What is a cell type? Conceptual and quantitative definitions in the single-cell omics era

Santiago Carmona

Reviews in Quantitative Biology

University of Lausanne

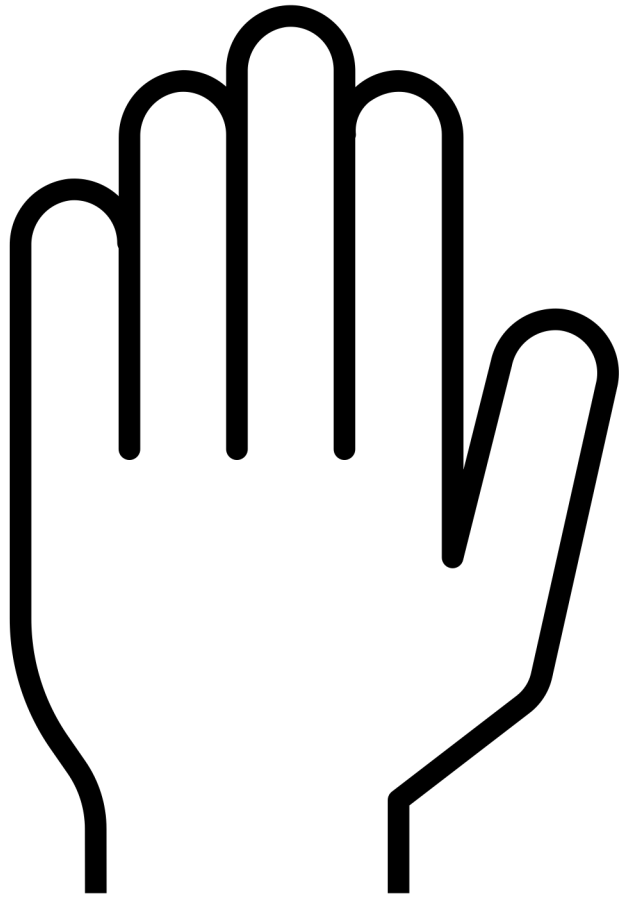
22 November 2024

Cells of the same cell type...

- ❑ have similar morphology
- ❑ have similar molecular composition
- ❑ have similar body location
- ❑ have the same functions
- ❑ share their developmental history
- ❑ change in evolution together
- ❑ are grouped together for biologists' convenience but cell types don't really exist in nature

Who has ever read a paper where they use **single-cell RNA-seq**?

Analyzed a scRNA-seq dataset?



Goal of this mini-review

Explore how single-cell genomics has changed the way biologists define cell types

(Mostly centered on humans as a model organism)

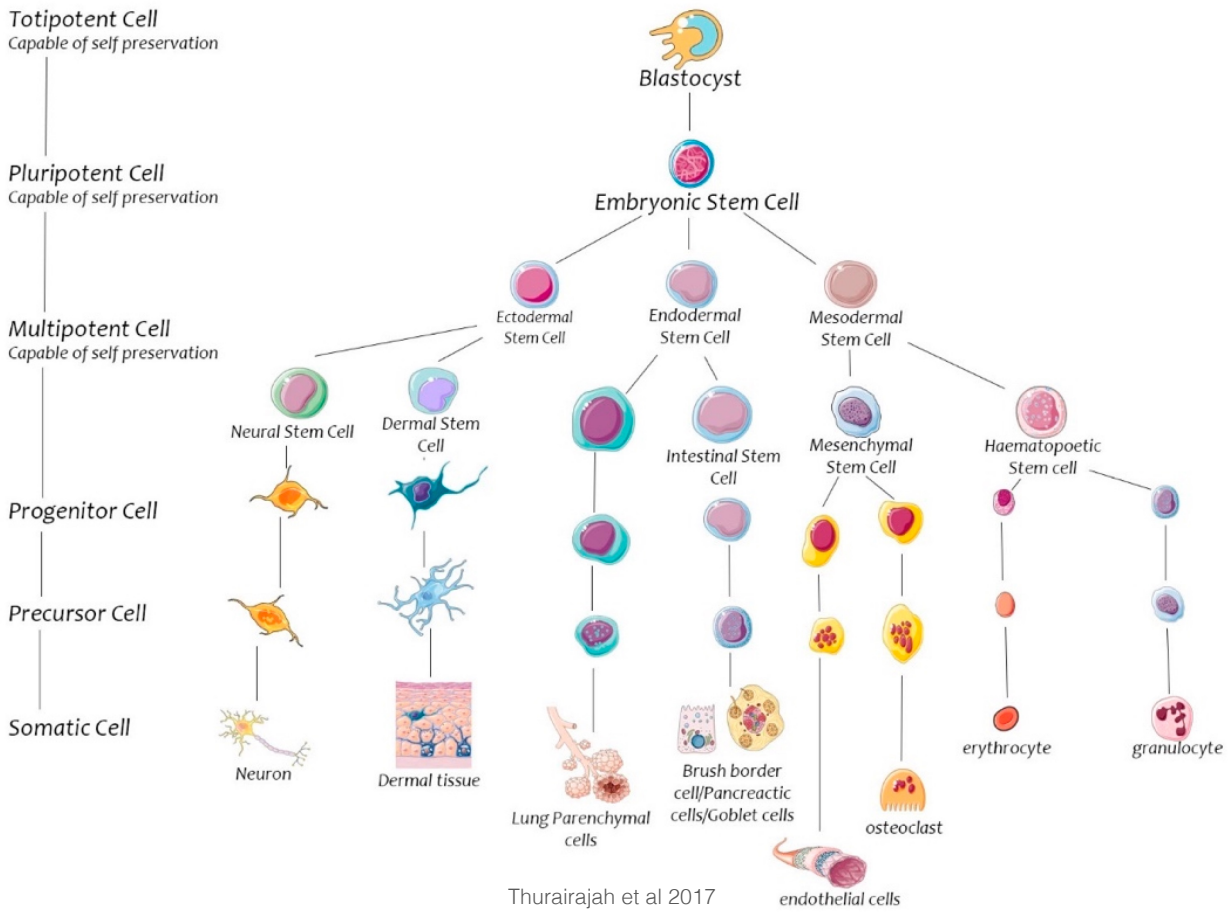
Multicellular organisms evolved with cell specialization

Potency

Asymmetric cell division and gene expression regulation

Functional specialization

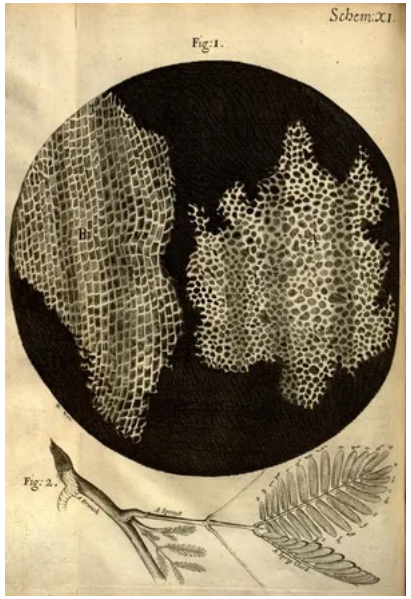
Mammalian development



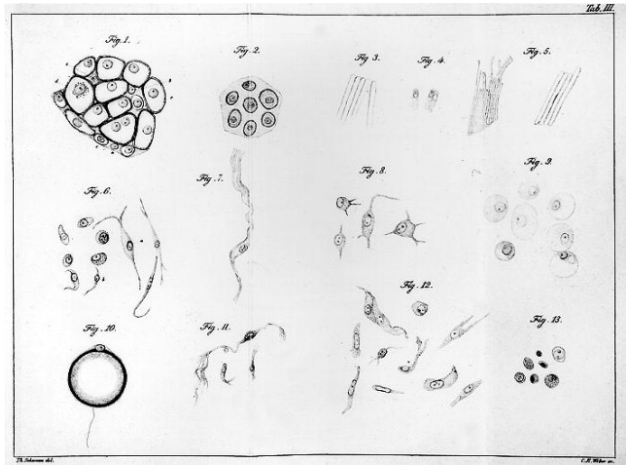
Thurairajah et al 2017

Humans develop from a single cell to ~30 trillion somatic cells with nearly the same genotype and a great diversity of phenotypes

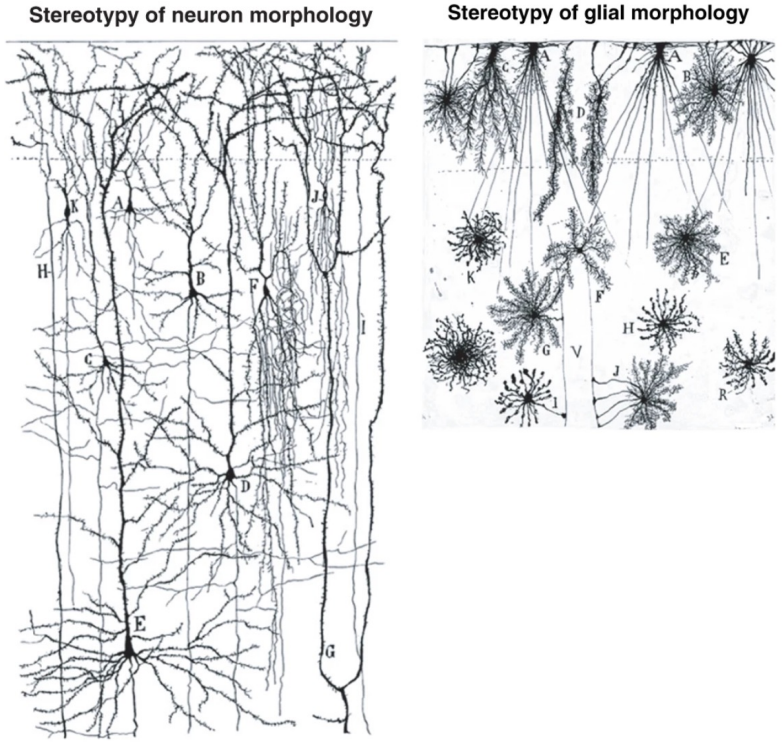
Since the discovery of the cell and the formulation of the cell theory, biologists have classified cells into types (i.e. created **cell type taxonomies**)



Robert Hooke's drawing of cork tissue (1665)



Theodor Schwann's drawings of animal cells (1839).
Led to classification of cells into epithelial, muscular, nervous, blood, and connective types



Different cell types in the nervous system
Ramón y Cajal (1904) using Golgi's silver staining

Morphology was the first criterion for cell type classification

Widely-used cell type taxonomies

Subjective but incredibly useful classification system

Bruce Alberts et al. (1989) described **210 human cell types** subdivided into 20 categories related to **functional** criteria

- Keratinizing Epithelial Cells
- Exocrine Secretory Epithelial Cells
- Hormone Secreting Cells
- Metabolism and Storage Cells
- Extracellular Matrix Secretion Cells
- Barrier Function Cells
- Contractile Cells
- Blood and Immune System Cells
- Sensory Transducer Cells
- Germ Cells
- etc.

Vickaryous & Hall (2006) estimated **411 human cell types** (of which 145 are types of neurons), organized into 34 categories based on commonly cited **cytological**, **histological**, and **functional** criteria

Do we need cell type taxonomies? Why?

Cell type classifications have been driven by technological advances

Key single-cell-resolution technologies

1660s	light microscope	Morphology & Tissue localization (e.g. red blood cell)
1900s	stains and synthetic dyes	
1930s	electron microscopy	Ultrastructure
1940s	immuno-staining	Marker cell surface antigens (e.g. Rh, CD3) & *ex vivo* functional assays
1970s	Fluorescence-Activated Cell Sorter	
1980s	RNA Fluorescence In Situ Hybridization	Marker RNA transcripts
2010s	mass cytometry & spectral flow cytometry	Multiple marker proteins (up to ~40)
2010s	**single-cell RNA sequencing**	**Sampling *all* mRNAs**
2020s	**highly-multiplexed immune-staining and RNA in situ hybridization**	**Multiple marker proteins (up to ~60) and expressed genes (up to 18'000)**

The single-cell RNA-seq revolution

"gene expression patterns have the potential to provide a unique description of each cell type. Until recently however, collecting the large amounts of expression data necessary to compare different cell types was prohibitively time-consuming. Fortunately, since the 1990s this particular obstacle has been largely overcome with the advent and widespread use of DNA microarrays"

Vickaryous & Hall (2006)

"A revolution in cellular measurement technology is under way: For the first time, we have the ability to monitor global gene regulation in thousands of individual cells in a single experiment. Such experiments will allow us to discover new cell types and states and trace their developmental origins."

C. Trapnell (2015)

"The recent advent of methods for high-throughput single-cell molecular profiling has catalyzed a growing sense in the scientific community that the time is ripe to complete the 150-year-old effort to identify all cell types in the human body." (the Human Cell Atlas Project)

A. Regev et al (2017)

10 years later….

The single-cell RNA-seq revolution

Comprehensive single-cell transcriptional profiling of a multicellular organism – Cao et al Science 2017

"We applied single-cell RNA-seq to profile nearly 50,000 cells from the nematode *Caenorhabditis elegans* at the L2 larval stage, which provided >50-fold "shotgun" cellular coverage of its somatic cell composition. From these data, we defined consensus expression profiles for 27 cell types"

Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly – Li et al Science 2022

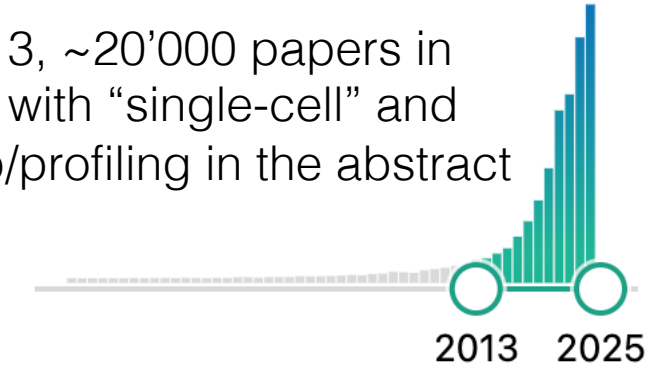
"we present a single-cell atlas of the adult fly that includes 580,000 nuclei from 15 individually dissected sexed tissues as well as the entire head and body, annotated to >250 distinct cell types"

A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain – Zizhen et al Nature 2023

"The atlas is hierarchically organized into 4 nested levels of classification: 34 classes, 338 subclasses, 1,201 supertypes and 5,322 clusters." (dataset of 4 million cells). Every subclass (and all supertypes within) has a unique and specific spatial localization pattern within the brain.

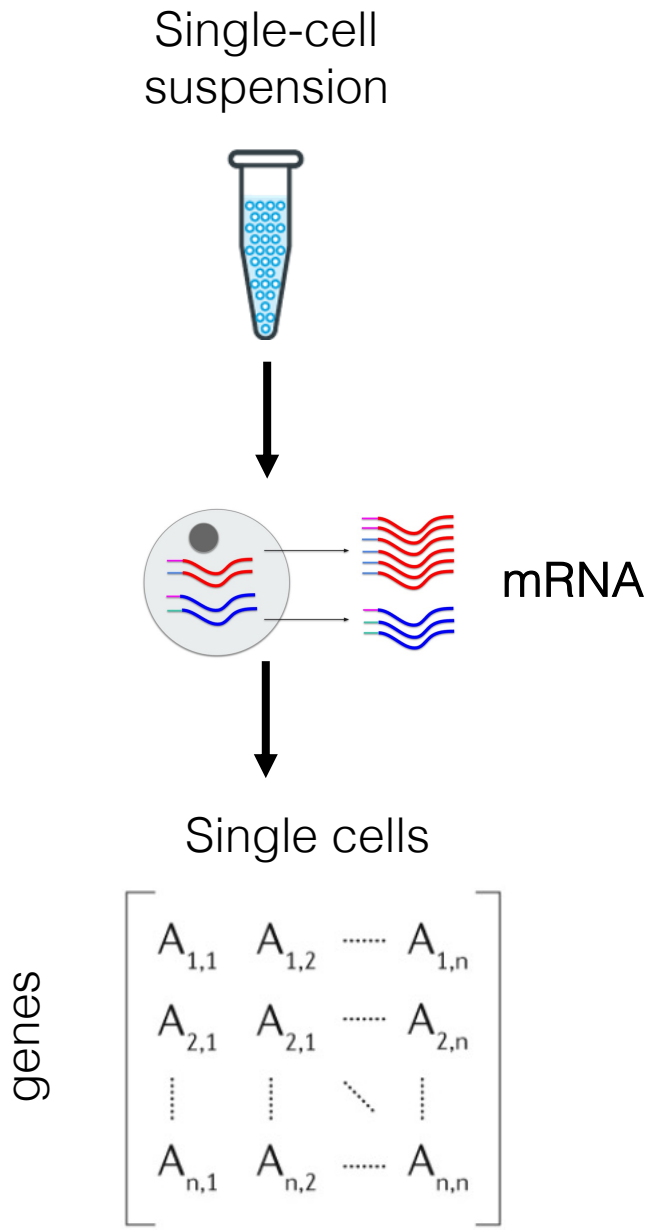
Single-cell sequencing 2013 "Method of the year"

Since 2013, ~20'000 papers in PUBMED with "single-cell" and atlas/map/profiling in the abstract



Single-cell RNA-seq

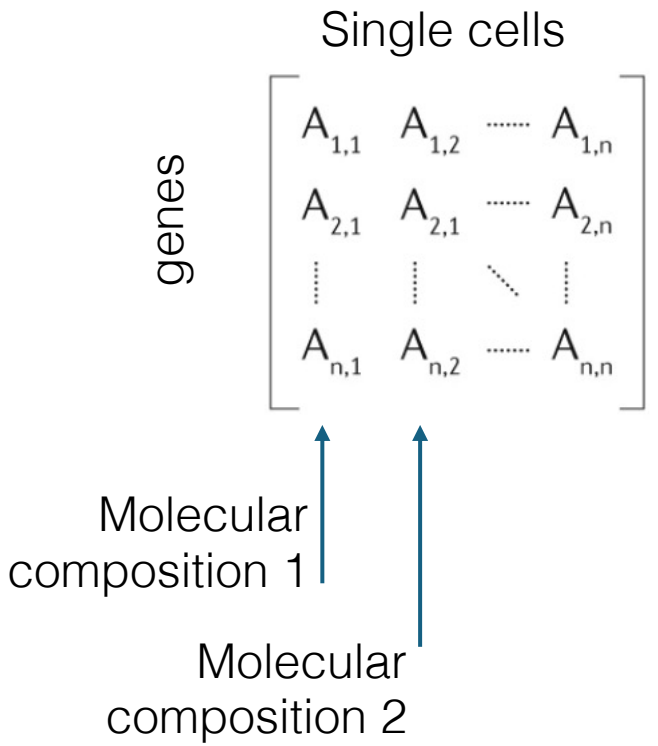
- Family of technologies that take dissociated cells (or nuclei) as input, and produce a gene expression matrix as output
- Provides a largely unbiased sample (~10%) of all mRNA in each cell in a given moment in time
- High-throughput (up to 10^5 cells per assay and sample multiplexing)
- Mature, reproducible, commercially available technologies
- Broadly applicable (protocols for ~any tissue; fresh, frozen, fixed)
- Some techniques allow for allele and isoform resolution but typically they average transcriptional output per gene
- Can be combined with other cell aspects/modalities (protein, chromatin accessibility, etc.)



Baysoy et al 2023 https://www.nature.com/articles/s41580-023-00615-w

Single-cell RNA-seq output

- scRNA-seq produces measurements of the molecular composition (of mRNA transcripts) of a cell in a given timepoint.
- At any given timepoint, the molecular composition of each cell unique.
- This instantaneous molecular composition of a cell is usually referred to as the cell *state* or *molecular state* and sometimes as the cell *identity* (Wagner et al 2017, Domcke & Shendure 2023).

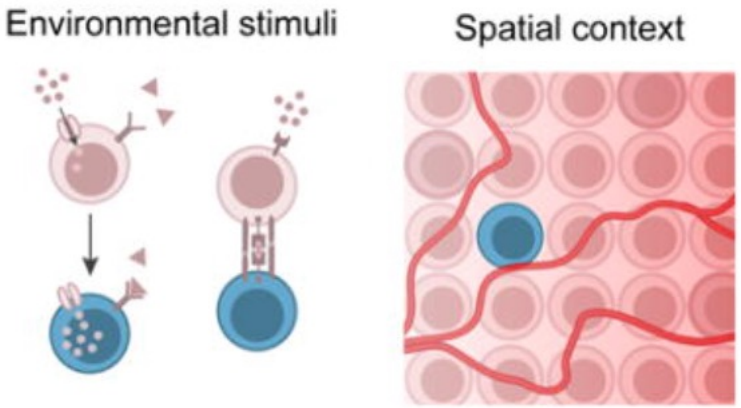


How are scRNA-seq measurements related to traditional cell types?

"The biological factors affecting the cell combine to create its unique, instantaneous identity, which is captured in the cell's molecular profile."

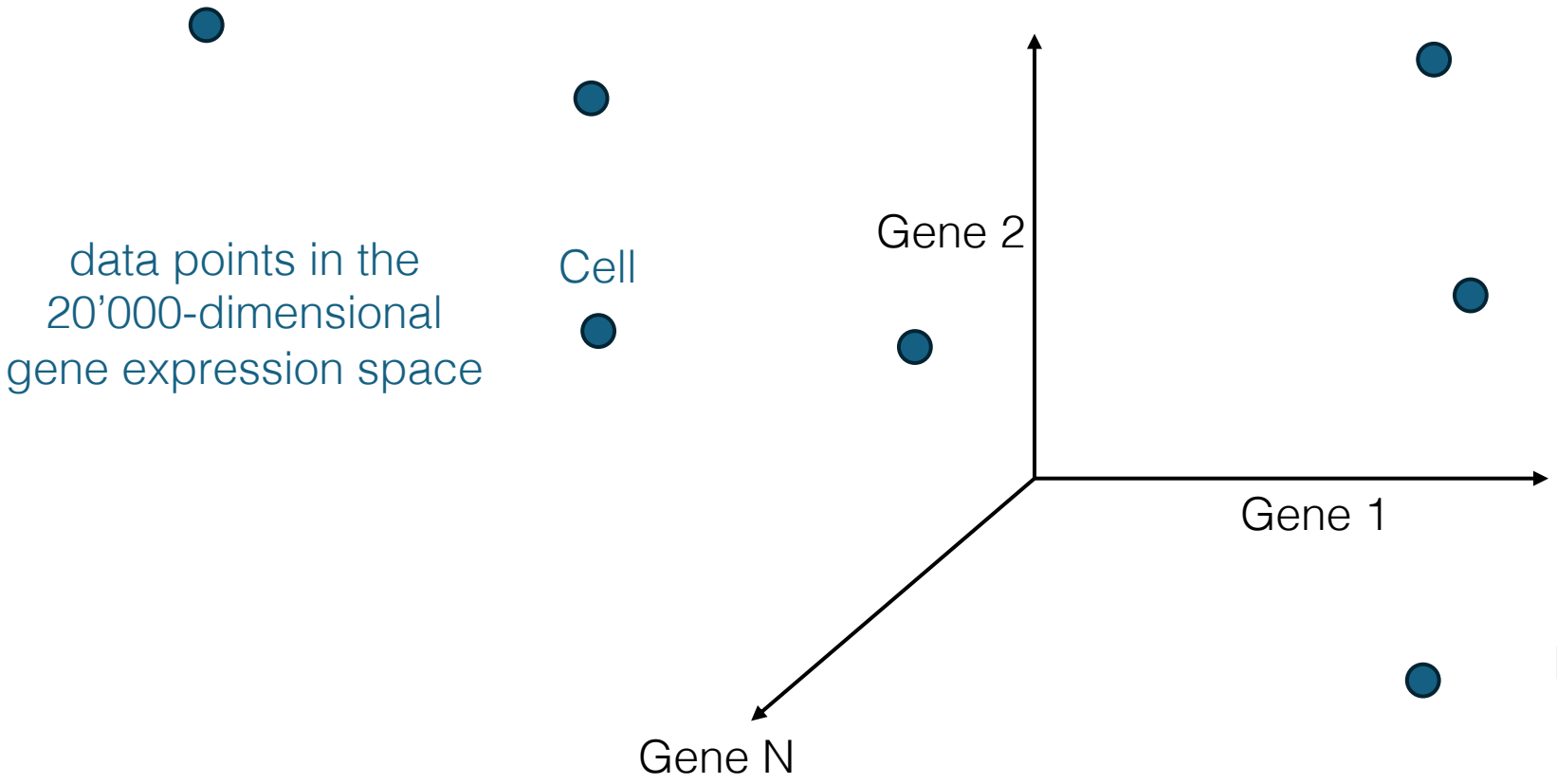
"We define a cell's identity as the outcome of the instantaneous intersection of all factors that affect it. We refer to the more permanent aspects in a cell's identity as its type (e.g., a hepatocyte typically cannot turn into a neuron) and to the more transient elements as its state."

Wagner, Regev, Yosef (2017)



Examples of factors affecting a cell state but not its type

- Single-cell RNA-seq produces snapshots of each cell's mRNA composition
  - A cell can be thought of as a vector in the gene expression space

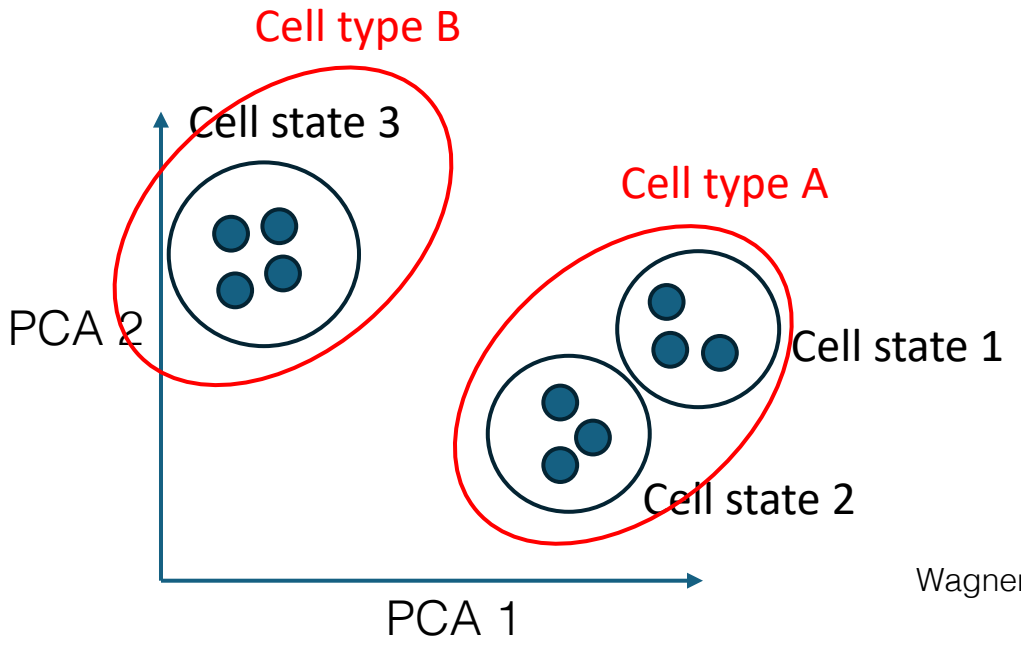


Data are sparse in higher dimensions, obscuring relationships

THE CURSE(S) OF DIMENSIONALITY

Cell states in lower-dimensional representations

Structure (relationships between datapoints or cells) tends to be clearer in lower dimensions

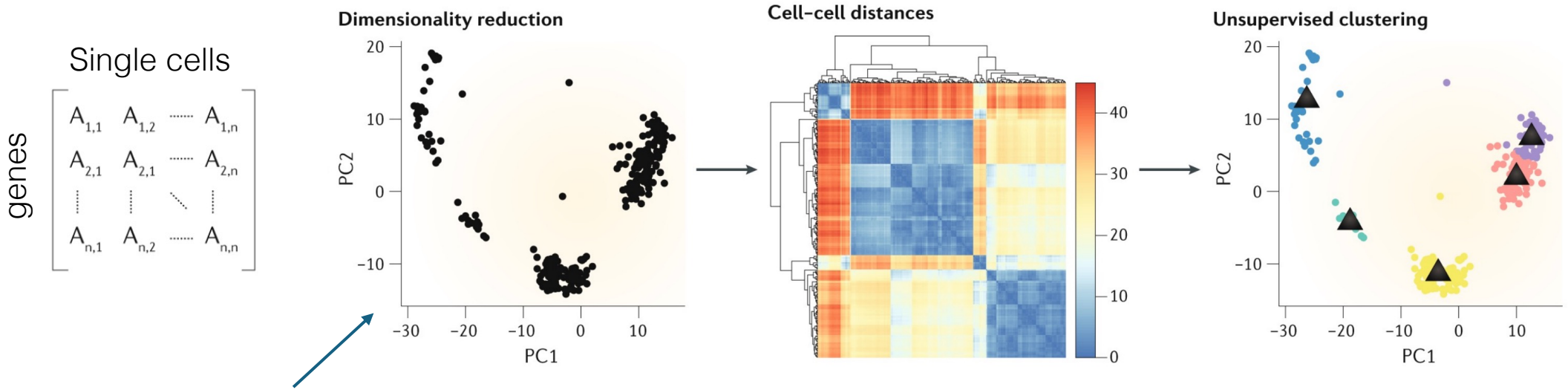


Trapnell (2015)
Wagner, Regev, Yosef (2017)

Low-dimension projection by Principal Components Analysis (PCA)

Altman & Krzywinski (2018) https://www.nature.com/articles/s41592-018-0019-x
Gunawan et al (2023) https://doi.org/10.1016/j.crmeth.2023.100547

How are transcriptional cell states quantitatively identified in practice?



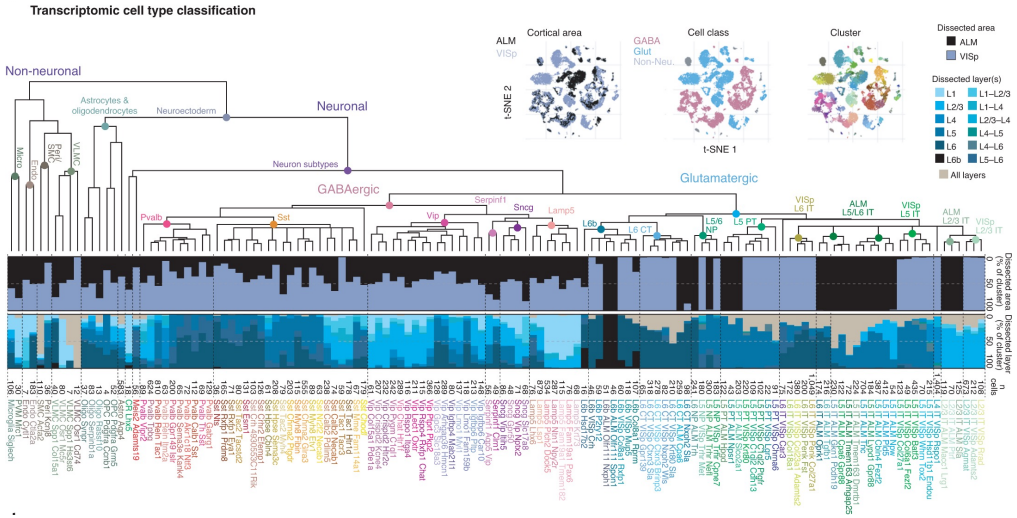
The manifold obtained by linear dimension reduction methods (e.g. PCA), are highly interpretable but *too rigid*. Non-linear (more flexible) methods are also used.

Clustering typically by community detection algorithm on K nearest neighbor (KNN) cell graph

- Many steps (data normalization, feature selection, dim reduction, kNN graph construction, clustering resolution, etc.) involving many parameters and (often subjective) choices!
- Typically iterative procedure including asessment of clusters robustness, and splitting/merging clusters according to biological interpretation (e.g. expression of marker genes of previously known cell types)
- Clustering results (cell states) are not definitive but hypotheses that require independent evidence for validation

Kiselev et al 2019

Proposal of a transcriptome-based taxonomy of cell types



A similarity tree of neocortical cell types

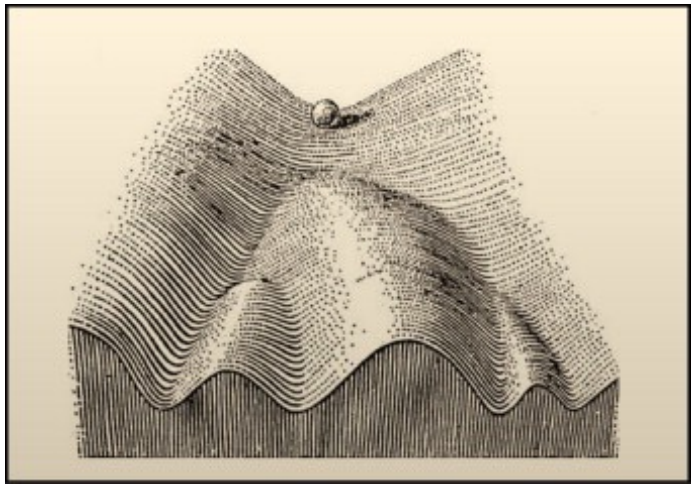
"Statistical analyses of these data reveal clusters that often correspond to cell types previously defined by morphological or physiological criteria and that appear conserved across cortical areas and species. To capitalize on these new methods, we propose the adoption of a transcriptome-based taxonomy of cell types for mammalian neocortex. This classification should be hierarchical and use a standardized nomenclature.

Yuste R et al 2020

From single-cells to landscapes

In Waddington's metaphor, cells are balls rolling down a landscape (of all possible cell states) with different possible paths, each representing a distinct cell differentiation trajectory. The basins in the landscape represent differentiated cell types.

Waddigton's metaphor also highlights the importance of cell development in our understanding of cell types.



1950s Waddington's landscape of epigenetic cell states

"These [scRNA-seq] measurements may finally make explicit the metaphor that C.H. Waddington posed nearly 60 years ago to explain cellular plasticity: Cells are residents of a vast "landscape" of possible *states*, over which they travel during development"

"Single-cell technology helps not only locate cells on this landscape, but illuminates the molecular mechanisms that shape the landscape itself."

C. Trapnell (2015)

From single-cells to landscapes (or manifolds)

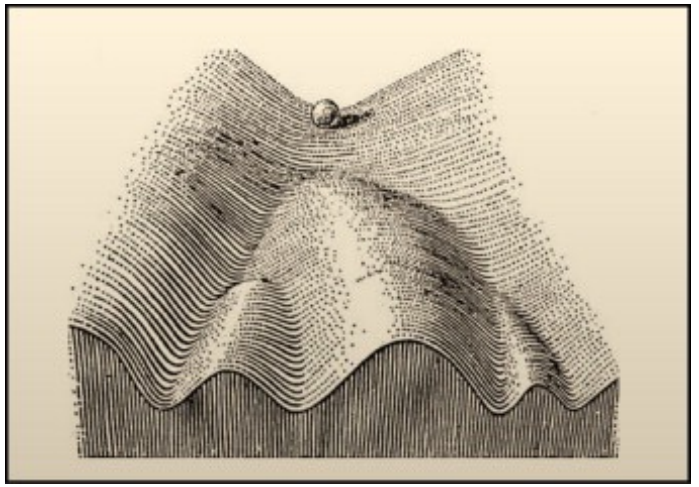
- A common and useful assumption is that cell states lie on a “manifold”, defined as a locally smooth continuum *landscape* that has lower dimensionality compared to the full gene expression (ambient) space

- Cell types are typically depicted as regions or probability distributions in the cell state manifold (i.e. in lower-dimensional projections of the full gene expression space).

- A cell type may be viewed as a collection of states among which cells can (often reversibly) transition by effect of transient fluctuations (e.g. such as circadian rhythm, stress responses, metabolic changes, disease or pharmacological intervention)

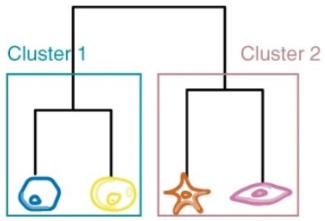
- Cell types are expected to be stable over time scales that exceed a cell cycle, while cell state transitions occur in much shorter time scales.

- During development, cells traverse the landscape through differentiation trajectories involving “intermediate” or “transitional” cell states



Regev et al (2017)
Xia & Yanai (2019
Wagner & Klein (2020)
Tanay & Sebé-Pedrós (2021)

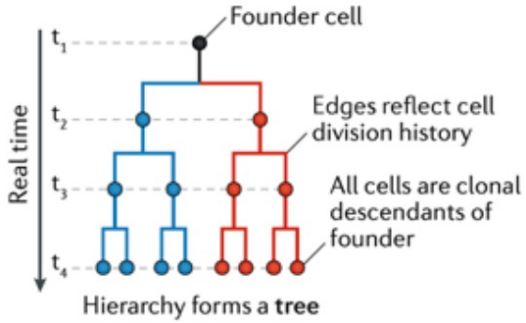
Organizing principles for cell type classification



Phenotypic

- Histological
- Functional
- Molecular

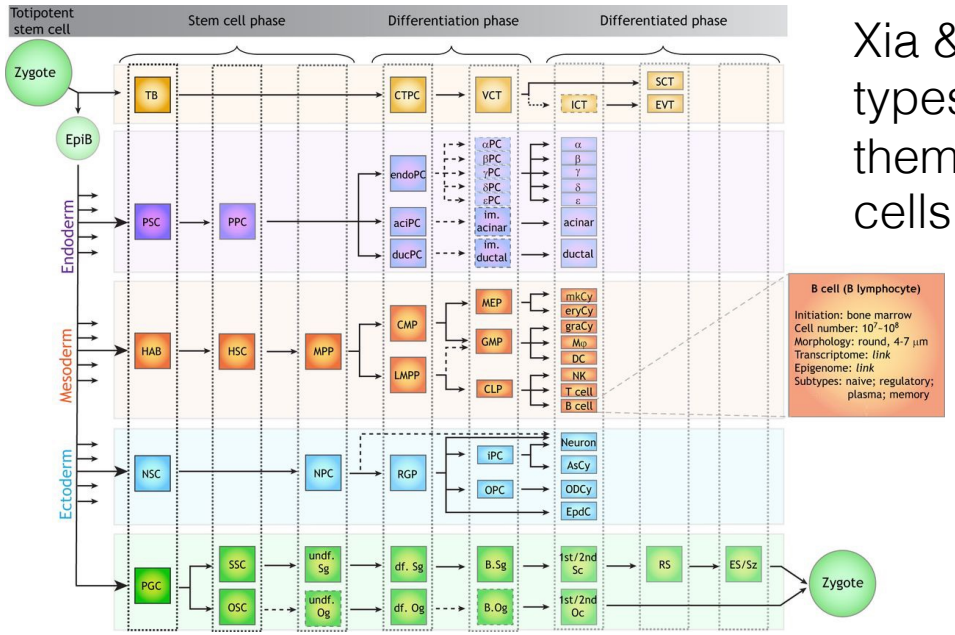
Based on global phenotypic similarity



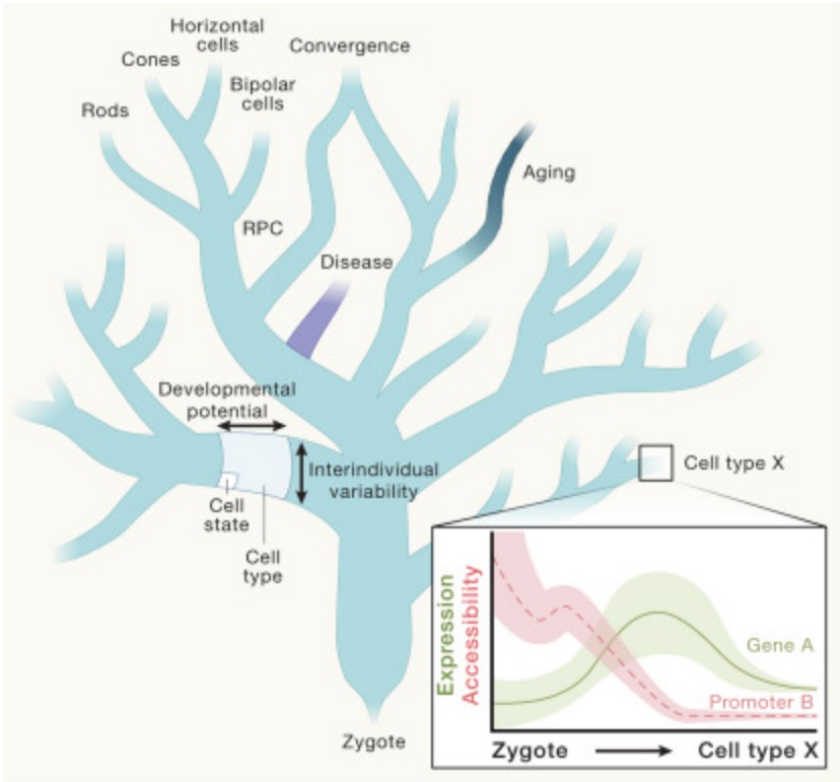
Developmental/lineage

Based on mitotic histories (ontogeny)

Beyond phenotypic similarity, developmental criteria as an organizing principle for cell type classification



Xia & Yanai (2019) proposed a periodic table that aligns cell types according to their <u>developmental stages</u>, connecting them to one another according to the universal axis from stem cells to differentiated cells.



Domcke & Shendure (2023) proposed a “consensus ontogeny” of cell types, integrating developmental and molecular cell type organizing principles

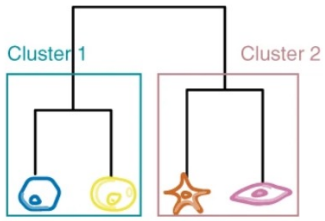
The metaphor of cell types as species

“Throughout the history of discussions on cell type, cell biologists have compared the problem of defining cell type with the interminable and often contentious debate over the definition of arguably the most important concept in systematics and evolutionary biology, ‘species’”

Parallels exhaustively explored by J. Doyle (2022)

The cladistic [phylogenetic] approach largely supplanted phenetics [global similarities] in eukaryotic systematics by the 1980s.

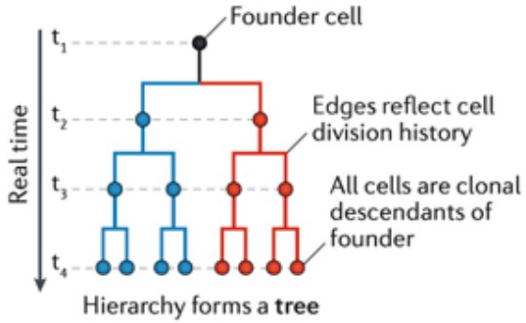
Three kinds of cell trees,
three possible organizing principles for cell type classification



Phenotypic

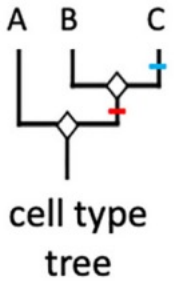
- Histological
- Functional
- Molecular

Based on global phenotypic similarity



Developmental/lineage

Based on mitotic histories (ontogeny)



Cladistic/phylogenetic

Based on evolutionary history

Beyond phenotypic and developmental criteria, evolution as a principle for cell type definition

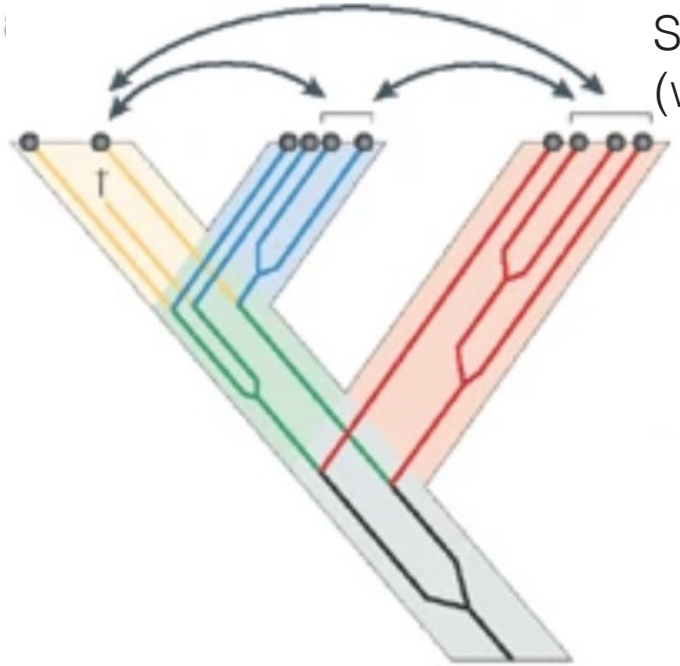
"Cell types are evolutionary units defined by common descent rather than phenotypic similarity, and characterized by their ability to evolve gene expression programmes independently of each other."

Each cell type is characterized by the presence of a unique Core Regulatory Complex (CoRC), defined as "A protein complex composed of terminal selector transcription factors that enables and maintains the distinct gene expression program of a cell"

Arendt et al 2016 & 2019

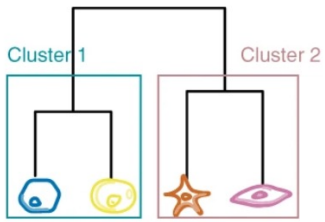
Homologous cell types
(between species)

Sister cell types
(within species)



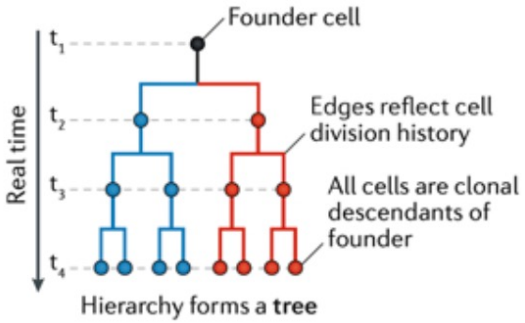
A species phylogenetic tree
with a cell typogenetic tree
superimposed

Three kinds of cell trees, three possible organizing principles for cell type classification

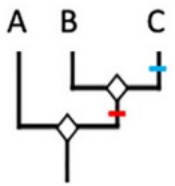


Phenotypic

- Histological
- Functional
- Molecular



Developmental/lineage



Phylogenetic

Main challenges

- Many subjective choices involved in clustering.
- Mostly operational, technology-specific, provisional classifications.

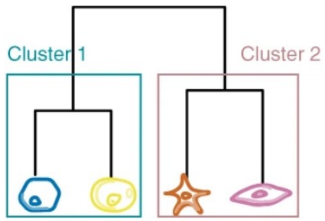
- Technically very challenging: so far complete only in C. elegans that has ~1000 cells and is transparent.
- Different cell lineages can produce phenotypically equivalent cells

- Requires well-annotated genomes, comparative gene ontologies, and consistently high-quality transcriptomic data generation from many species
- Challenges of identifying orthologous cell types across large evolutionary distance
- Not clear how to identify cell type-defininig CoRCs

Arendt et al (2016), J. Doyle (2022), Domcke & Shendure (2023)

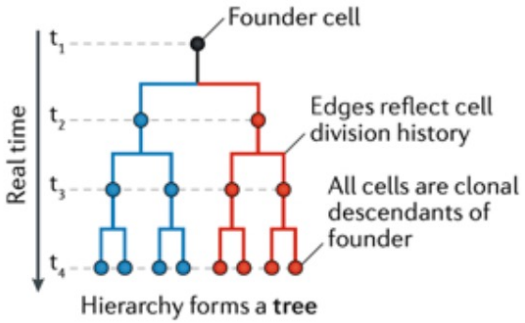
Single-cell technologies are profoundly changing the understanding and operational definitions of cell type.

And this is just the beginning.



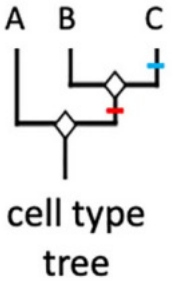
Phenotypic
- Histological
- Functional
- Molecular

multi-modal profiling
- RNA + Chromatin Accessibility
- RNA + Proteins
- RNA + Protein + xyz coordinates
- RNA + functional readouts (PATCH-seq, live-seq, etc.)



Developmental/lineage

in vivo DNA writers – lineage tracing in complex organisms



Phylogenetic

Increasing throughput and multiplexing enabling profiling of thousands of species to model evolutionary processes

Tanay & Sebé-Pedrós (2021), Hongkui Zeng (2022), Domcke & Shendure (2023)


Despite having developed the most perfect technologies,
humans were never able to define what is a cell type