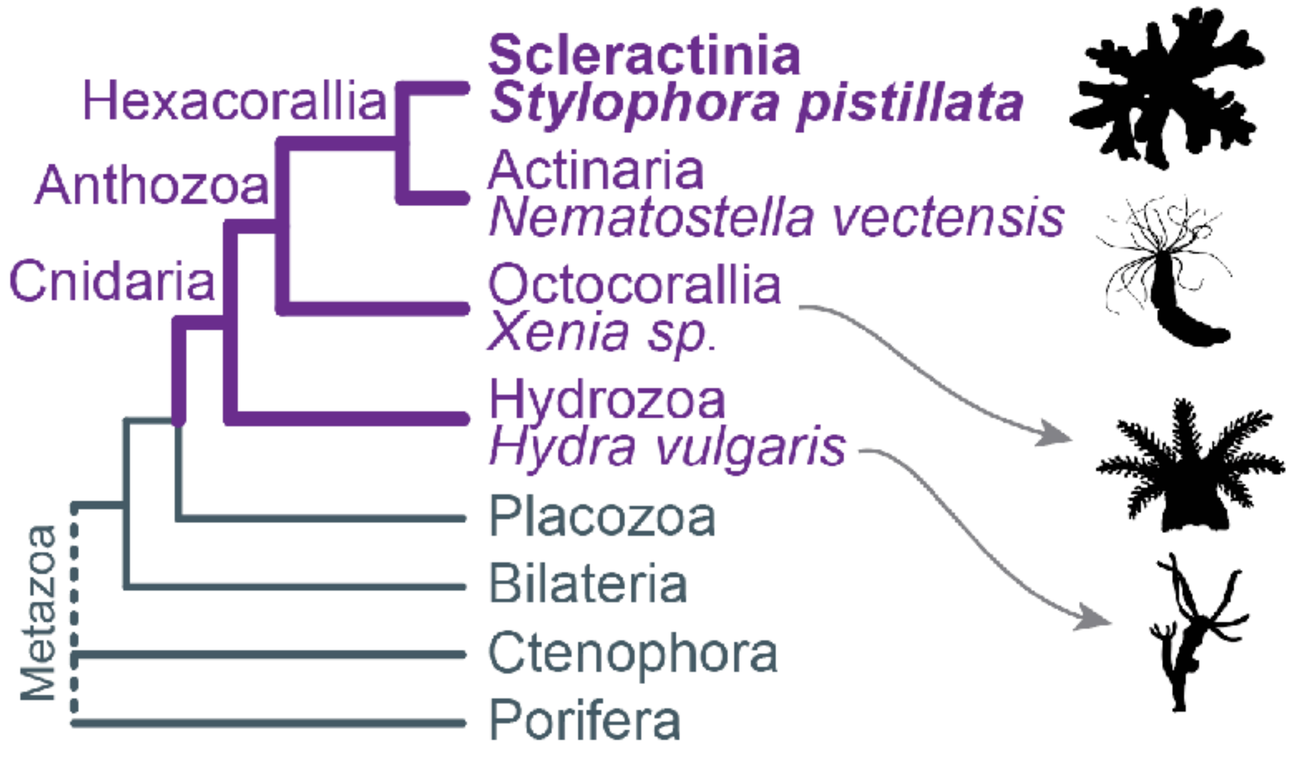


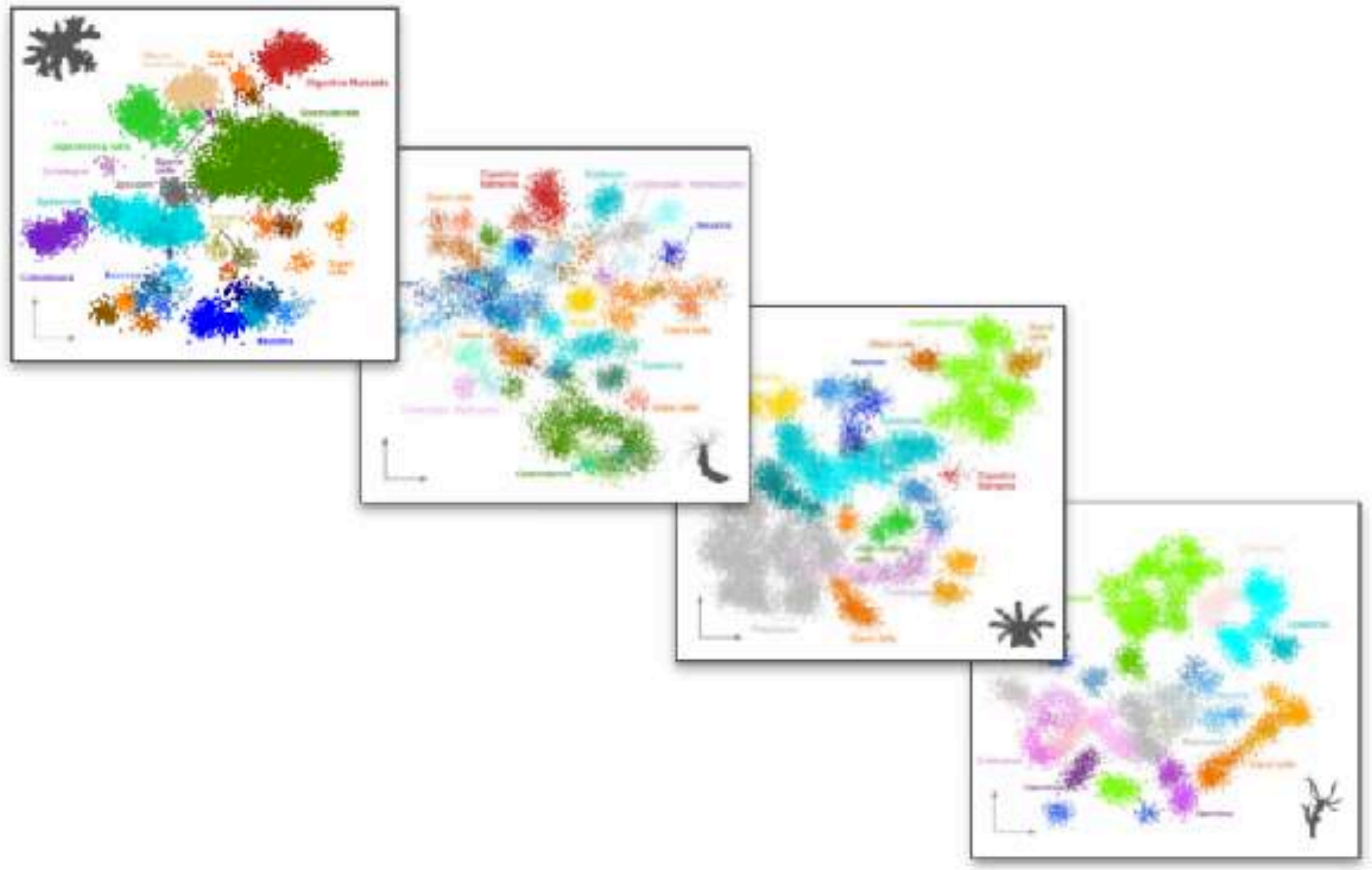


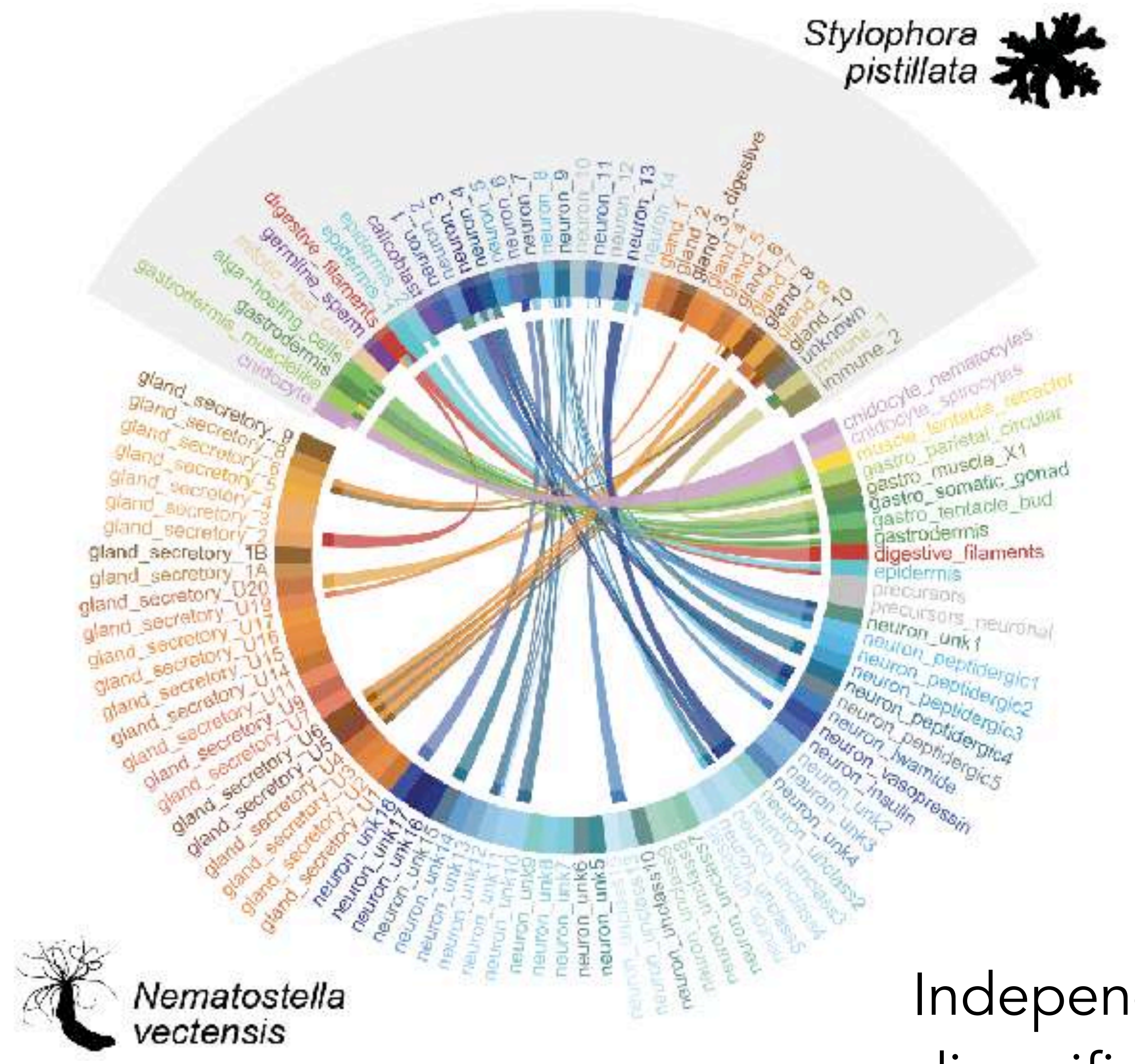


Robertson et al. *Nat Comm* 2024

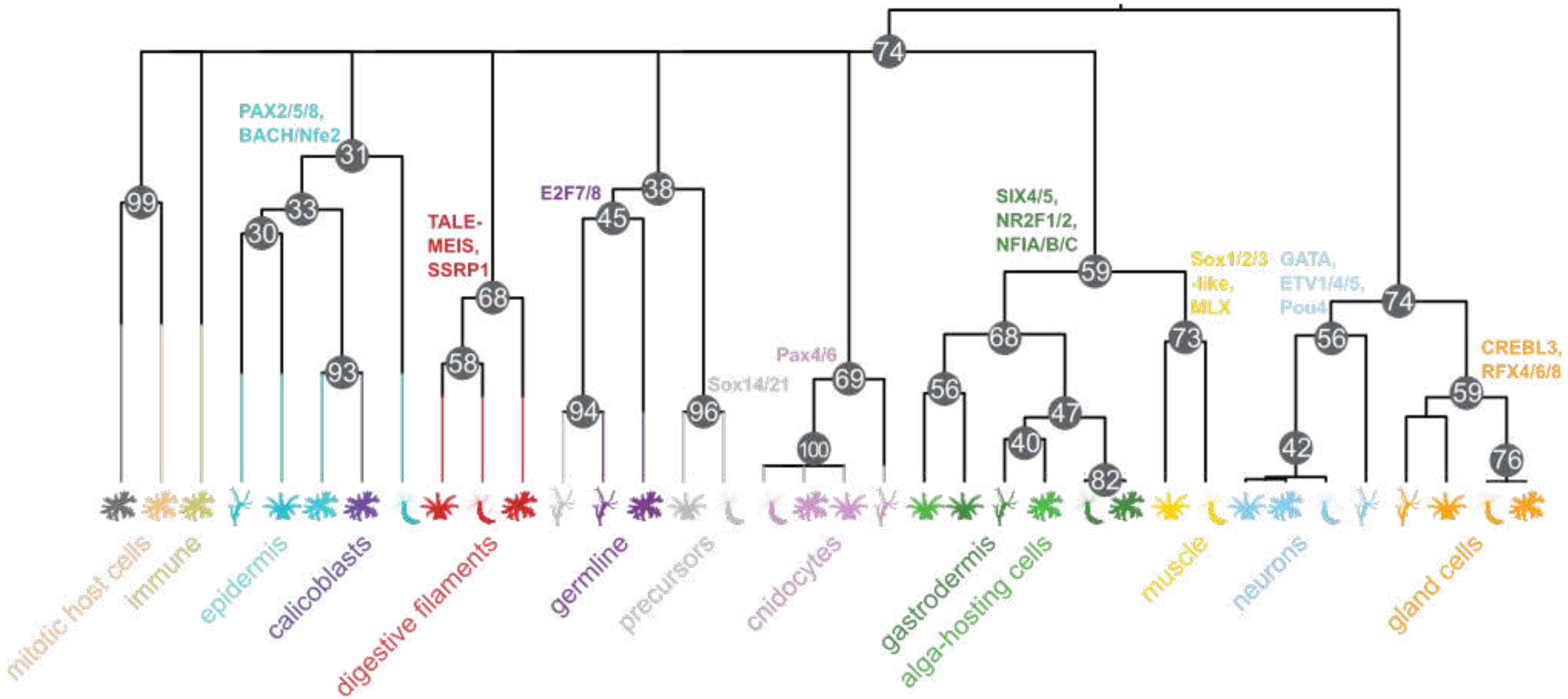
Cnidarian cell type evolution





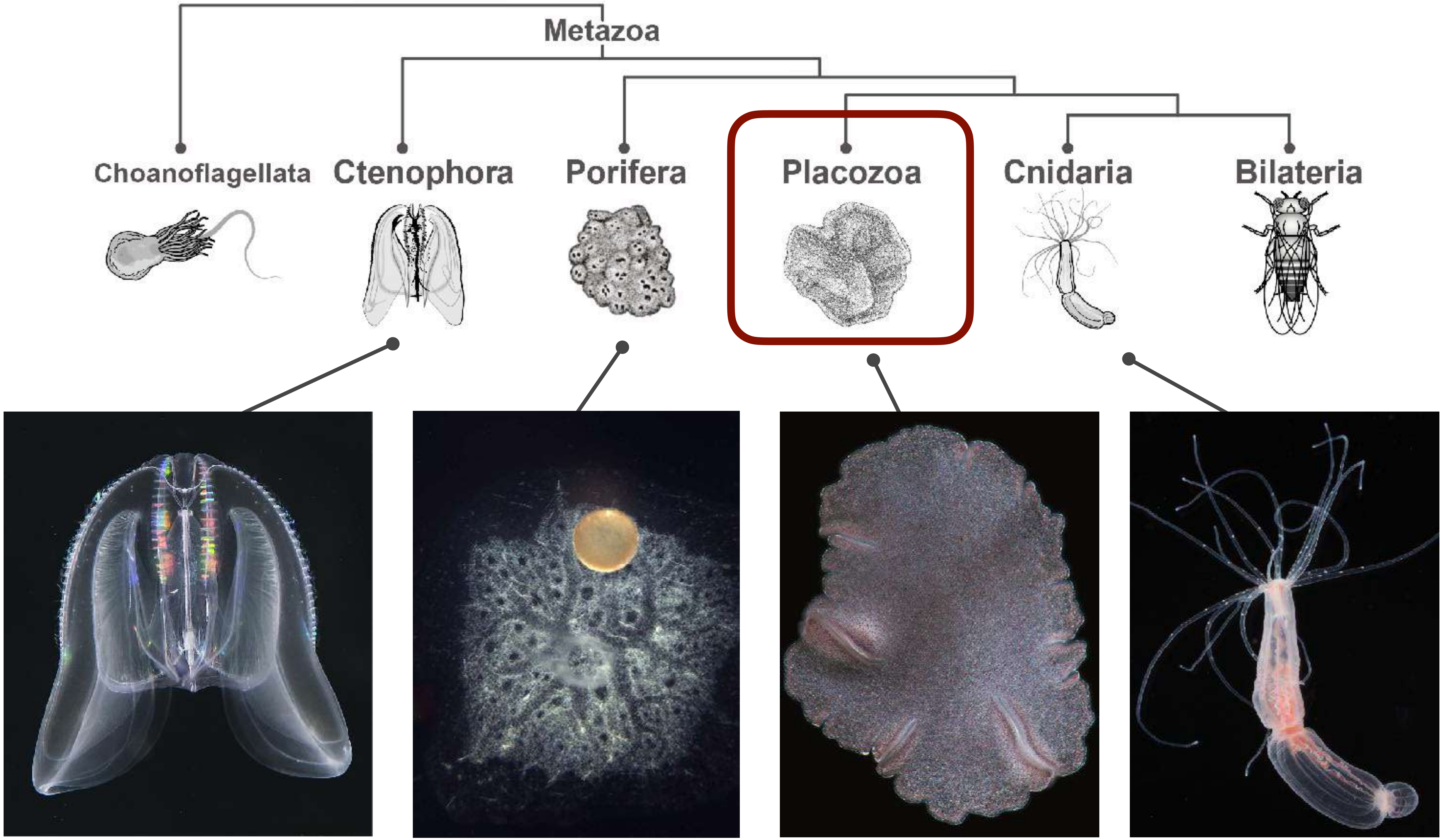


Independent diversification of neurosecretory cell types



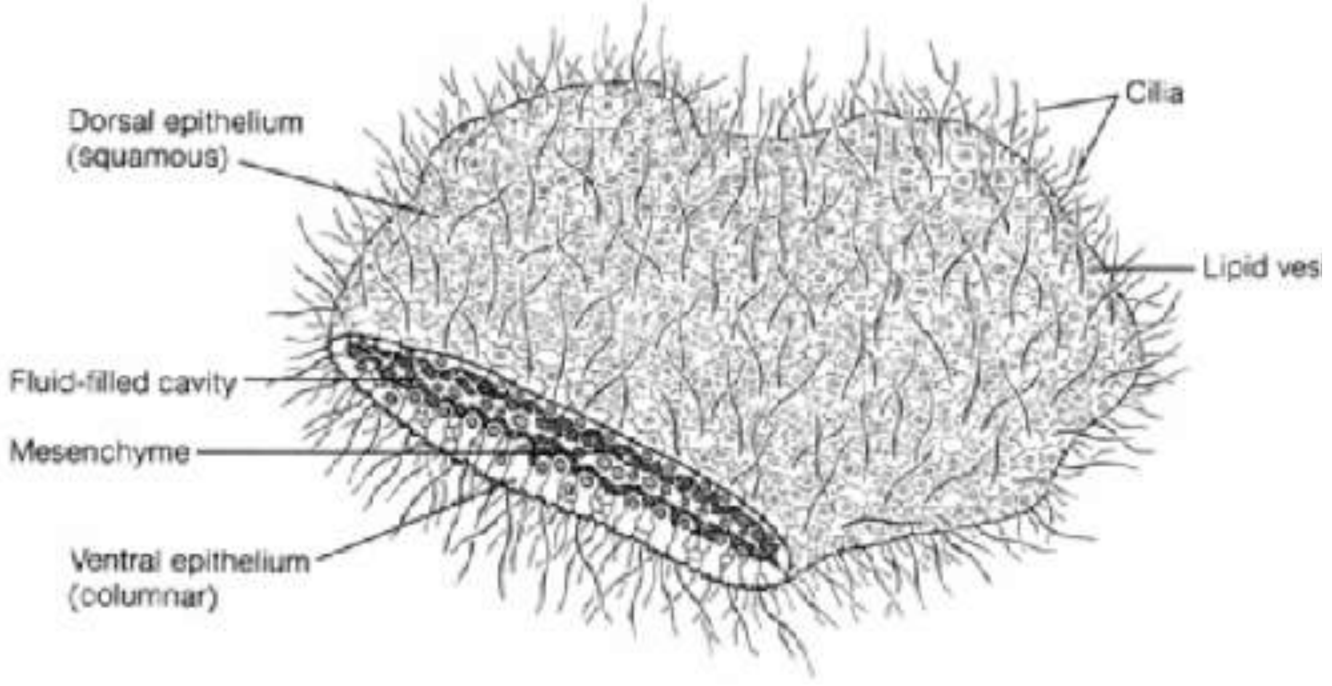
Conserved major cell types across 550 million years of evolution

Story 2: The evolution of the neuronal gene expression program



Phylogenetic framework: **placozoans**

Simple bodyplan and six/nine cell types



Eibildung und Furchung
von *Trichoplax adhaerens* F. E. Schulze (Placozoa)

Karl G. Grell
Zoologisches Institut der Universität Tübingen

Eingegangen am 2. Juli 1972

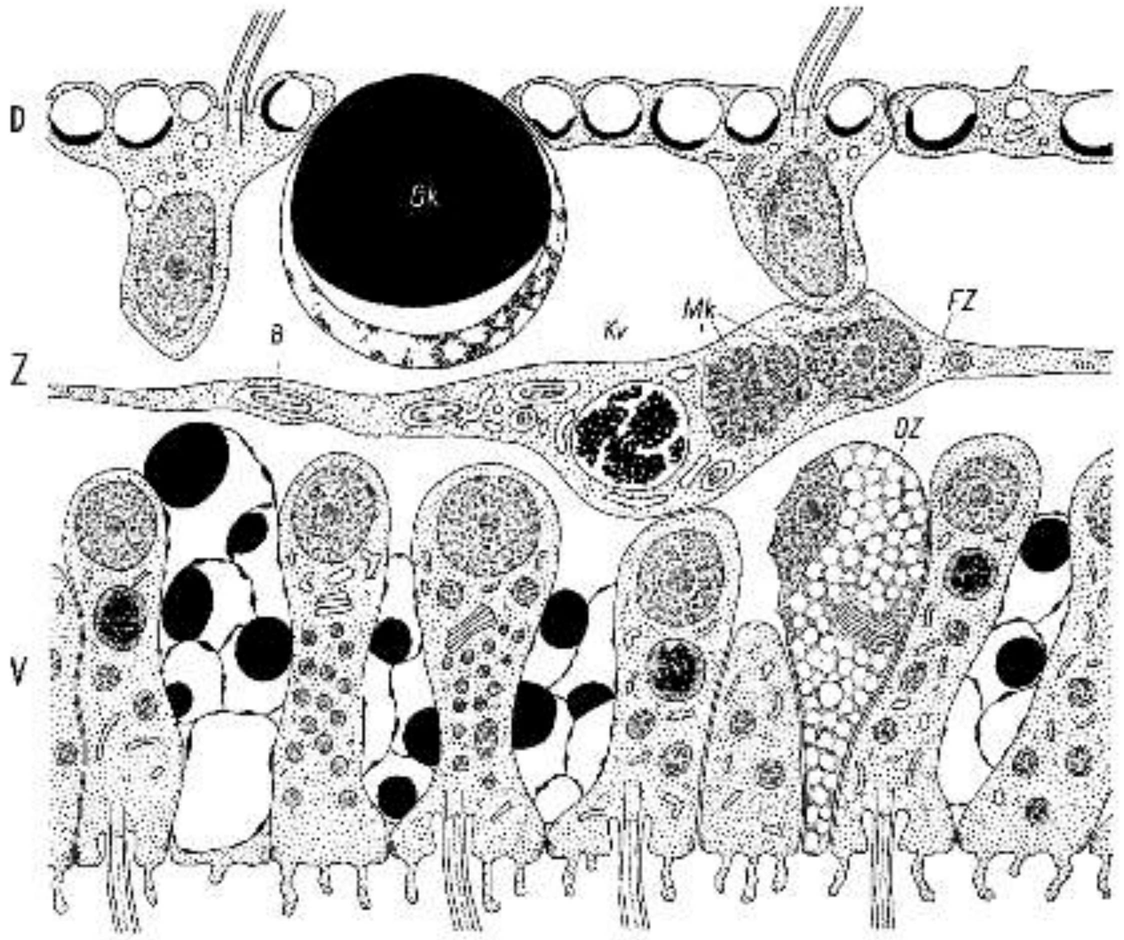
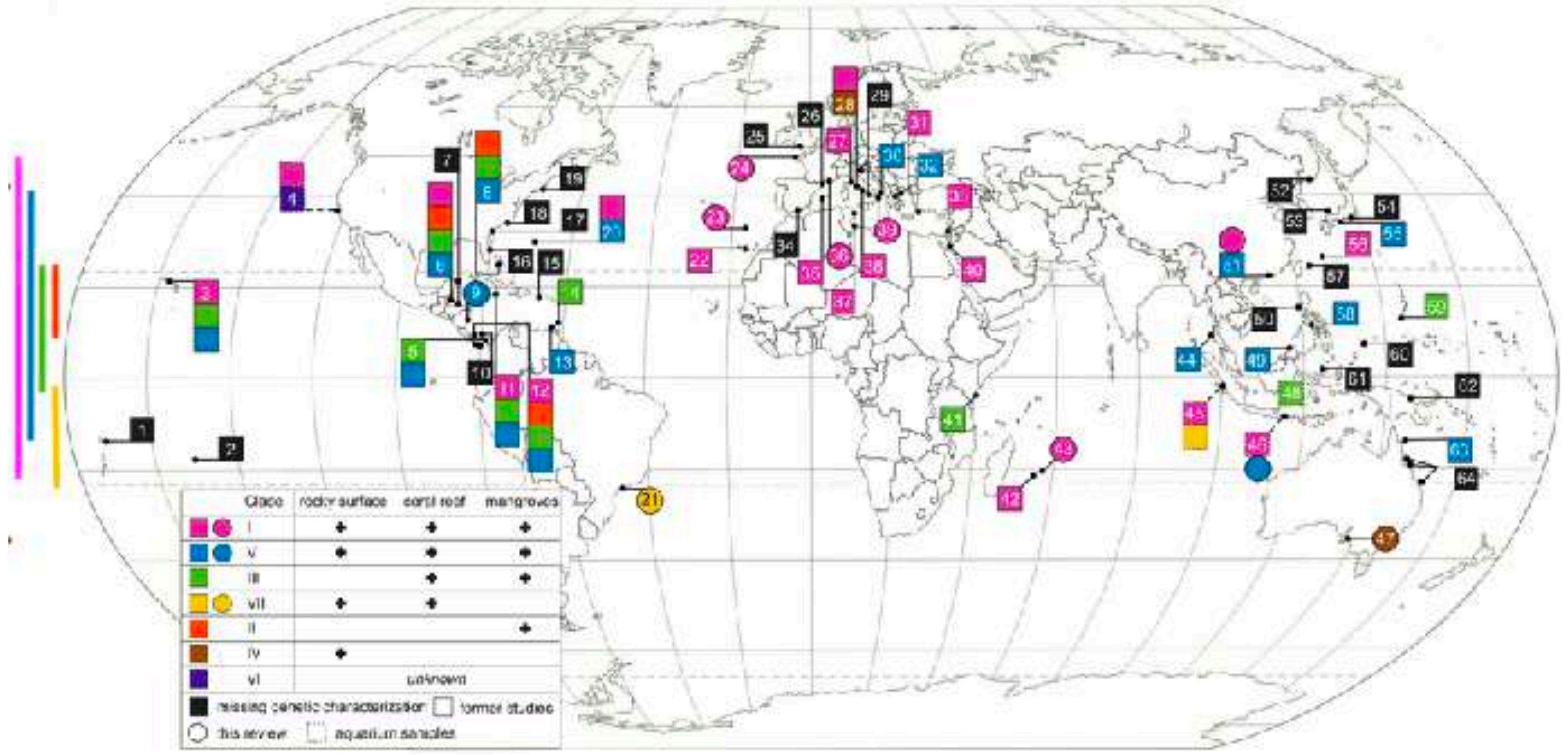


Abb. 1. *Trichoplax adhaerens*. Schema des histologischen Aufbaues. *D* Dorsalepithel. *Z* Zwischenschicht. *V* Ventralepithel. *Gk* Glanzkugel. *FZ* Faserzelle mit Mitochondrienkomplex (*Mk*), Konkrementvakuole (*Kv*) und Bakterien (*B*). *DZ* Drüsenzelle

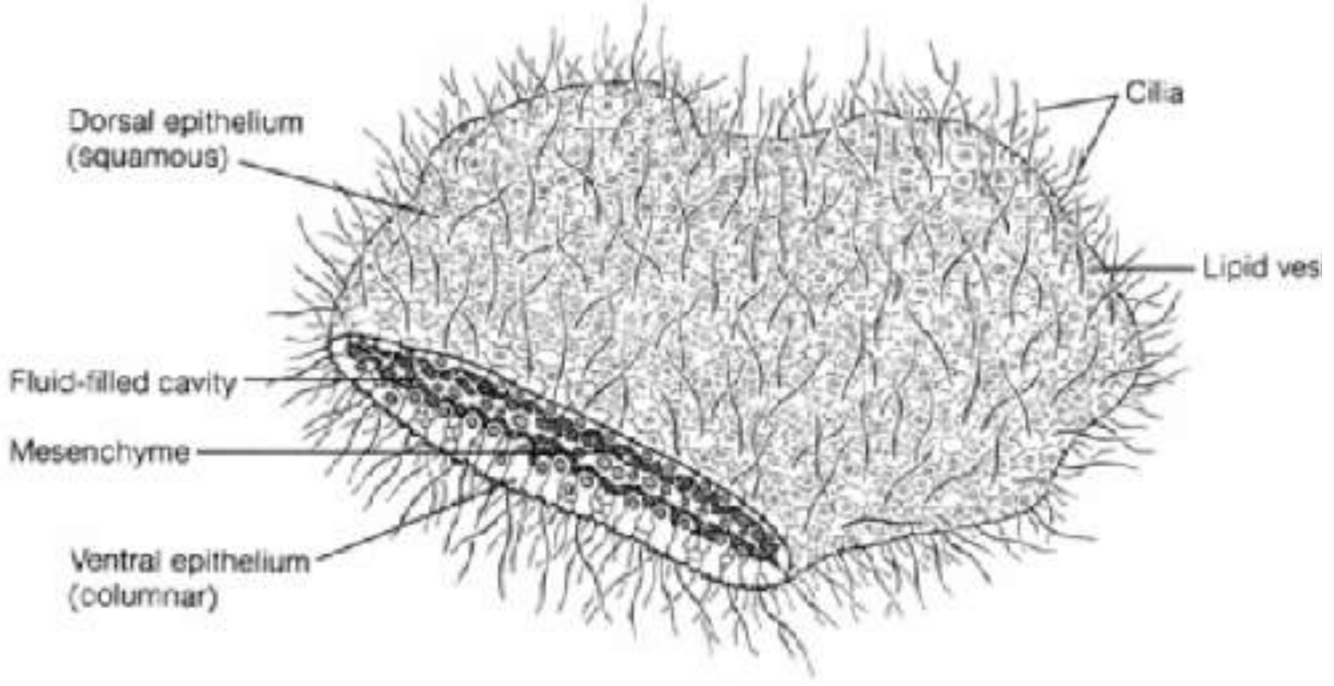
Biogeography - tropical and subtropical seas



Eitel et al., *PLOS One*, 2013

Phylogenetic framework: **placozoans**

Simple bodyplan and six/nine cell types



Eibildung und Furchung
von *Trichoplax adhaerens* F. E. Schulze (Placozoa)

Karl G. Grell
Zoologisches Institut der Universität Tübingen

Eingegangen am 2. Juli 1972

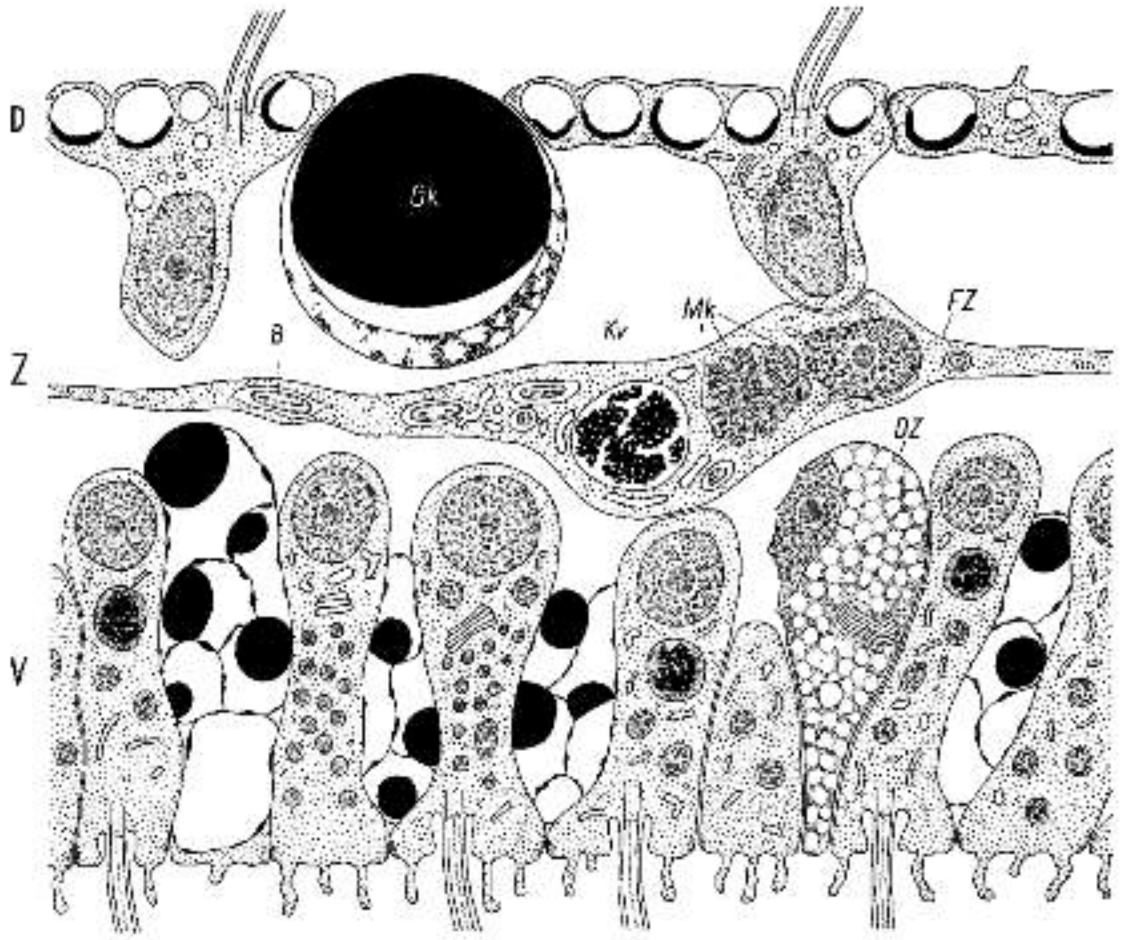
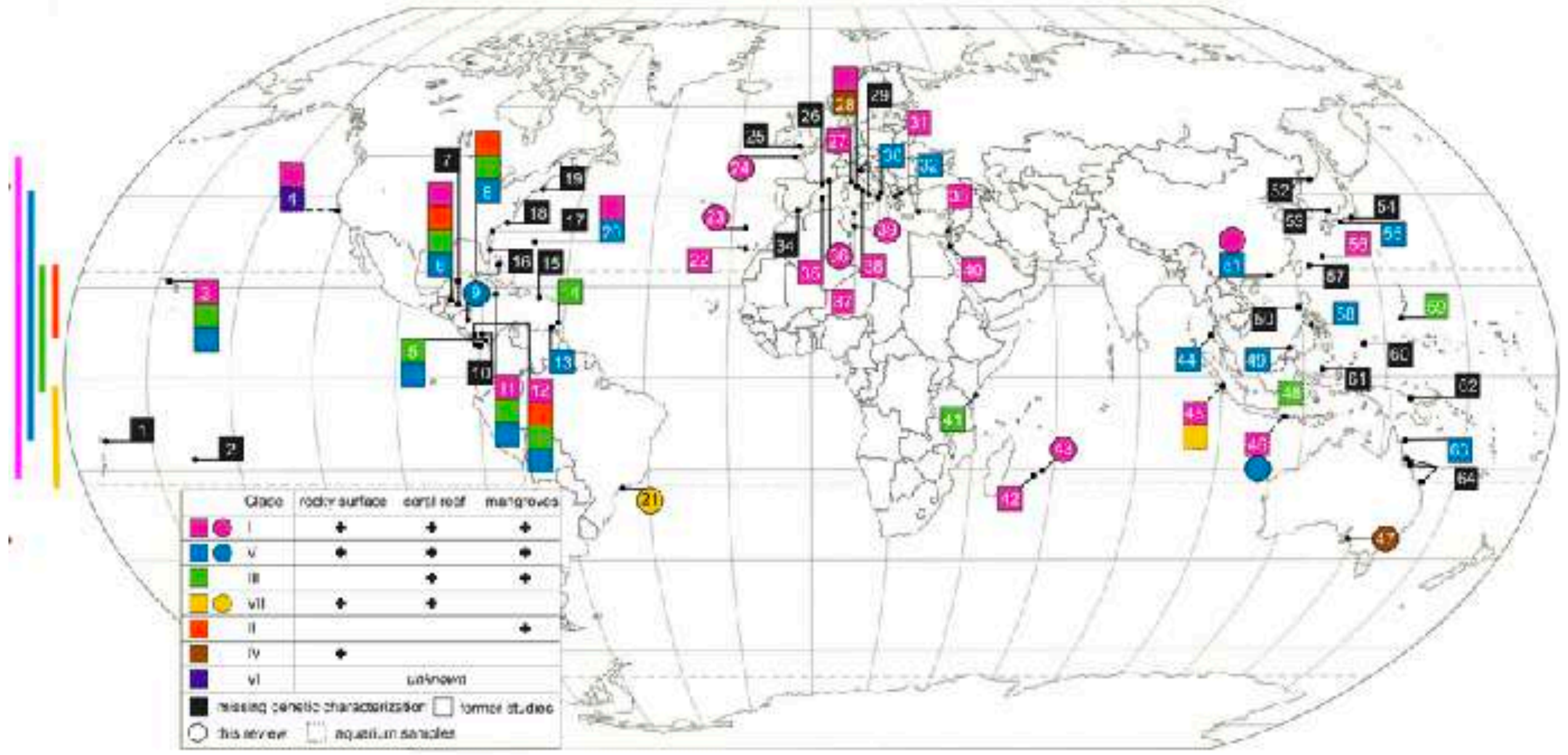


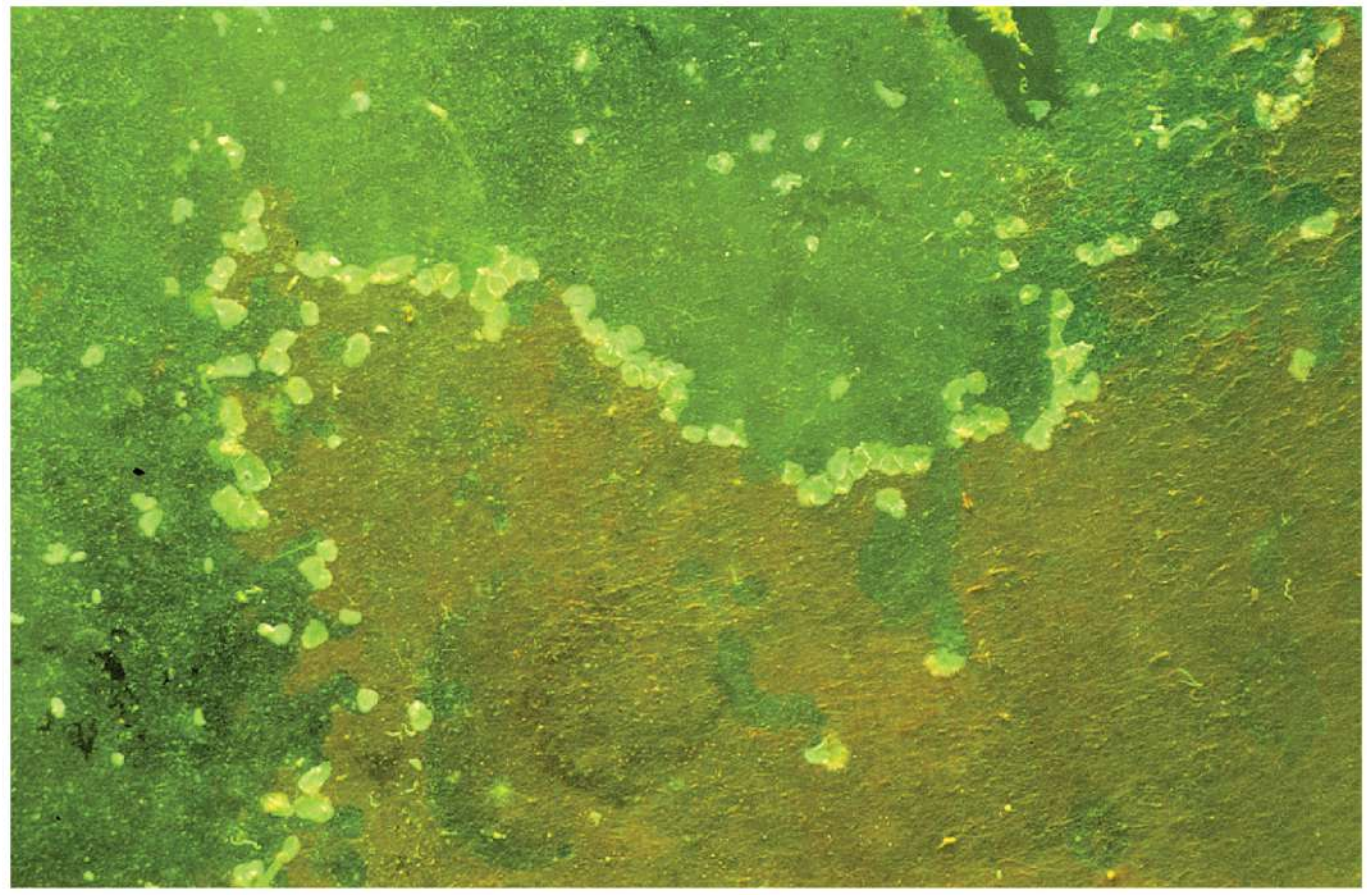
Abb. 1. *Trichoplax adhaerens*. Schema des histologischen Aufbaues. *D* Dorsalepithel. *Z* Zwischenschicht. *V* Ventralepithel. *Gk* Glanzkugel. *FZ* Faserzelle mit Mitochondrienkomplex (*Mk*), Konkrementvakuole (*Kv*) und Bakterien (*B*). *DZ* Drüsenzelle

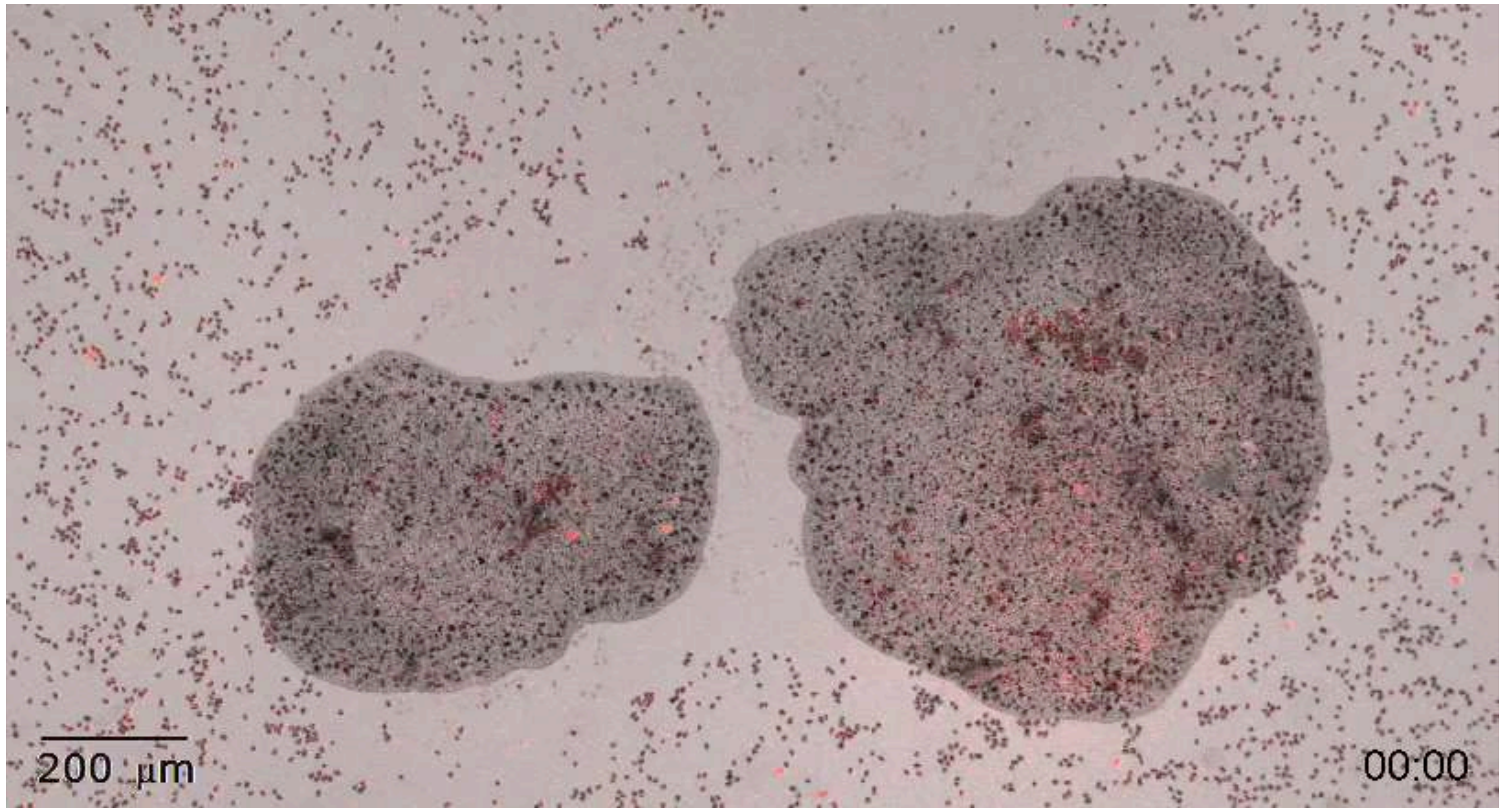
Biogeography - tropical and subtropical seas



Eitel et al., *PLOS One*, 2013

Habitat - microbial mats, feeding by extracellular digestion

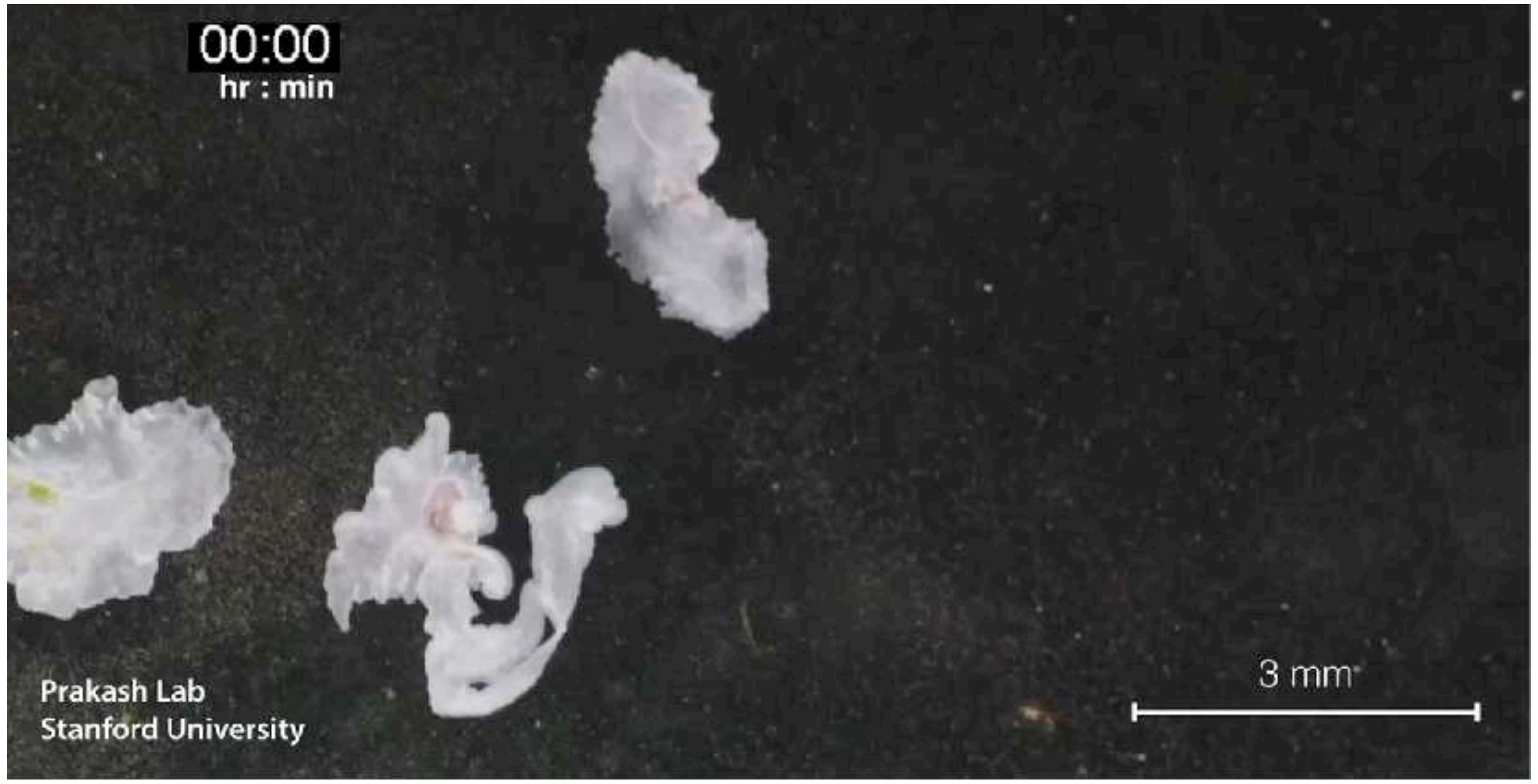




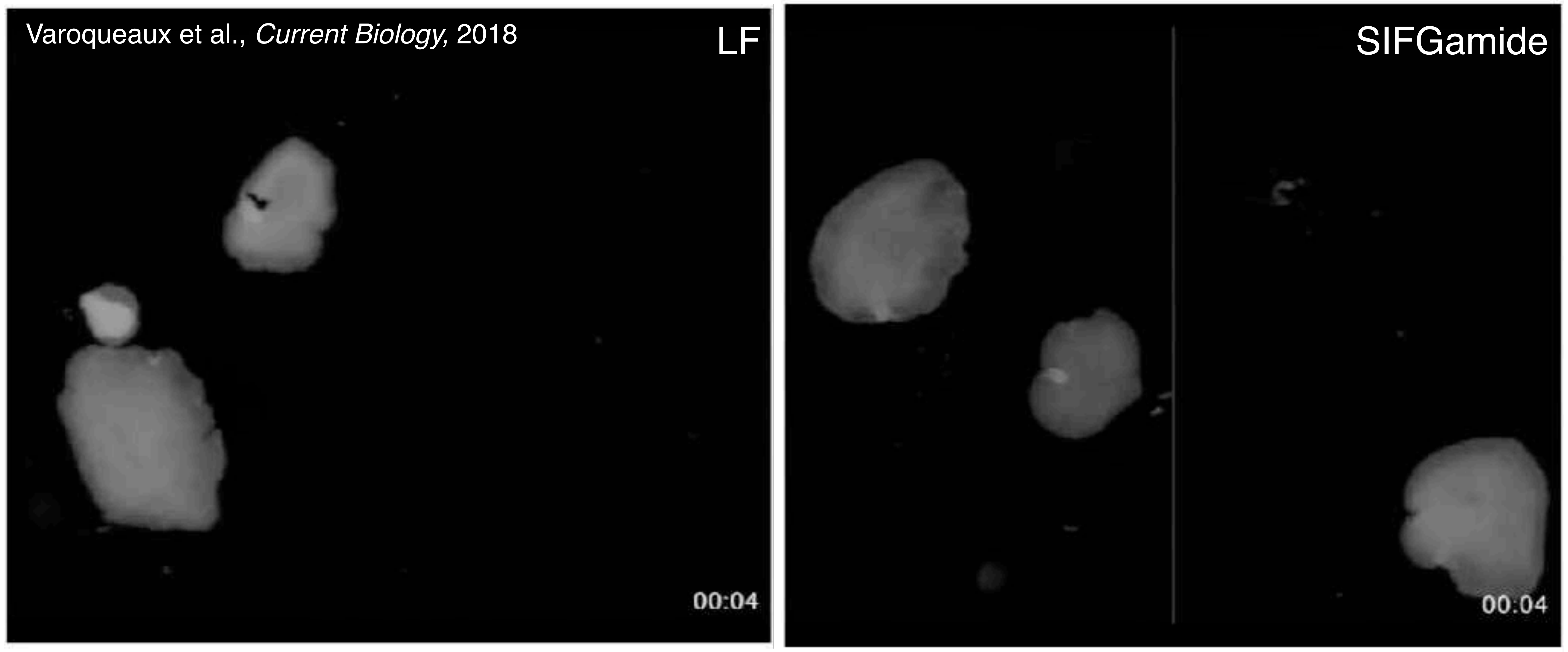
Senatore et al., *The Journal of Experimental Biology*, 2017

Phylogenetic framework: **placozoans**

Asexual reproduction by fission

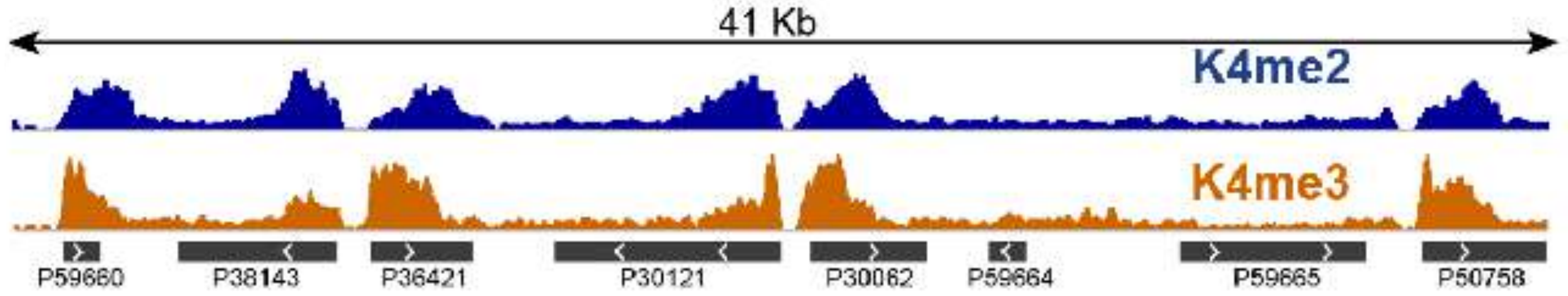


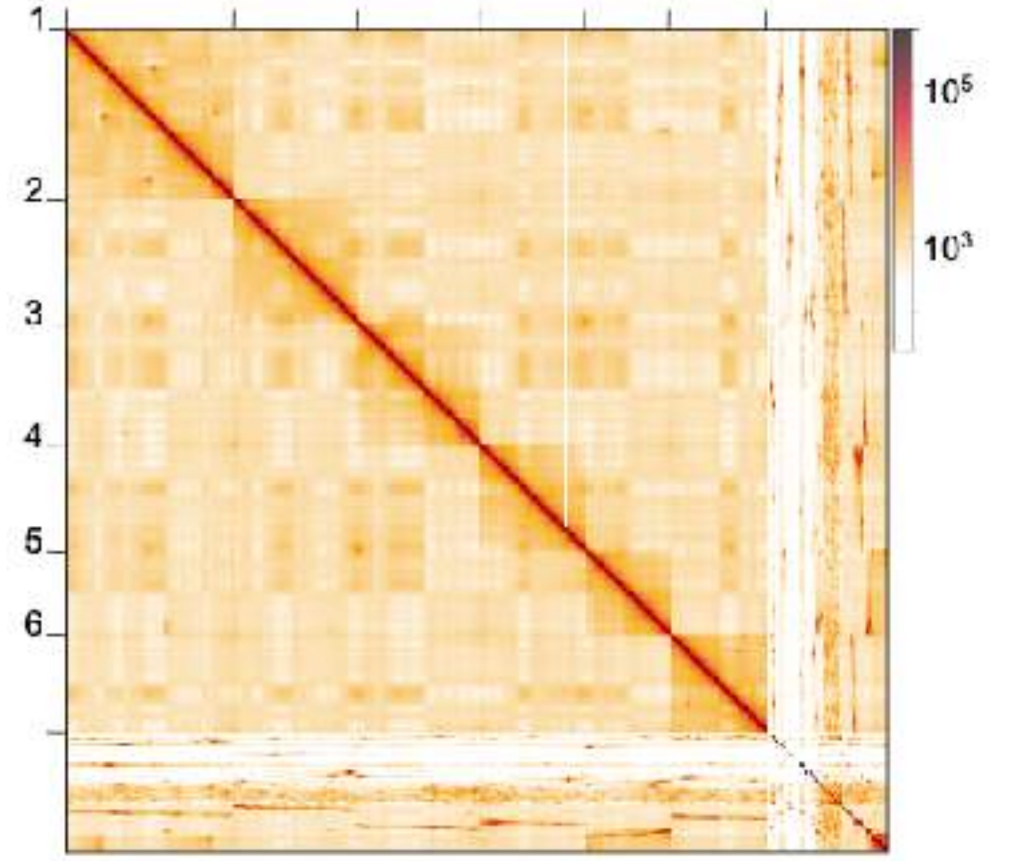
Collective cell behaviors controlled by small peptides

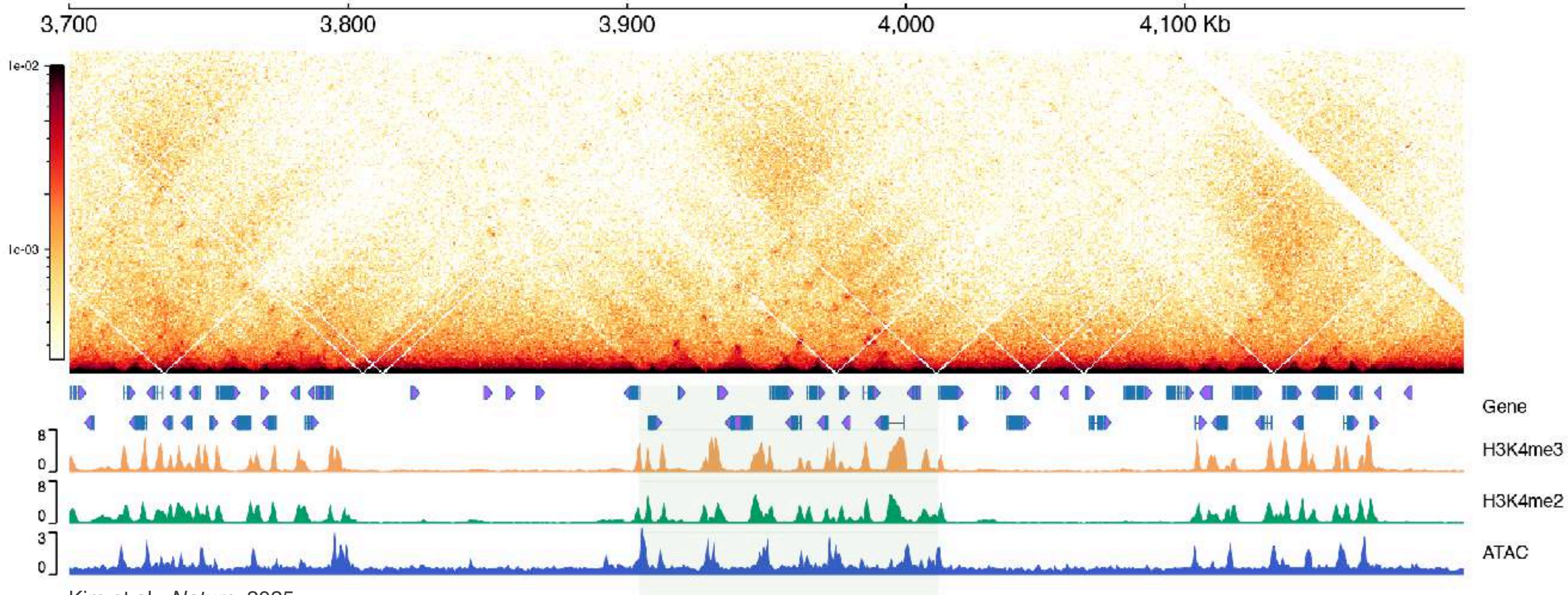


Phylogenetic framework: Placozoa genomes

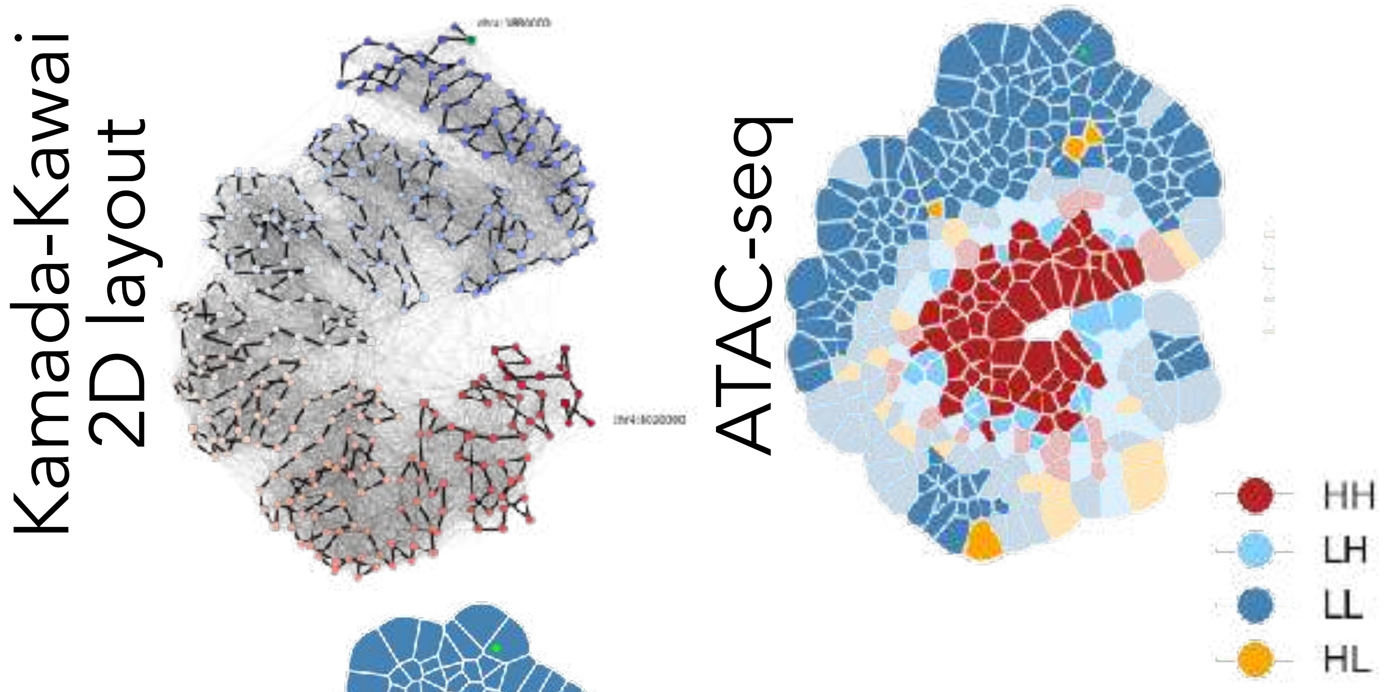
- *Trichoplax adhaerens* (H1) in 2008 + 6 others in recent years
- 87-108Mb
- 6 chromosomes
- ±12,000 genes
- highly-conserved gene repertoire
- proximal promoter gene regulation

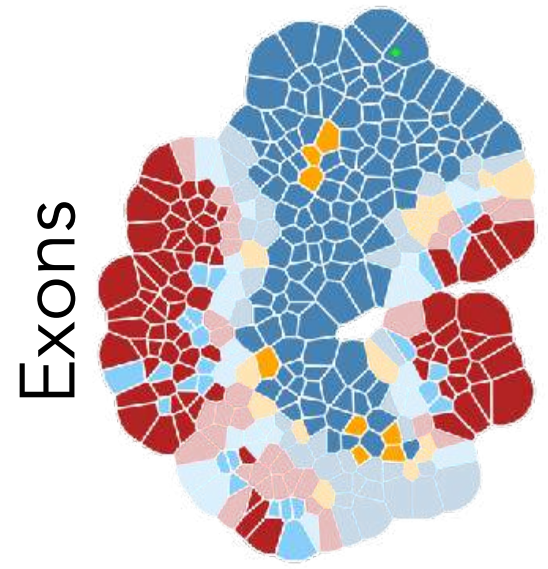






Kim et al., *Nature*, 2025





Iana Kim

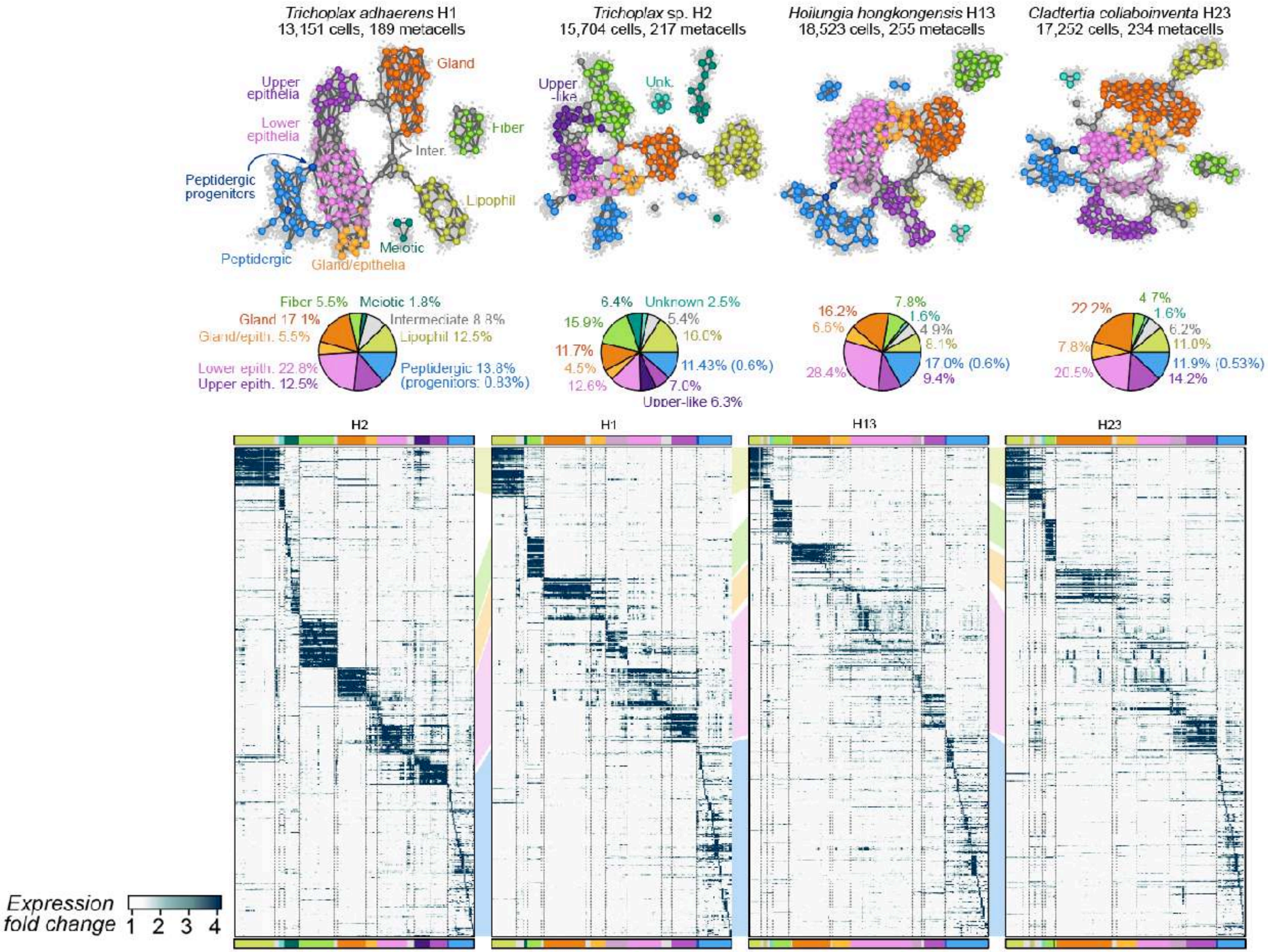
A multi-species **placozoan cell type atlas**

Bernd Schierwater

Harald Gruber-Vodicka

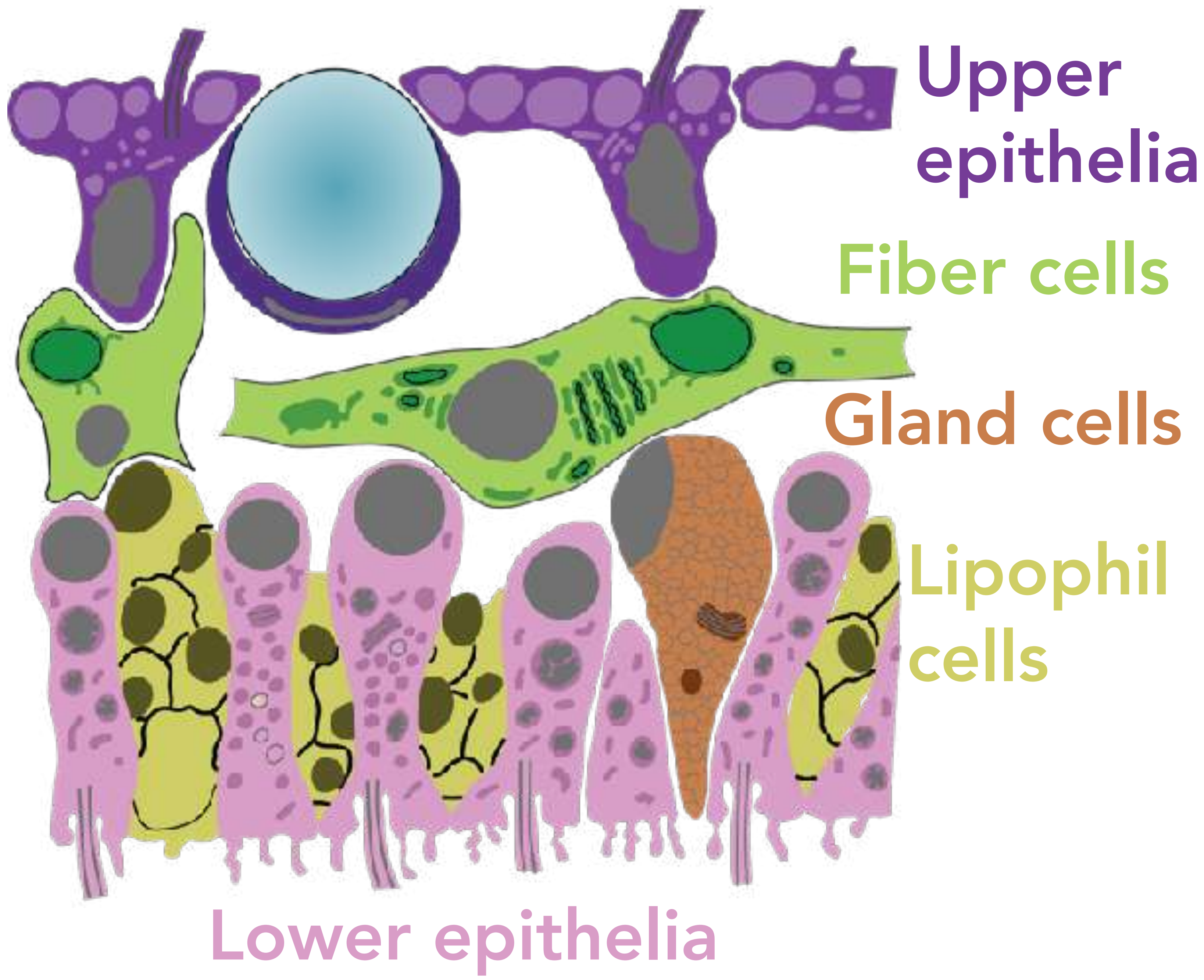
Sebastian Najle

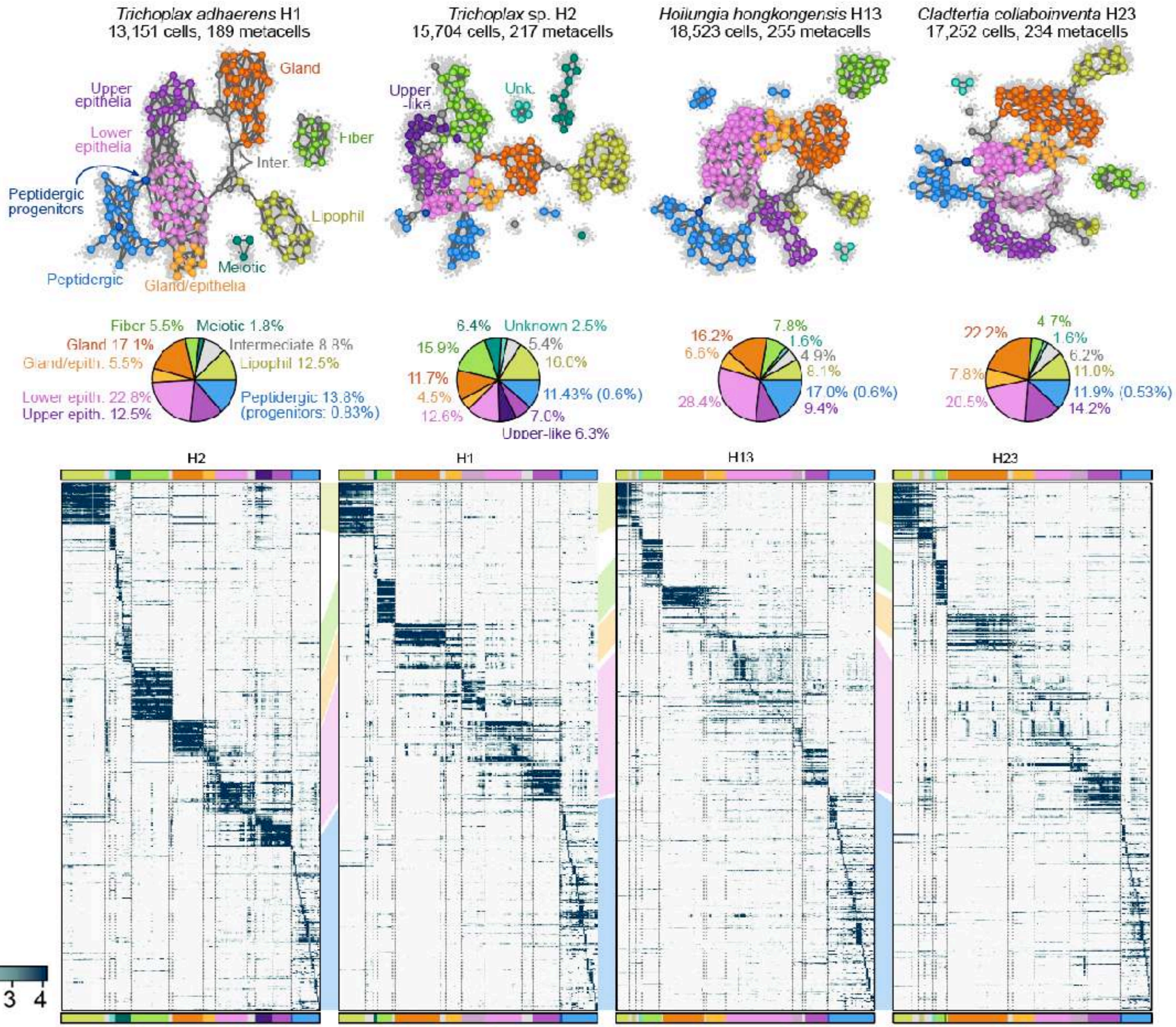
Xavier Grau-Bove



Najle, Grau-Bové et al., *Cell* 2023

A multi-species **placozoan cell type atlas**

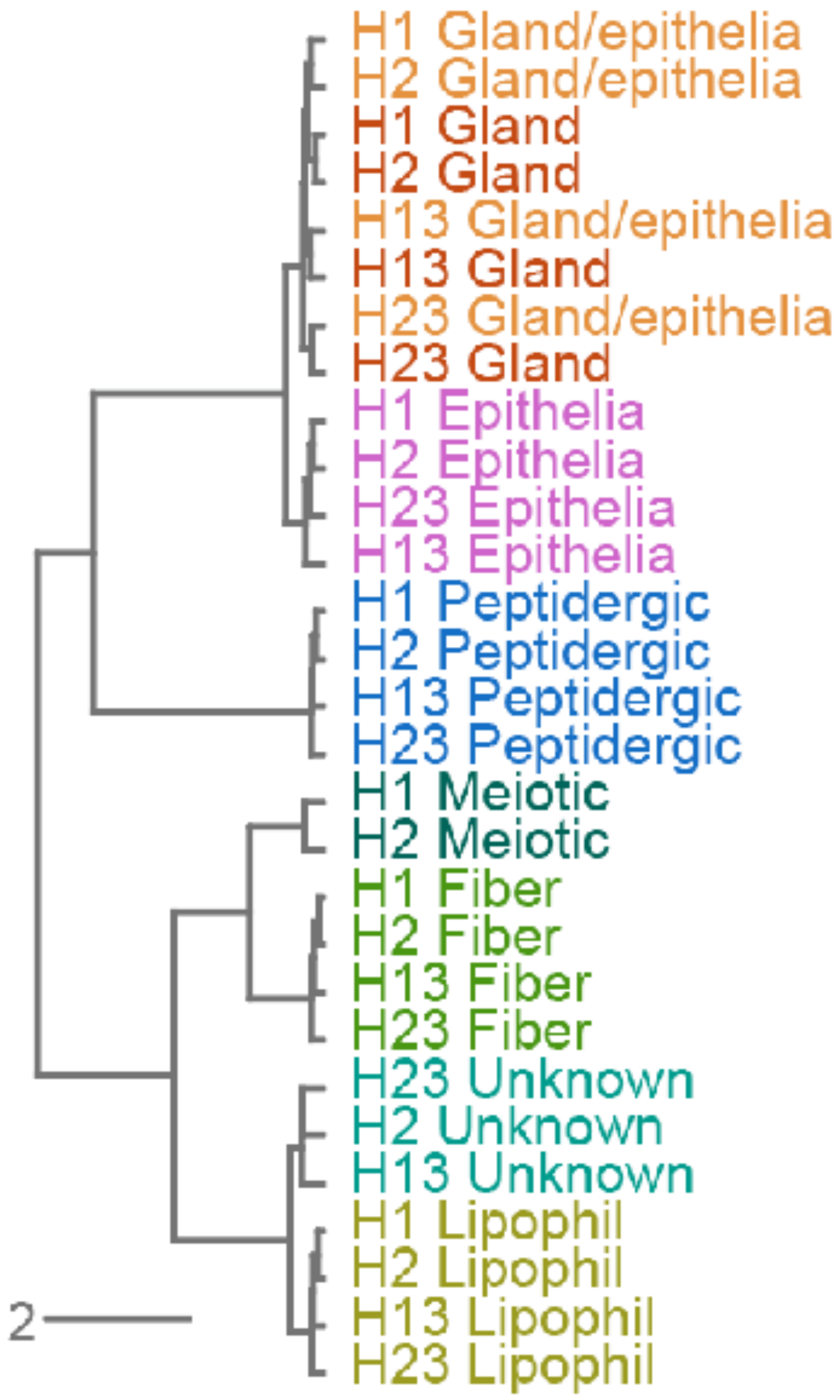


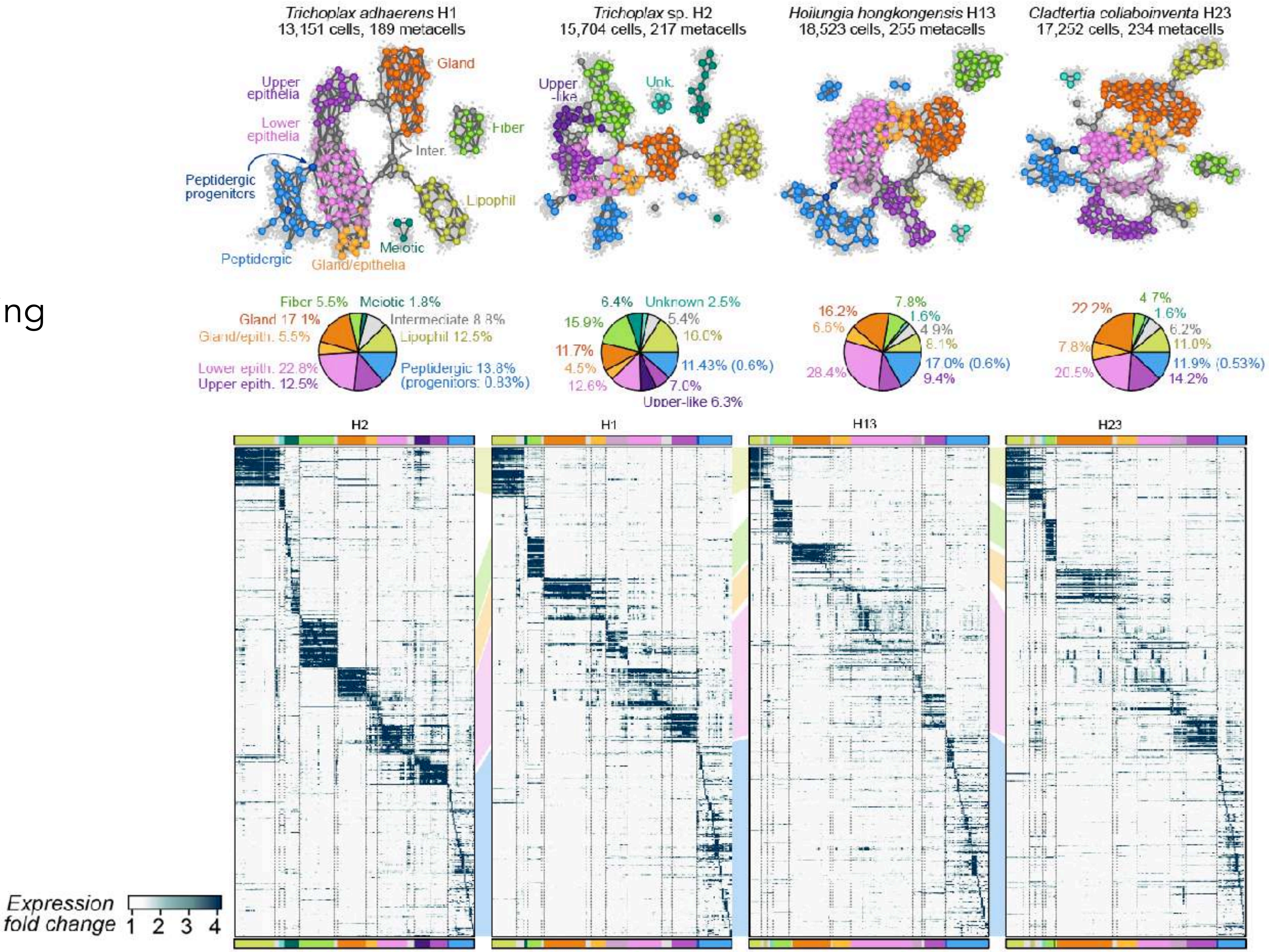


Najle, Grau-Bové et al., *Cell* 2023

Conserved broad cell types across Placozoa

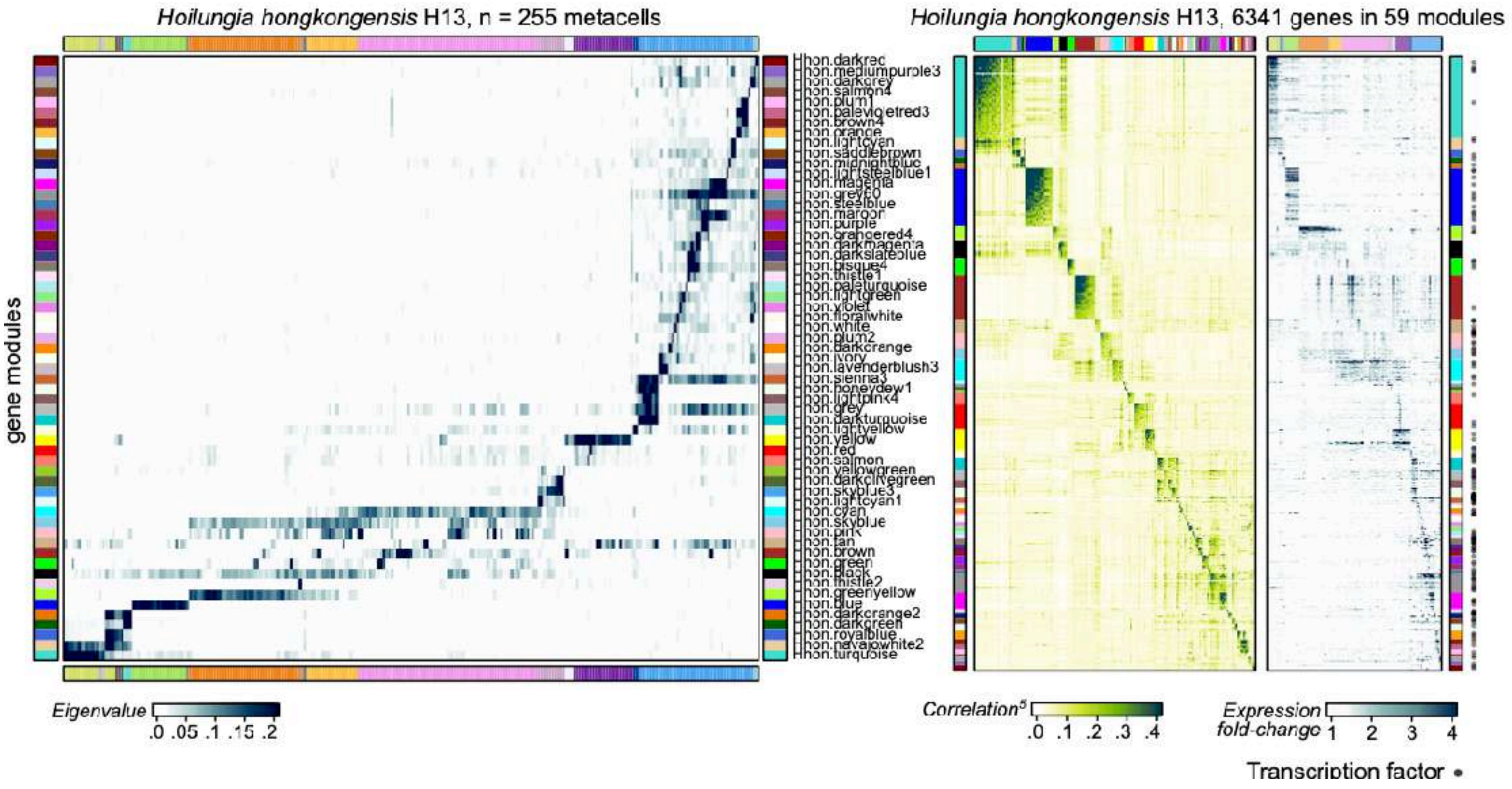
Multi-species cell type clustering



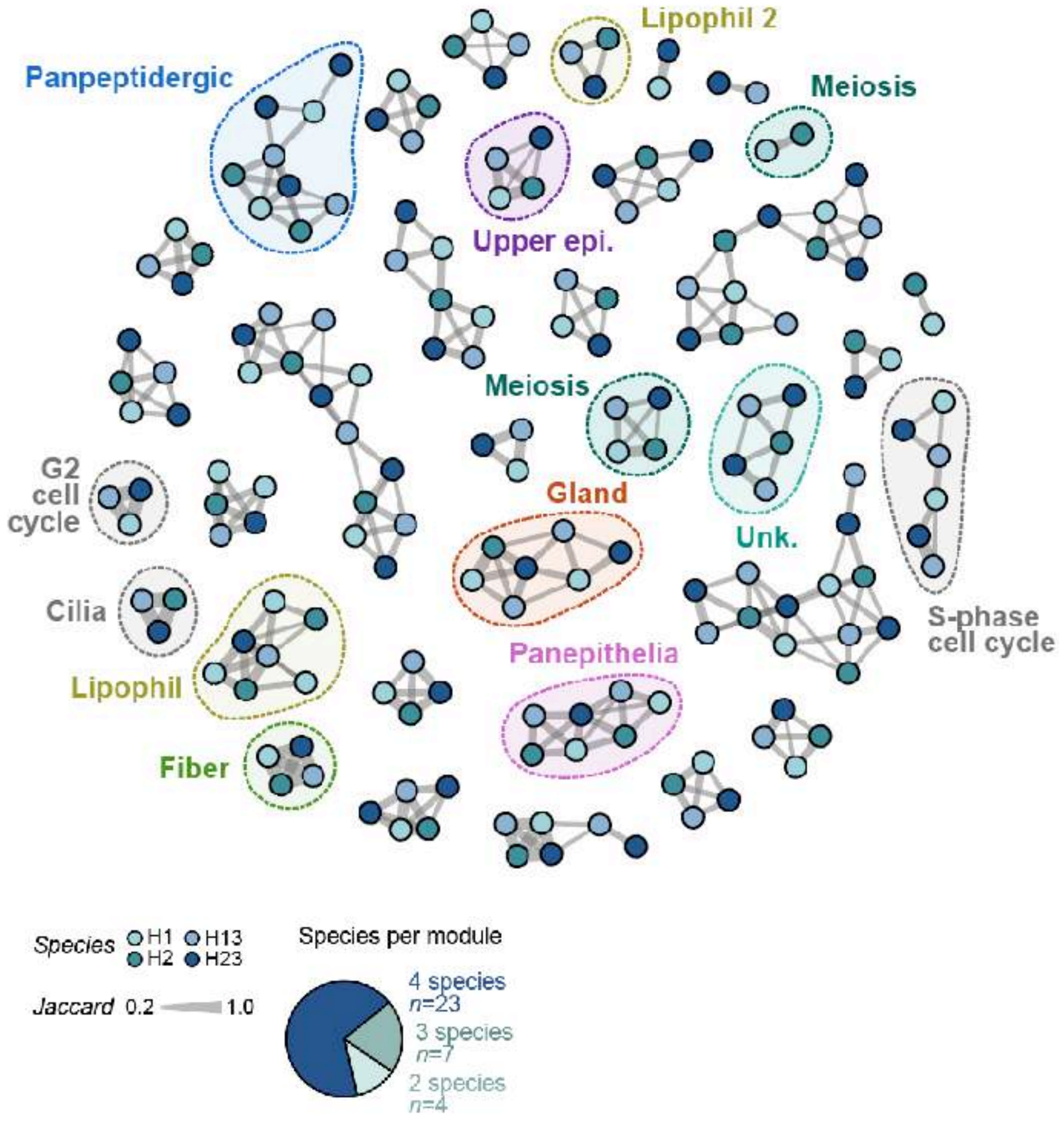


Highly conserved gene modules across Placozoa

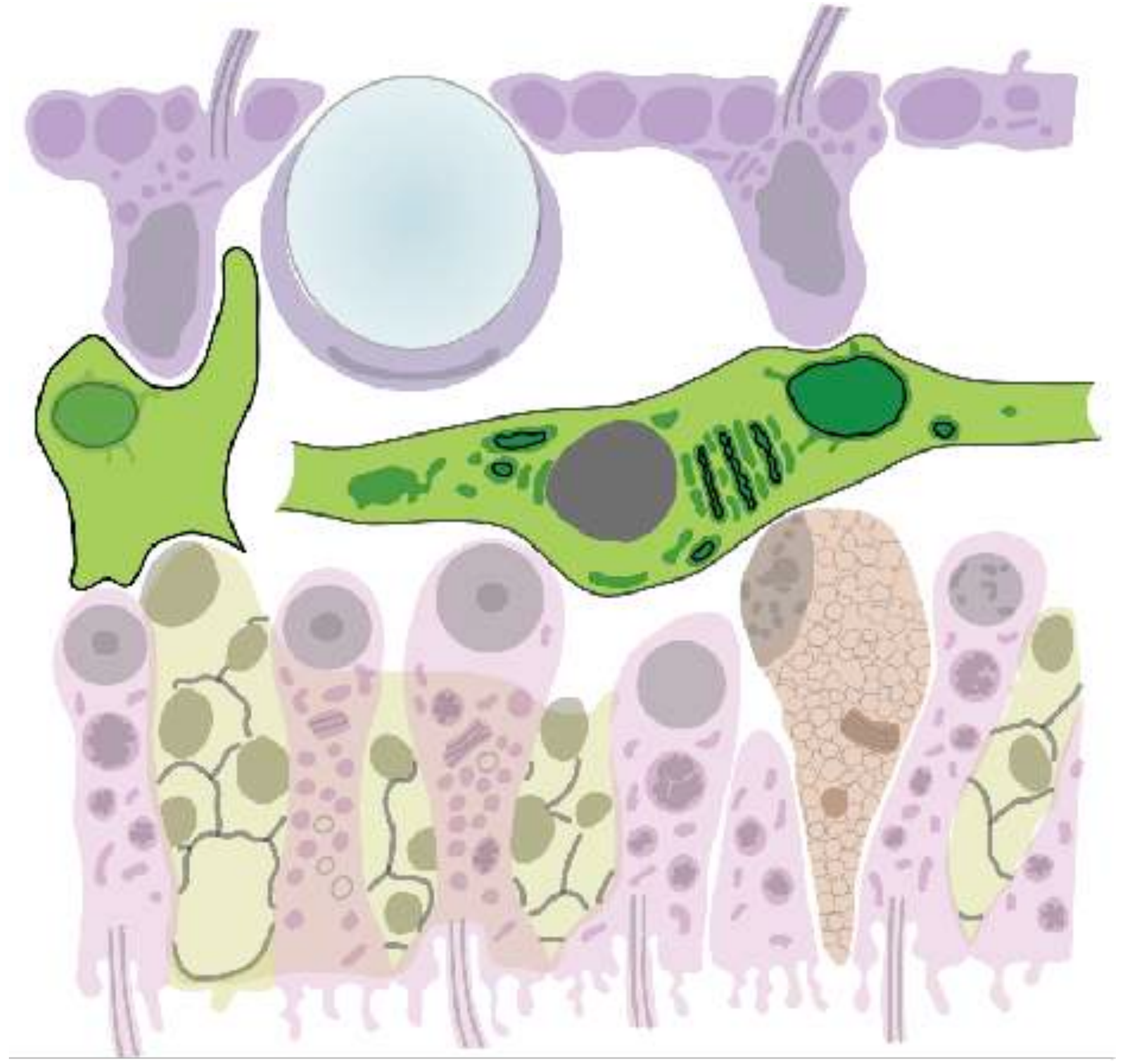
Single-species gene modules
(based on metacell-level gene-gene correlations)



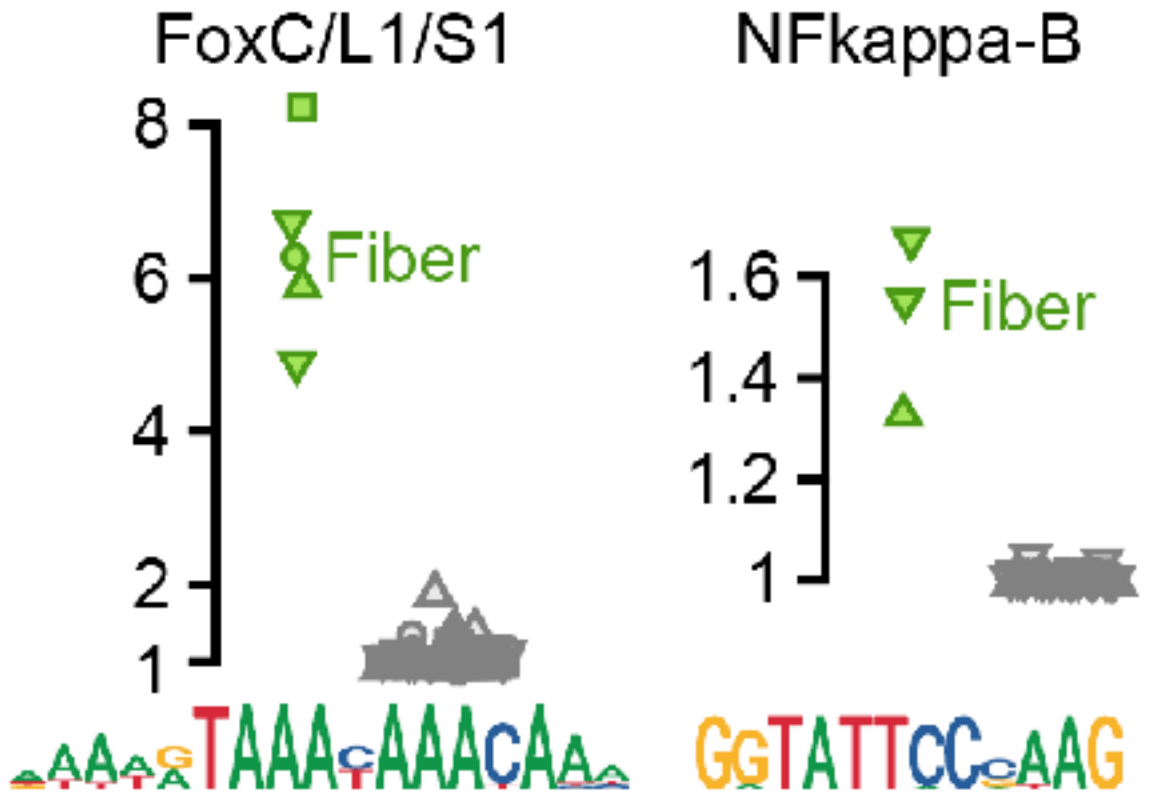
Multi-species gene module clustering



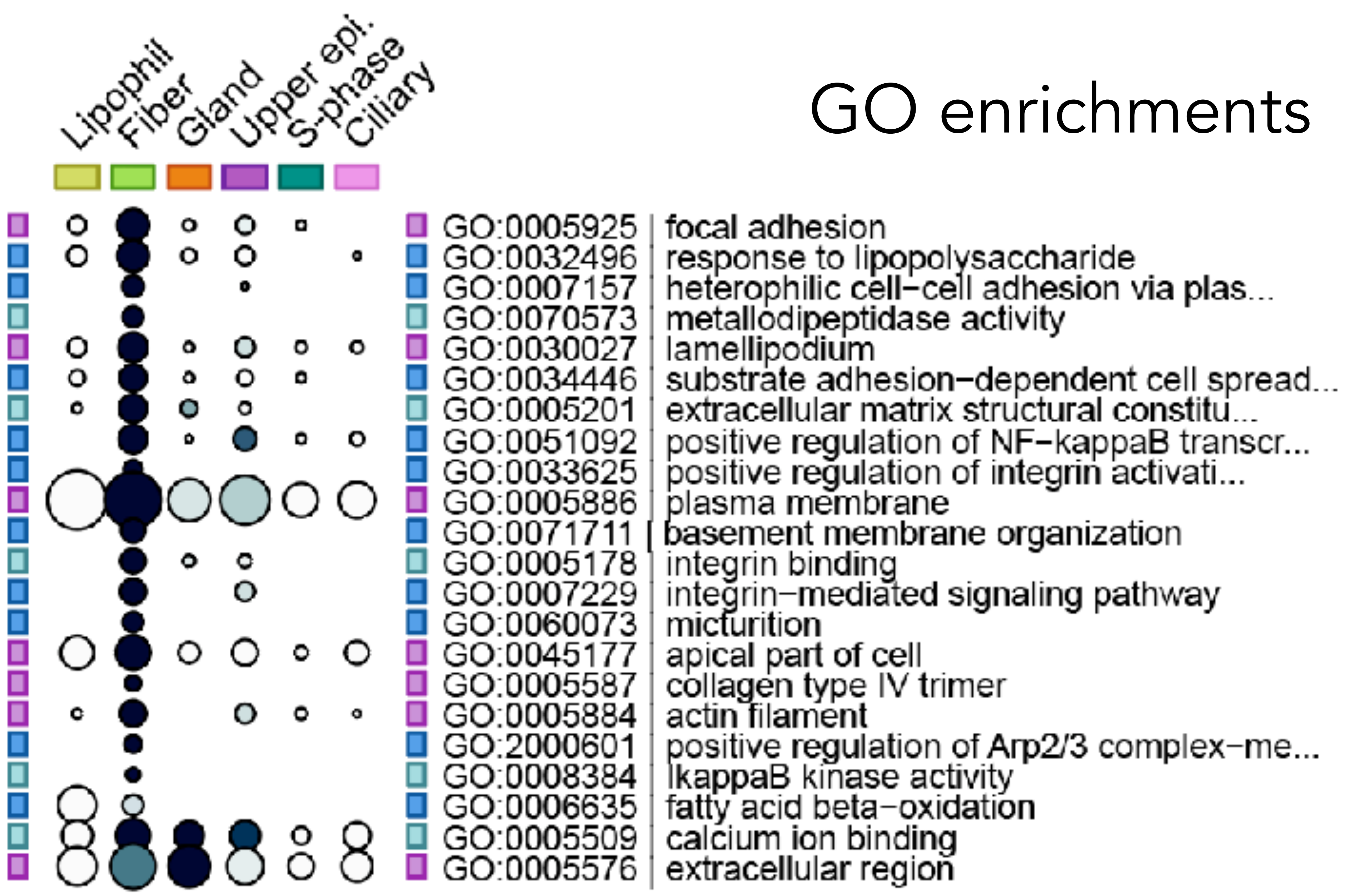
Functional enrichments in cross-species gene modules: fiber cells



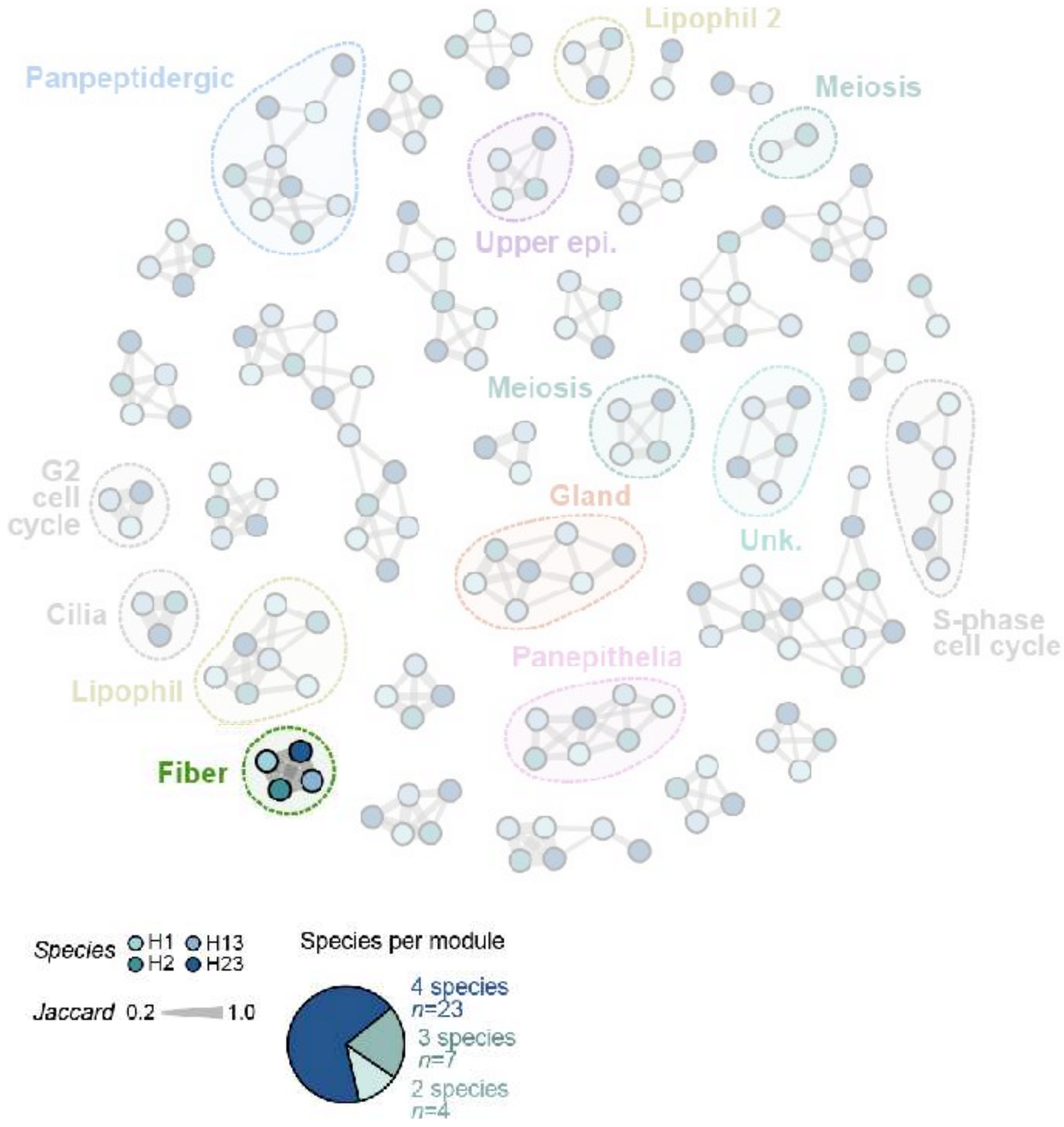
Transcription factors



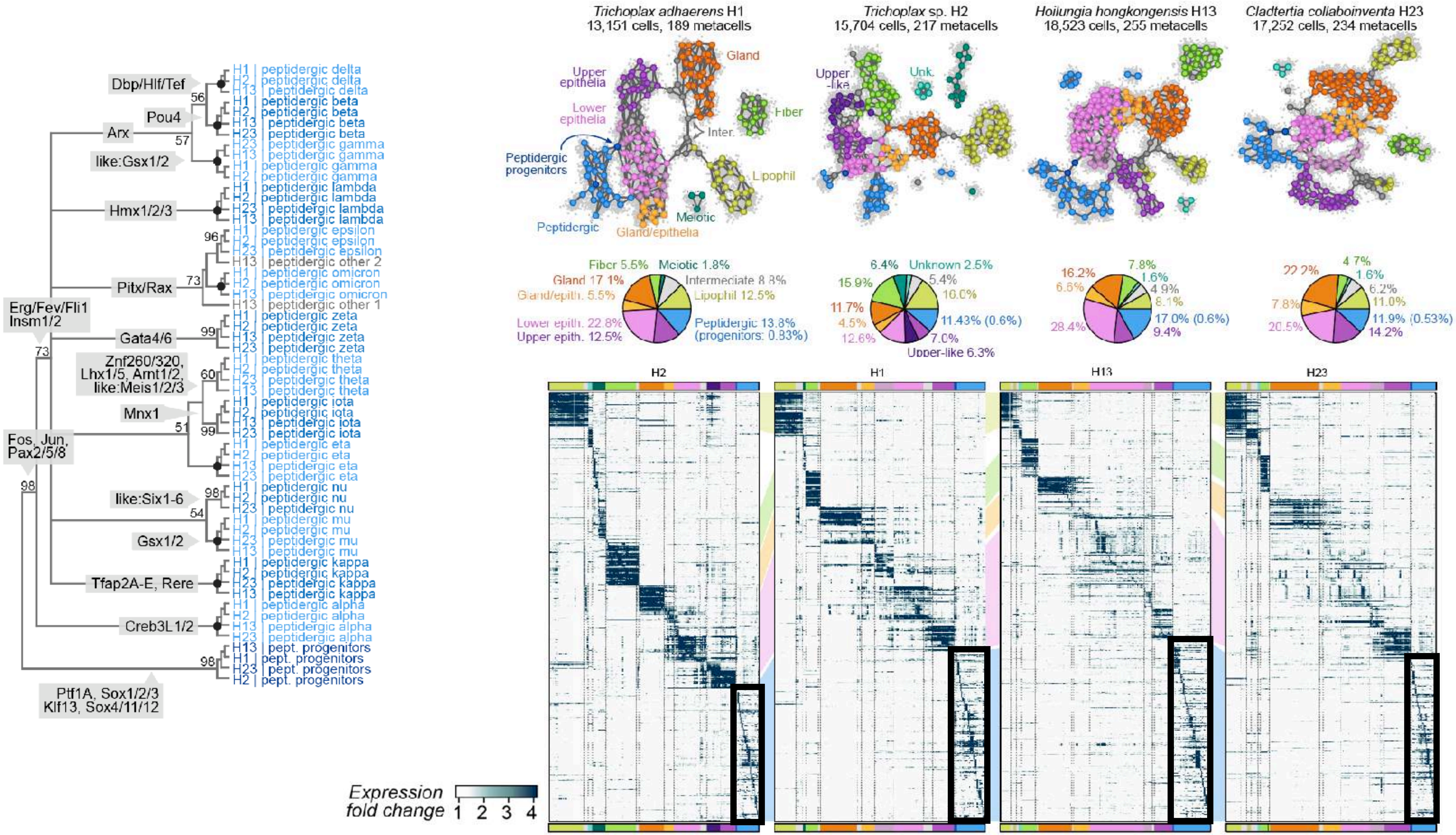
GO enrichments



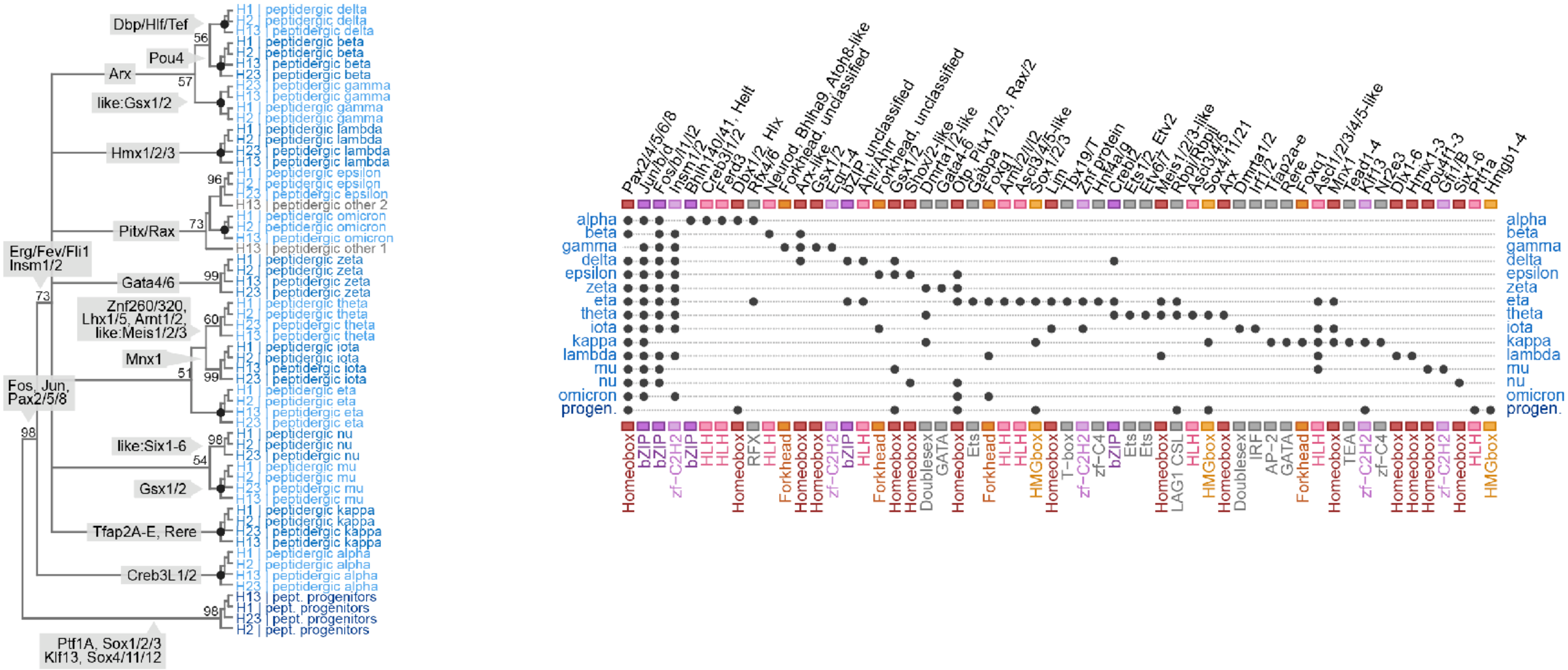
Multi-species gene module clustering



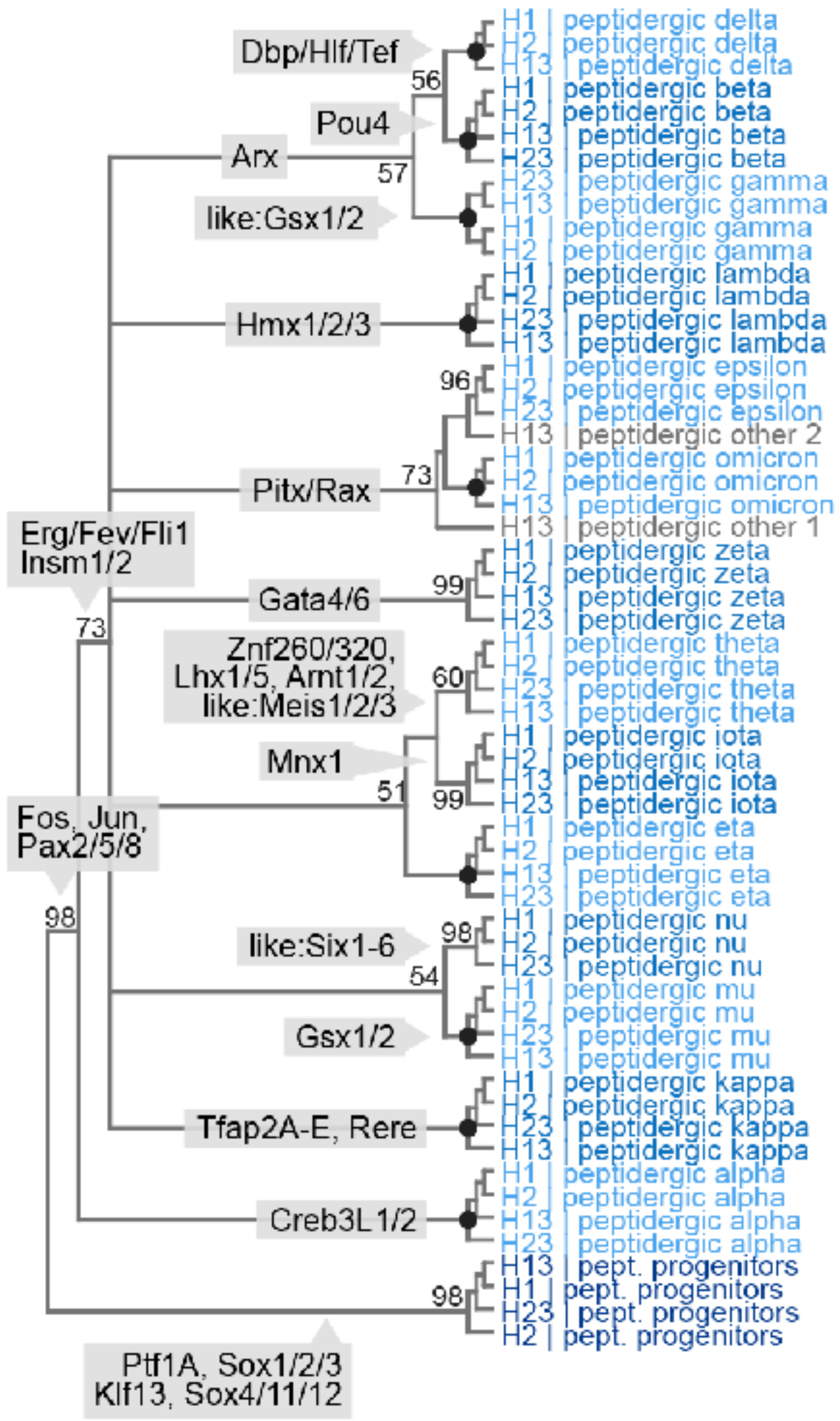
Unexpected diversity of peptidergic cell types



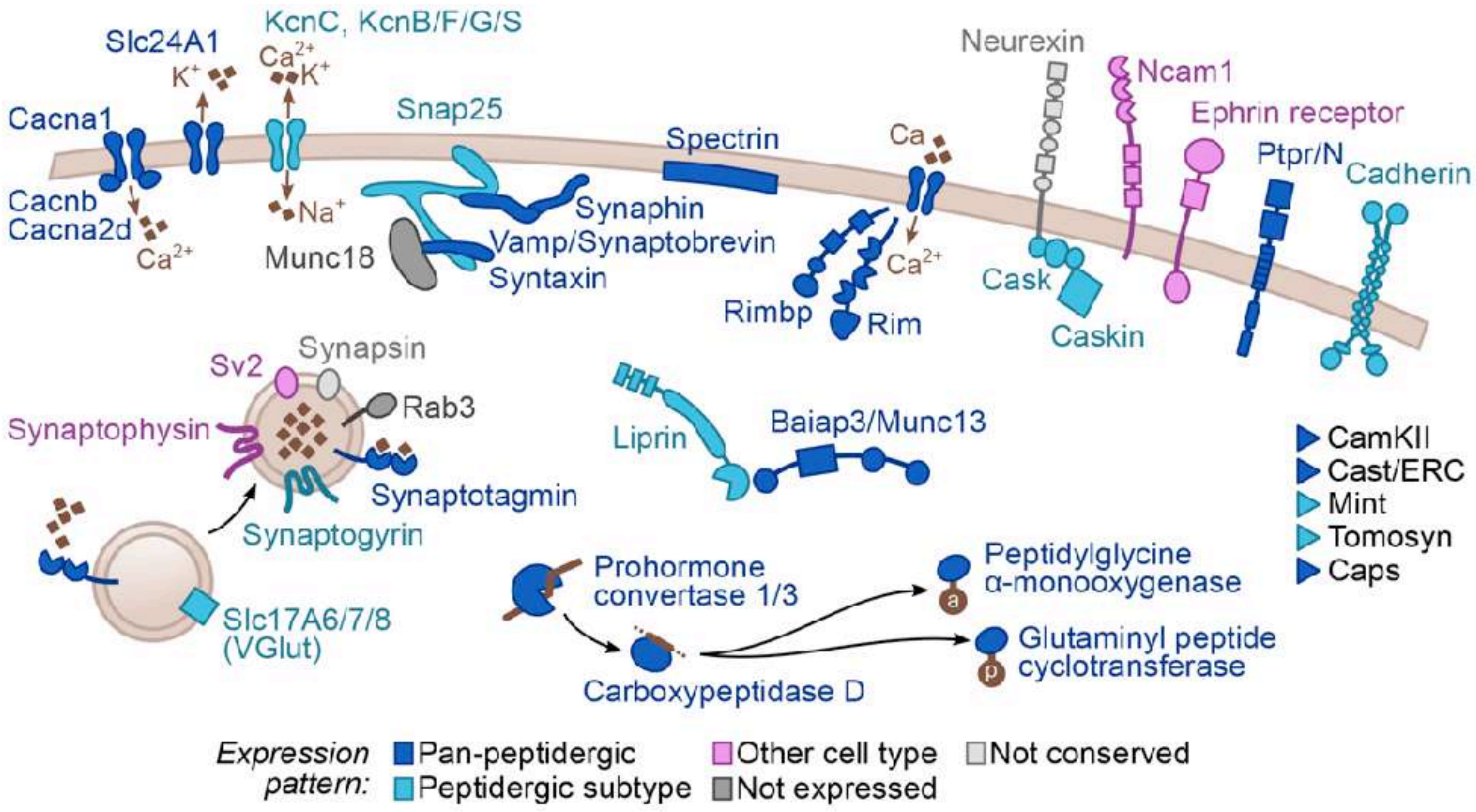
Peptidergic cell types **transcription factor code**



Shared, pan-peptidergic gene modules

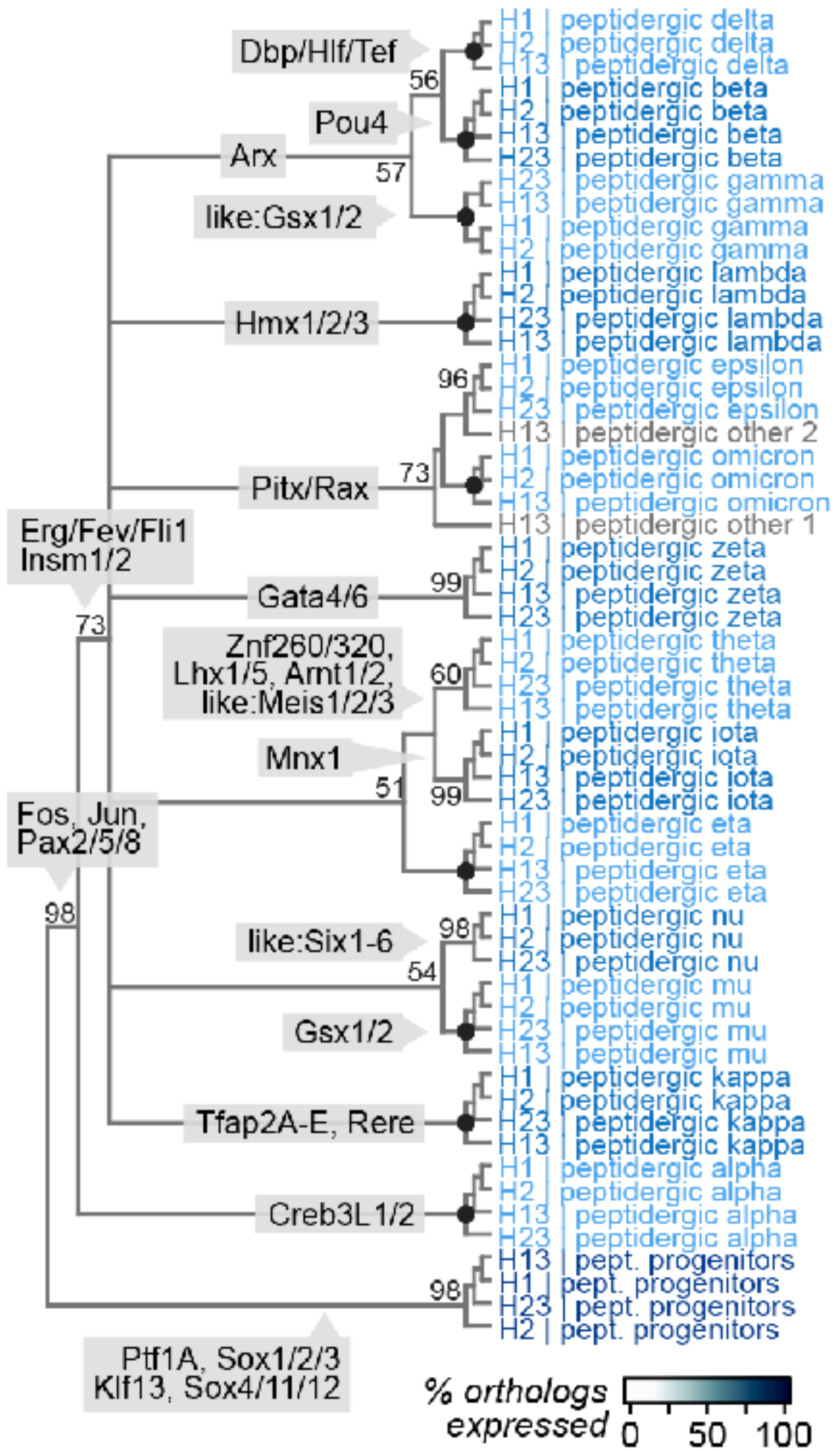


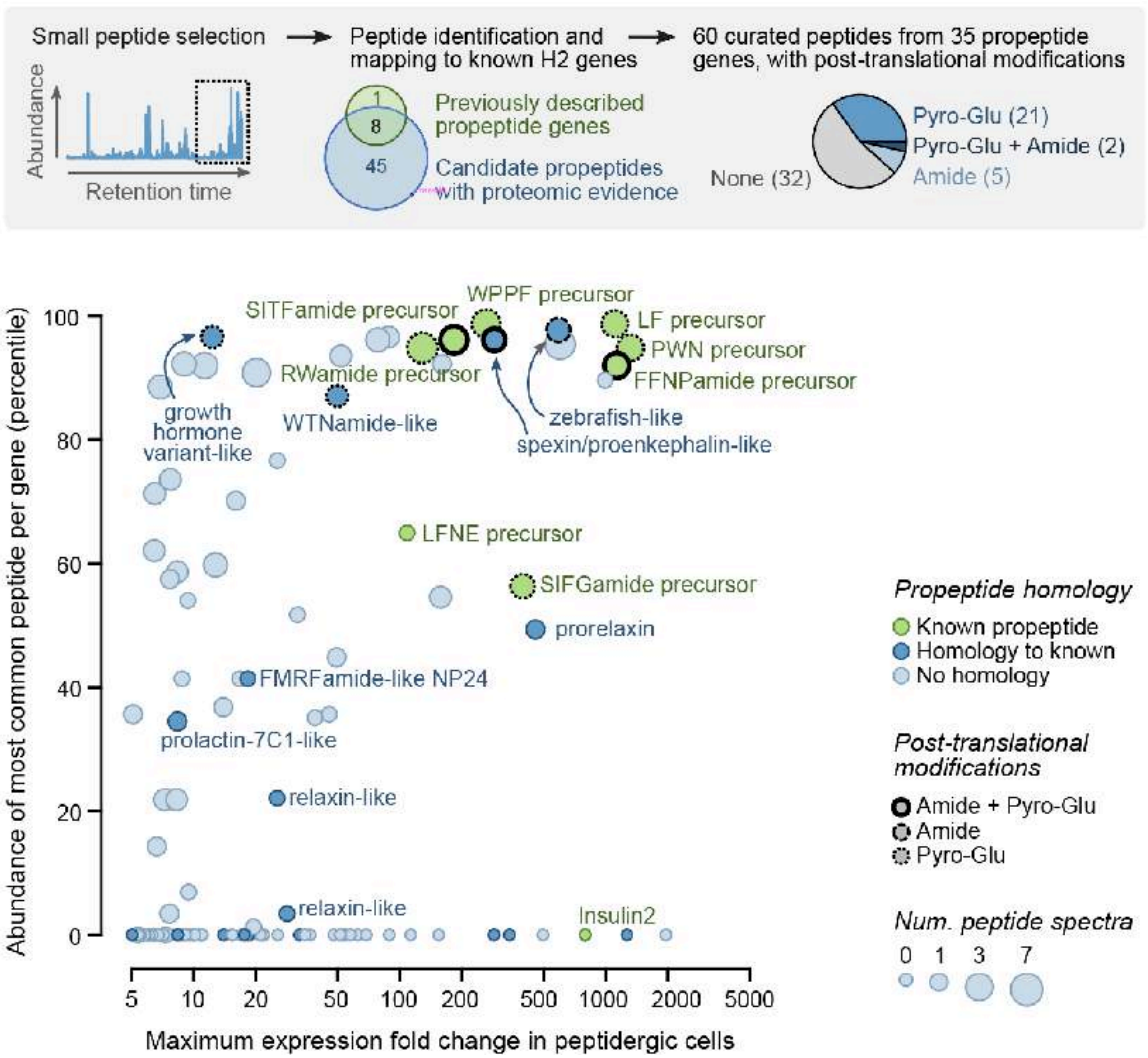
Peptidergic expression of presynaptic scaffold genes



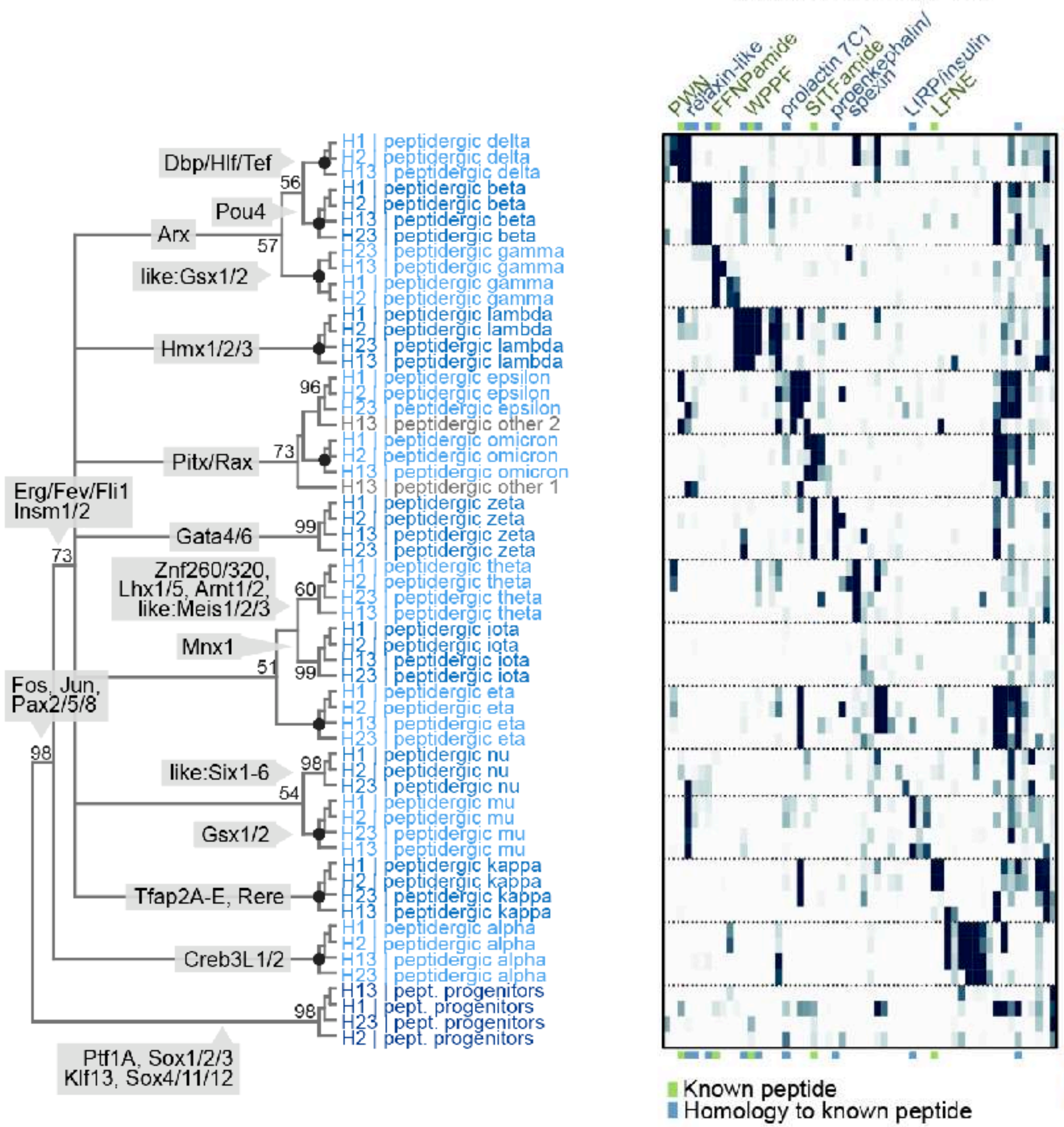
… and neuropeptide processing genes

Identifying small peptides and their **post-translational modifications**

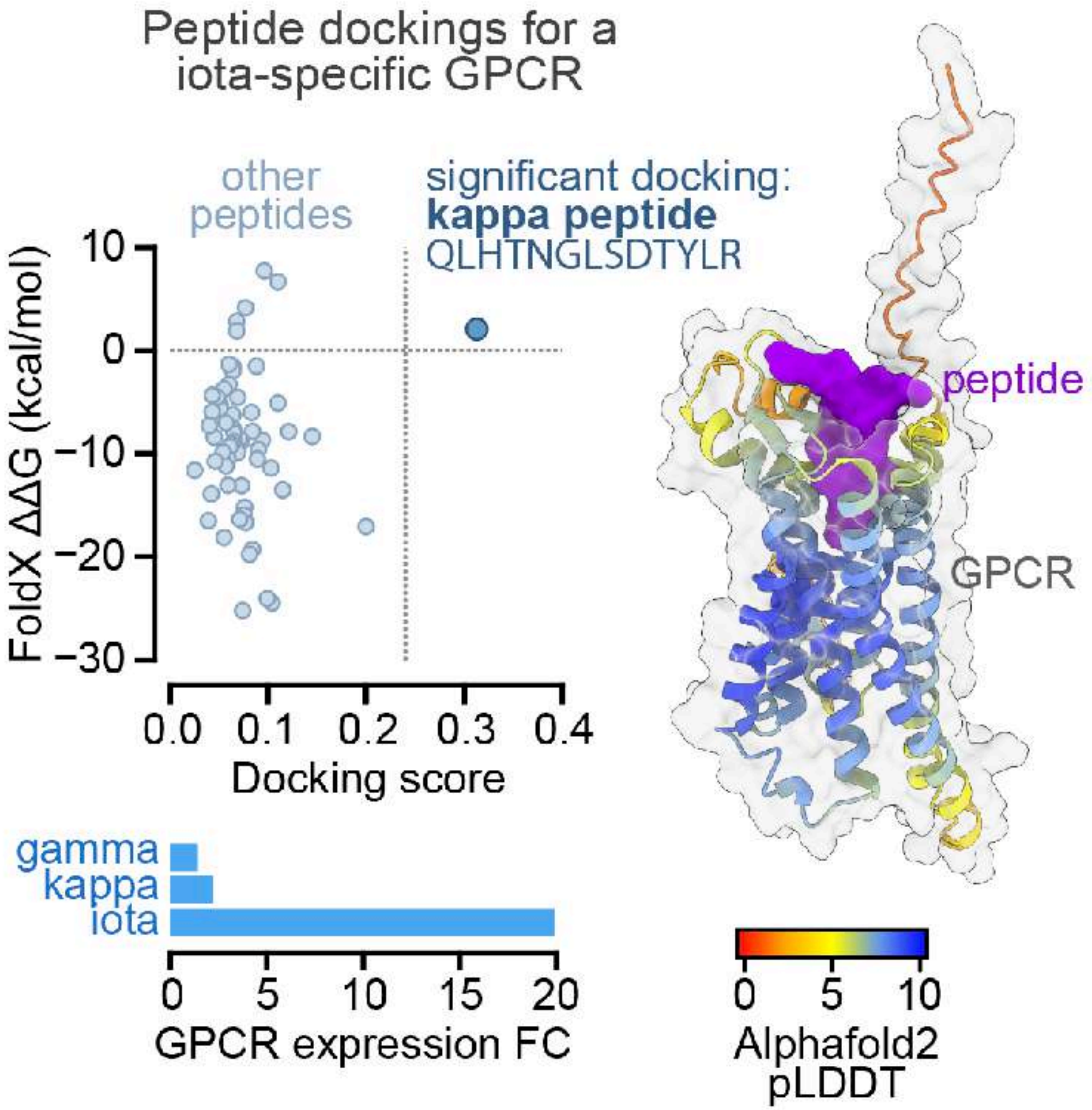




Unique combinations of peptides and GPCRs across peptidergic cell types



Predicting **peptidergic cell-cell communication** in Placozoa

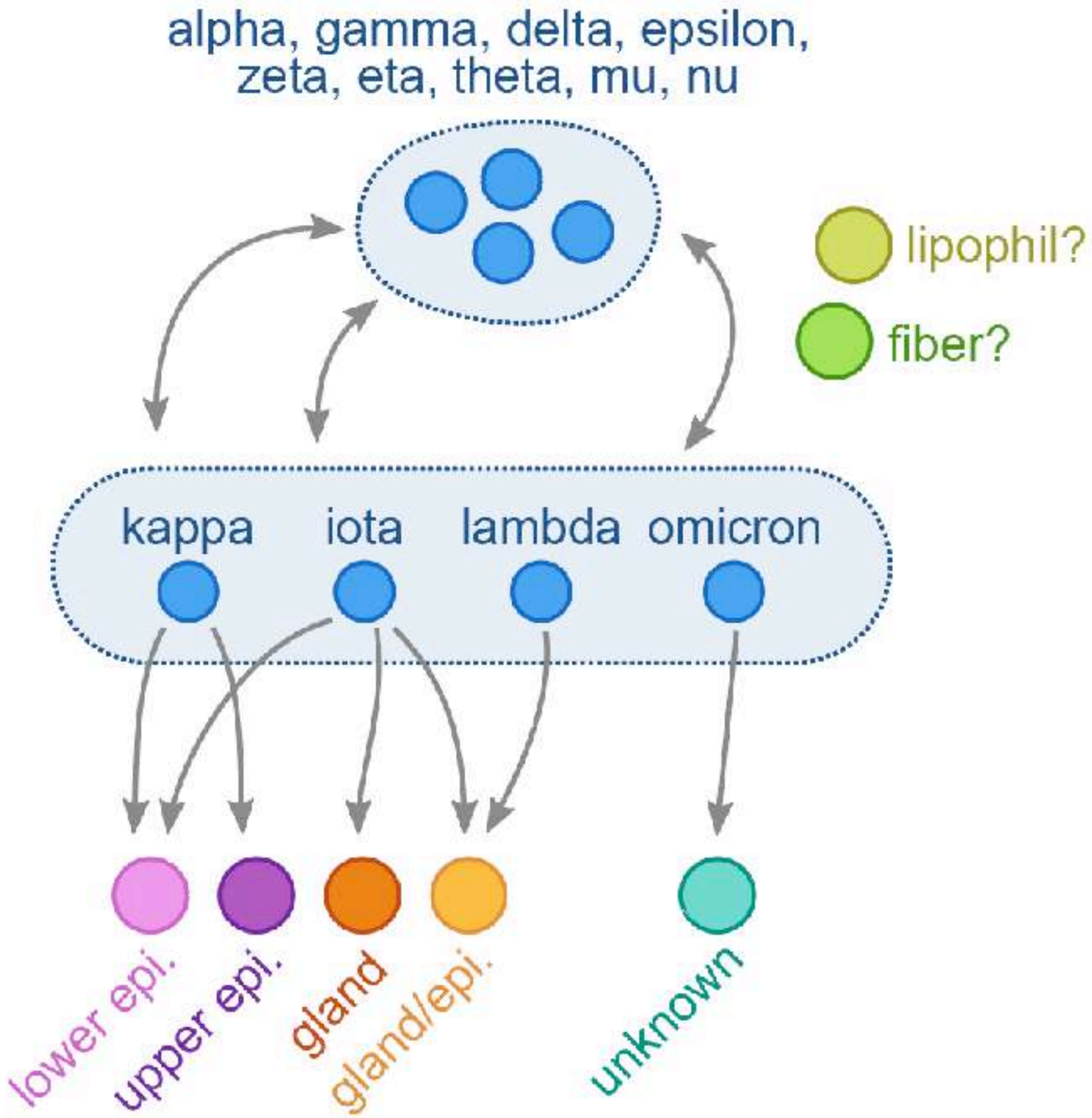


Damiano Cianferoni

Luis Serrano

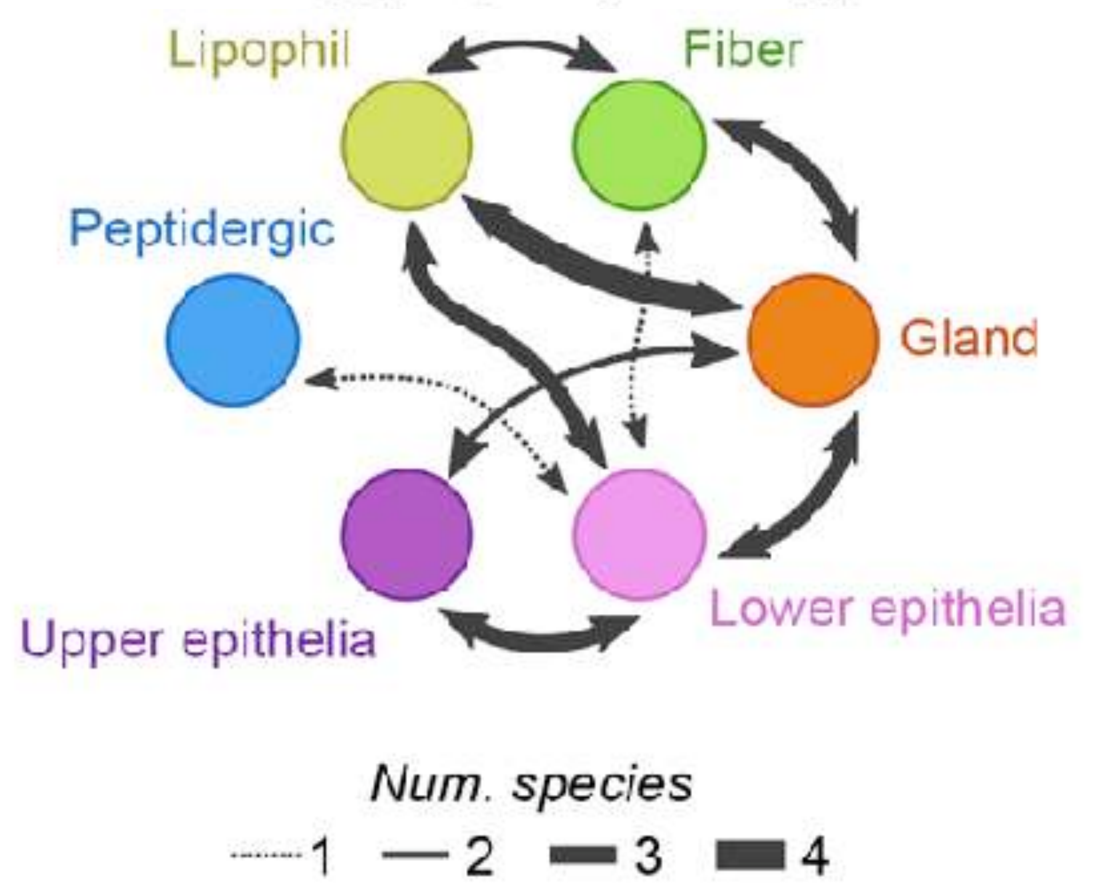
Neuropeptide+GPCR structural modeling (AlphaFold2), and docking analysis to predict peptide-receptor pairs

Hypothetical peptidergic signalling network

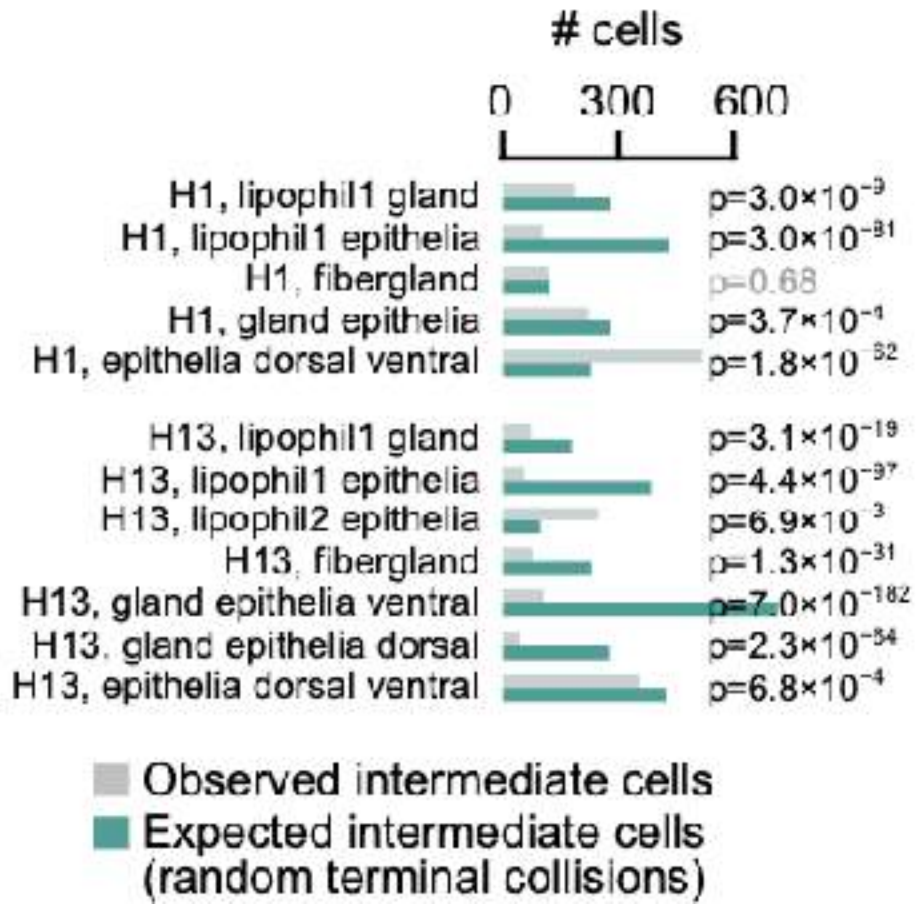


30 peptide-receptor pairs +
cell type-specific expression patterns

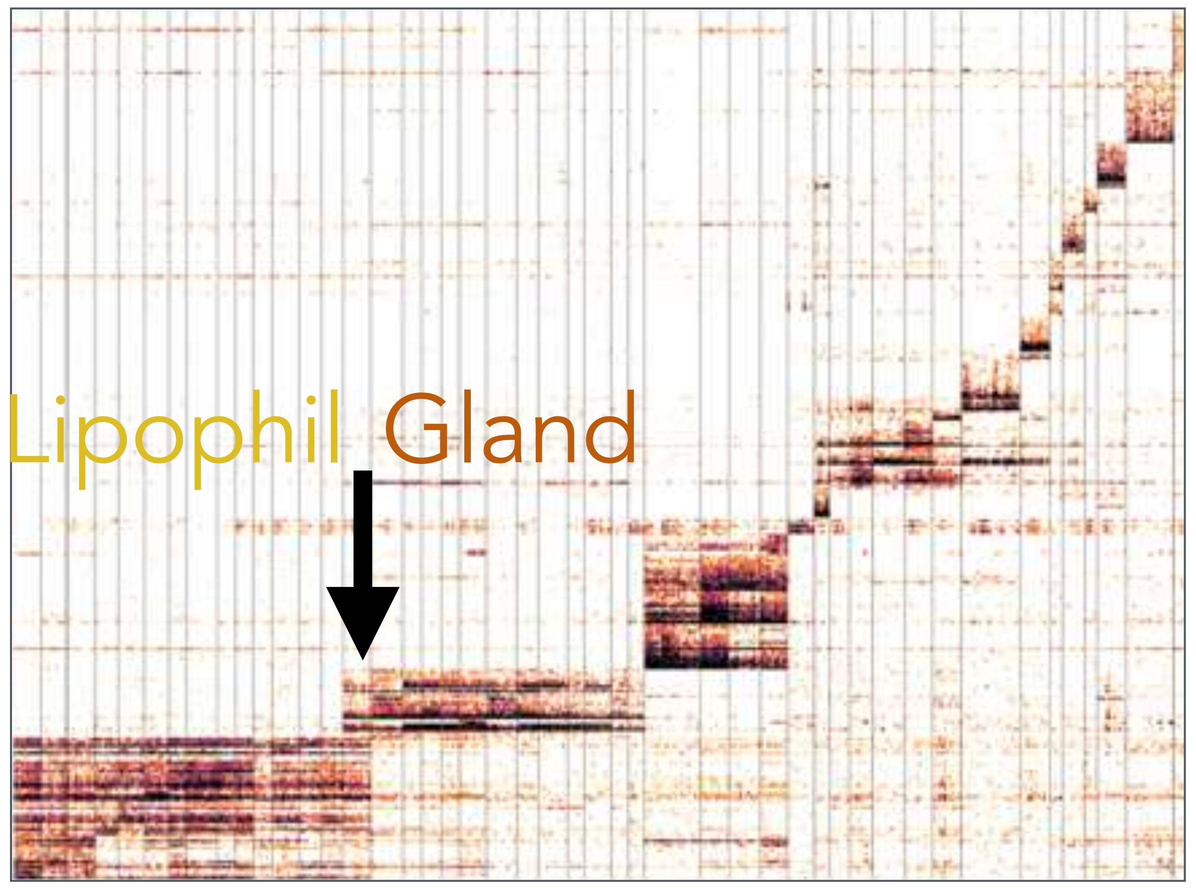
Placozoa intermediate cell states: transdifferentiation?



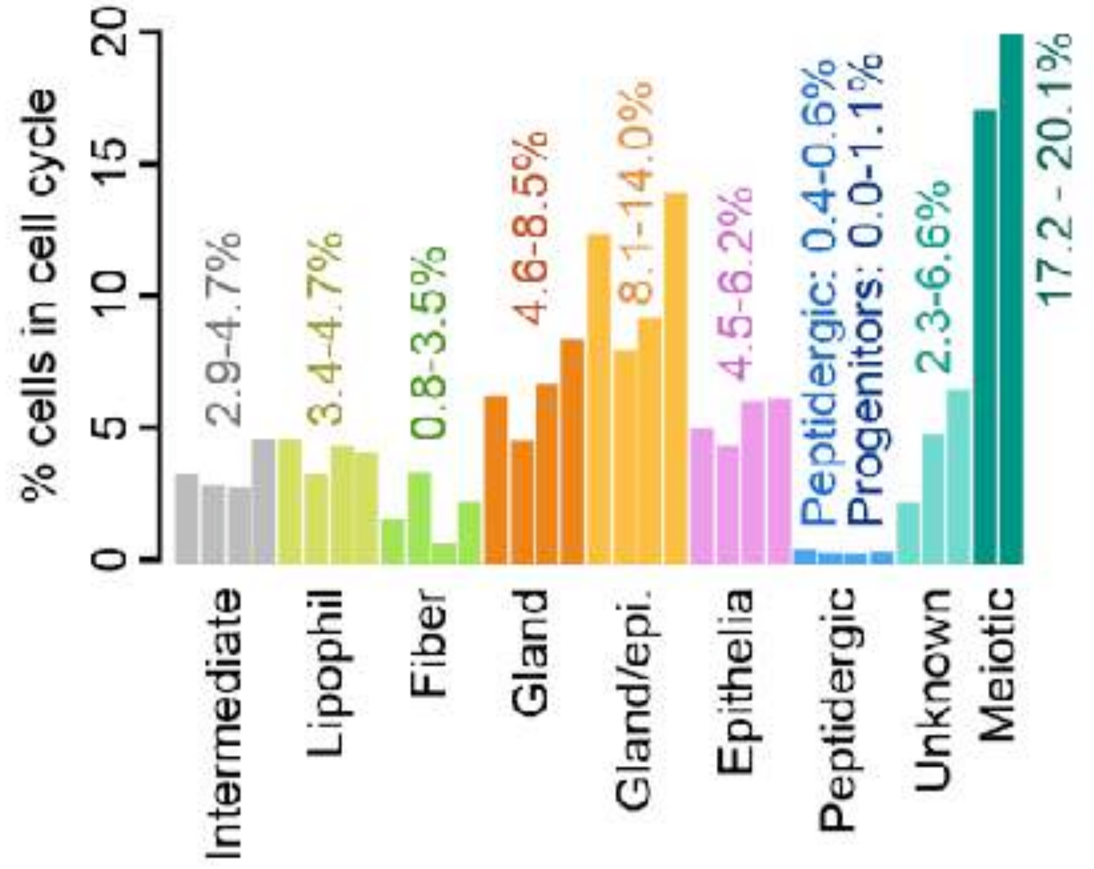
Observed in multiple species



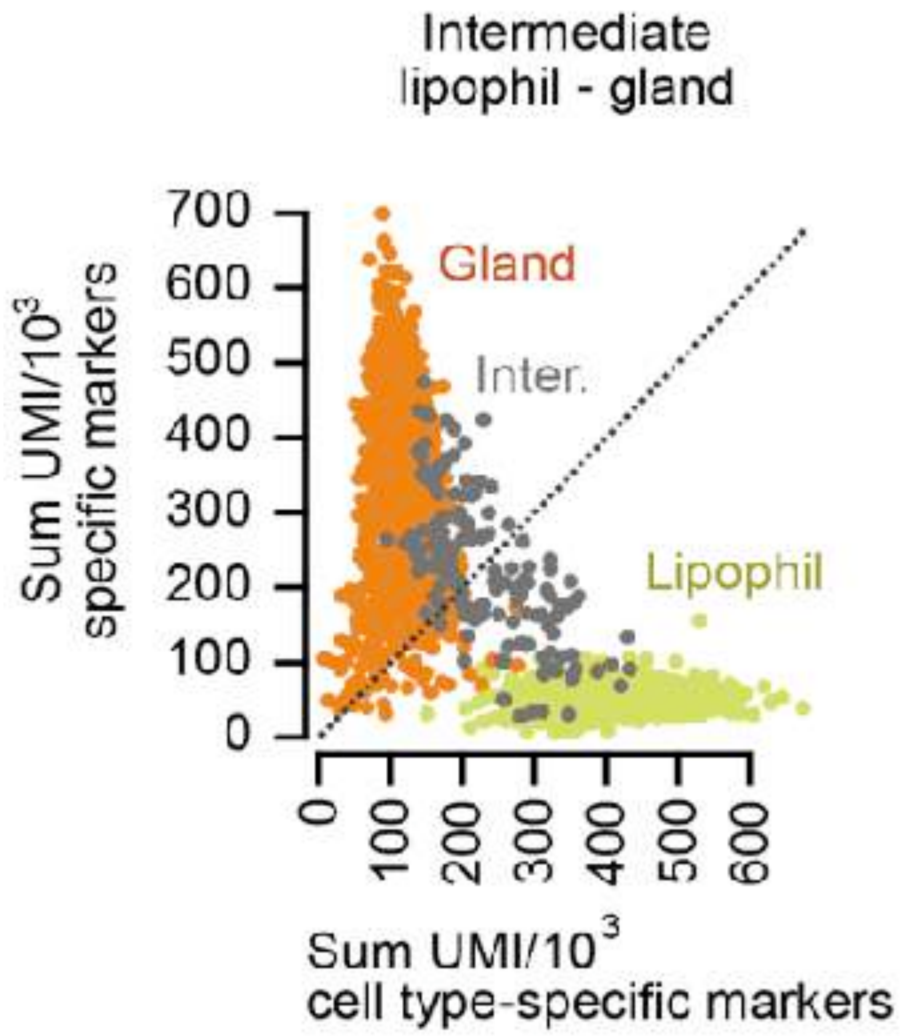
Not explained by random co-encapsulation ("doublets")

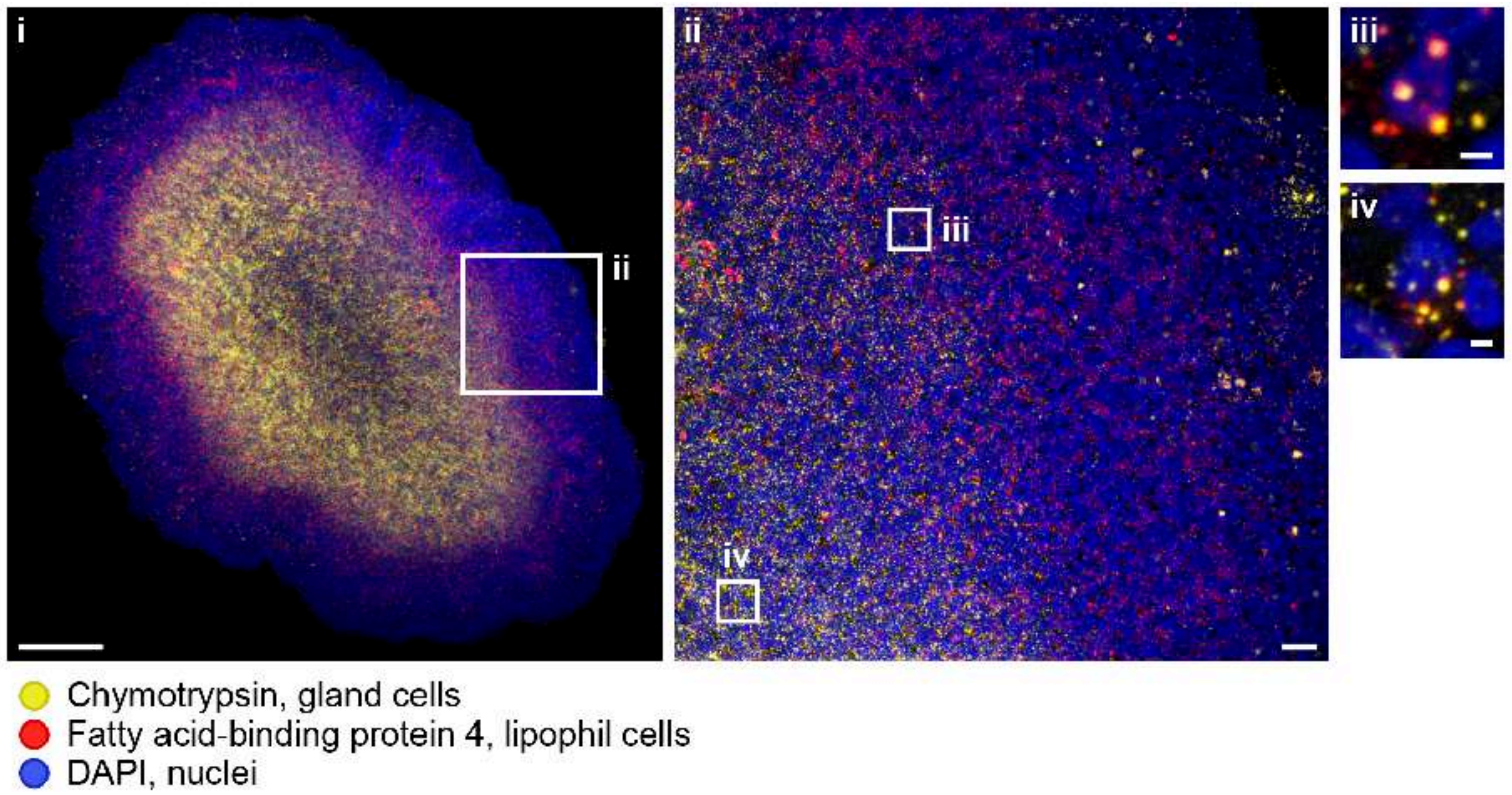


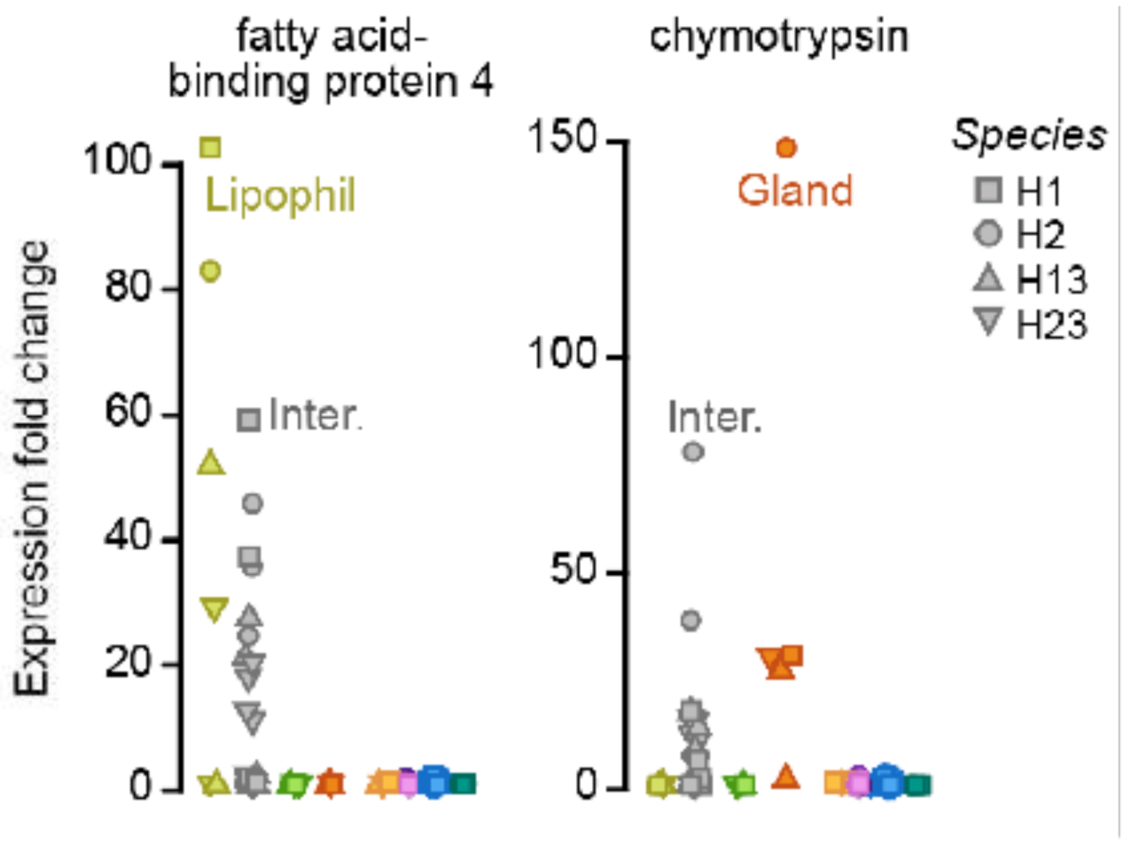
Also observed in our 2018 MARS-seq *Trichoplax* sc atlas



Many differentiated cells express cell cycle genes

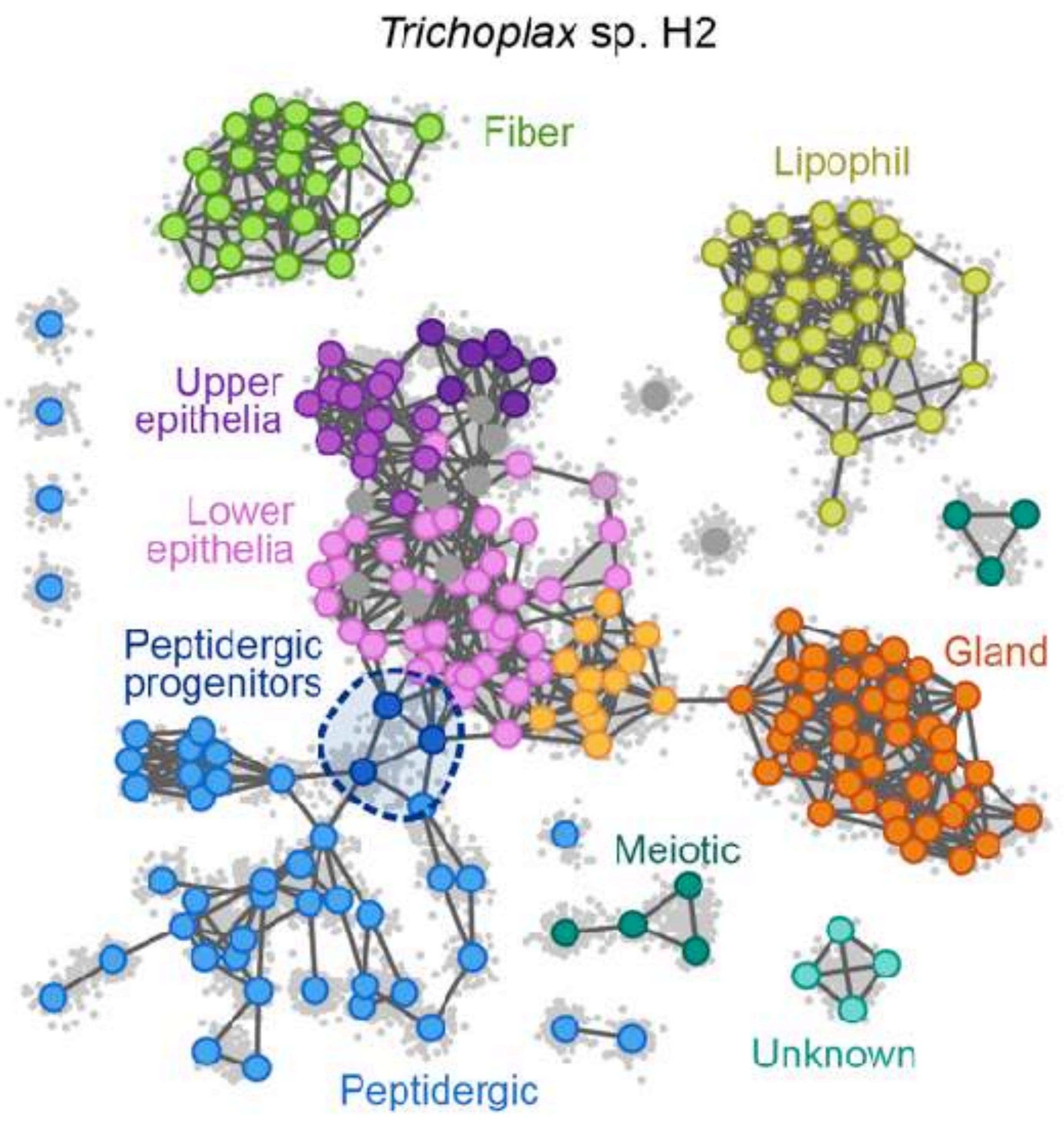


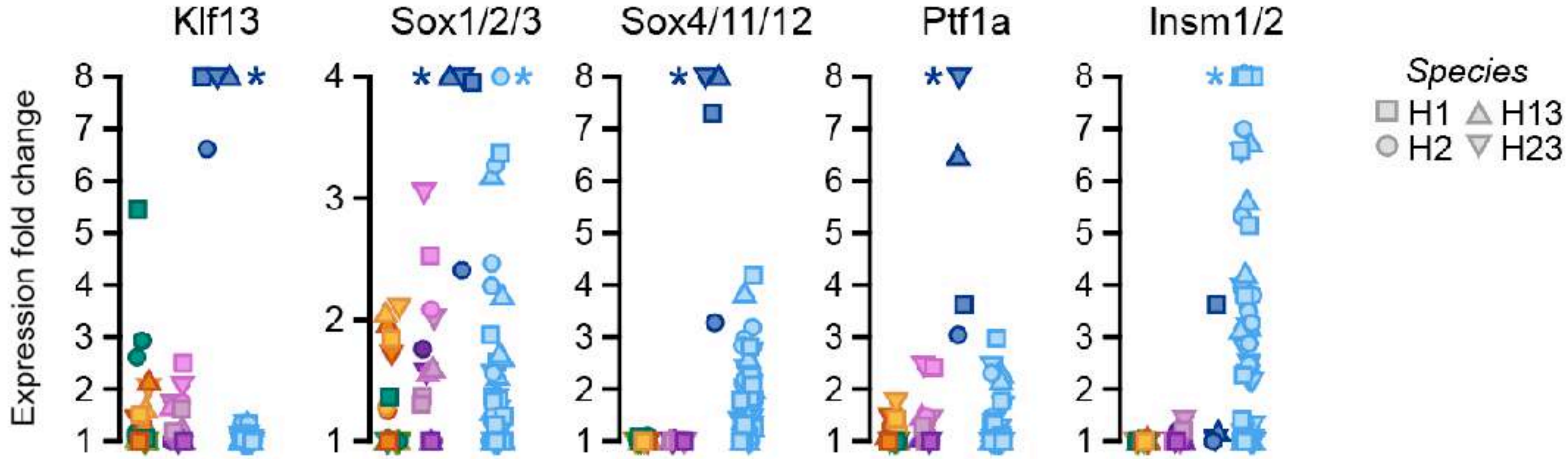


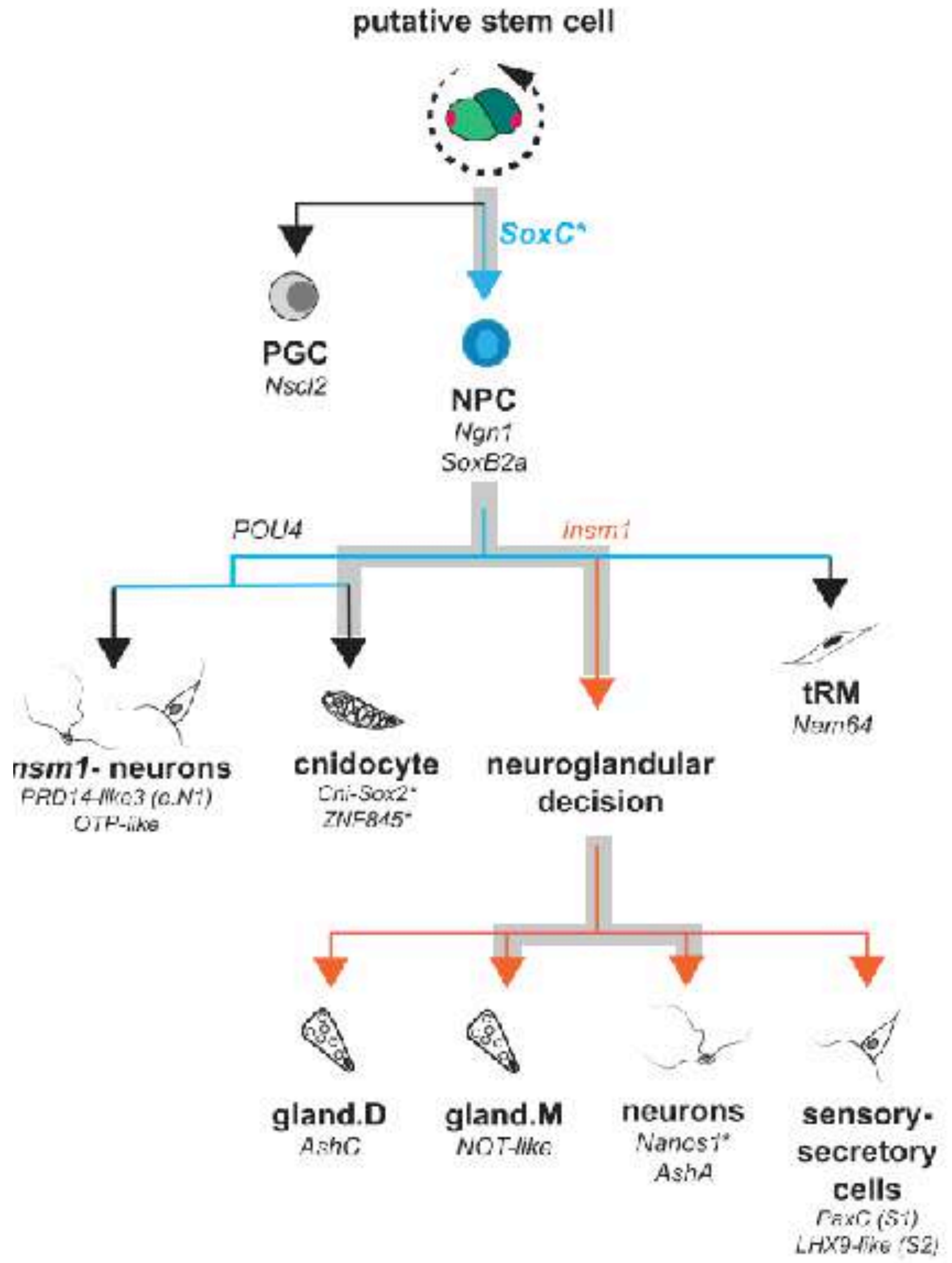


Observed by ISH (and FACS-ISH)

Peptidergic progenitors express **TFs involved in neurogenesis** in other animals







Cell Reports
CellPress
OPEN ACCESS

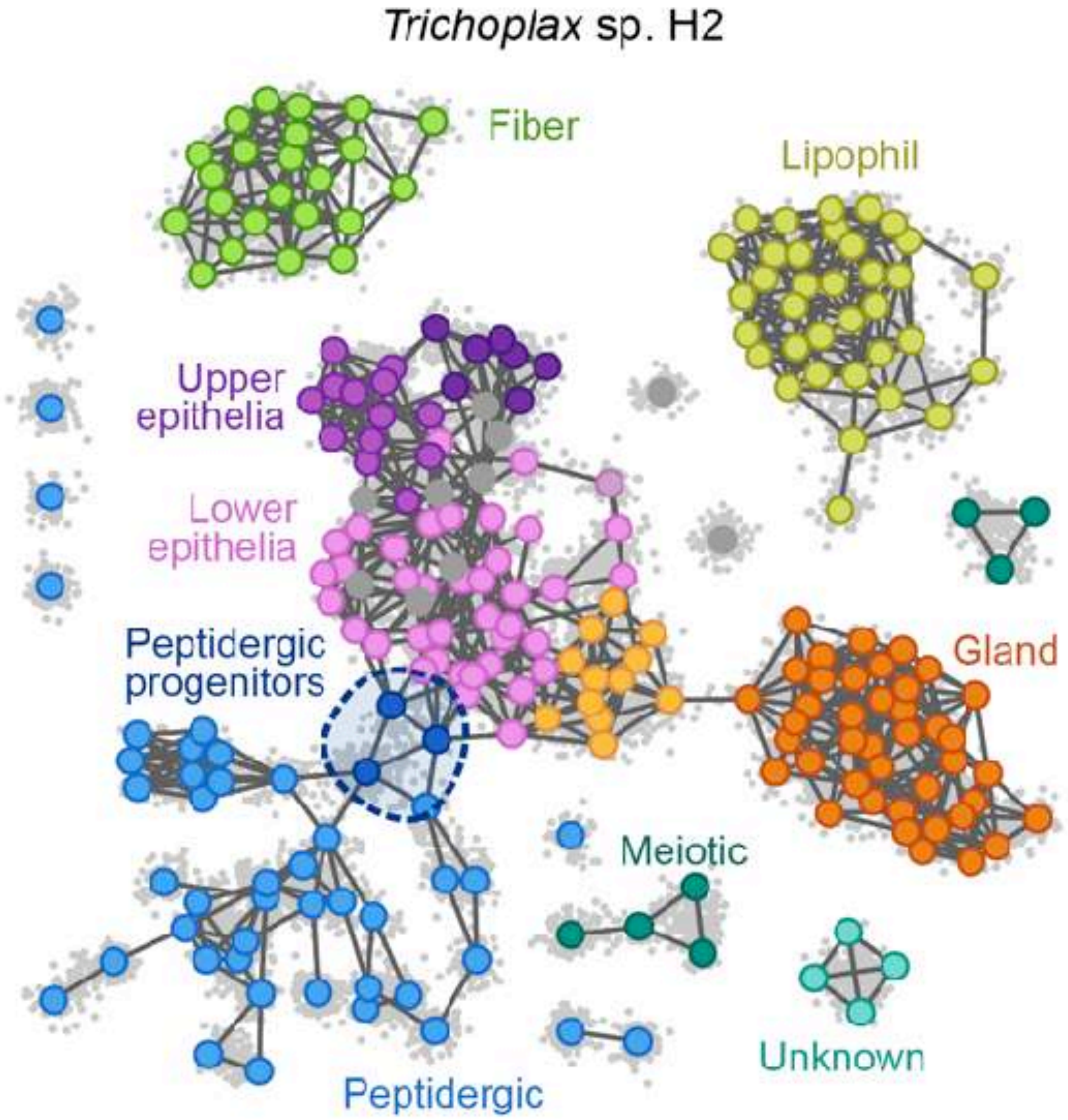
Resource

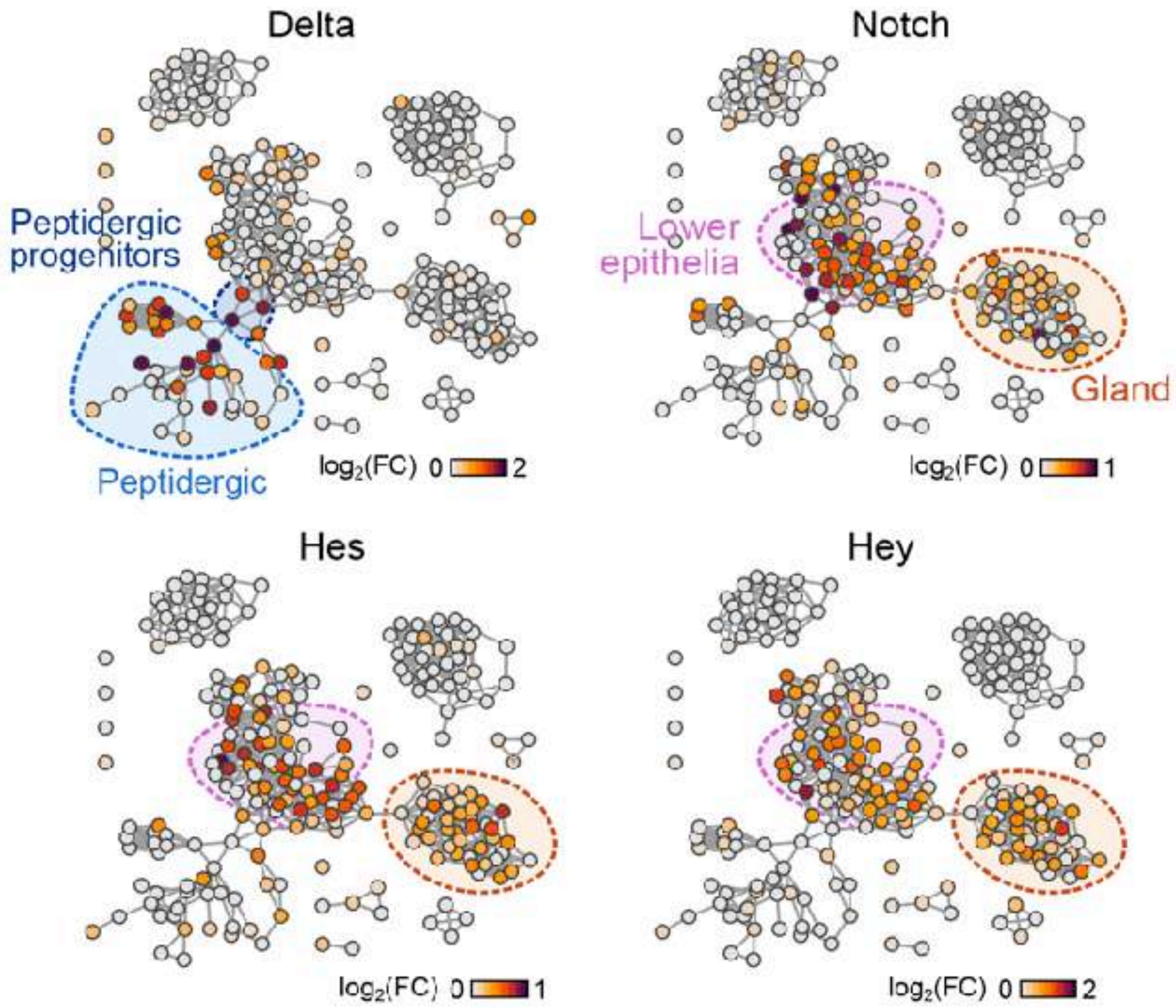
Single-cell transcriptomics identifies conserved regulators of neuroglandular lineages

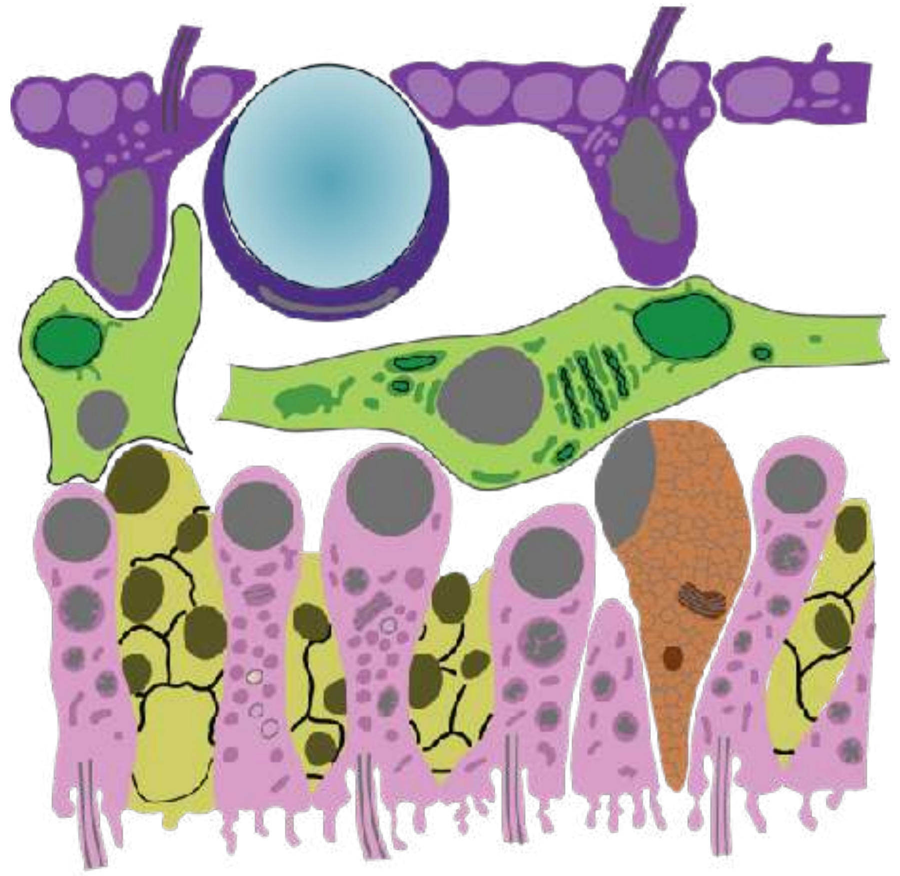
Julia Steger,[1,4] Alison G. Cole,[1,3,4,*] Andreas Denner,[1,4] Tatiana Lebedeva,[1] Grigory Genikhovich,[1] Alexander Ries,[1] Robert Reischl,[1] Elisabeth Taudes,[1] Mark Lassnig,[1] and Ulrich Technau[1,2,3,5,*]
[1]Department of Neurosciences and Developmental Biology, Faculty of Life Sciences, University of Vienna, 1030 Vienna, Austria
[2]Max-Perutz Labs, Dr.-Bohr-Gasse 9, 1030 Vienna, Austria
[3]Research Platform "SinCeReSt: Single Cell Regulation of Stem Cells," University of Vienna, 1030 Vienna, Austria
[4]These authors contributed equally
[5]Lead contact
*Correspondence: alison.cole@univie.ac.at (A.G.C.), ulrich.technau@univie.ac.at (U.T.)

https://doi.org/10.1016/j.celrep.2022.111370

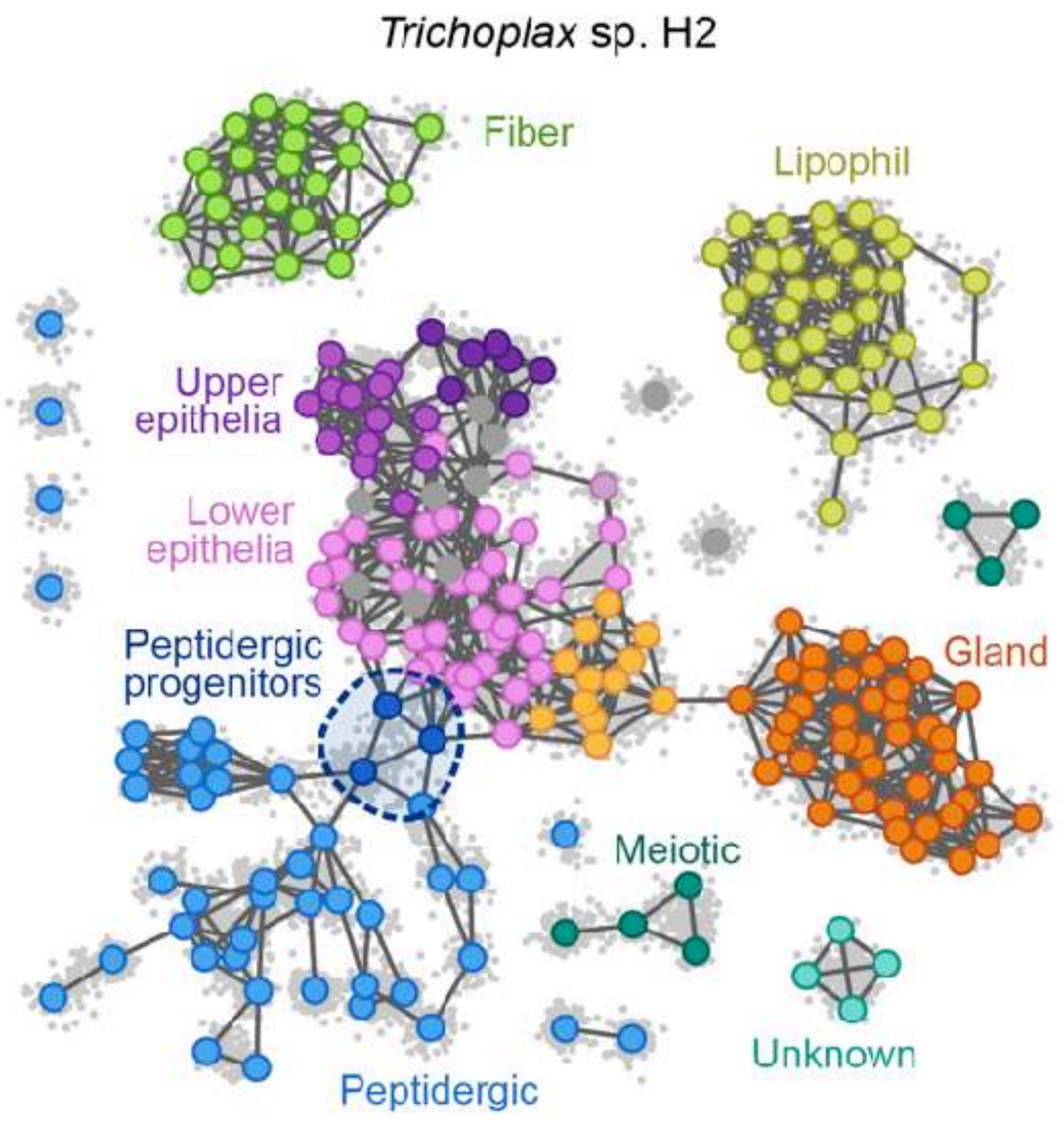
Peptidergic progenitors are **specified by Notch-Delta** signaling

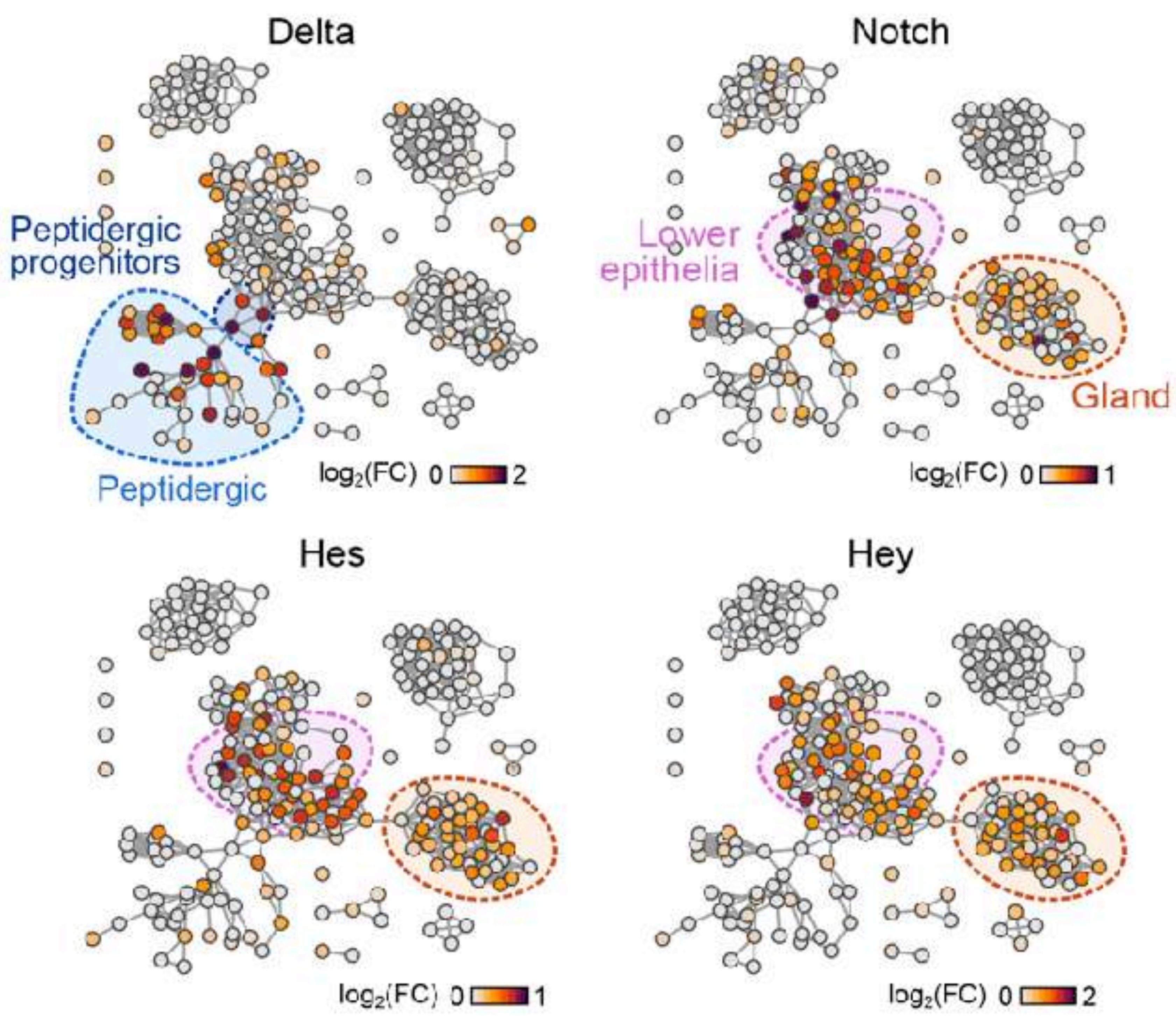


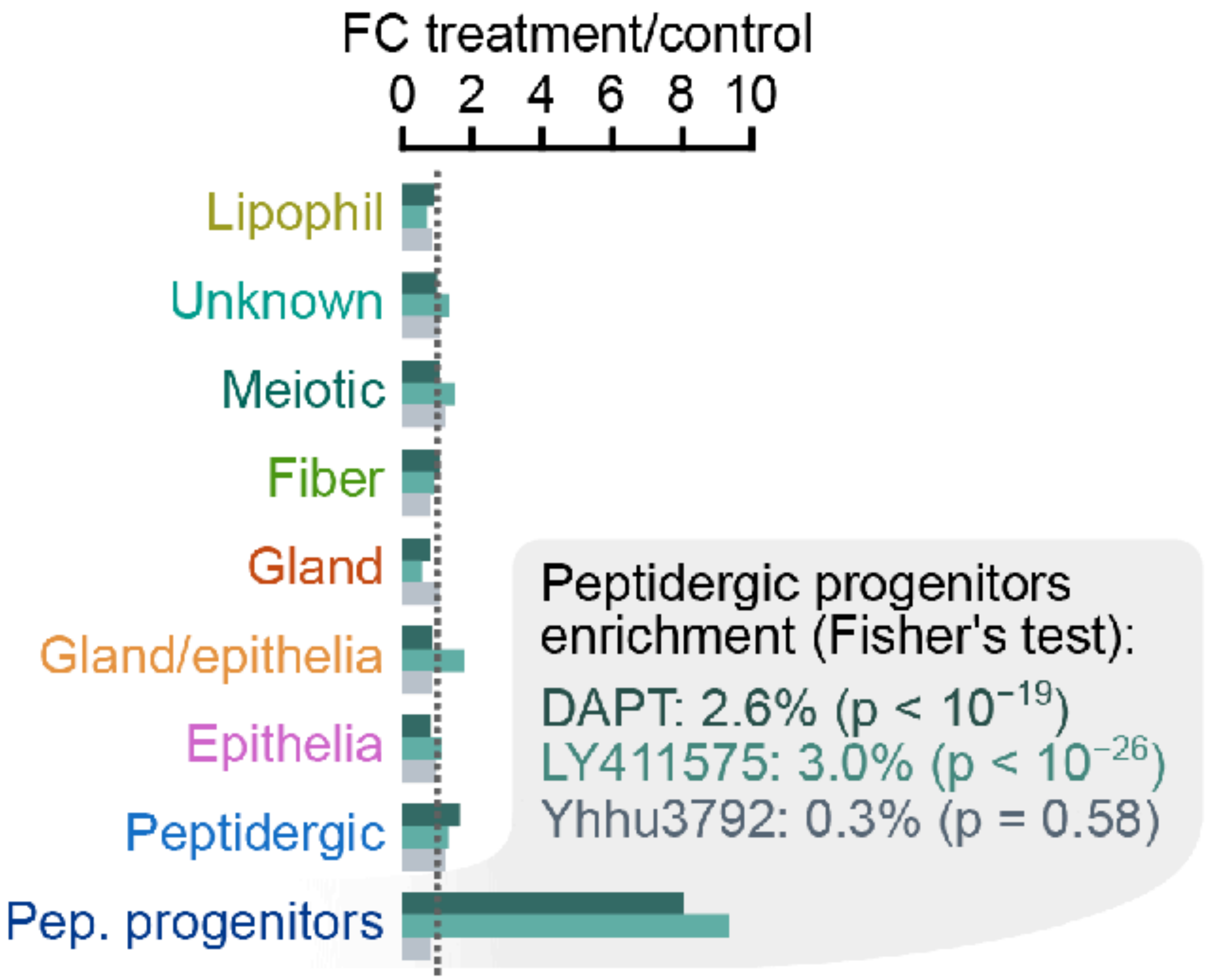


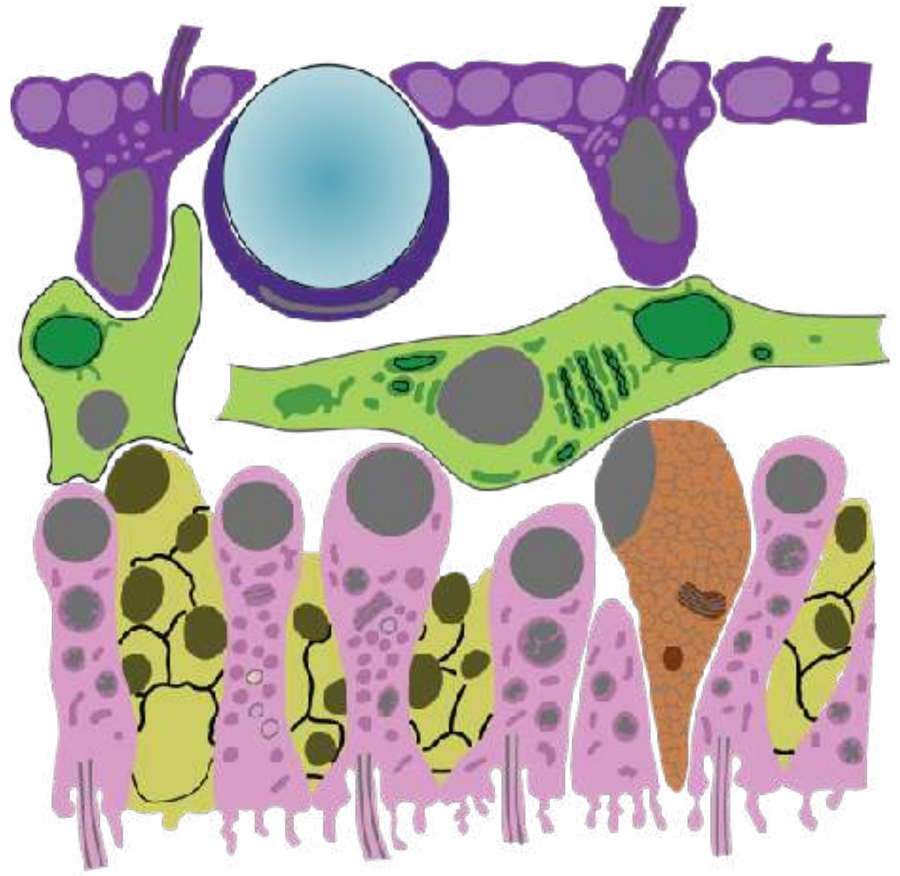


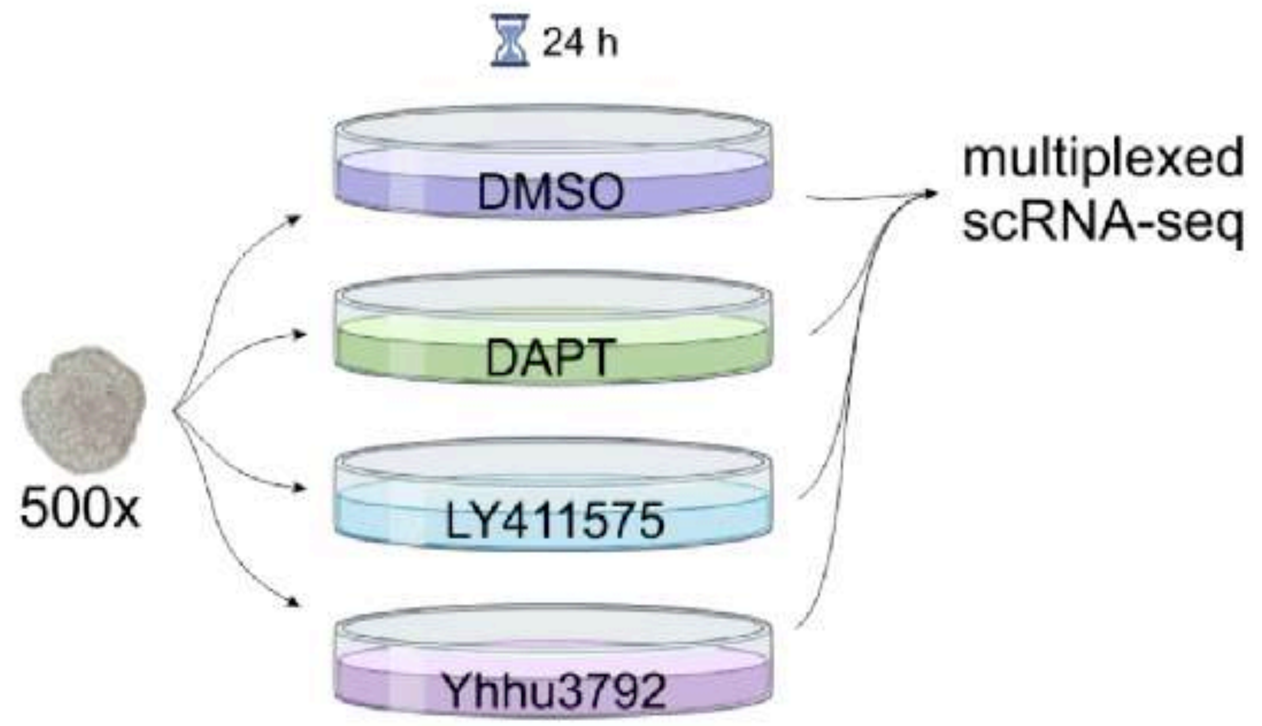
Peptidergic progenitors are **specified by Notch-Delta** signaling





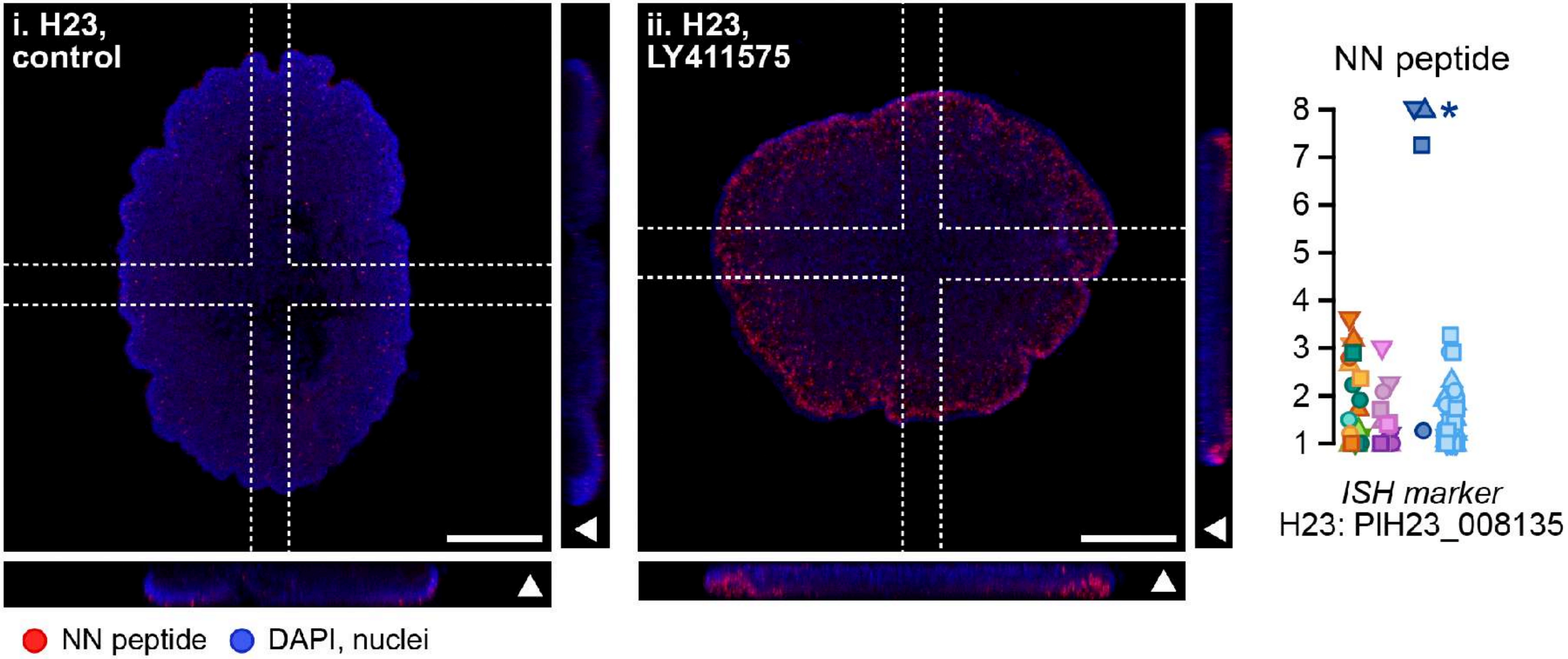






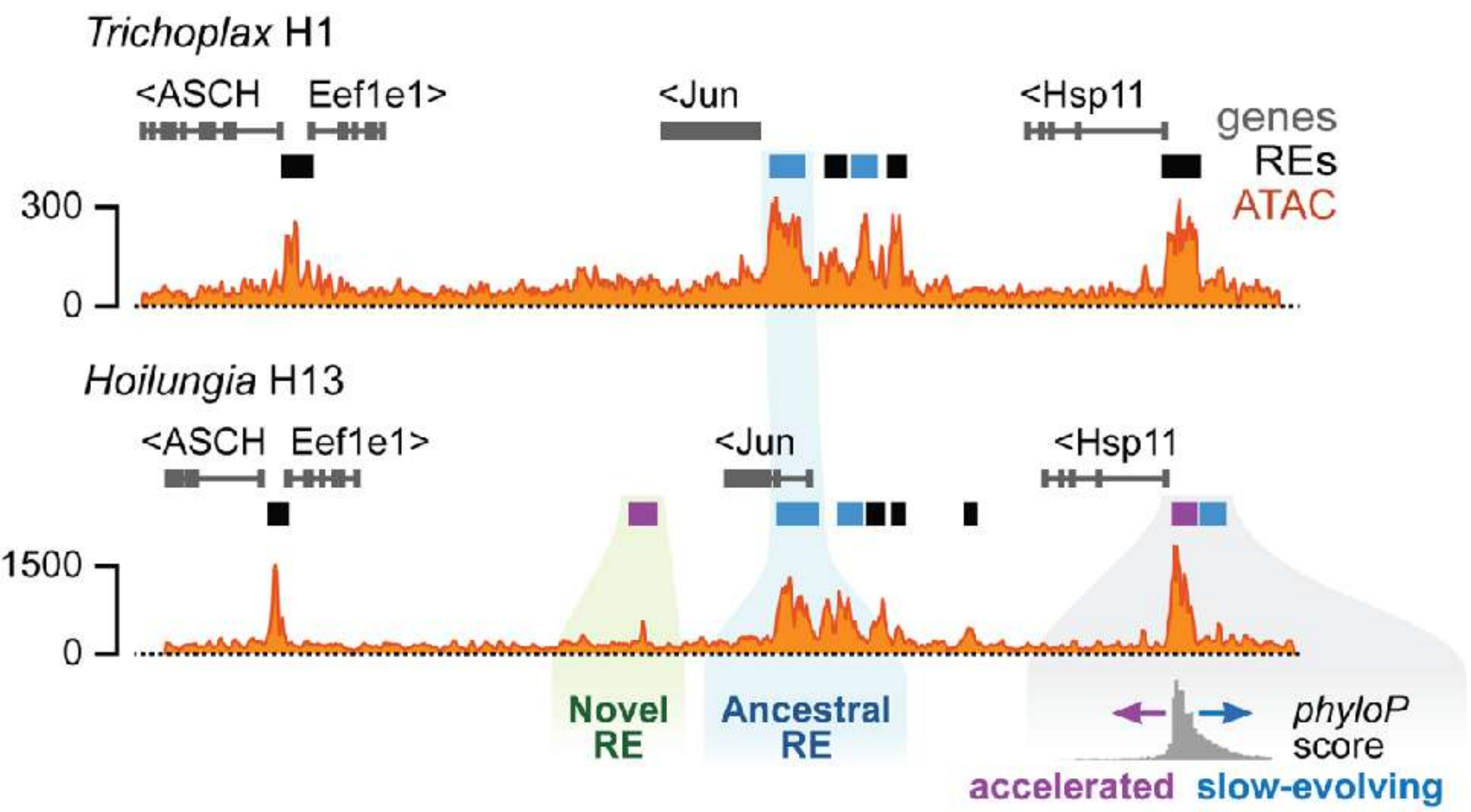
Notch antagonists increase the relative abundance of peptidergic progenitor cells

Peptidergic progenitors are located **in the peripheral lower epithelium**



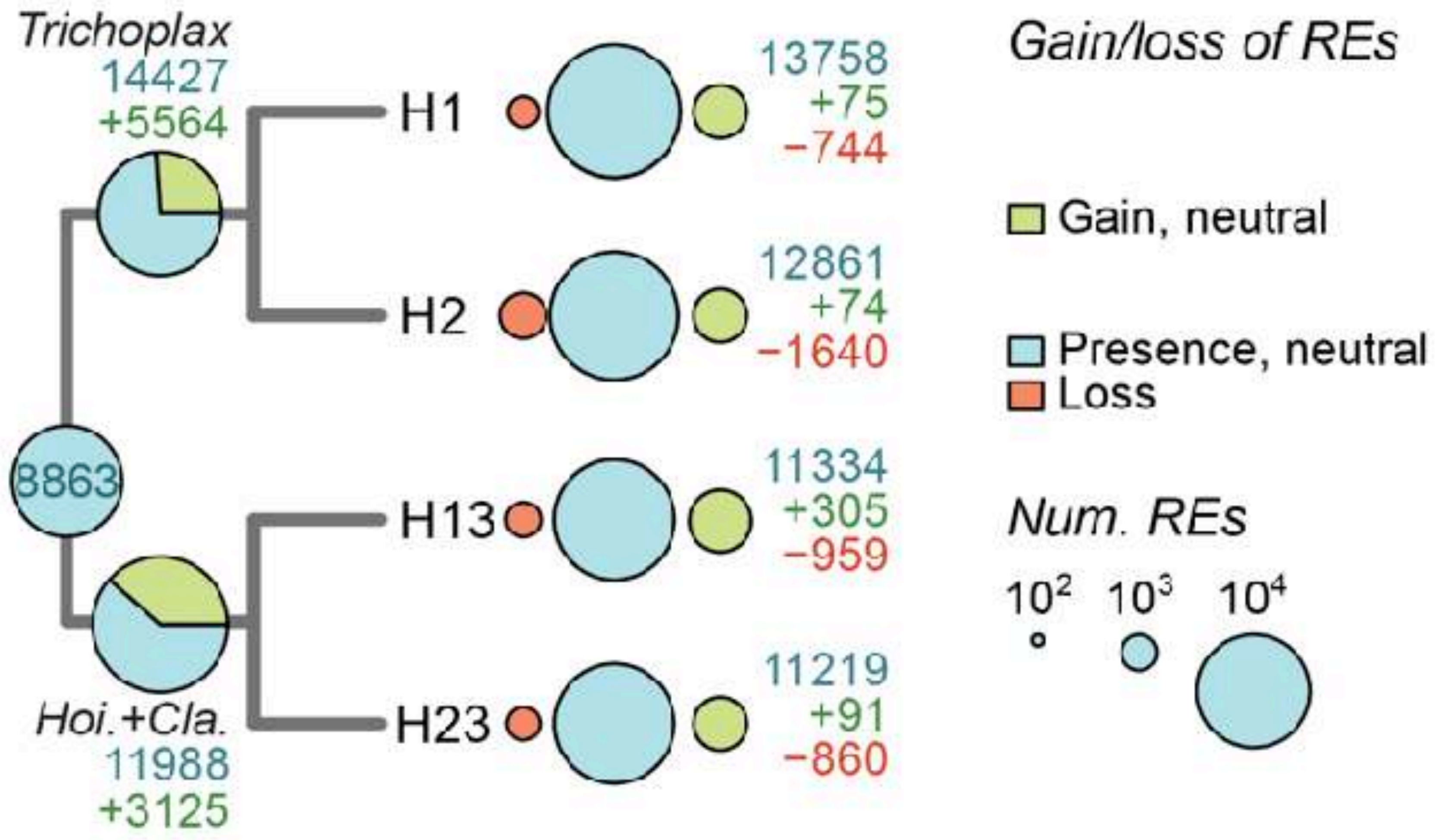
The genetic basis of placozoan cell type gene expression evolution

Mapping *cis*-regulatory elements in four placozoans (ATAC, H3K4me3, H3K4me2)

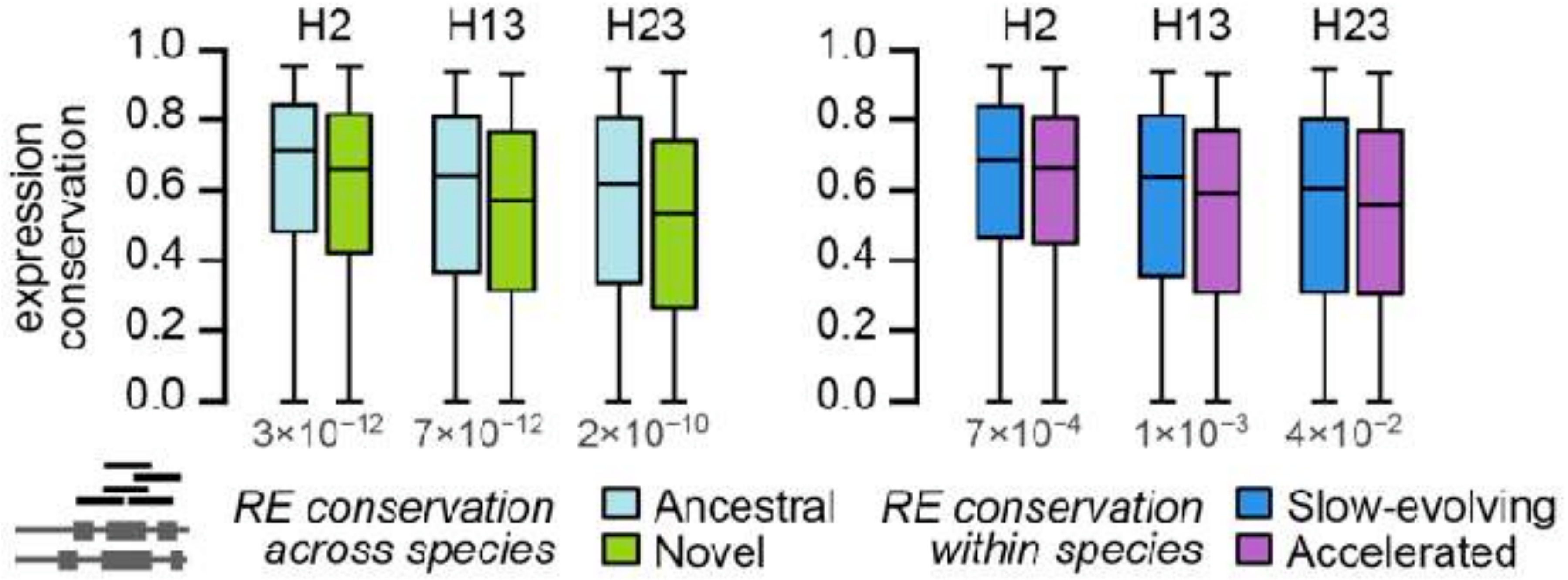


Classifying REs into novel and ancestral and into fast and slow-evolving

CRE gains and losses across placozoan phylogeny

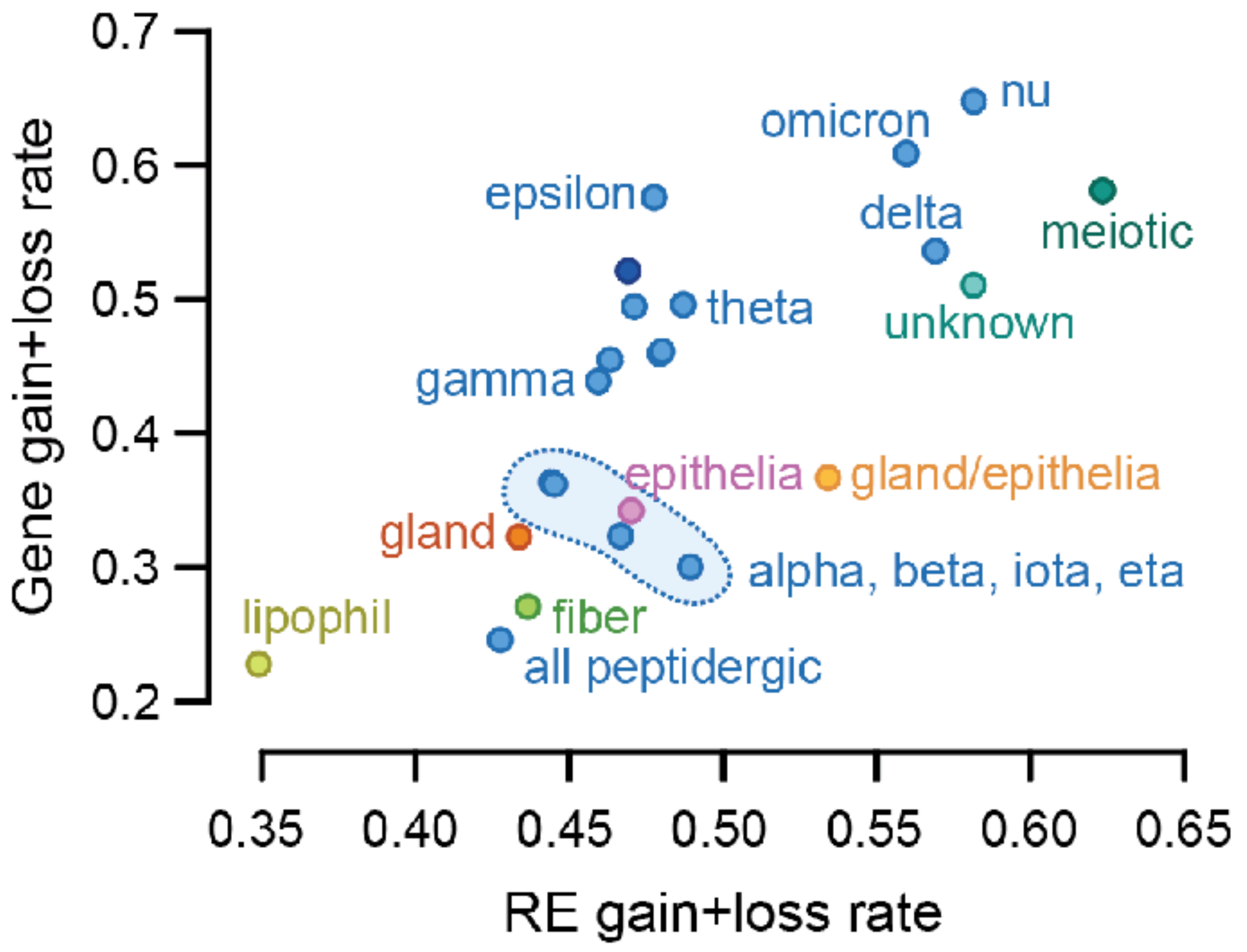


CRE evolution linked to gene expression divergence

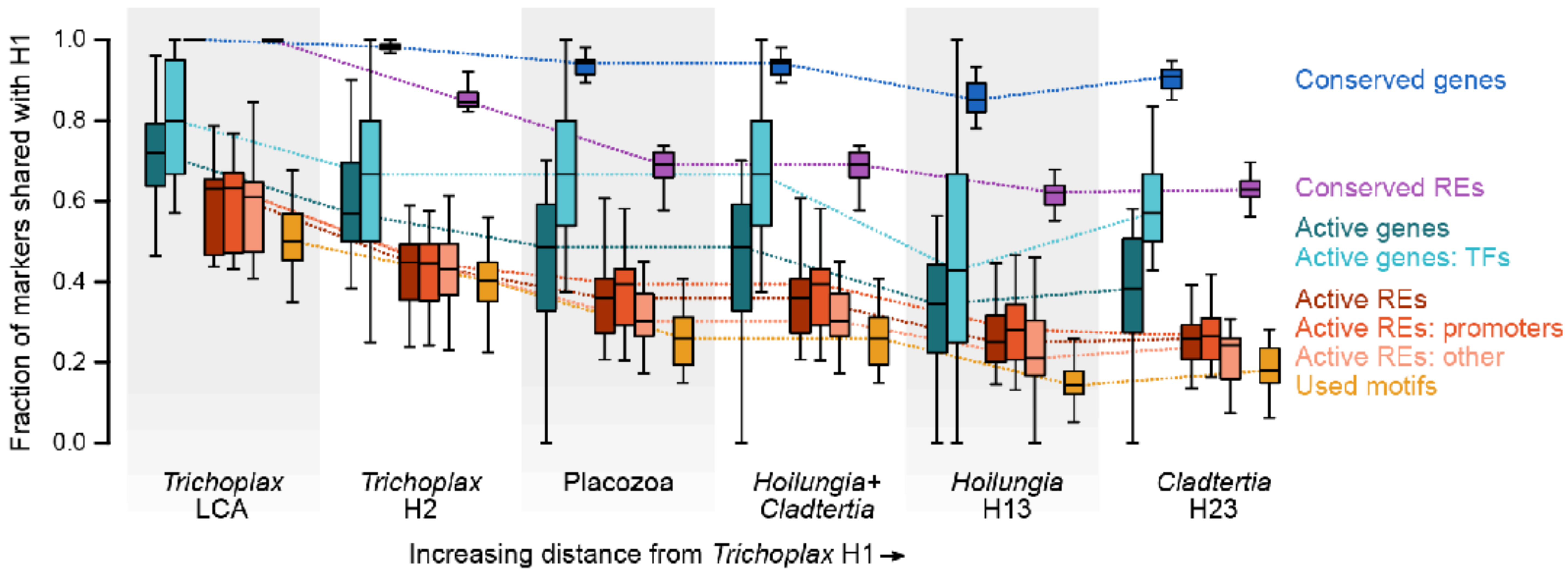


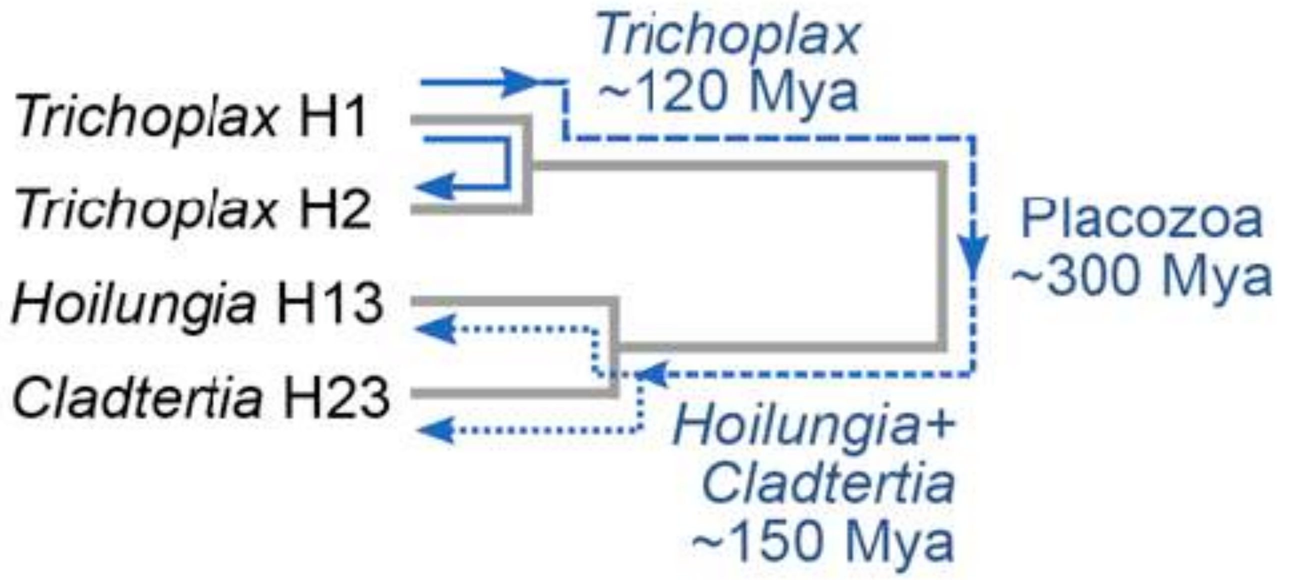
The genetic basis of placozoan cell type gene expression evolution

Some cell types evolve faster than others (gene and RE gains/losses are correlated)



Degree of conservation of cell identity determinants with phylogenetic divergence





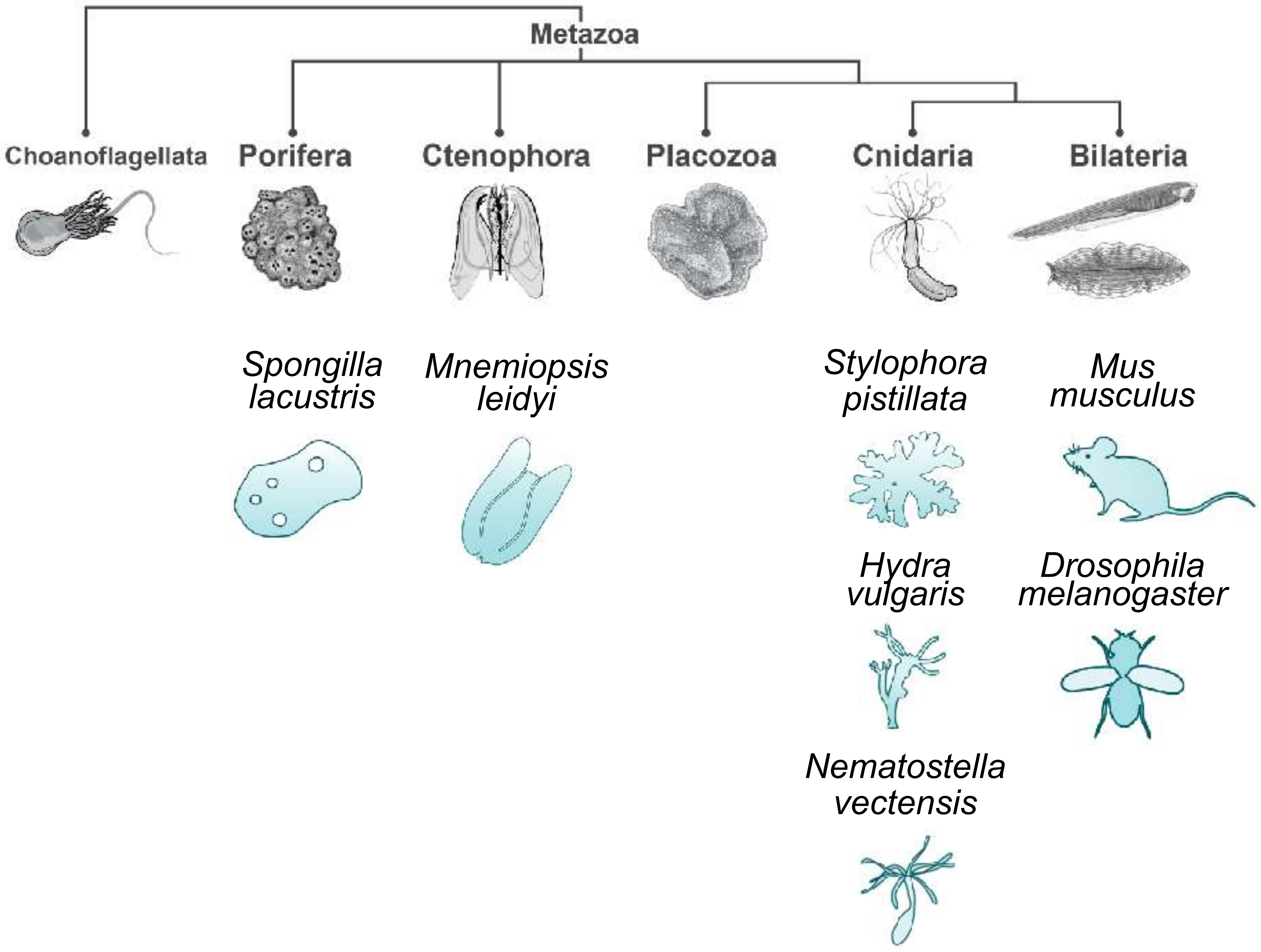
Evidence for a common evolutionary rate in metazoan transcriptional networks

Anne-Ruxandra Carvunis†, Tina Wang†, Dylan Skola†, Alice Yu, Jonathan Chen, Jason F Kreisberg, Trey Ideker*

Department of Medicine, University of California, San Diego, La Jolla, United States

Cell type transcriptome **macroevolutionary comparisons**

Cross-phyla cell type comparisons
using published whole-organism cell atlases



Expression Conservation scores (EC) via Iterative Comparison of Coexpression (ICC)

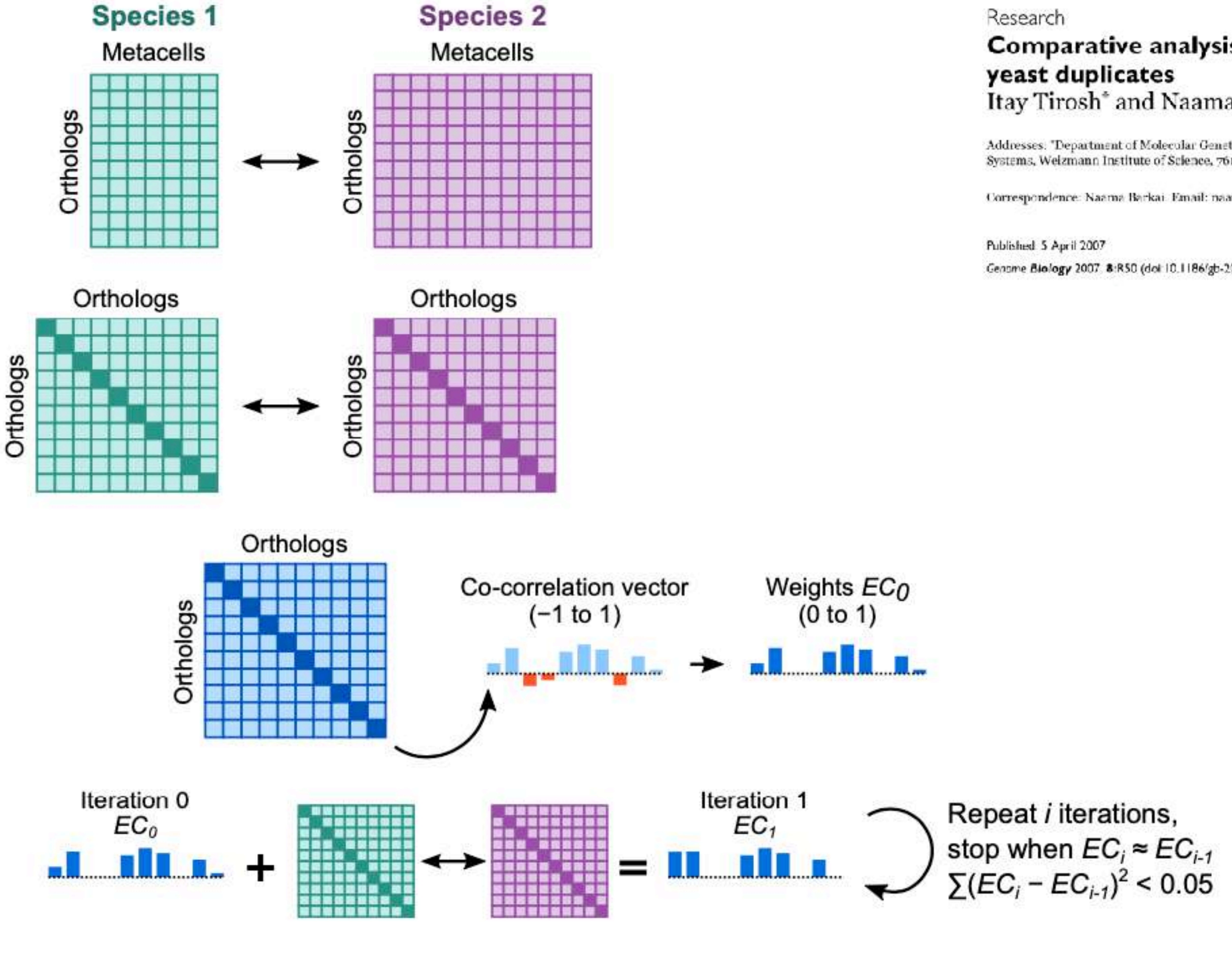
A

1. Expression matrices
Conditions are unmatched
Orthologs are matched
Only one-to-one orthologs

2. One-to-one ortholog correlation matrices
Orthologs are matched
Pearson correlation

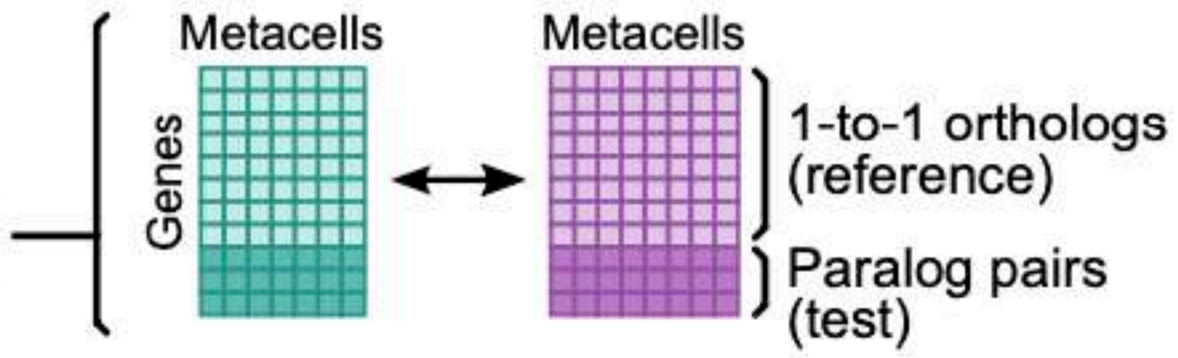
3. Correlation between correlation values for ortholog pairs, to obtain pair-specific weight vector (range: 0 to 1), or expression conservation vector EC_0

4. Weighted Pearson correlation between species correlation matrices, using weights from initial iteration (EC_0), resulting in a new vector (EC_1). Repeat for i iterations until $EC_i \approx EC_{i-1}$



5. Expression conservation scores for each ortholog pair correspond to the final EC_i vector.

6. Best paralog selection: repeat steps one through five adding sets of paralogous gene pairs to the one-to-one ortholog matrices, and selecting the pair of paralogs with the highest pairwise EC score.



Research

Open Access

Comparative analysis indicates regulatory neofunctionalization of yeast duplicates

Itay Tirosh* and Naama Barkai†

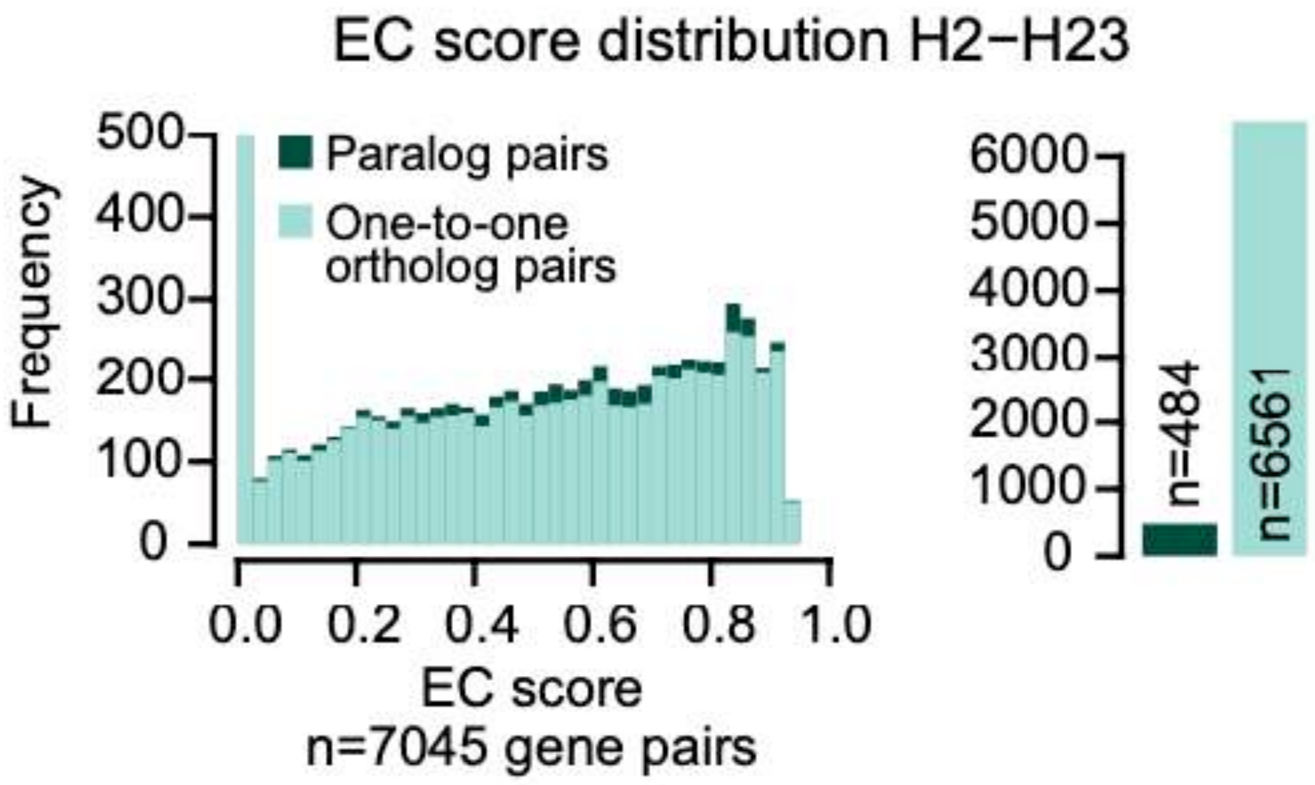
Addresses: *Department of Molecular Genetics, Weizmann Institute of Science, 76100 Rehovot, Israel. †Department of Physics of Complex Systems, Weizmann Institute of Science, 76100 Rehovot, Israel.

Correspondence: Naama Barkai. Email: naama.barkai@weizmann.ac.il

Published: 5 April 2007

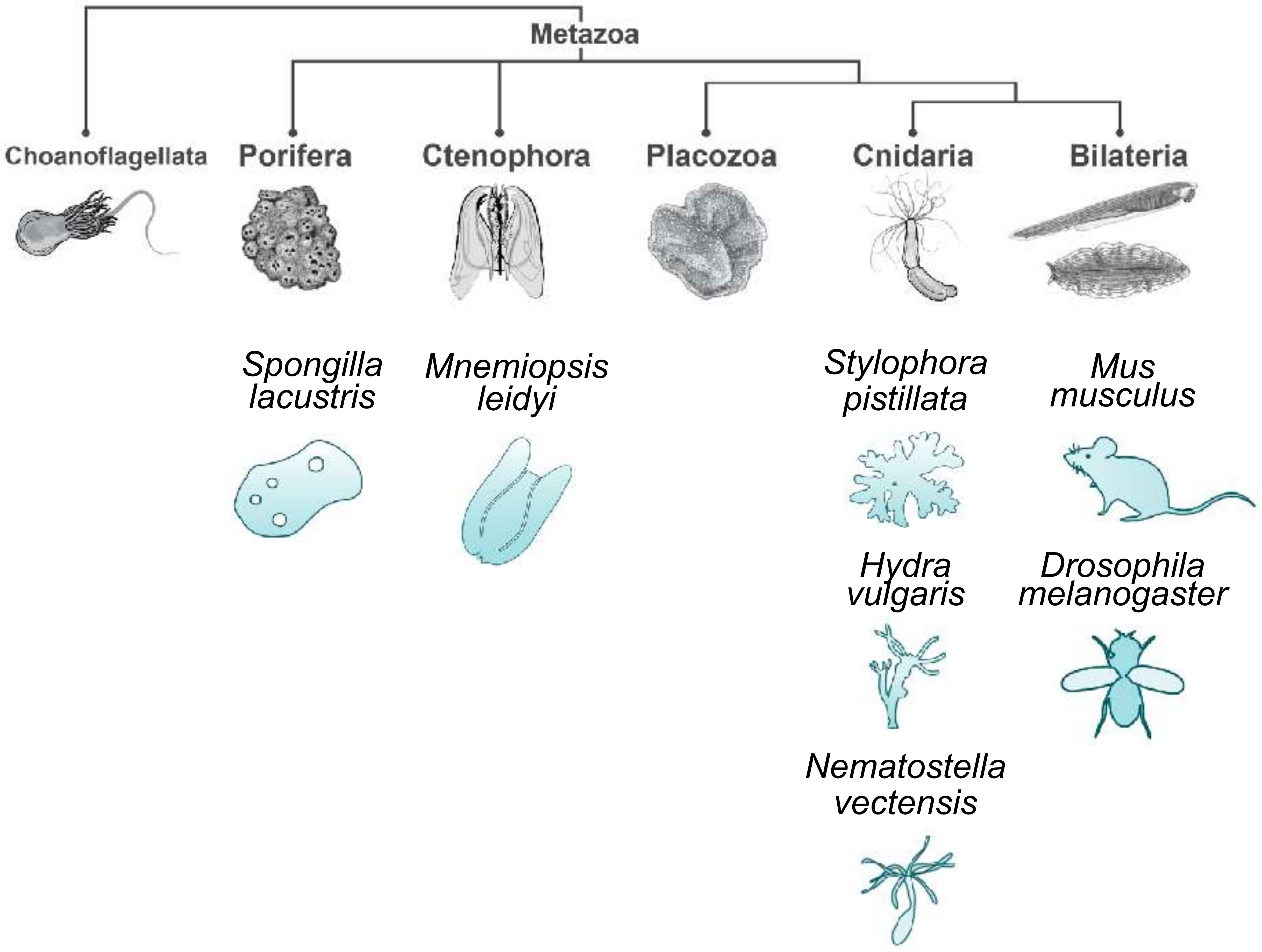
Genome **Biology** 2007, **8**:R50 (doi:10.1186/gb-2007-8-4-r50)

Received: 21 December 2006
Revised: 15 February 2007
Accepted: 5 April 2007

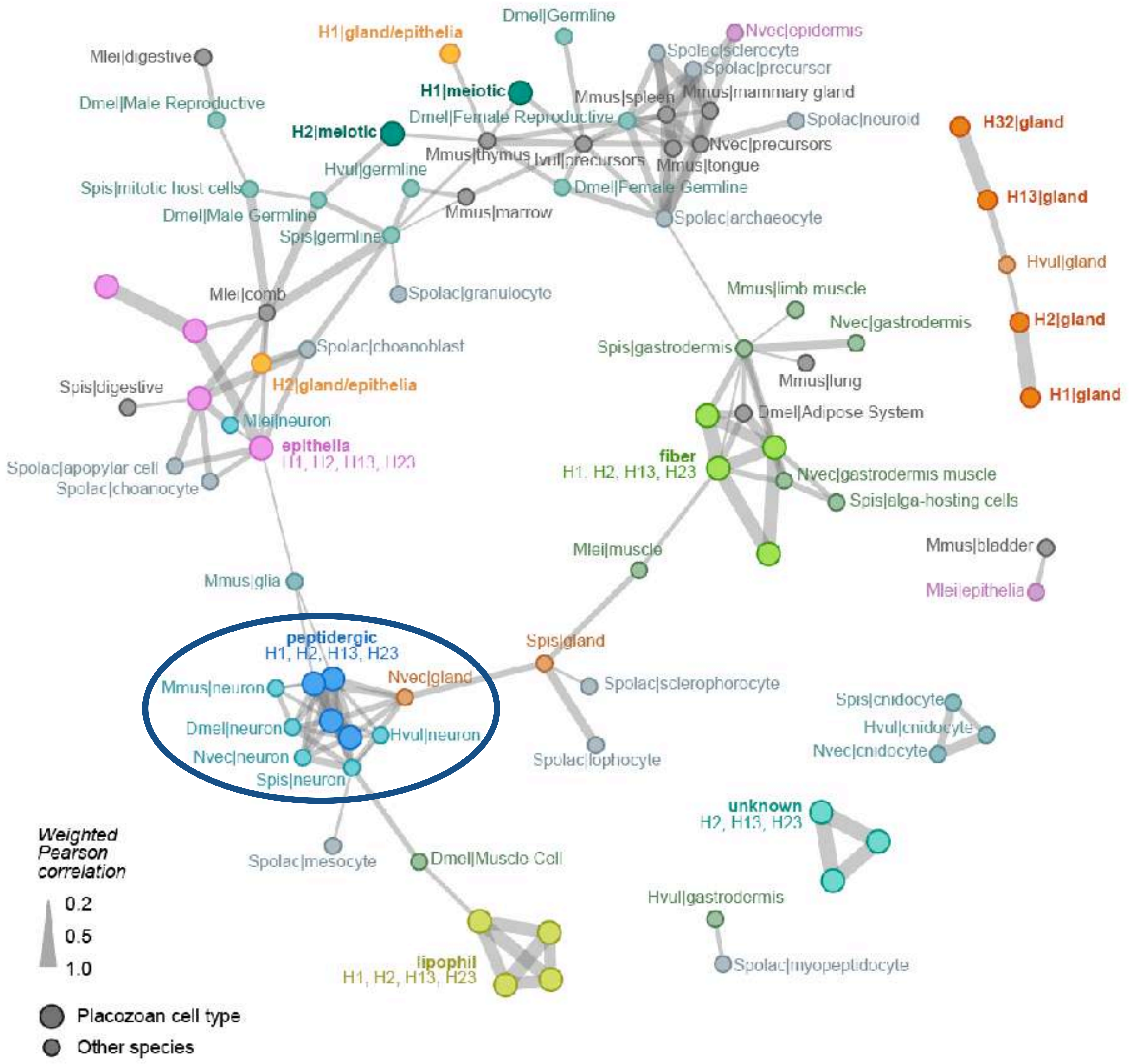


Cell type transcriptome **macroevolutionary comparisons**

Cross-phyla cell type comparisons using published whole-organism cell atlases

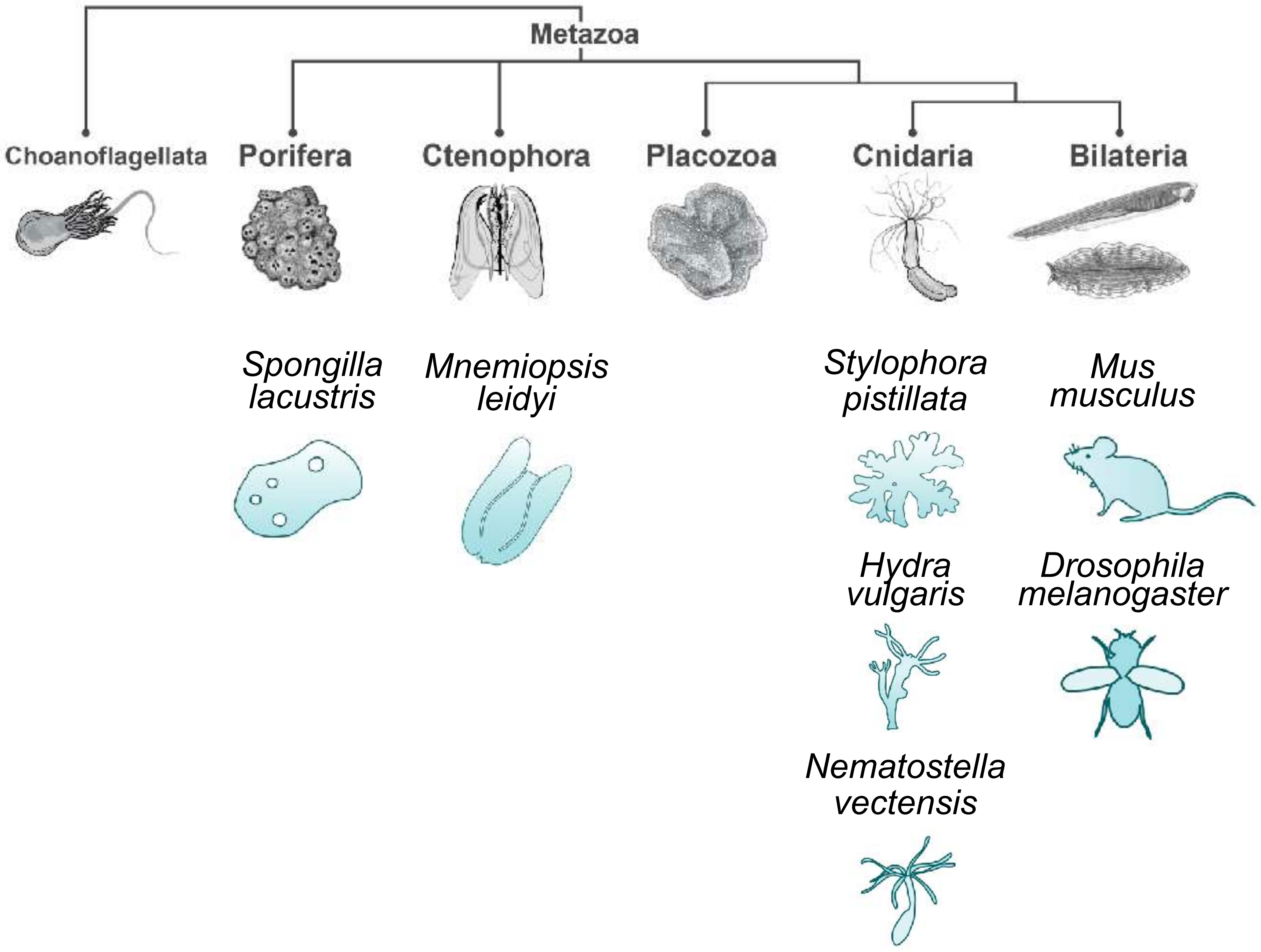


Peptidergic cells transcriptionally resemble neurons

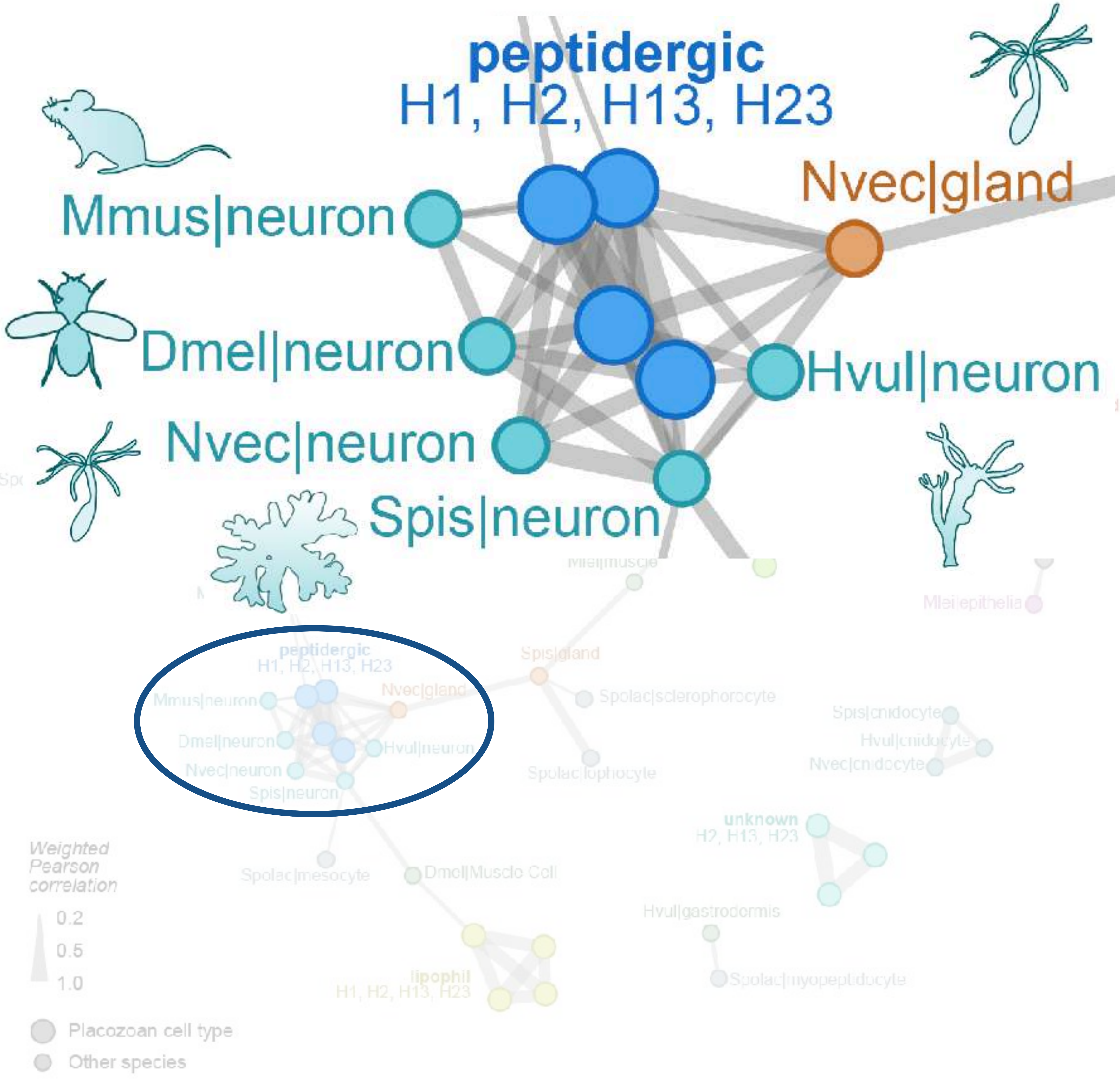


Cell type transcriptome **macroevolutionary comparisons**

Cross-phyla cell type comparisons
using published whole-organism cell atlases

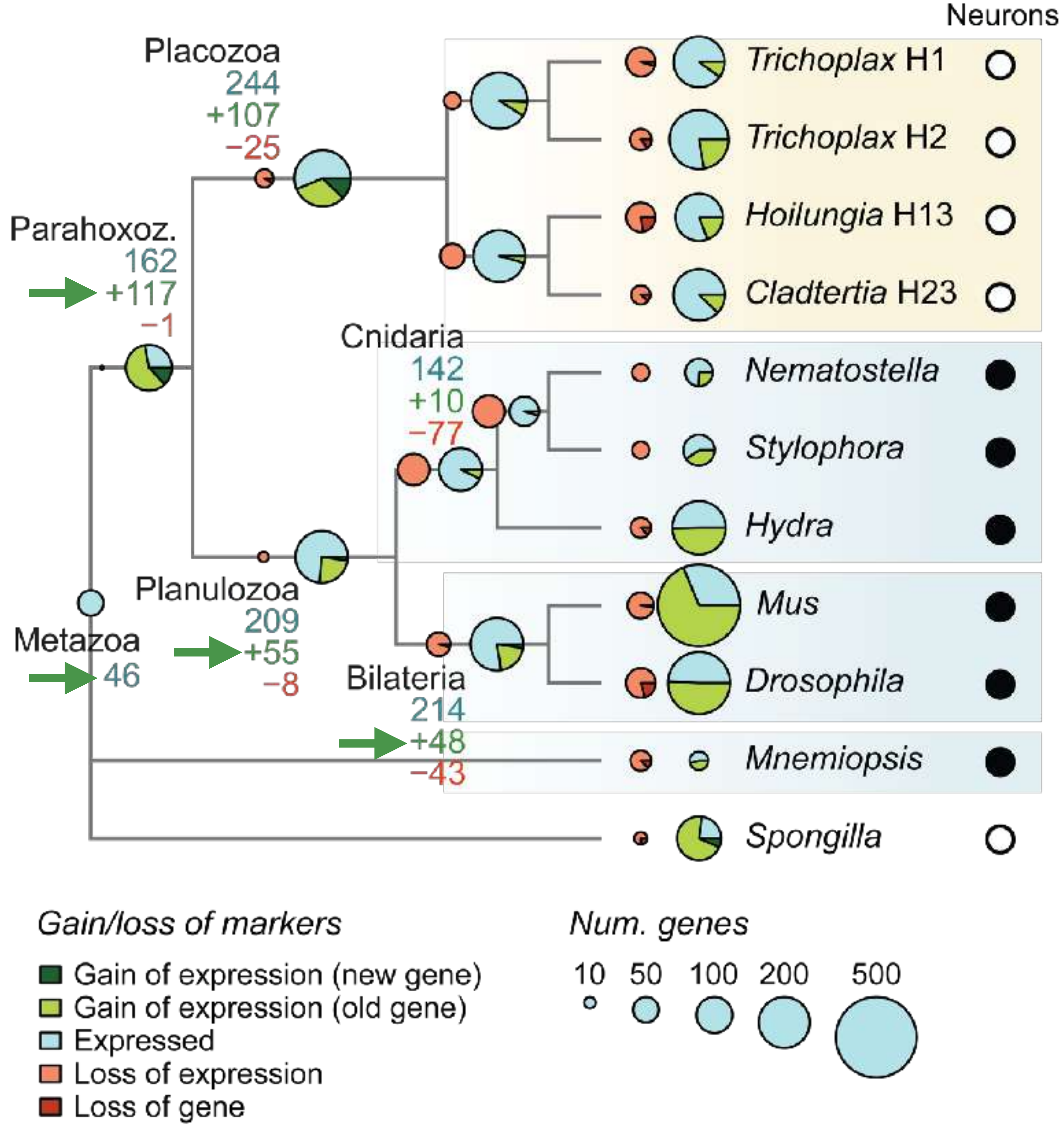


Peptidergic cells transcriptionally resemble neurons

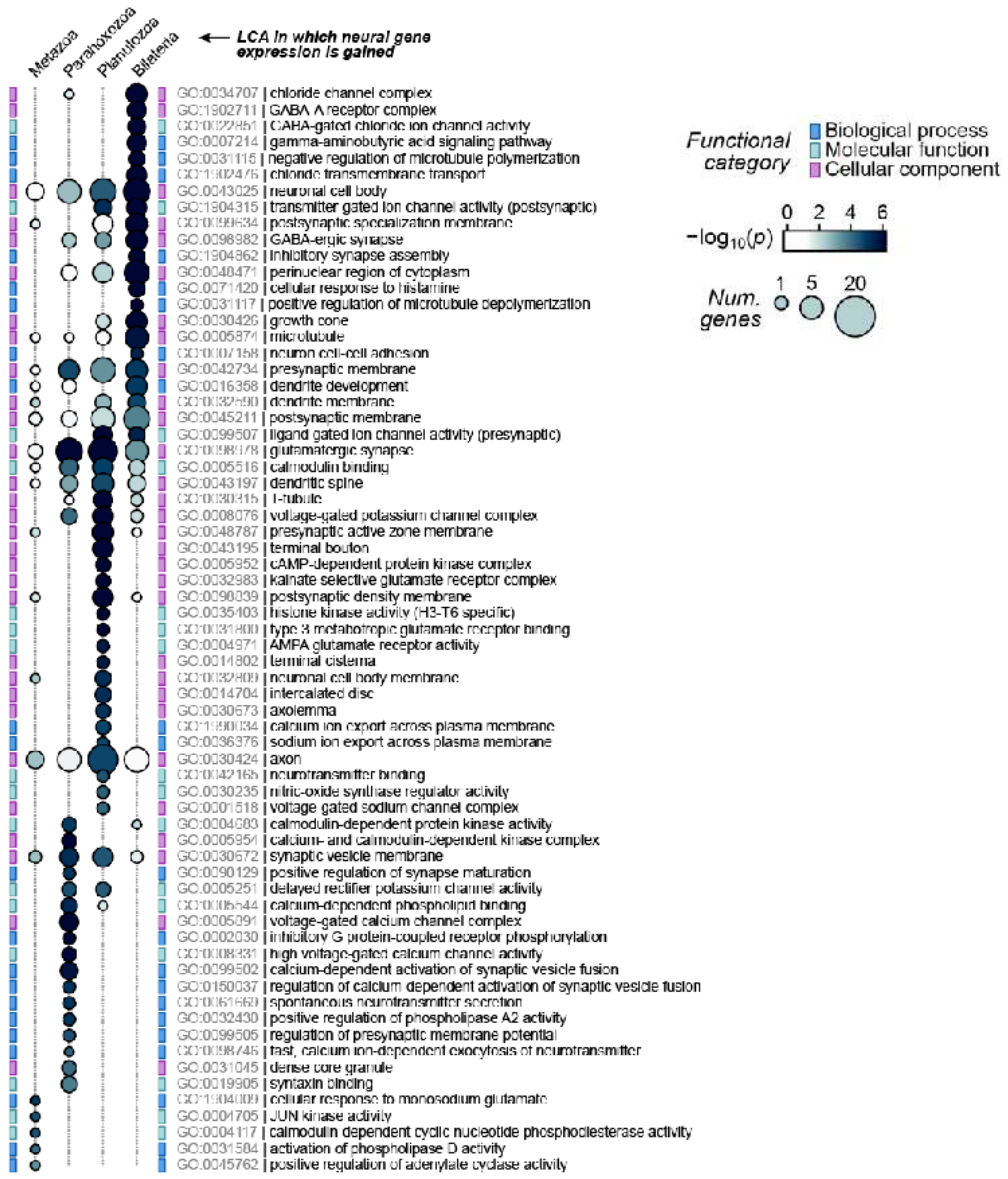


Evolution of the neuronal gene expression program

Reconstruction of gene expression ancestral states, losses and novelties in neurons/neuronal-like cells



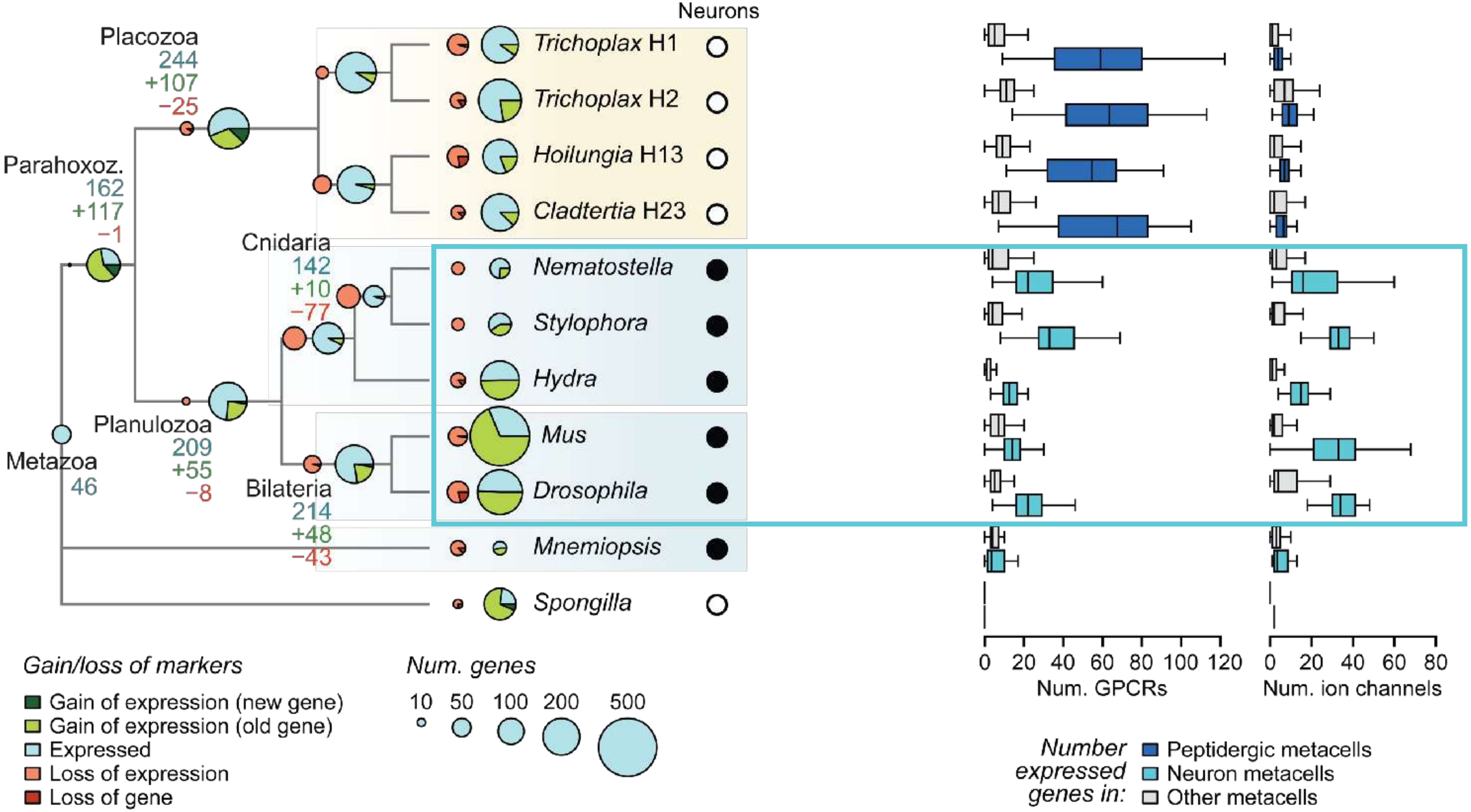
Gene gains enriched functional categories



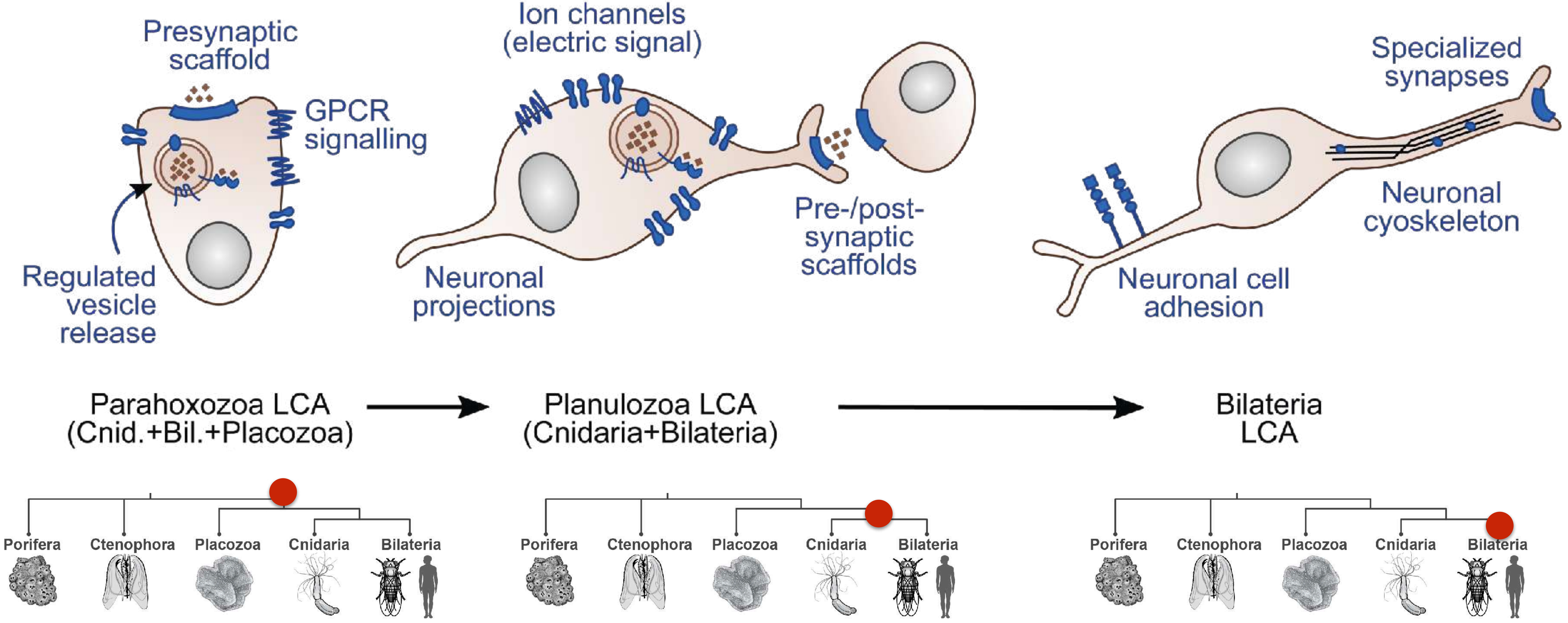
Stepwise evolutionary emergence of the neuronal gene expression program

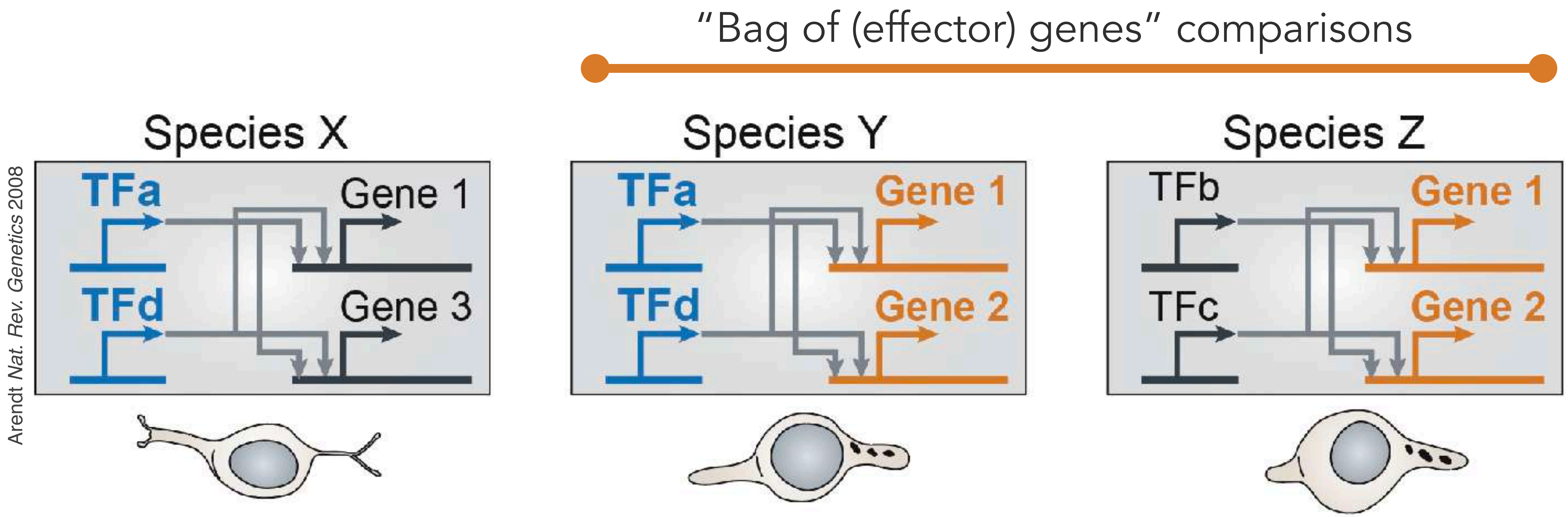
Reconstruction of gene expression ancestral states, losses and novelties in neurons/neuronal-like cells

High GPCR and Ion Channel gene counts is a hallmark of cnidarian and bilaterian neurons



Evolution of the neuronal gene expression program

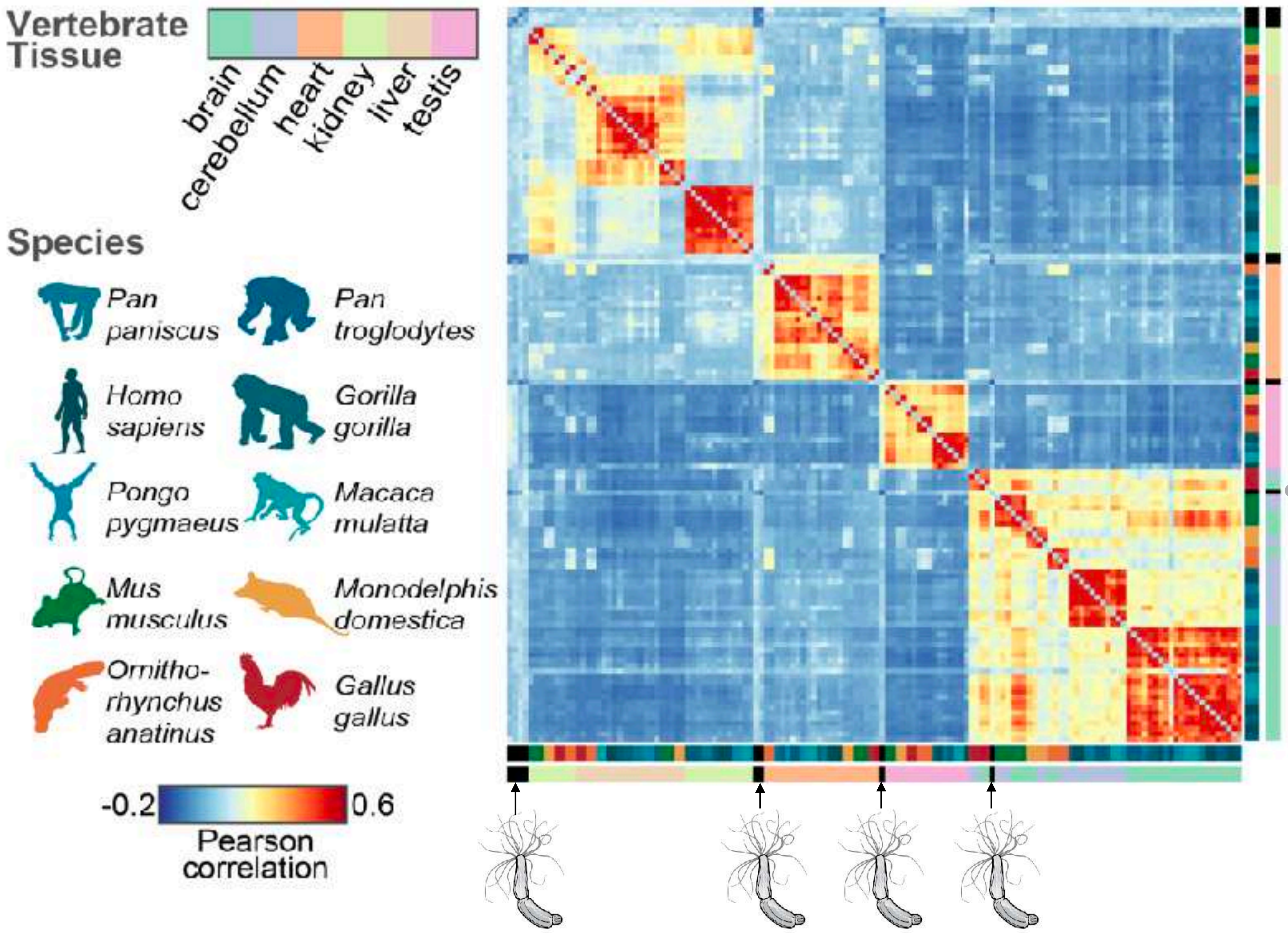


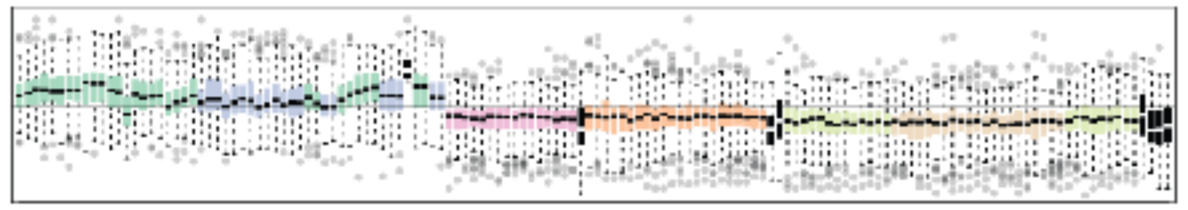


1. Functional constraints -> convergent (and divergent) gene usage.
2. We don't apply explicit evolutionary models for gene expression characters.
3. Genes are not independent characters.

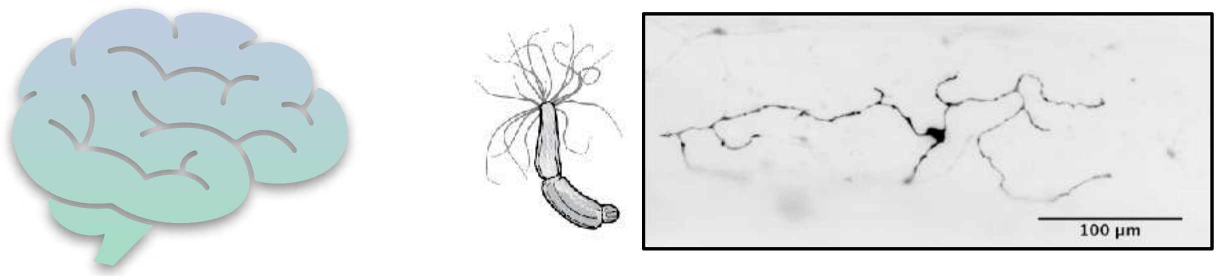
Cell type macroevolution, **similarity beyond form and function**

Sebe-Pedros et al., *Cell* 2018





Term	Description	Population	Set	log2FC
GO:0005244	voltage-gated ion channel activity	22	12	2.5
GO:0046058	cAMP metabolic process	17	9	2.5
GO:0022839	ion gated channel activity	28	14	2.4
GO:0070382	exocytic vesicle	26	12	2.3
GO:0098794	postsynapse	46	17	1.9
GO:0004930	G-protein coupled receptor activity	69	23	1.8
GO:0098793	presynapse	56	18	1.7
GO:0030425	dendrite	61	19	1.7
GO:0045202	synapse	112	33	1.6
GO:0005216	ion channel activity	51	15	1.6
GO:0007268	chemical synaptic transmission	55	16	1.6
GO:0036477	somatodendritic compartment	87	24	1.5
GO:0043005	neuron projection	146	33	1.2
GO:0097458	neuron part	188	41	1.2
GO:0030182	neuron differentiation	164	34	1.1
GO:0048699	generation of neurons	181	35	1.2
GO:0022008	neurogenesis	191	36	1.0
GO:0007399	nervous system development	284	54	1.0

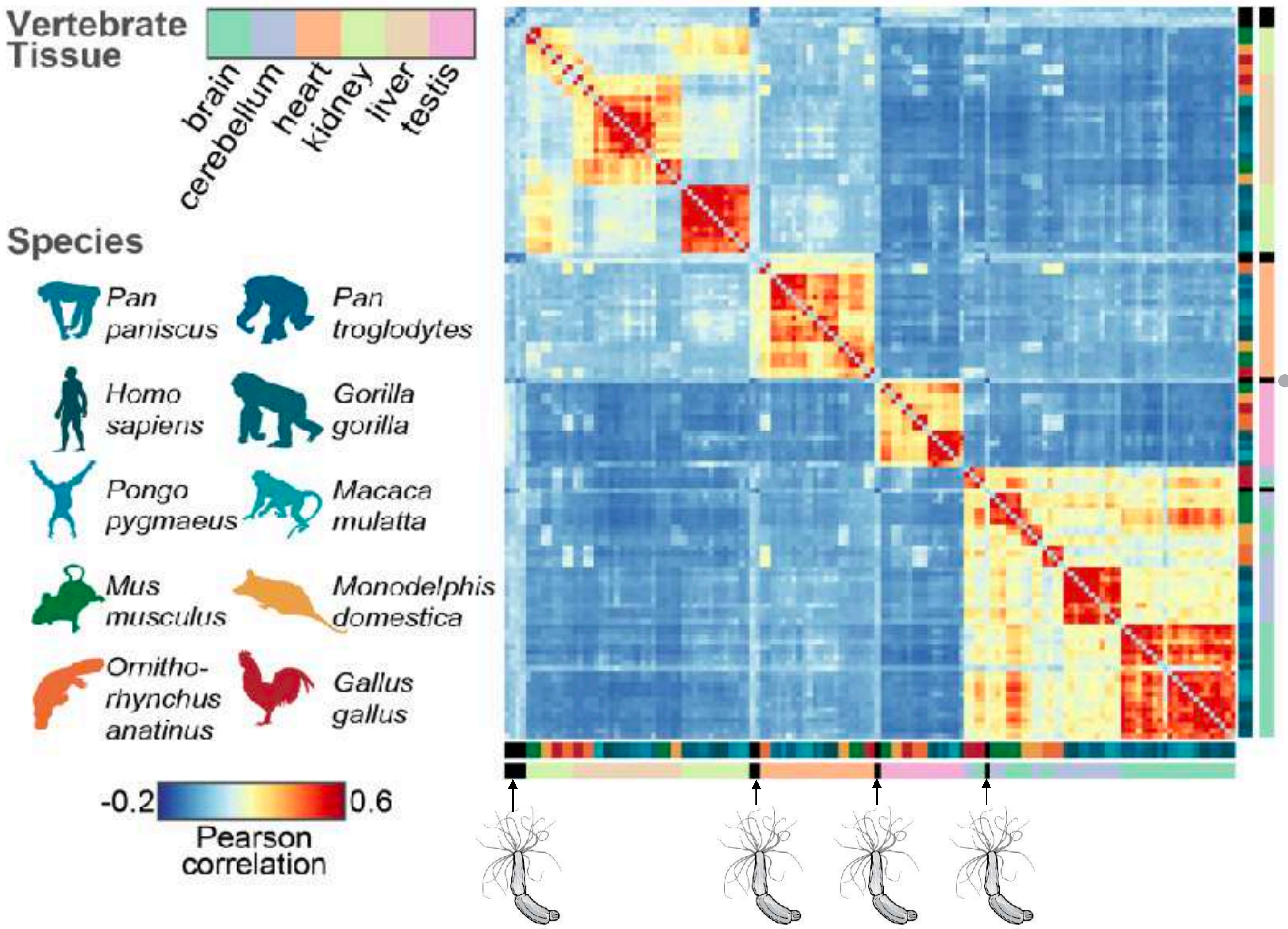


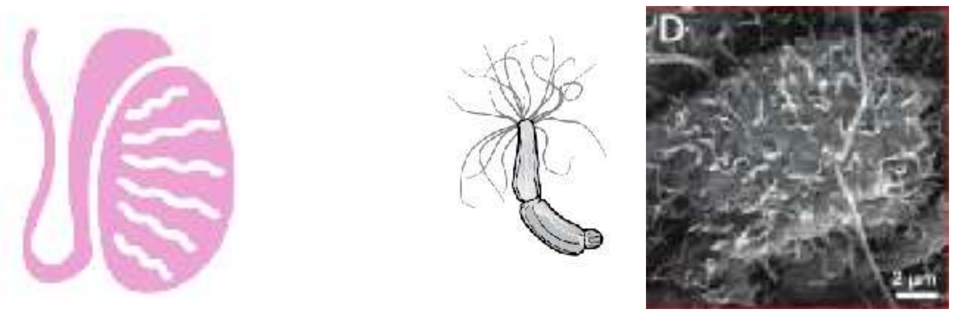
Brain + *Nematostella* neurons
(similarity driven by neuronal genes)

Cnidarian neurons (transcriptionally) resemble vertebrate brain/cerebellum

Cell type macroevolution, **similarity beyond form and function**

Sebe-Pedros et al., *Cell* 2018

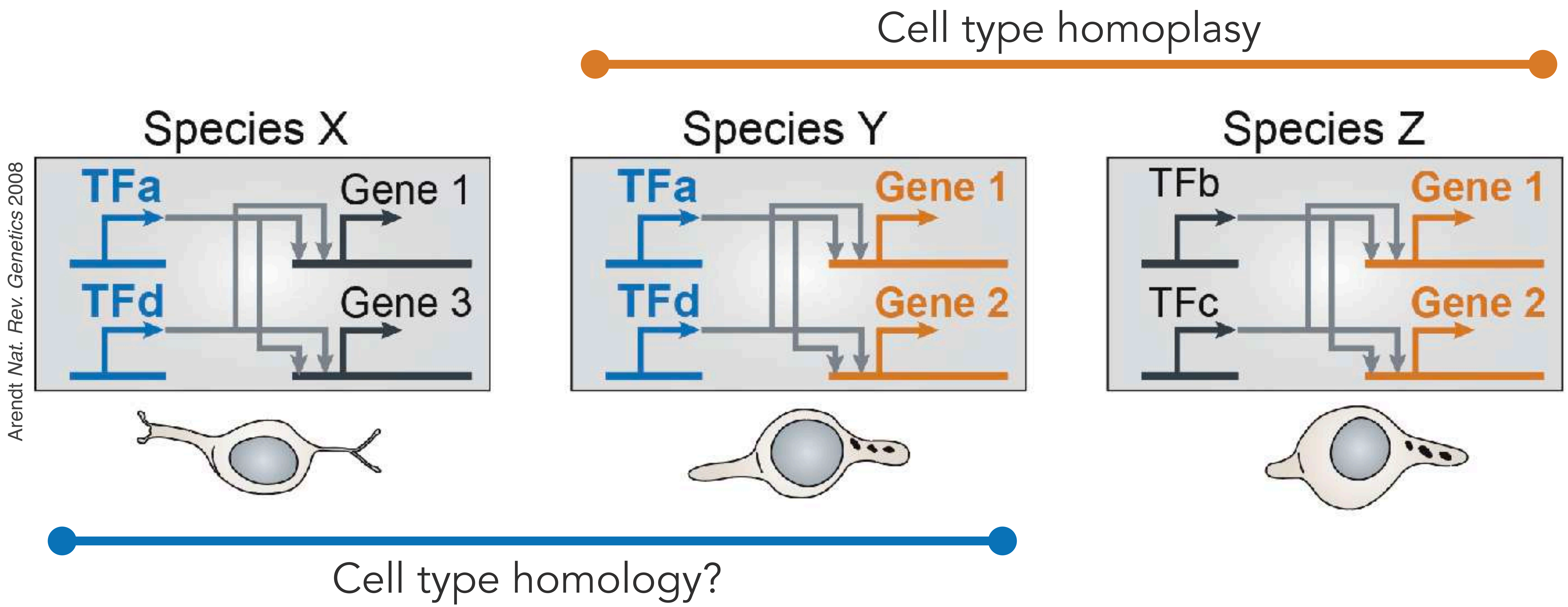




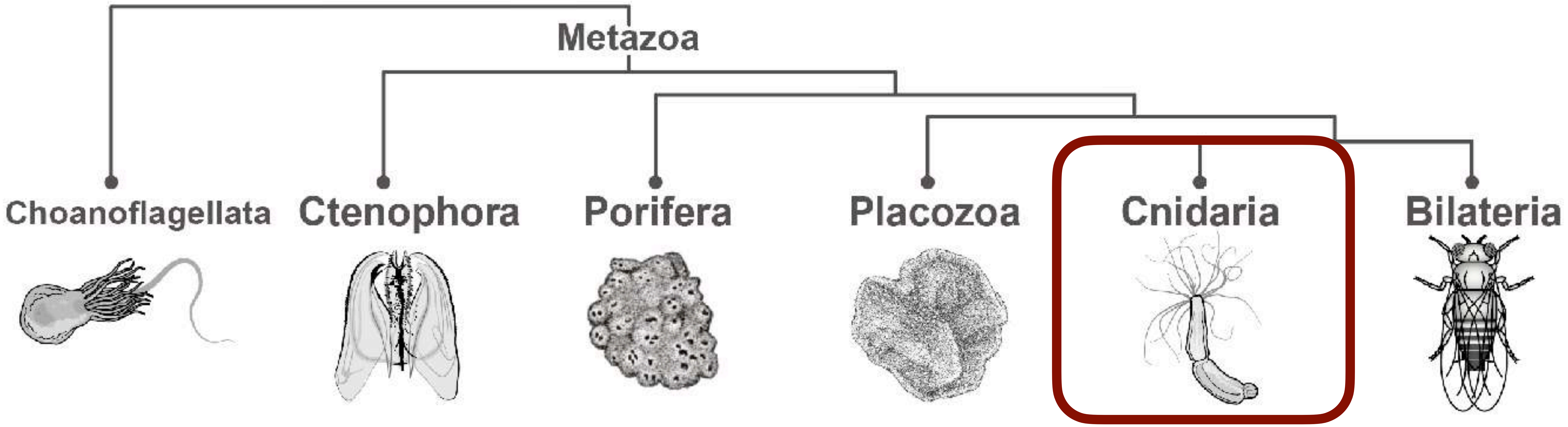
Testis + *Nematostella* gut cells
(similarity driven by ciliary genes)

Direct comparisons of cell type transcriptomes are confounded
by convergent effector gene usage
(and divergent gene usage, and TF replacement, and more)

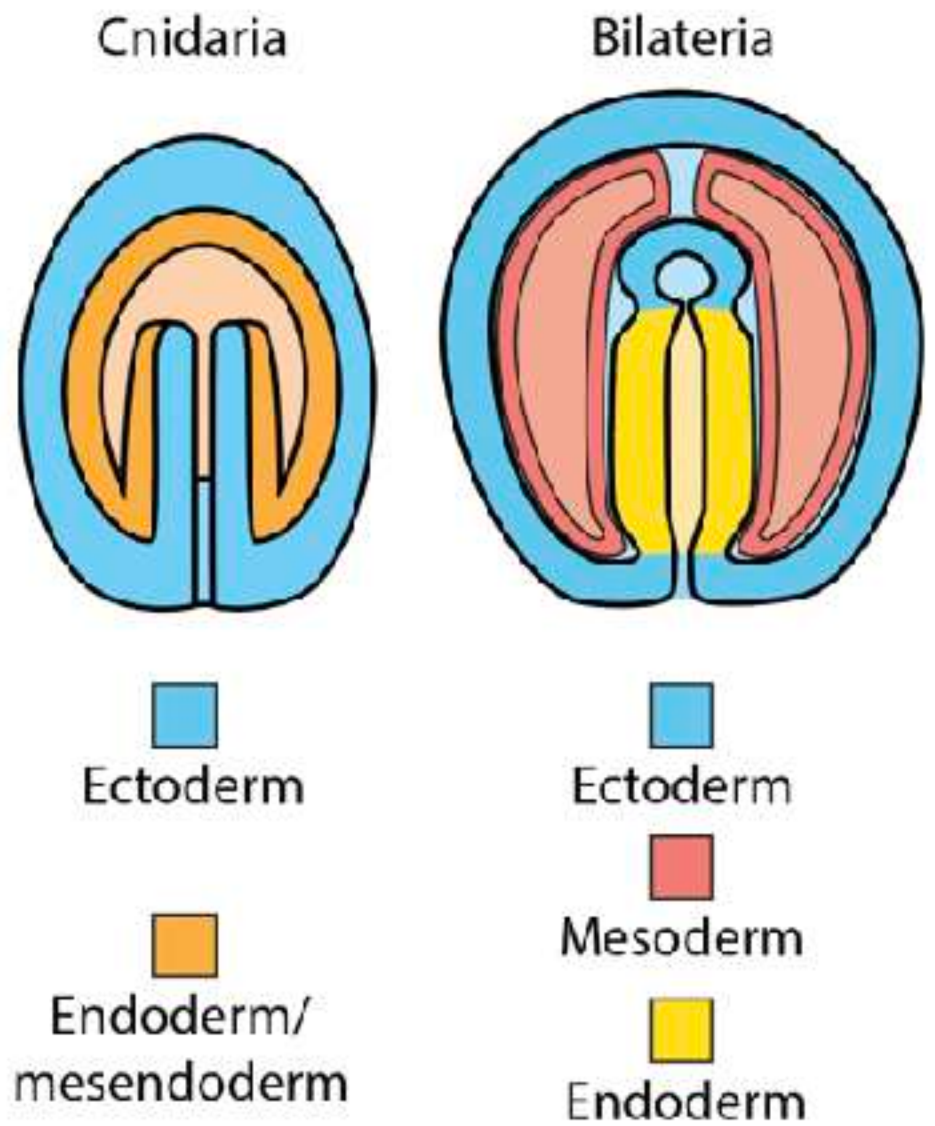
Cell type macroevolution, **similarity beyond form and function**



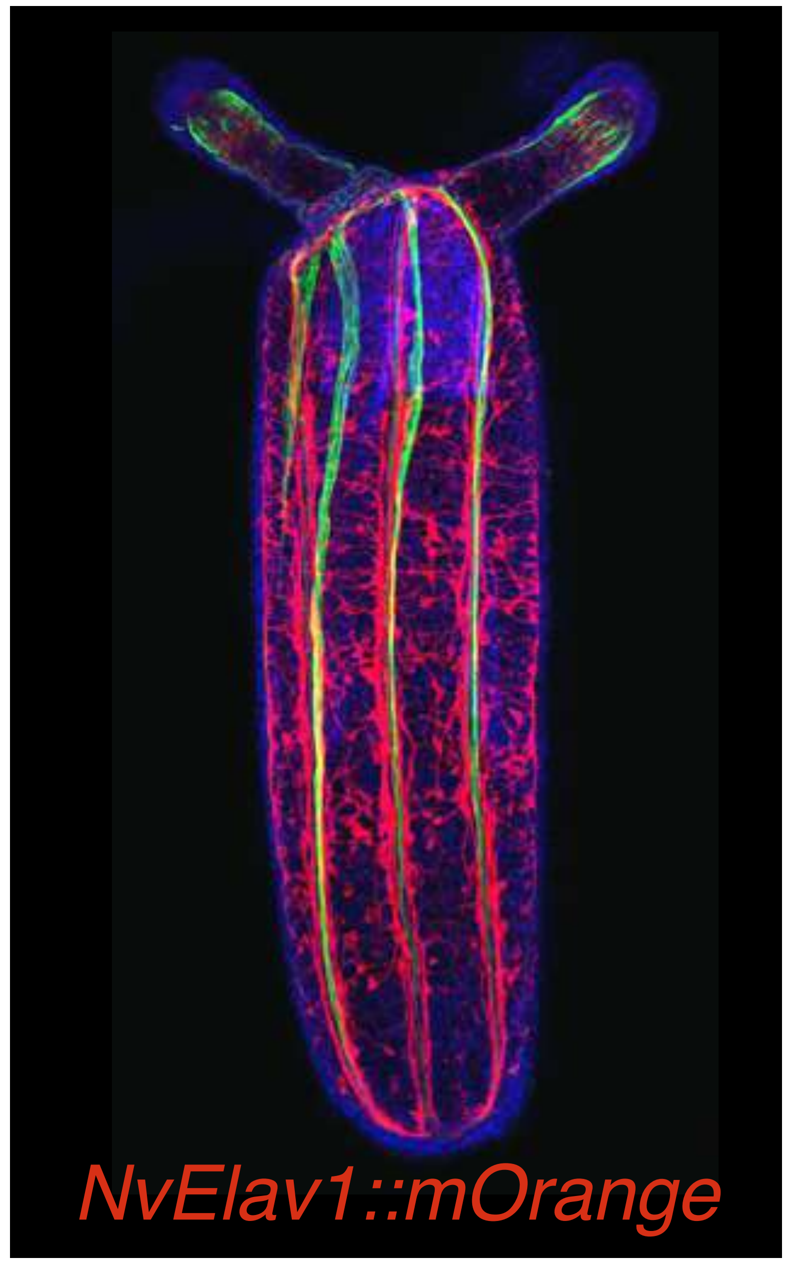
Story 3: Decoding cnidarian cell type regulatory identities



Diploblastic (no mesoderm)

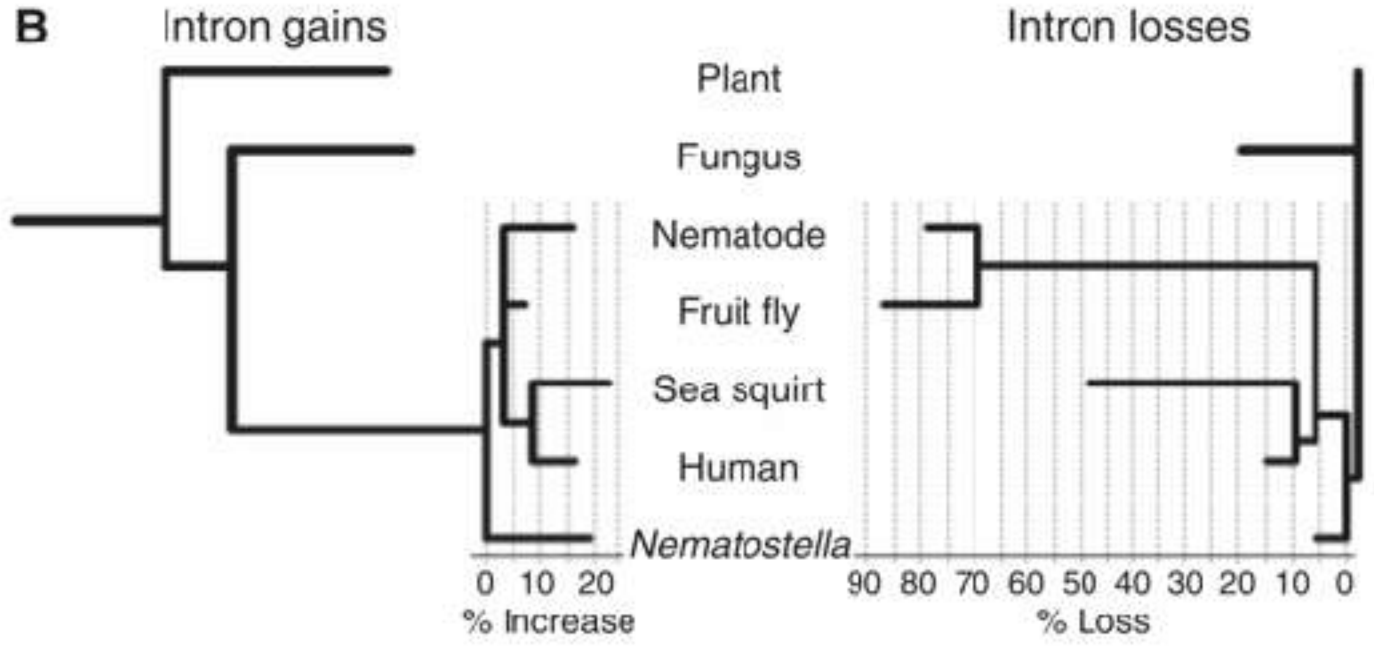


Neurons, but no CNS



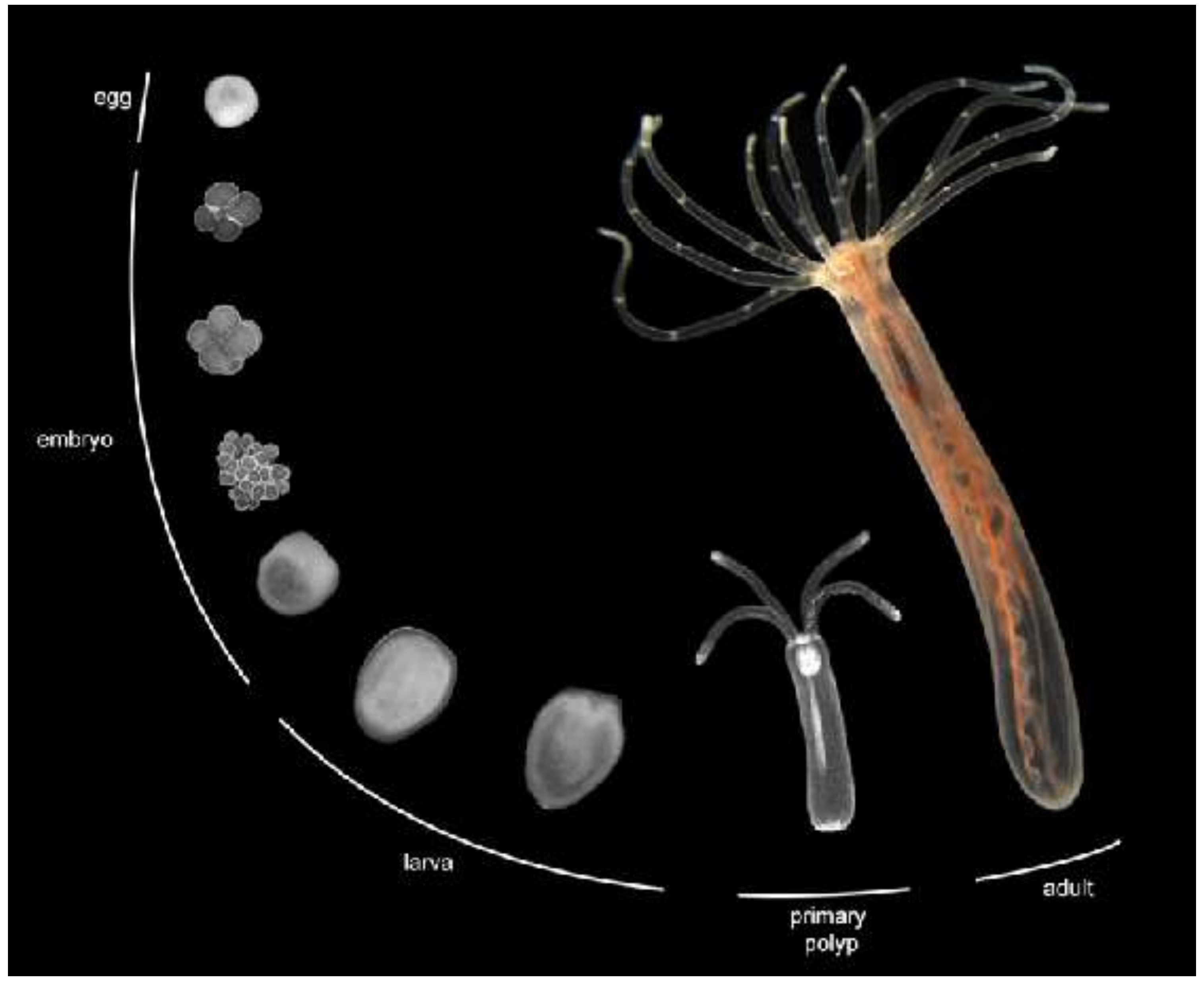
Nakanishi et al. 2012

Slow-evolving genome:
conserved intron positions, syntenic blocks, gene repertoire

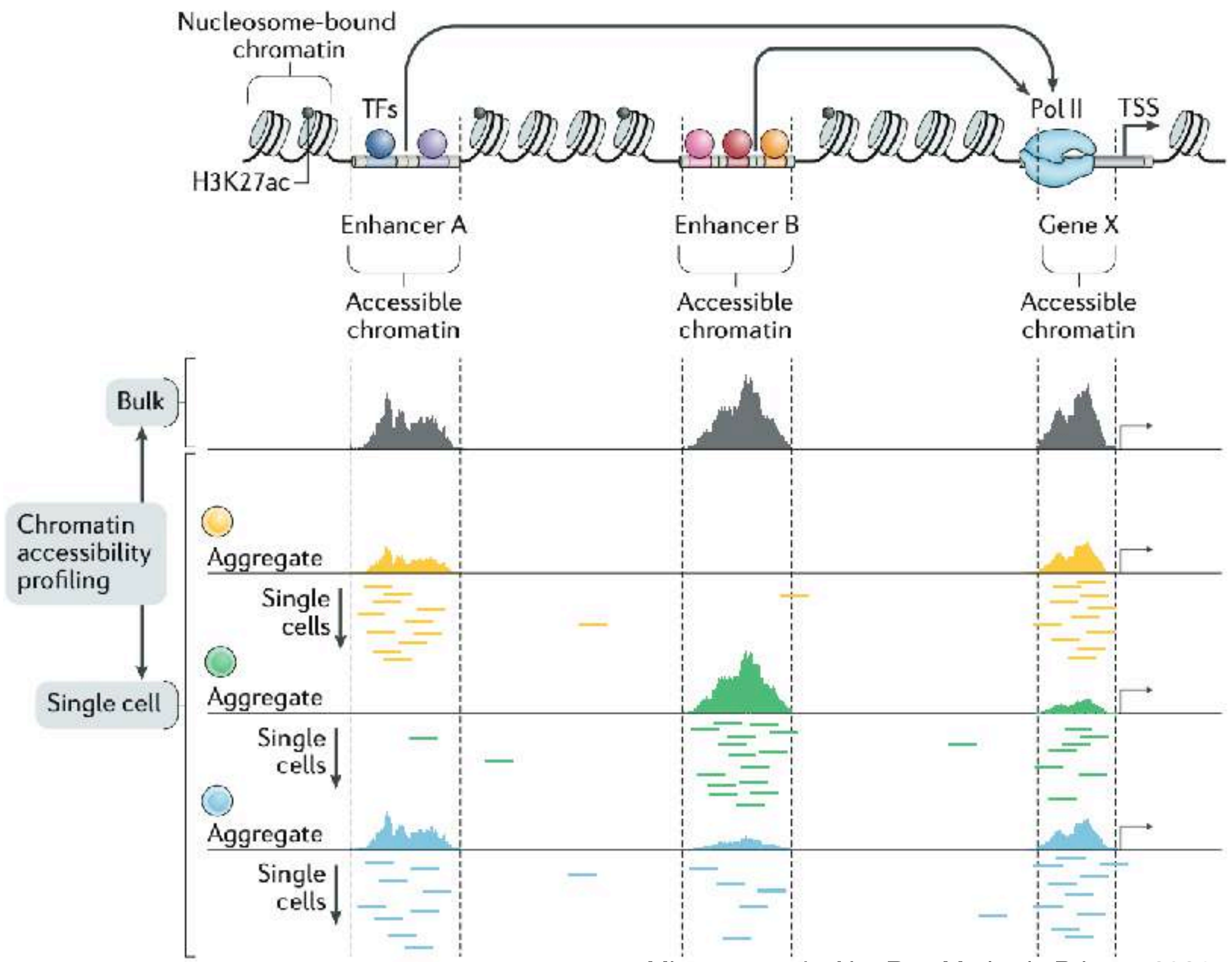


Putnam et al. 2007

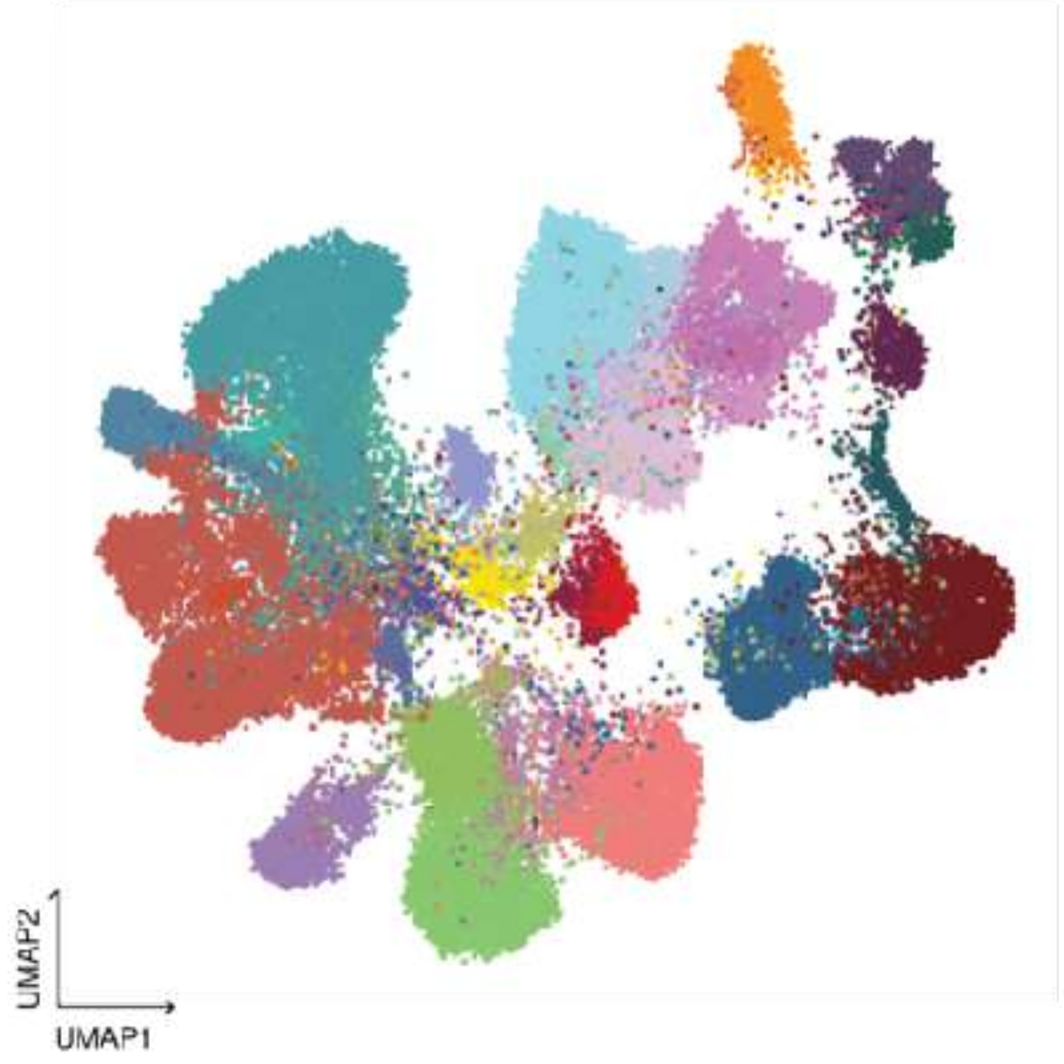
Indirect development (dispersive larva)



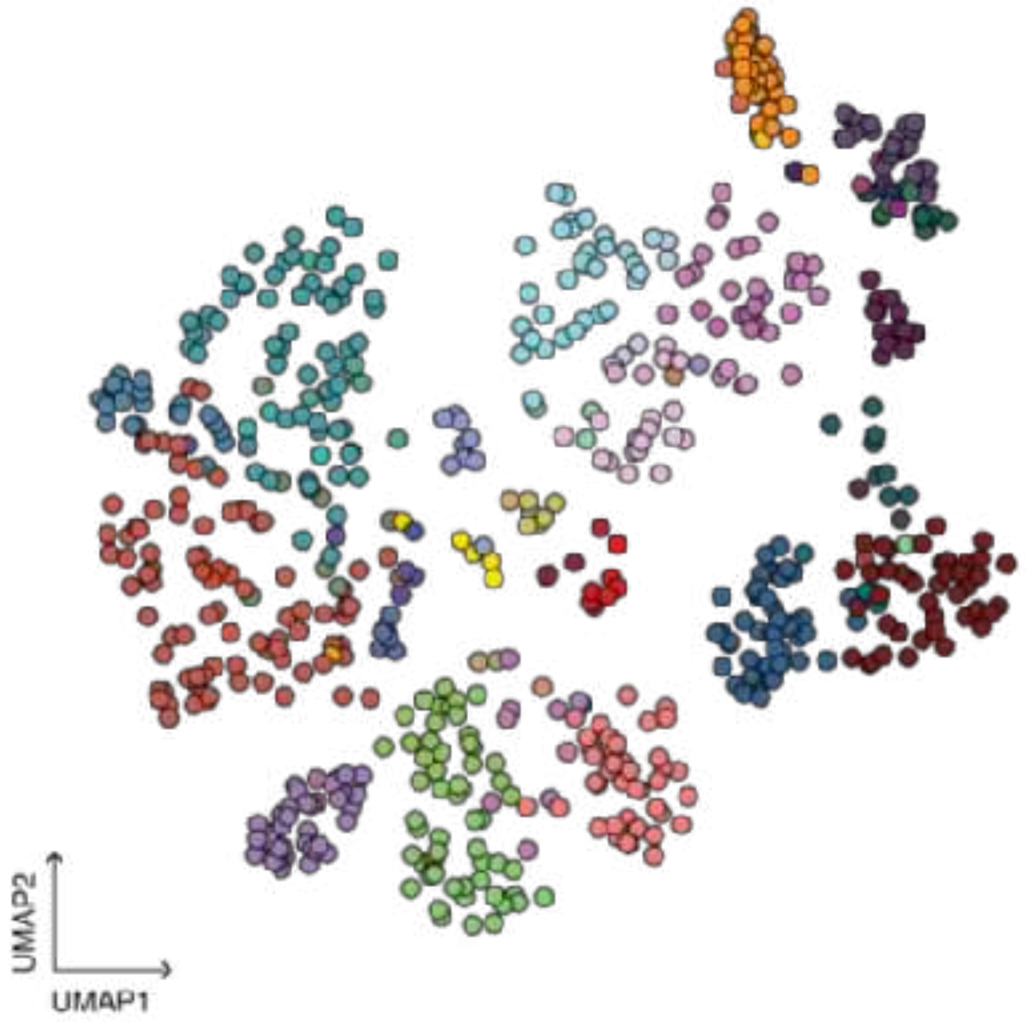
Nematostella single-cell chromatin accessibility atlas



Minnoye et al., *Nat Rev Methods Primers* 2021



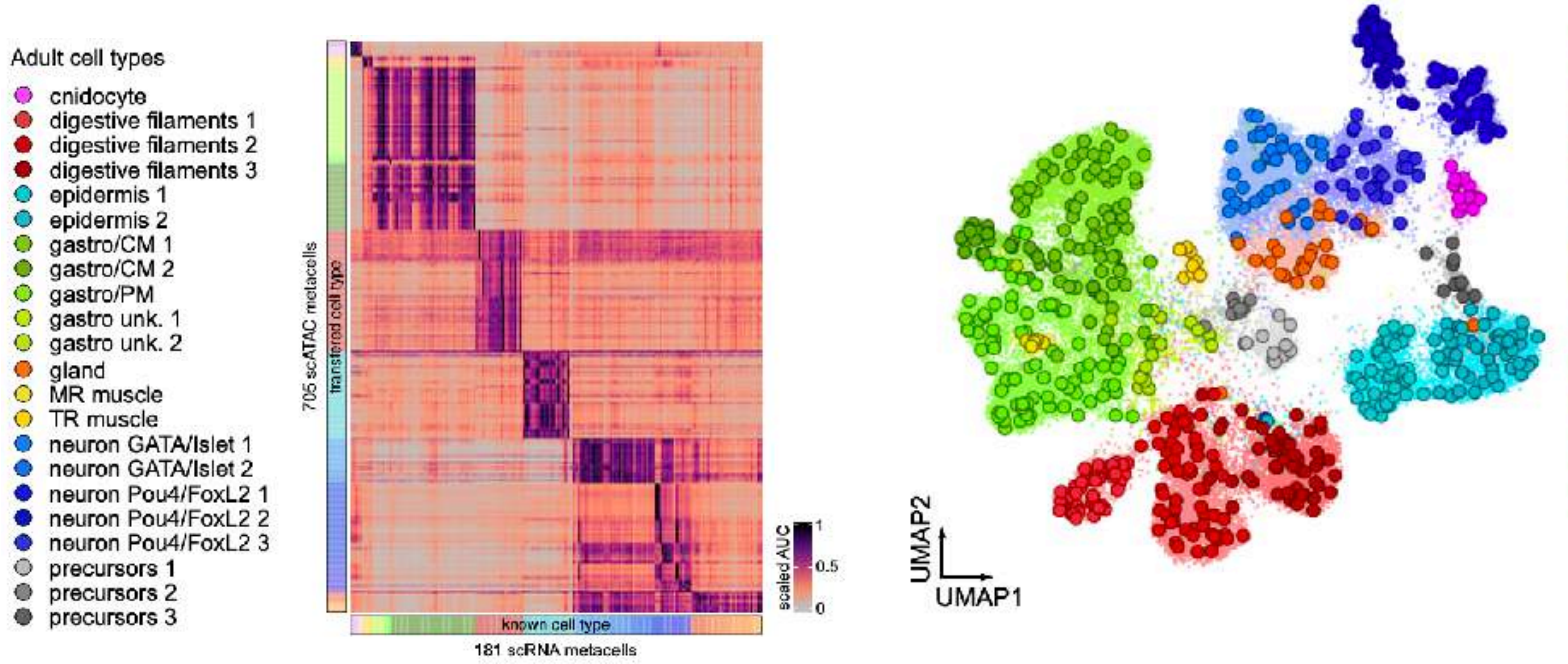
>65,000 scATAC-seq profiles



Reduced into 705 metacells

Anamaria Elek

Marta Iglesias



Annotation transfer from scRNA-seq atlas

Elek, Iglesias, et al., *Nature Ecol&Evol* 2026

Cell type-specific **gene regulatory landscapes**

Cell type specific accessibility, known marker

Different landscape in gastrula and adult

Complex landscape

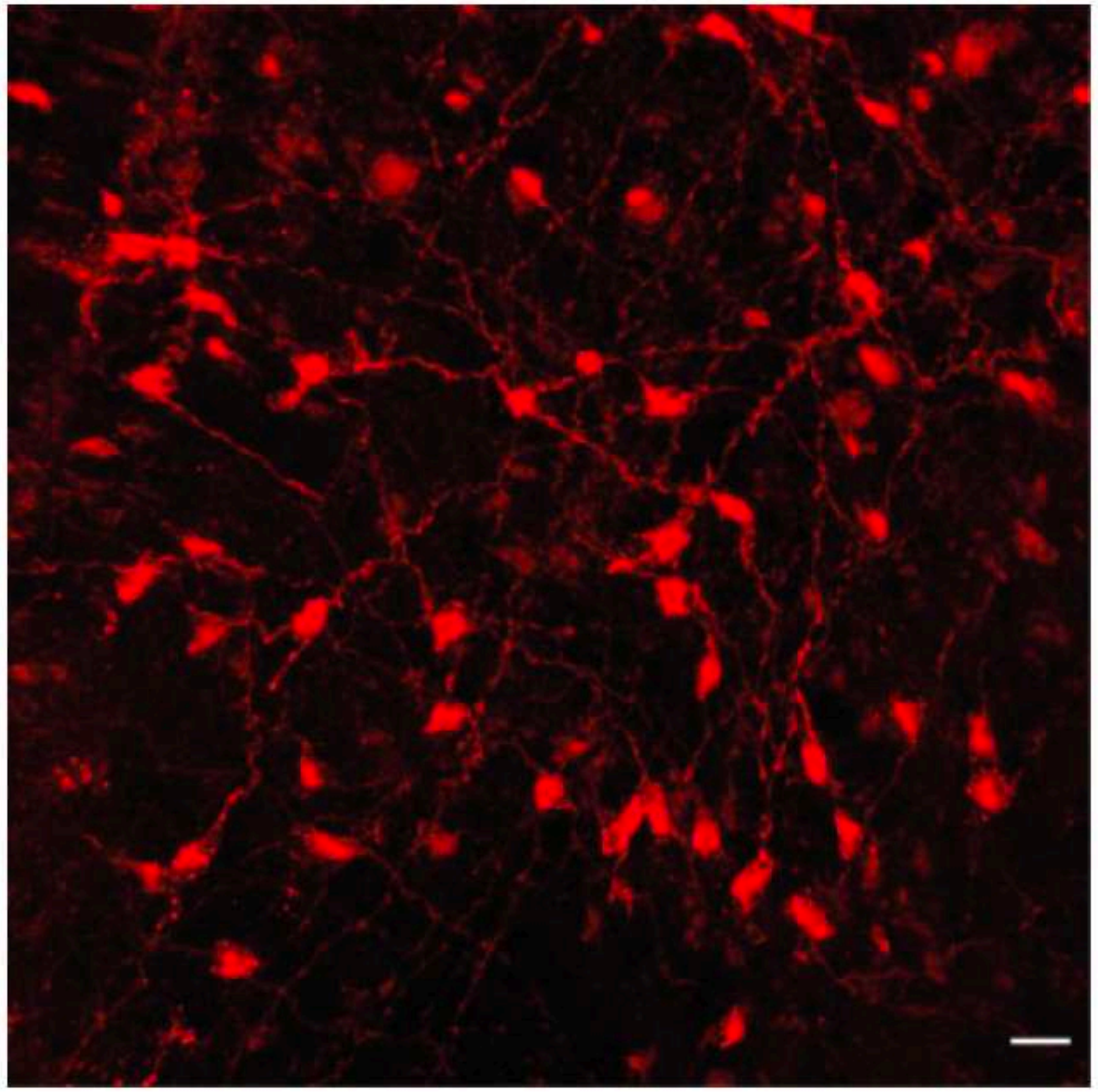
Constitutive promoter (CP)

Alternative promoter (AP)

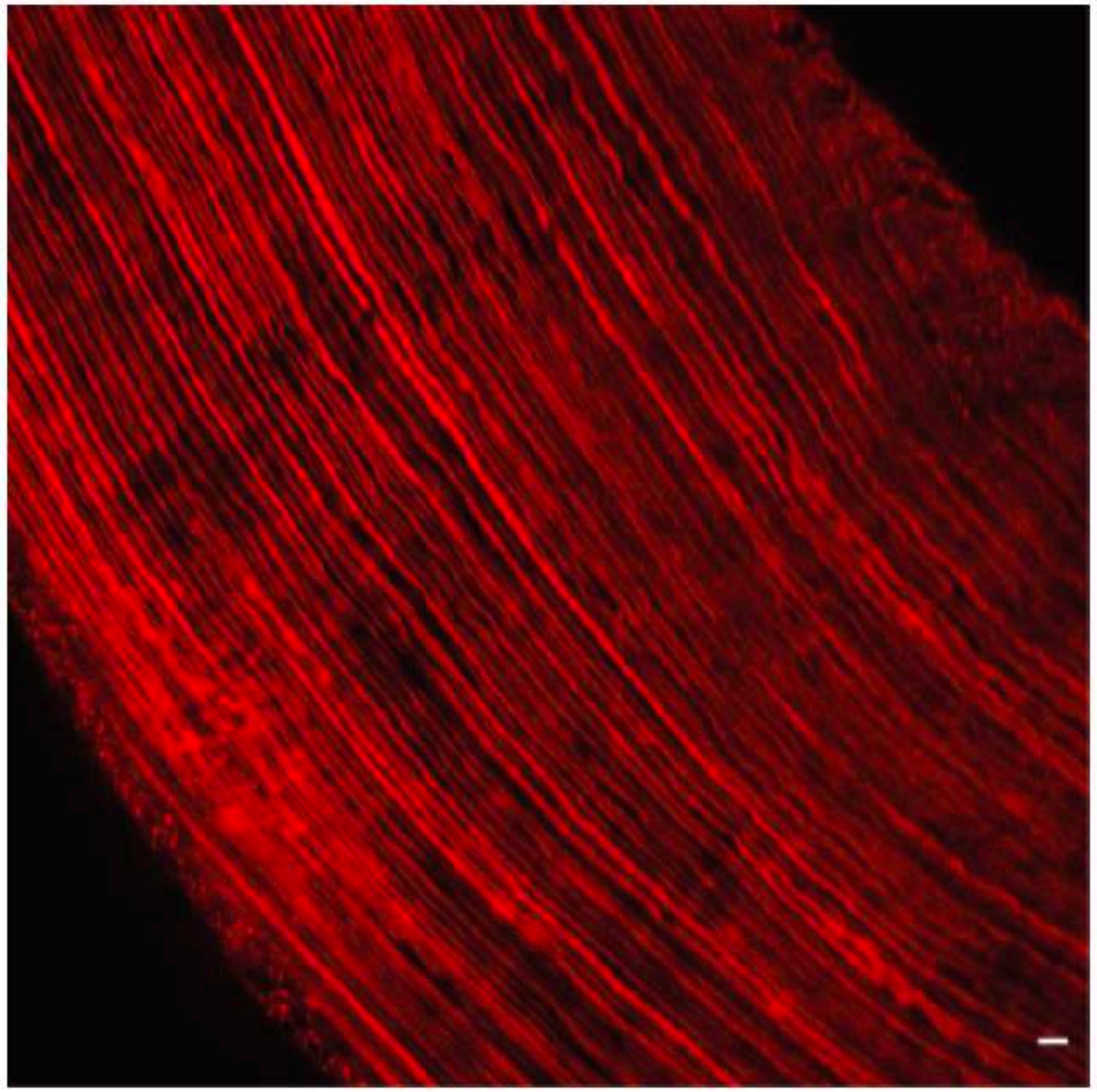
Elek, Iglesias, et al., *Nature Ecol&Evol* 2026

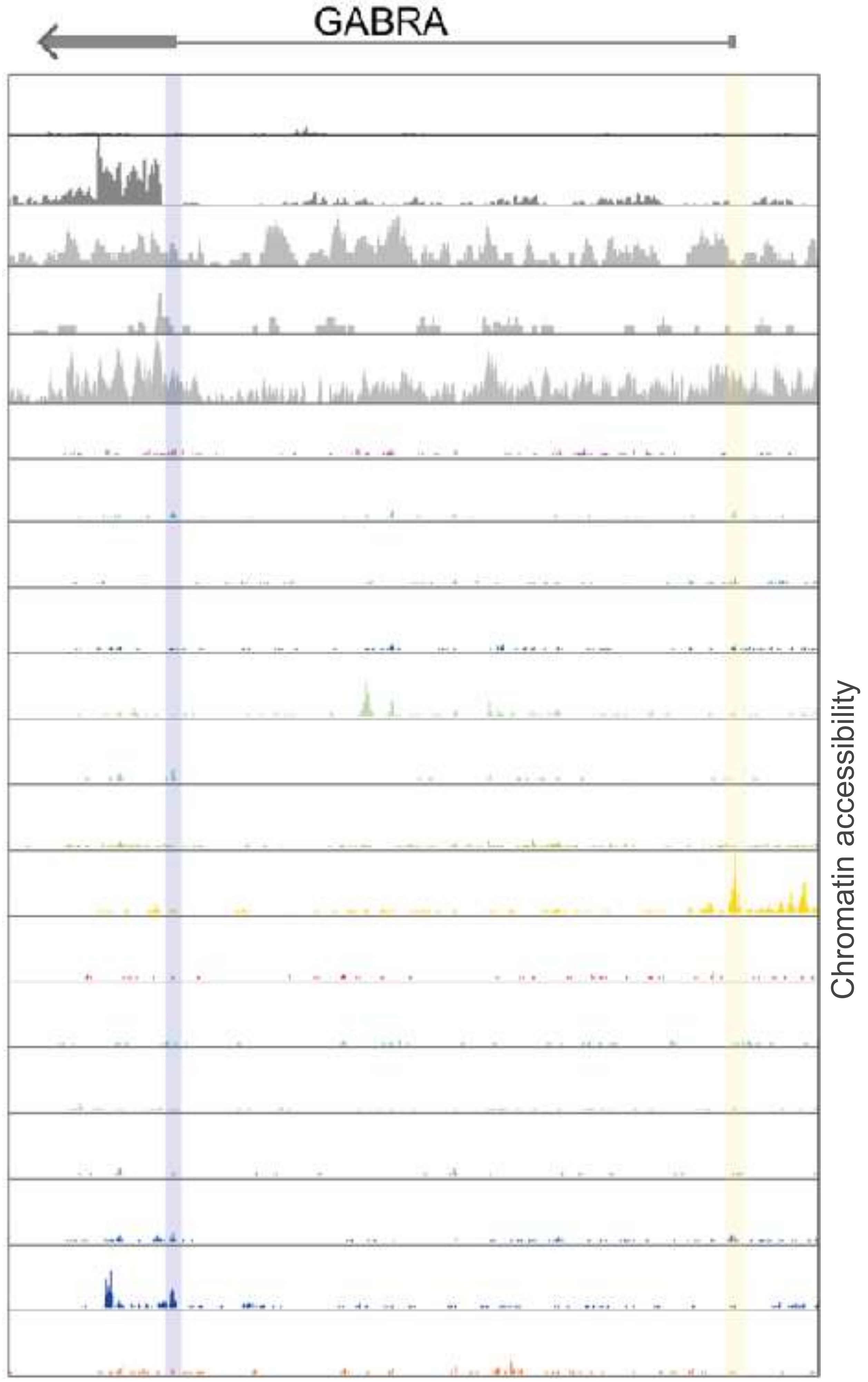
Cell type-specific **gene regulatory landscapes**

(1) *NeuroPou4/FoxL2-AP::mOrange*



(2) *tRM-AP::mOrange*

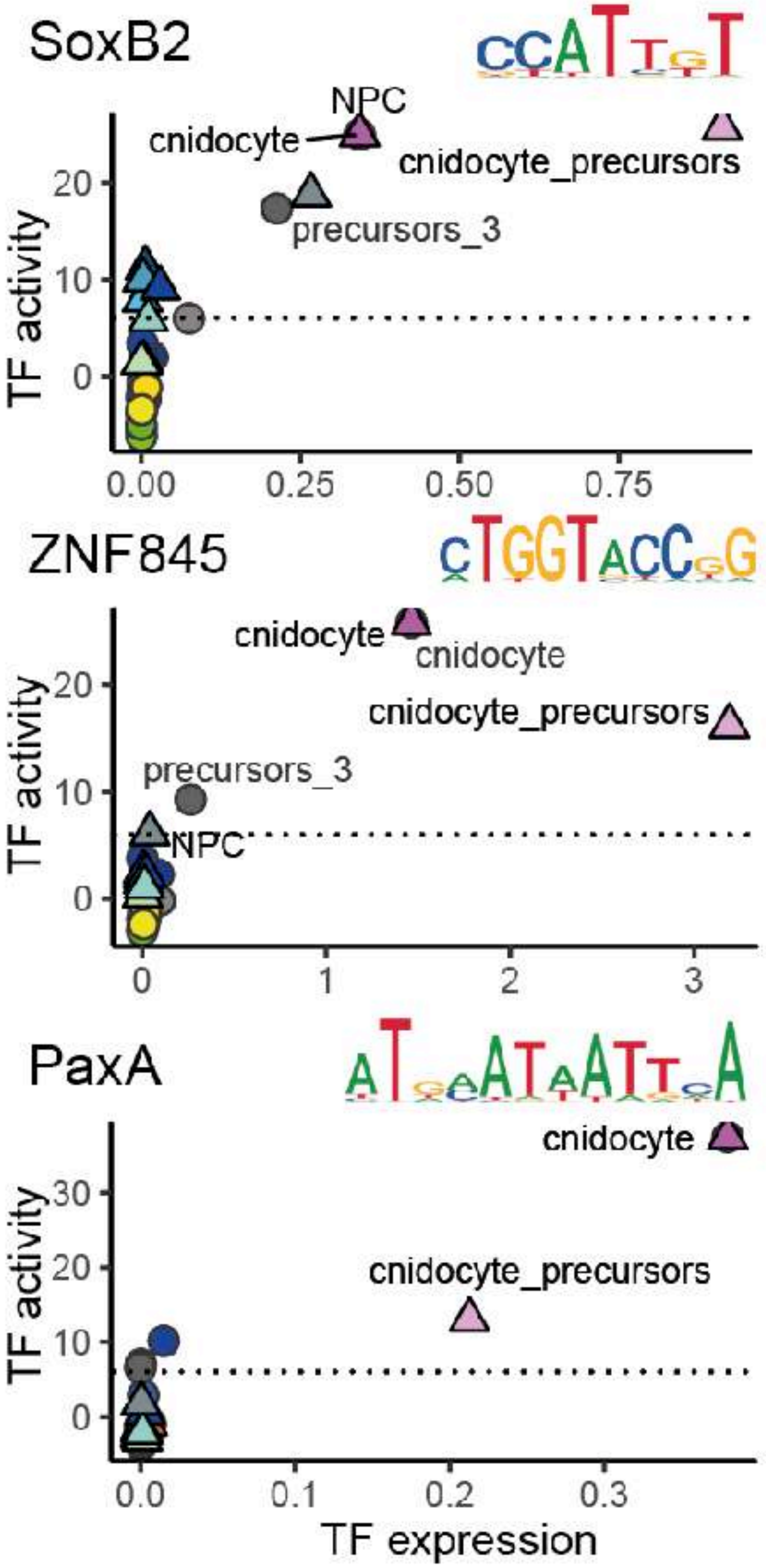


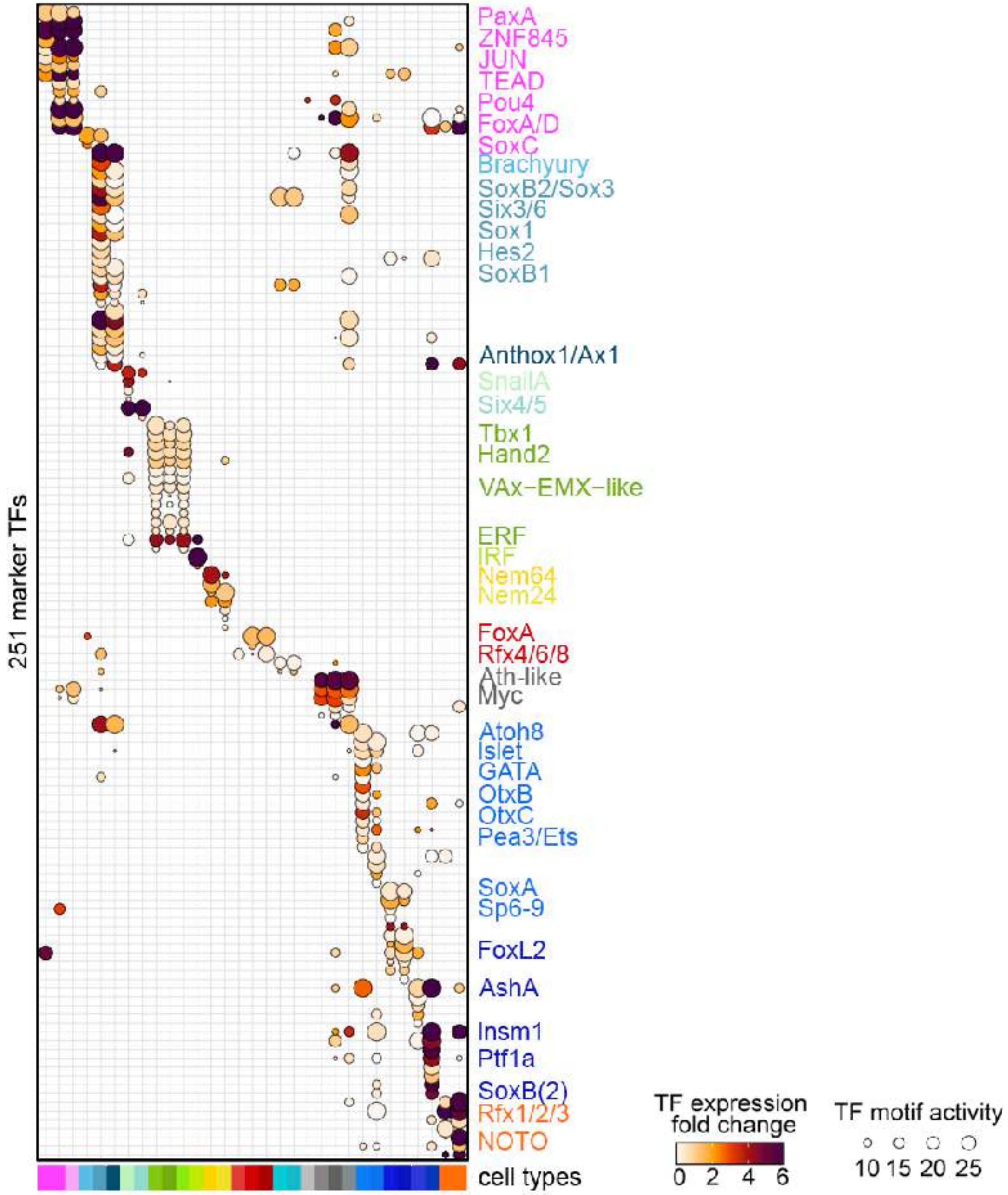


Alternative promoter
(AP)

Cell type regulatory identity 1: **Transcription Factor activity**

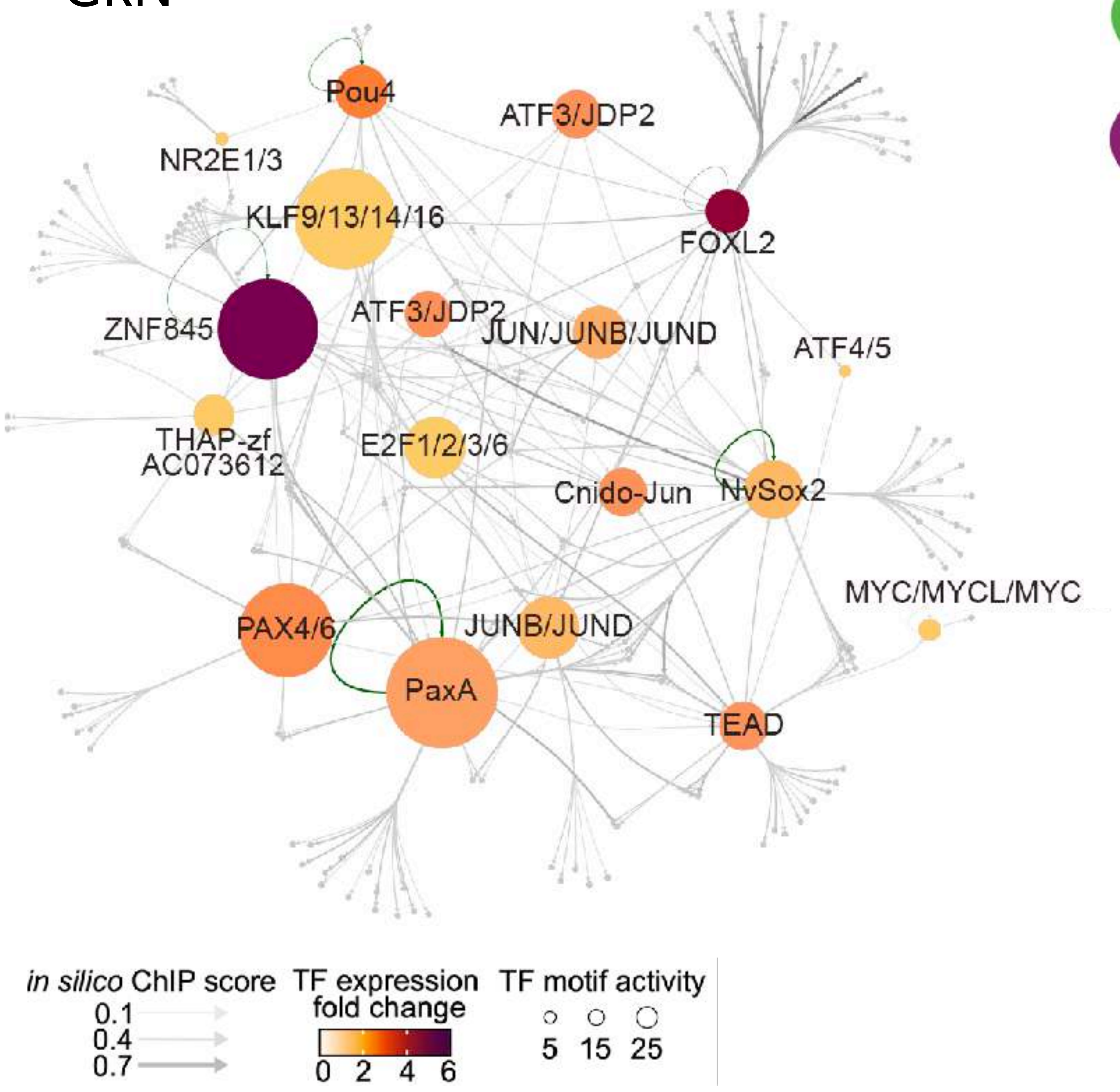
Cnidocyte
TF activity

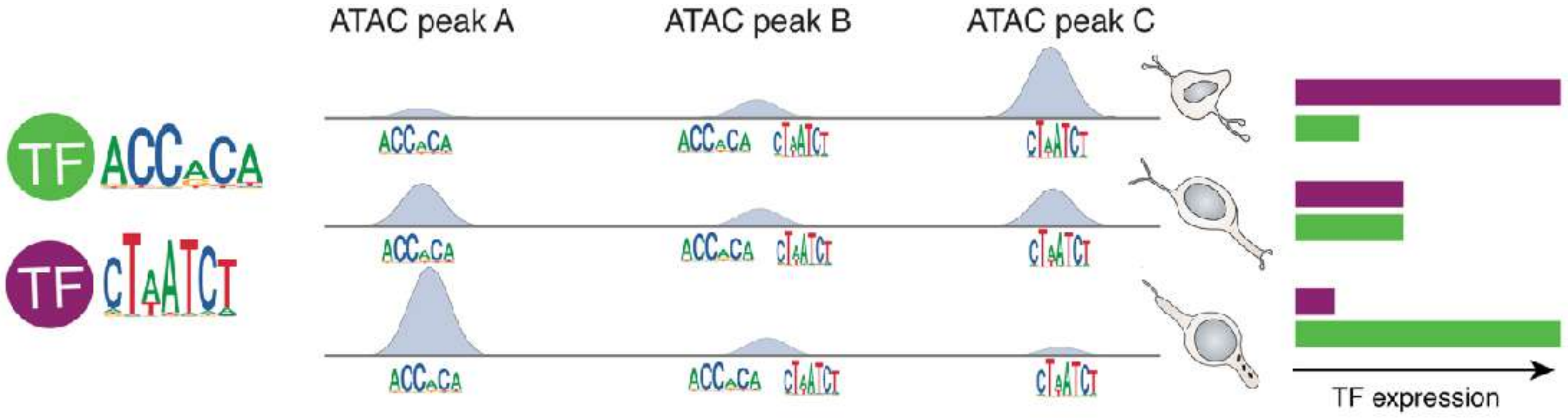


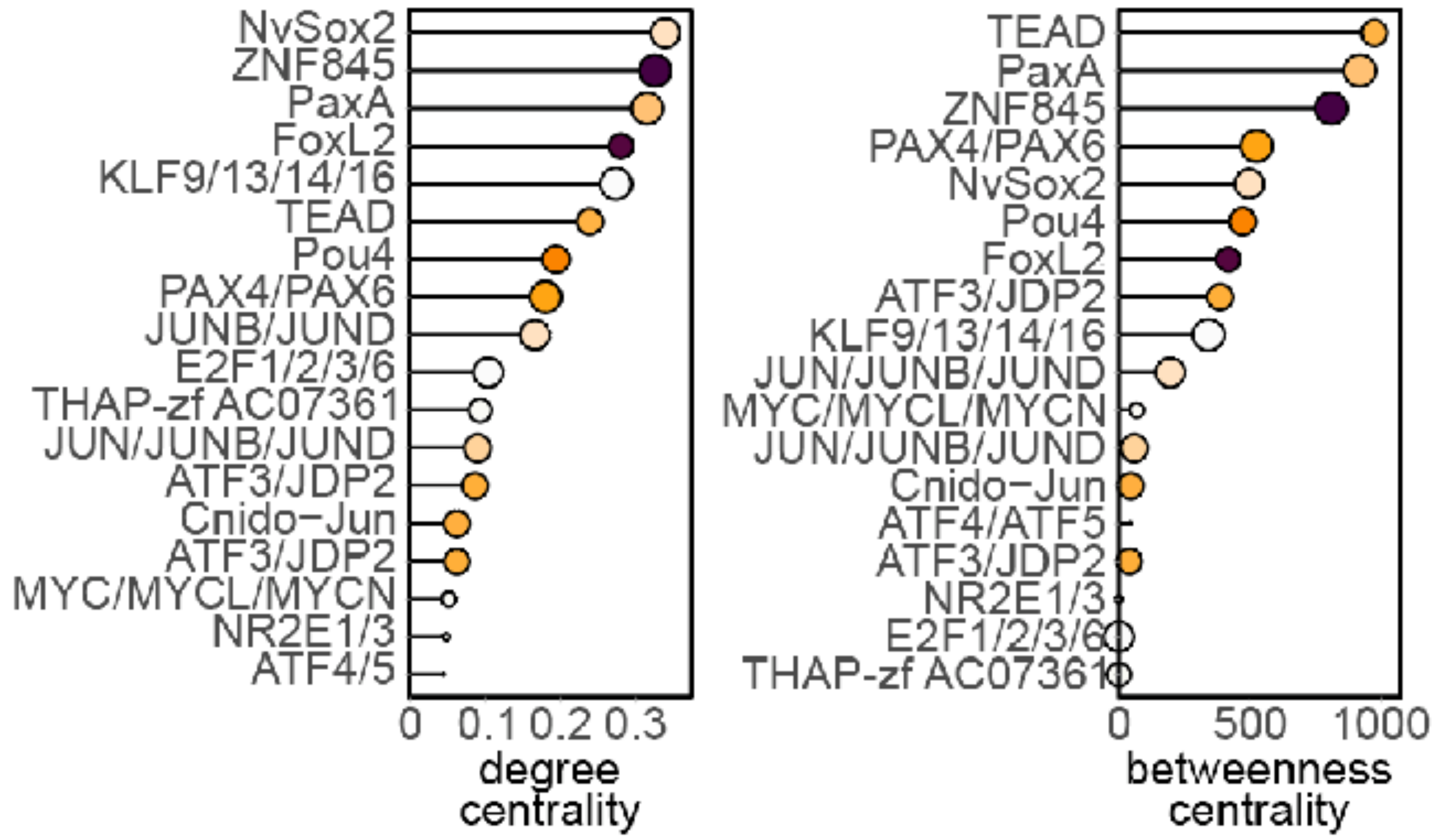


Cell type regulatory identity 2: **Gene Regulatory Networks**

Cnidocyte GRN



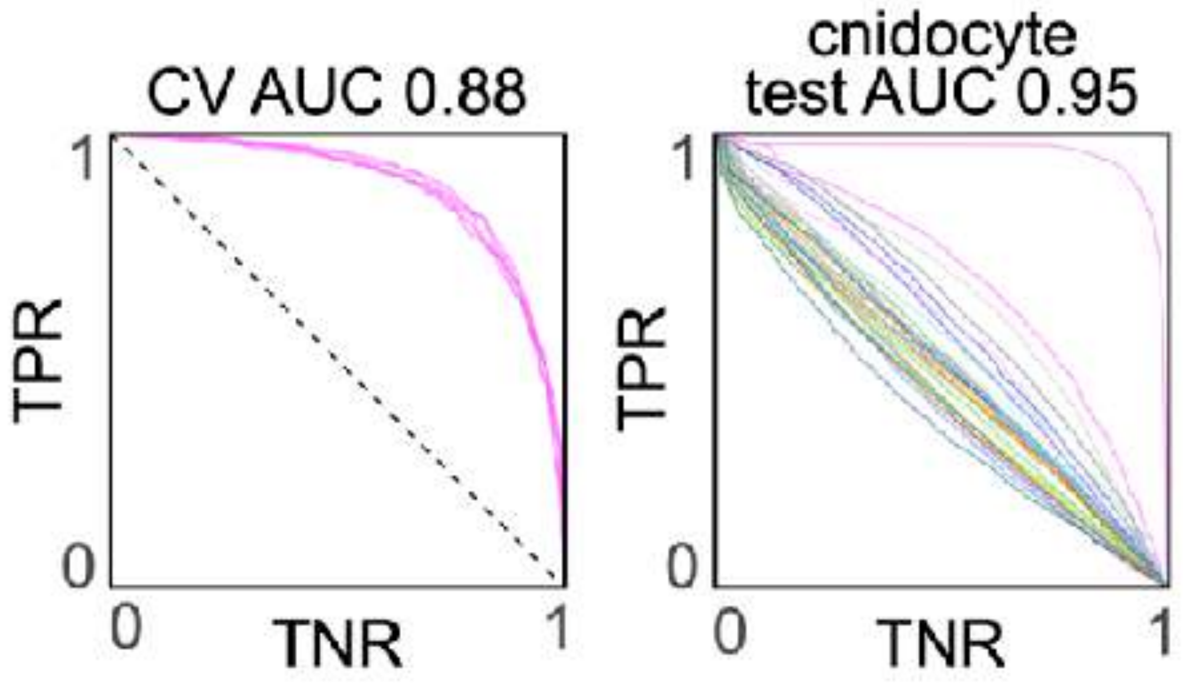




TF network connectivity

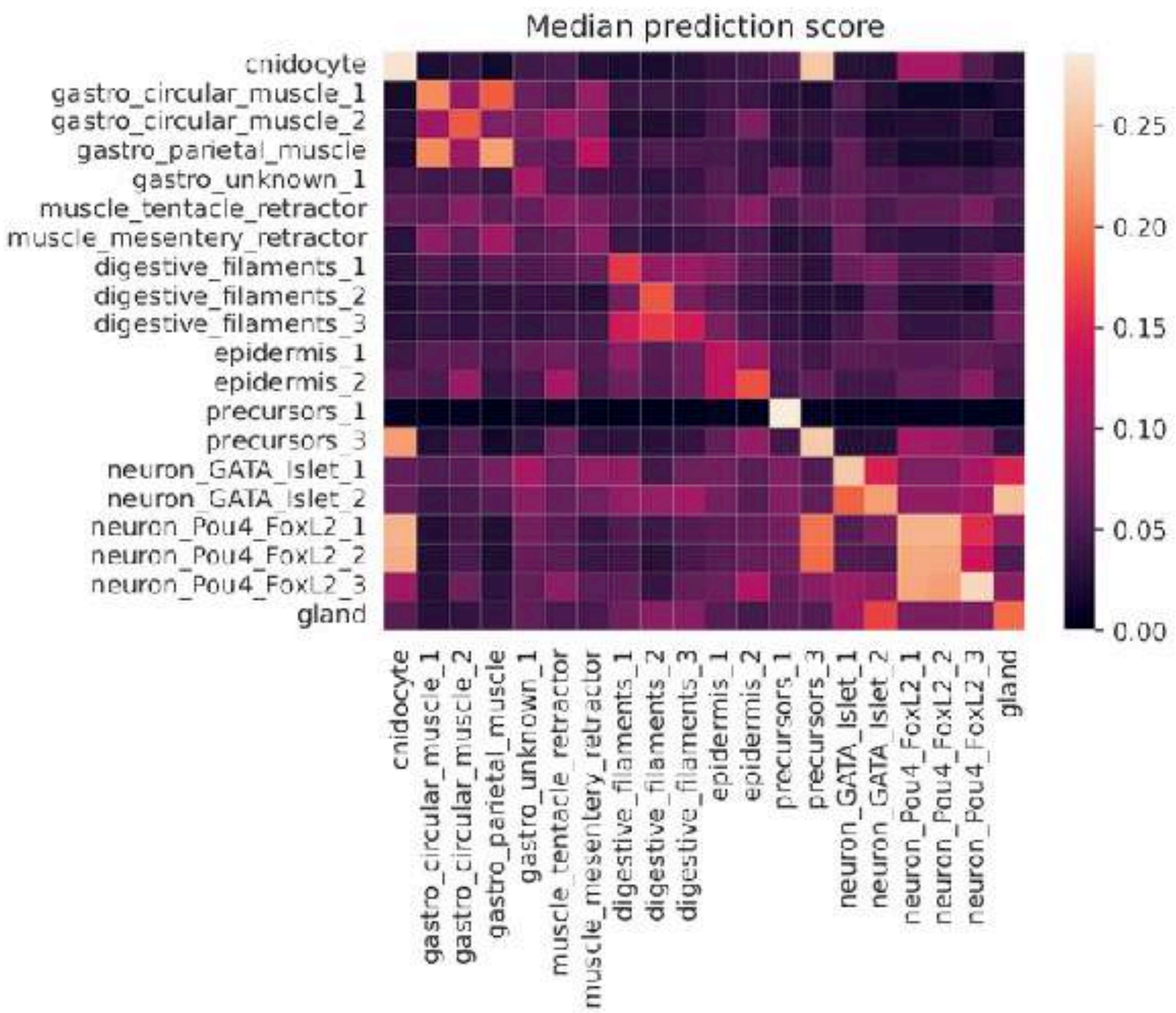
Cell type regulatory identity 3: **Sequence motif grammars**

Cnidocyte CRE sequence model

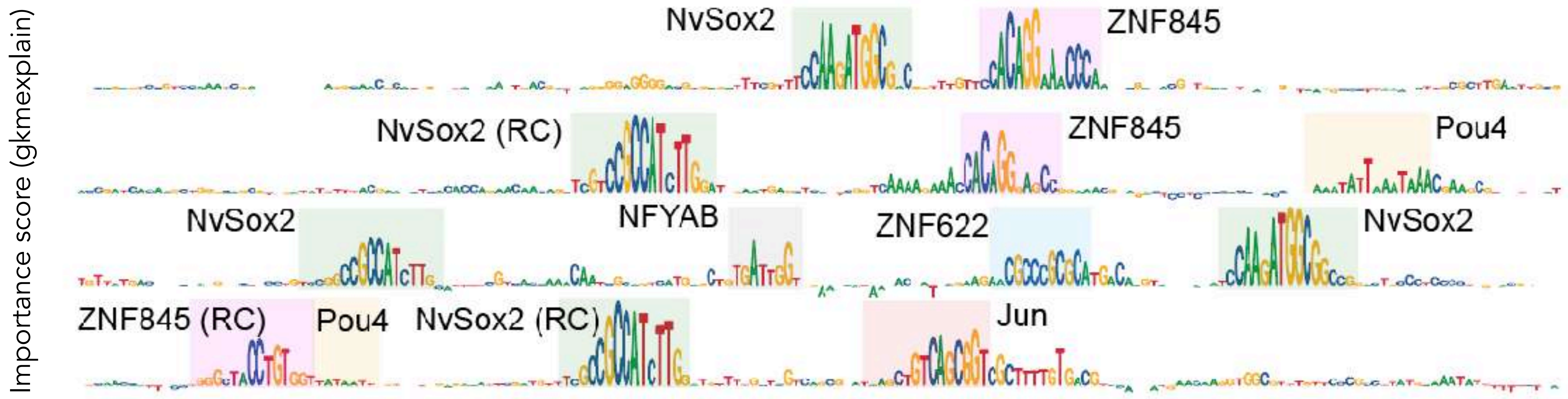


gkmSVM cnidocyte model

Apply sequence model classifiers across cell types



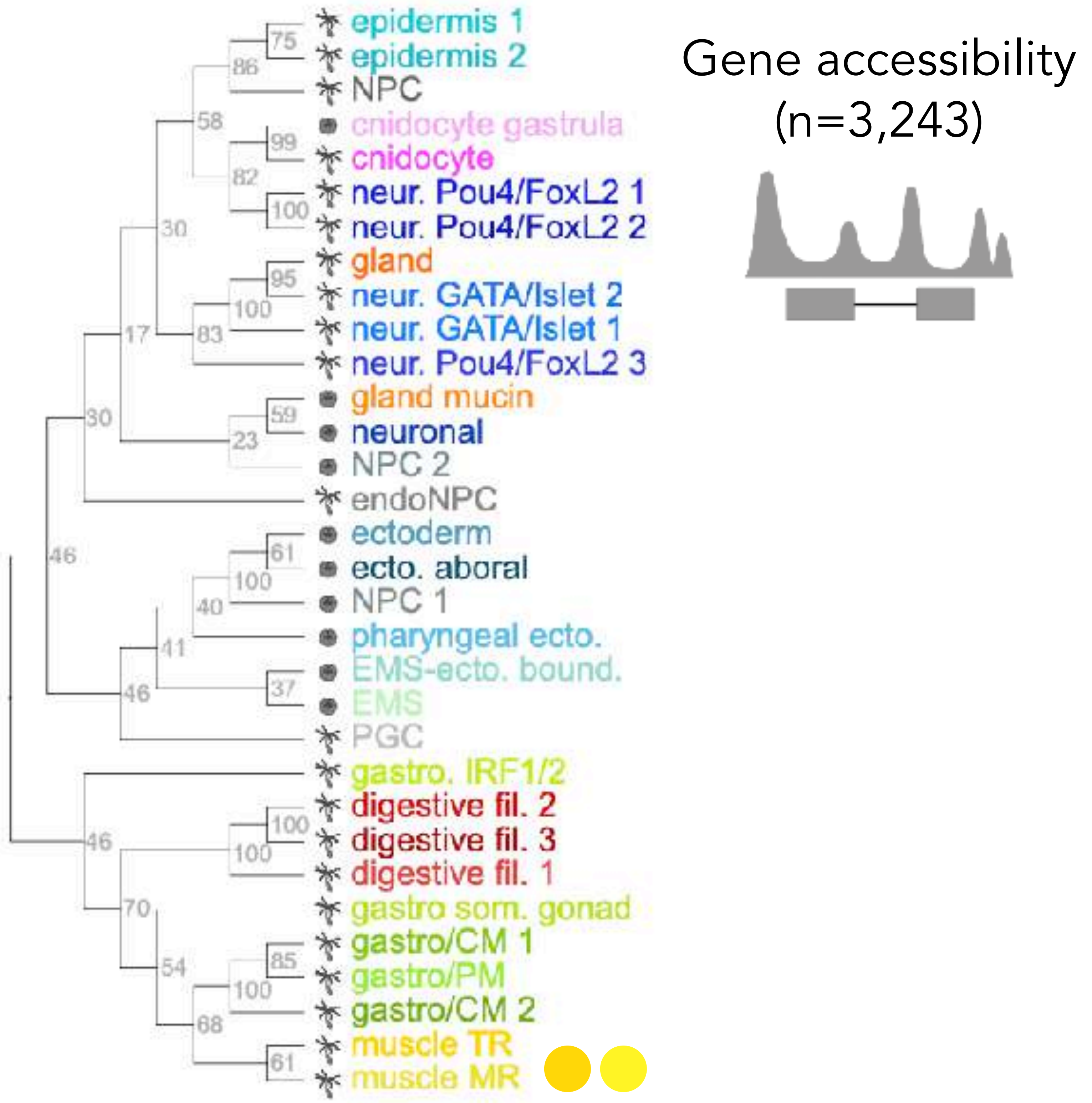
Most common motif lexicons in cnidocyte CREs



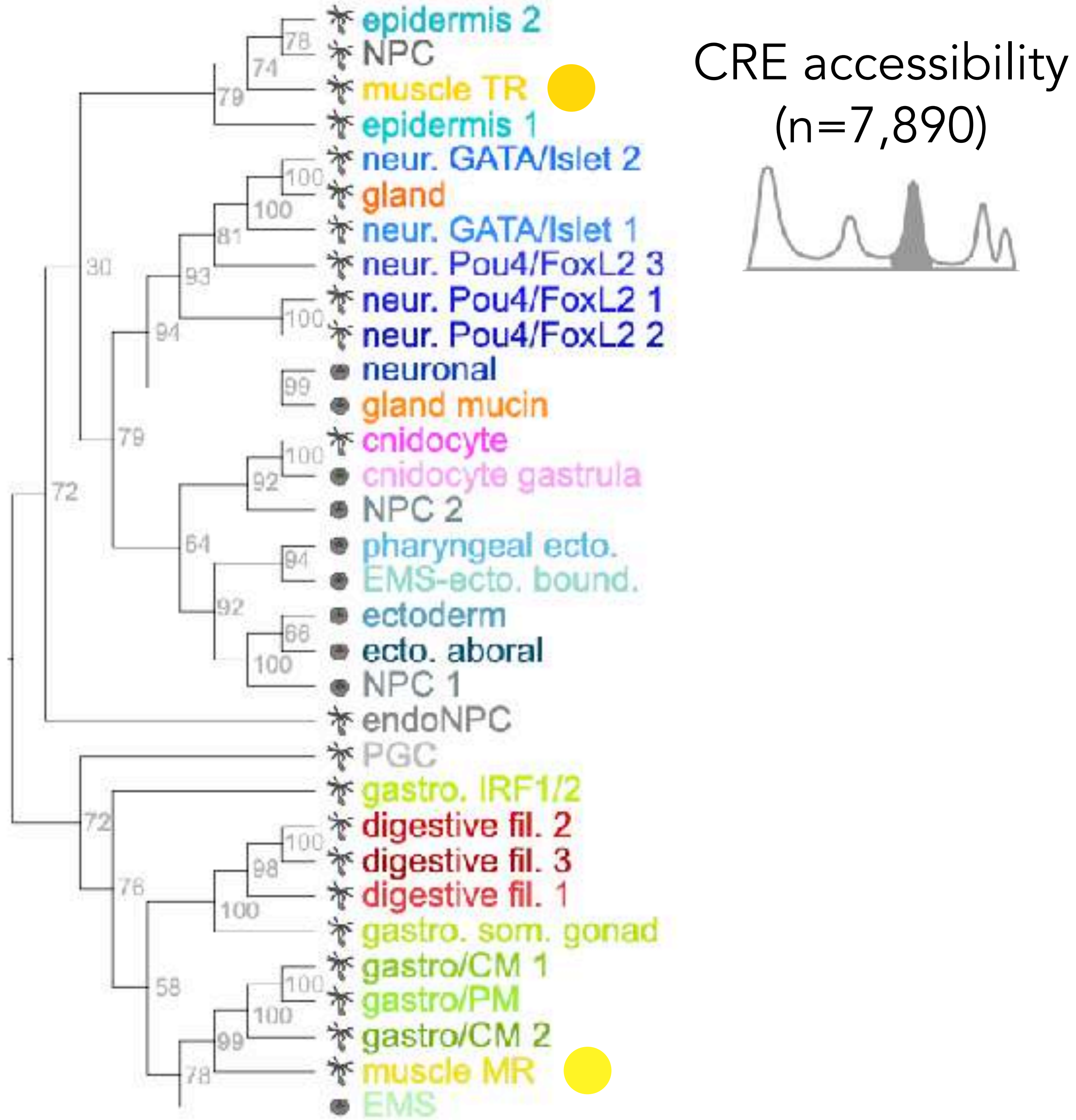
Postion (250bp CREs)

Cell type relationships inferred from effector gene usage *versus* **regulatory characters**

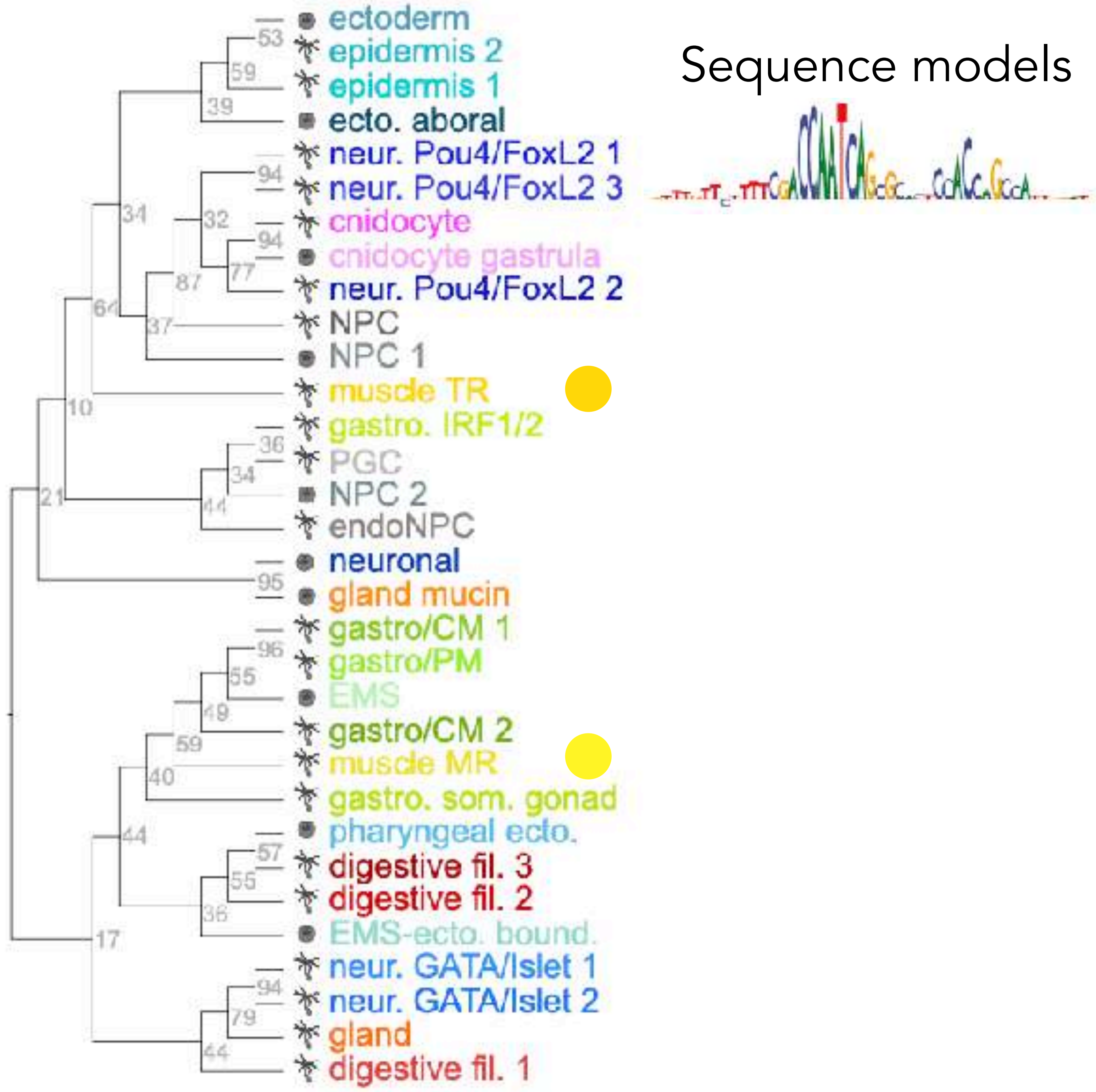
Gene accessibility (n=3,243)

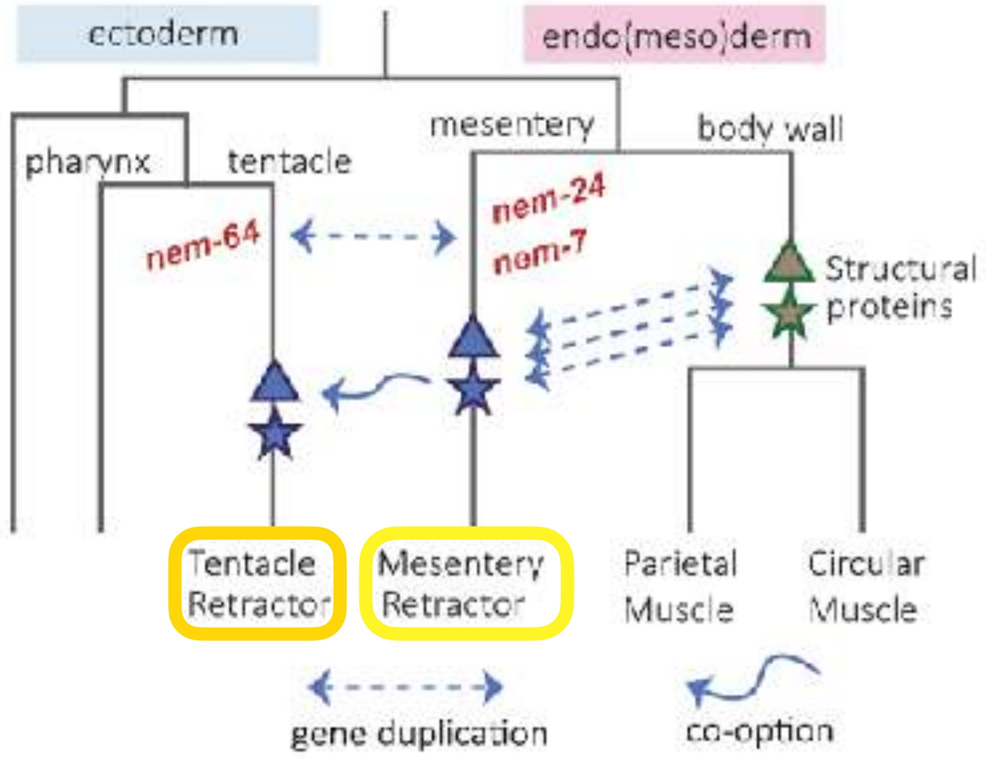


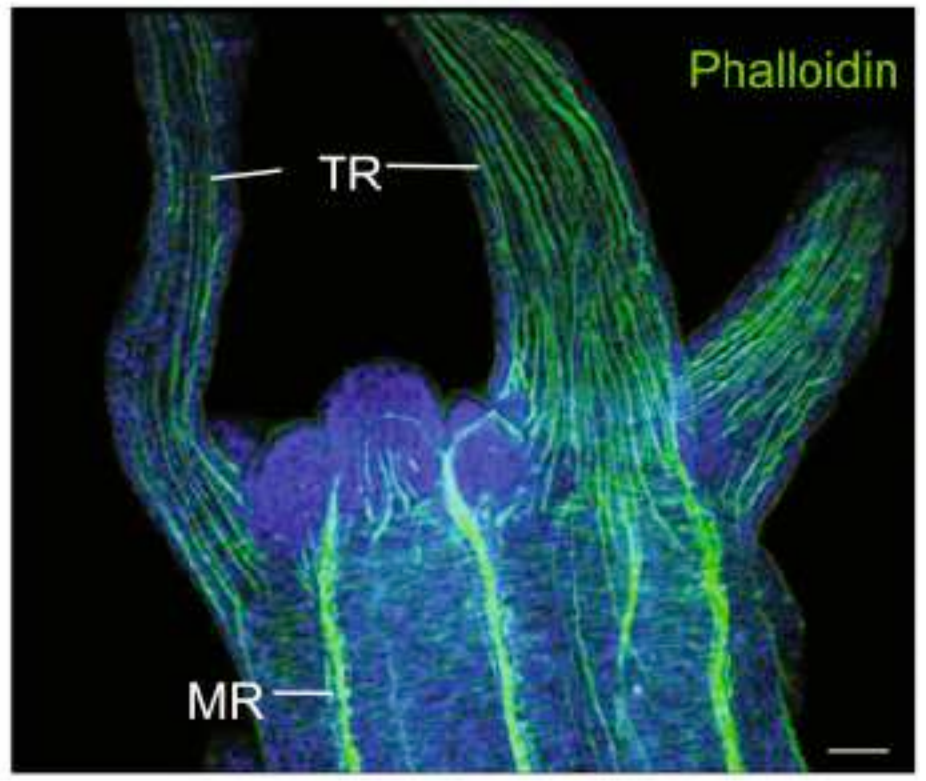
CRE accessibility (n=7,890)

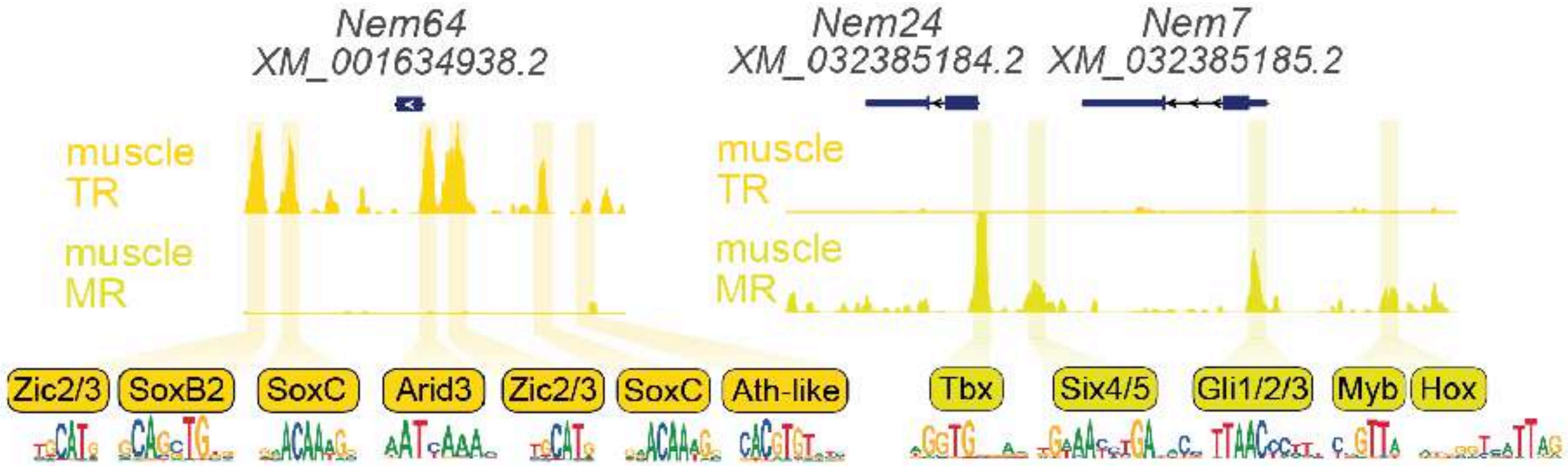


Sequence models









Cole et al., *Nat. Comm.* 2023