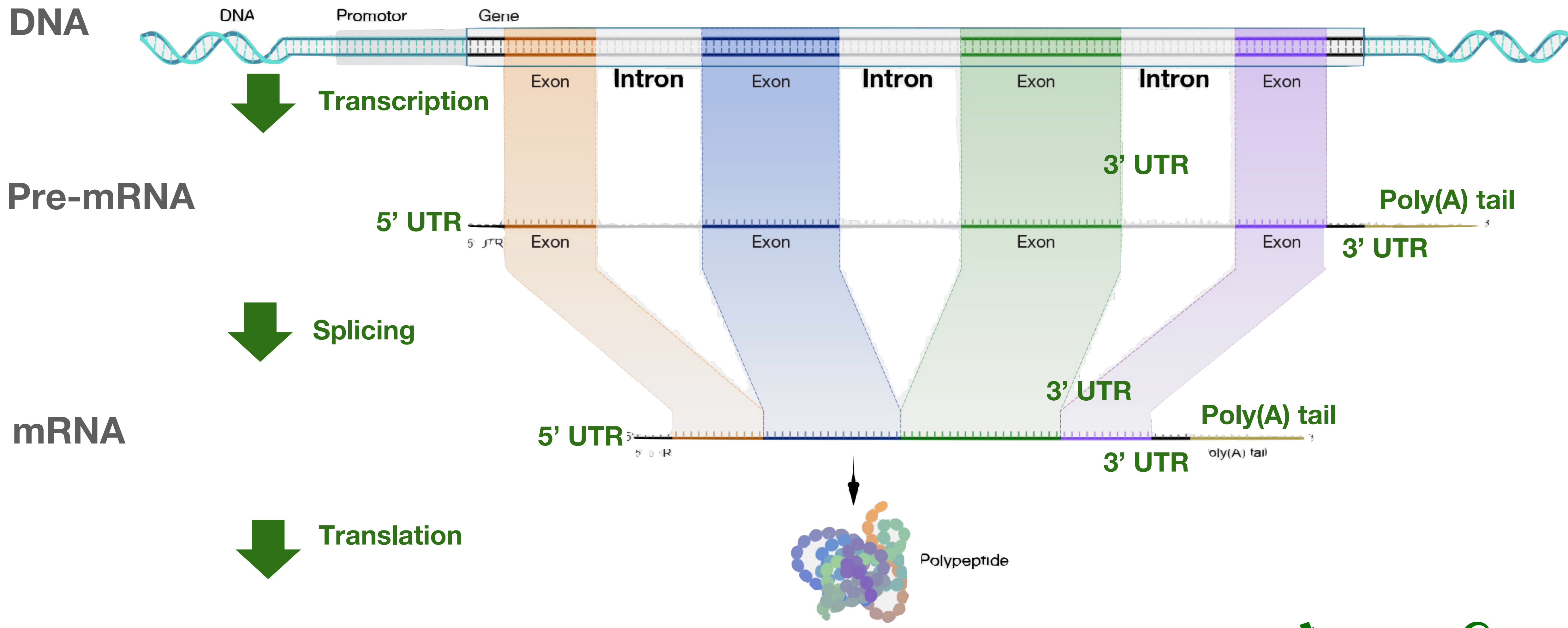




Finding Targets

mRNA expression



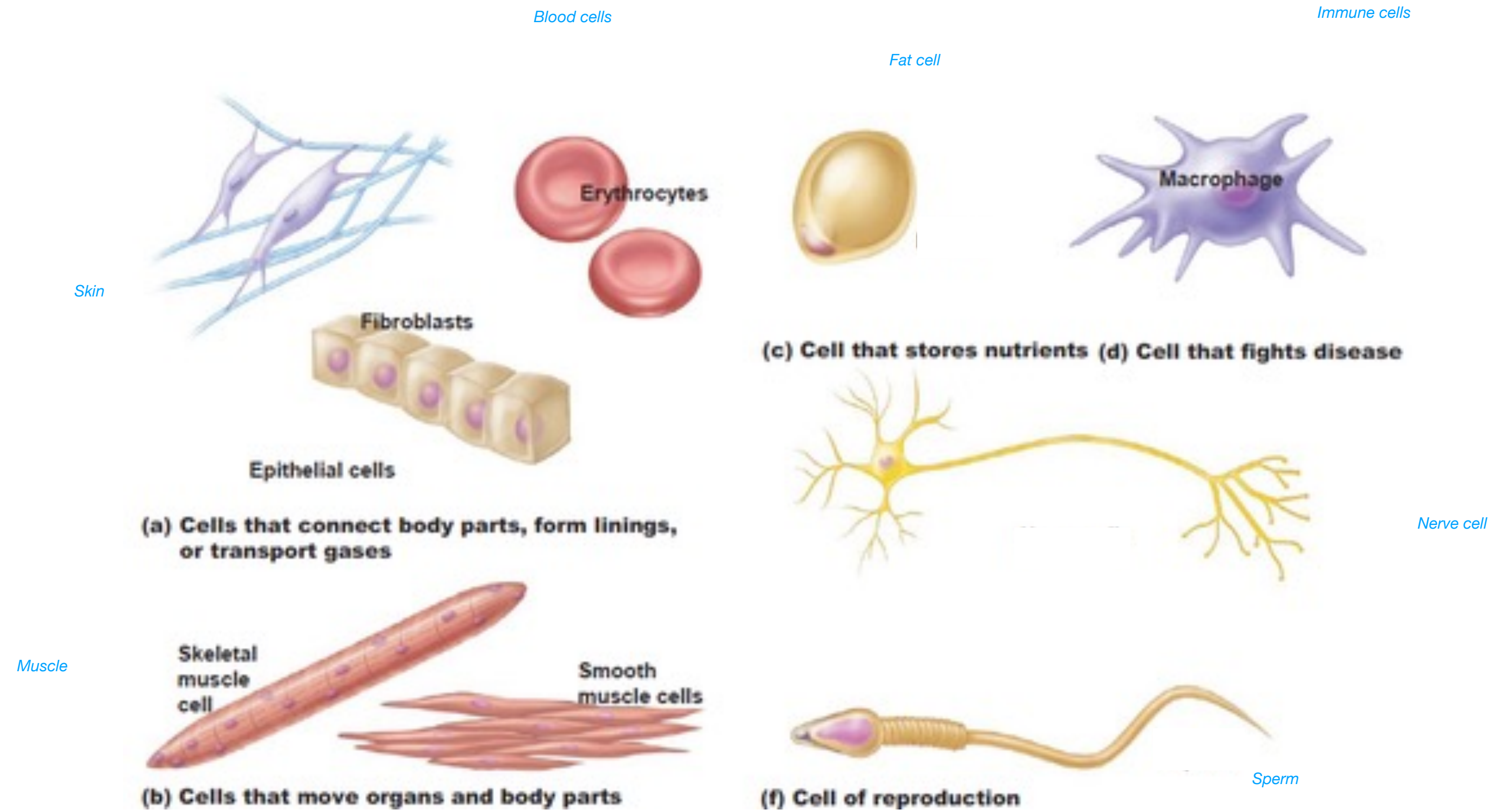
One genome, many outputs

In human body there are:

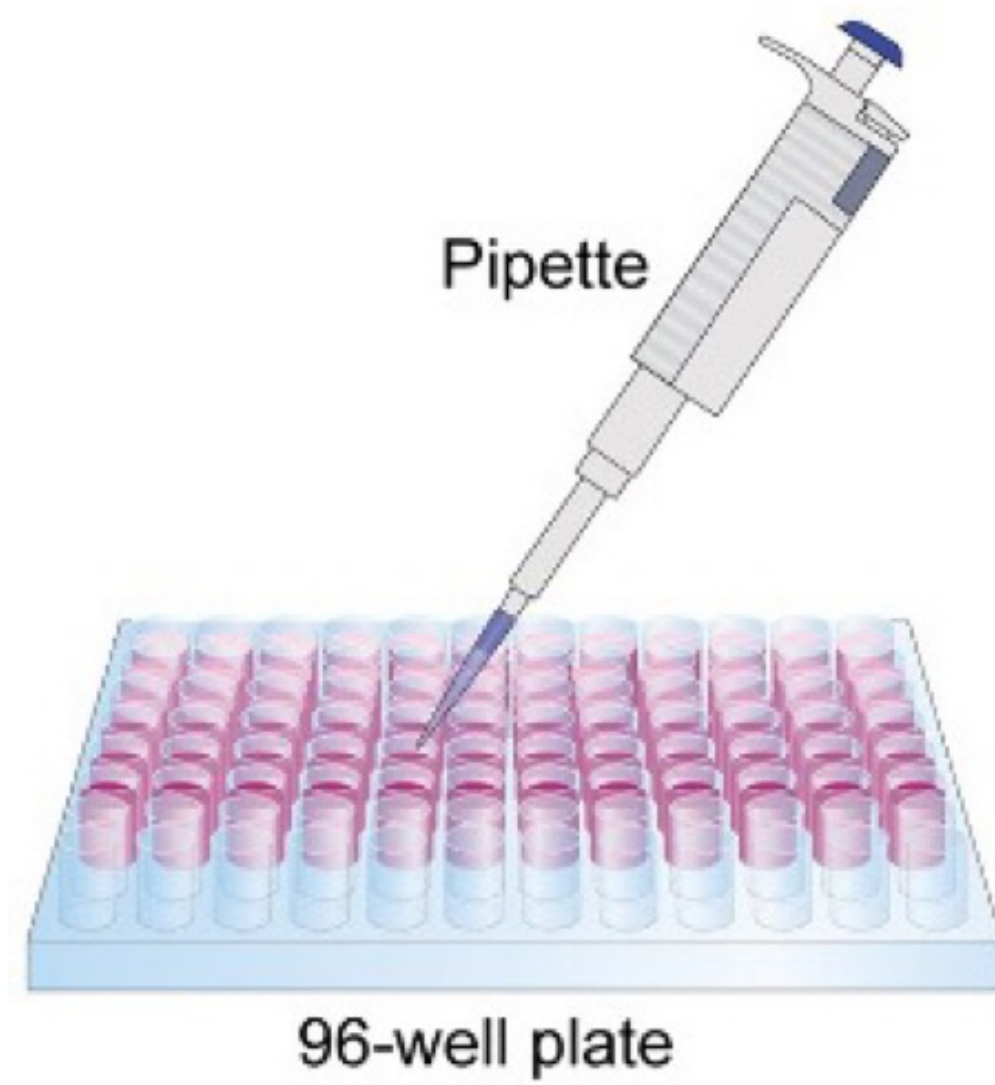
~ 100 trillion cells

~ 200 types of cells

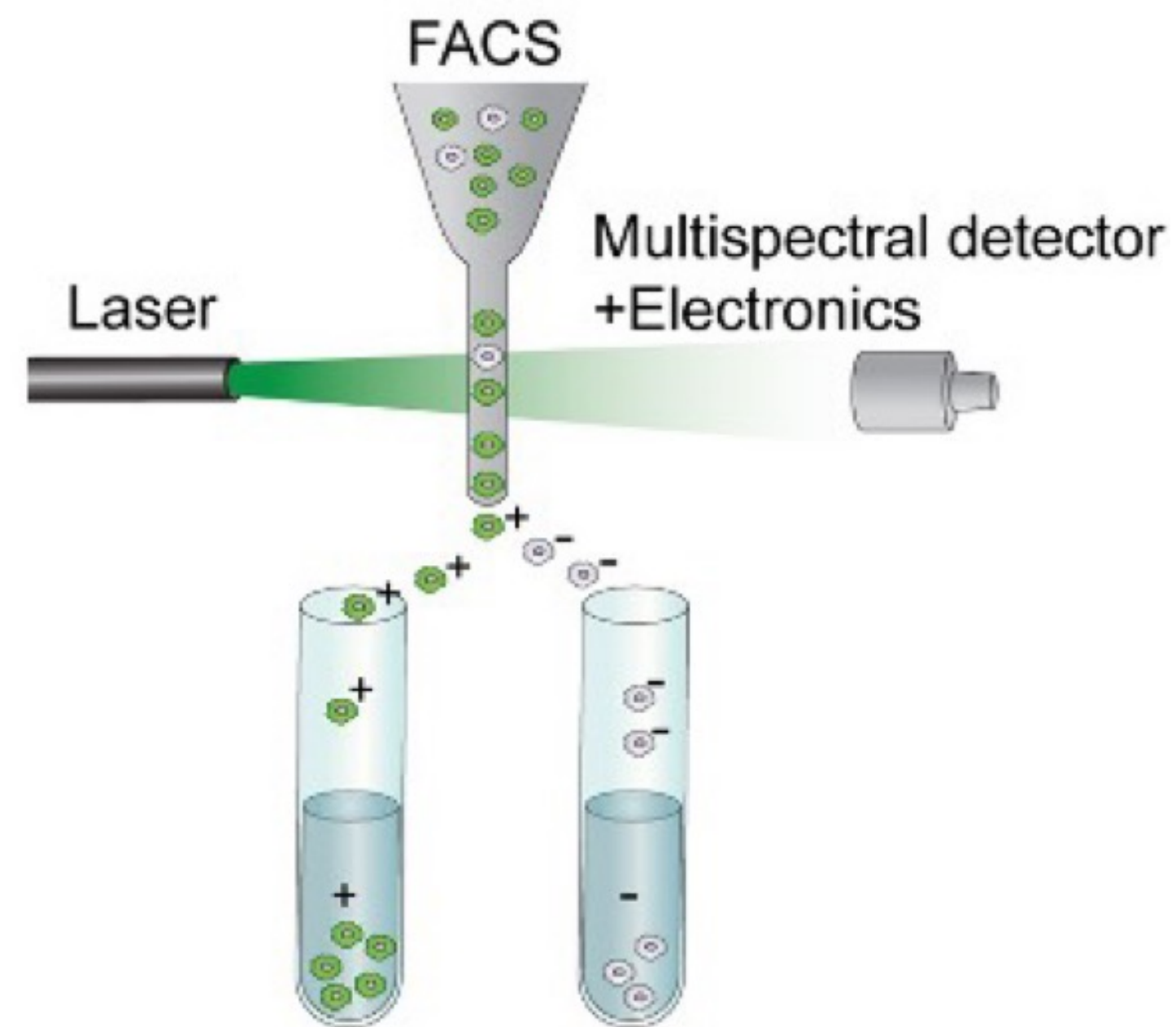
Individual cell, tissue and organ outputs from the genome are very complex



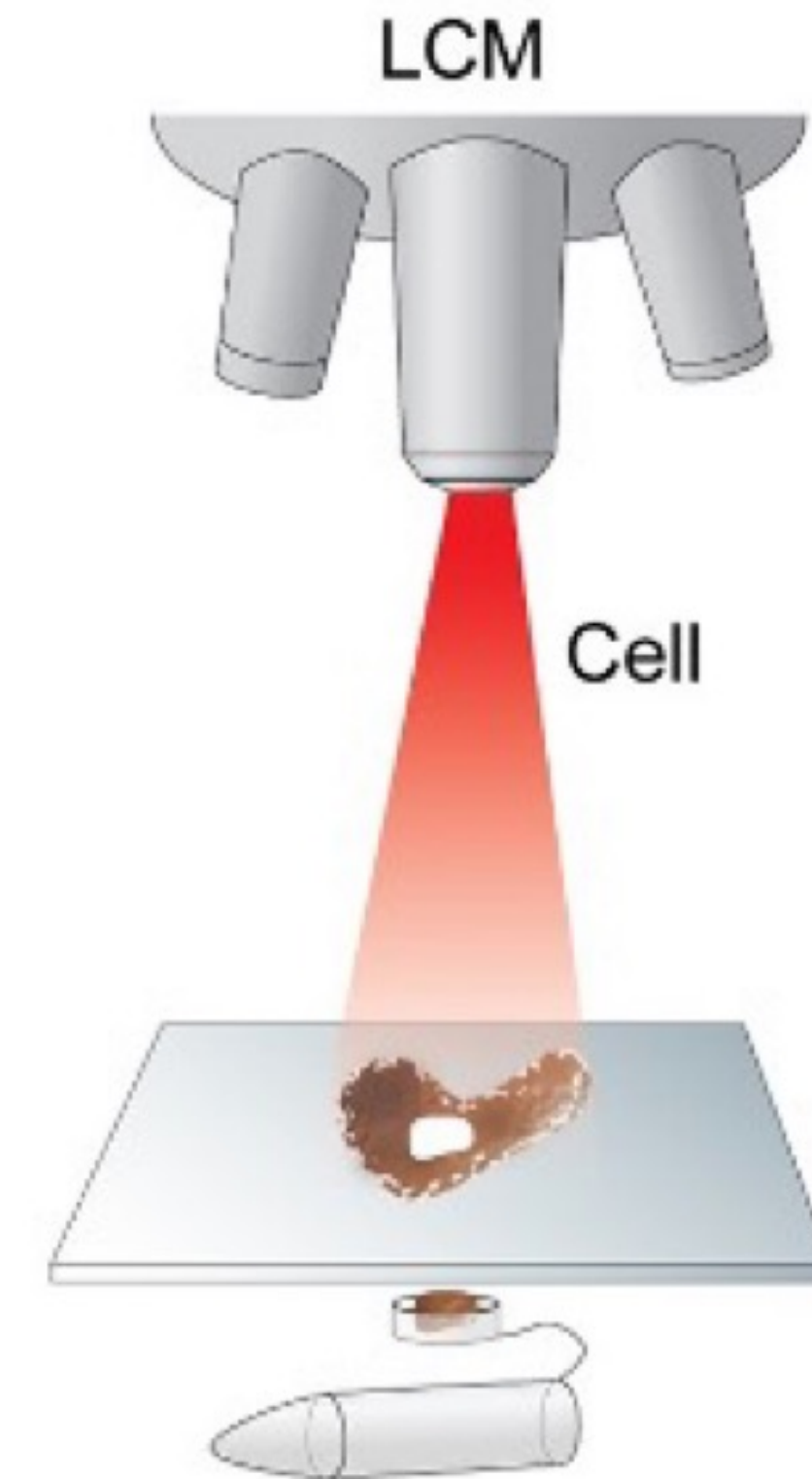
Getting Samples to Sequence



Cell lines

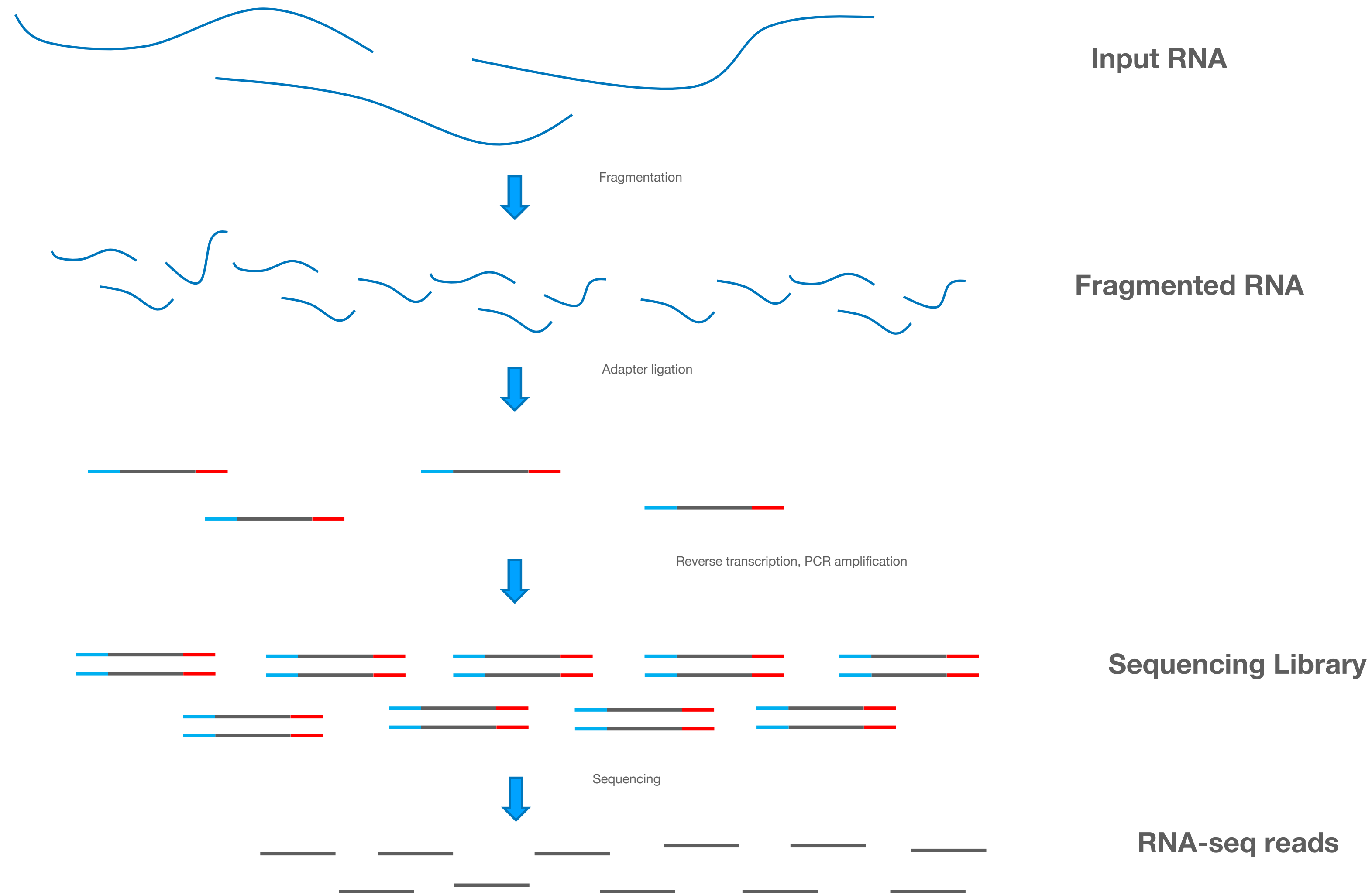


Fluorescence-activated cell sorting



Laser Capture Microdissection

Bulk RNA sequencing

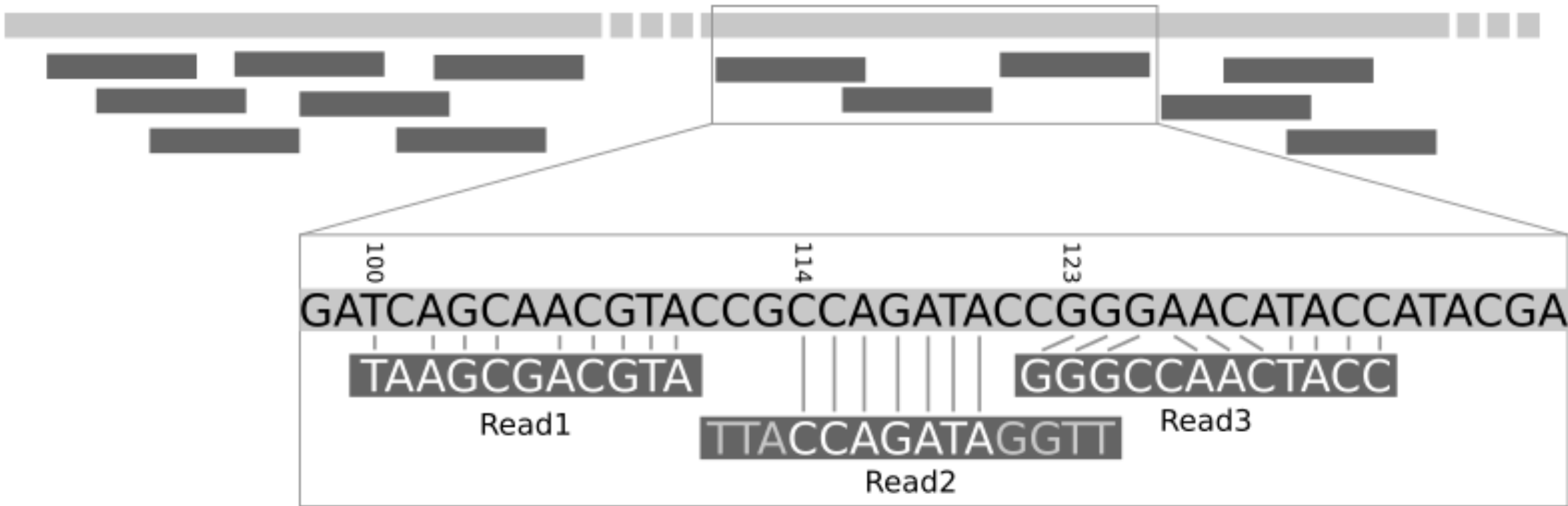


RNA sequencing



RNA-seq raw reads

Mapping to reference genome



Mapped reads



Quantify gene expression as read abundance, per gene

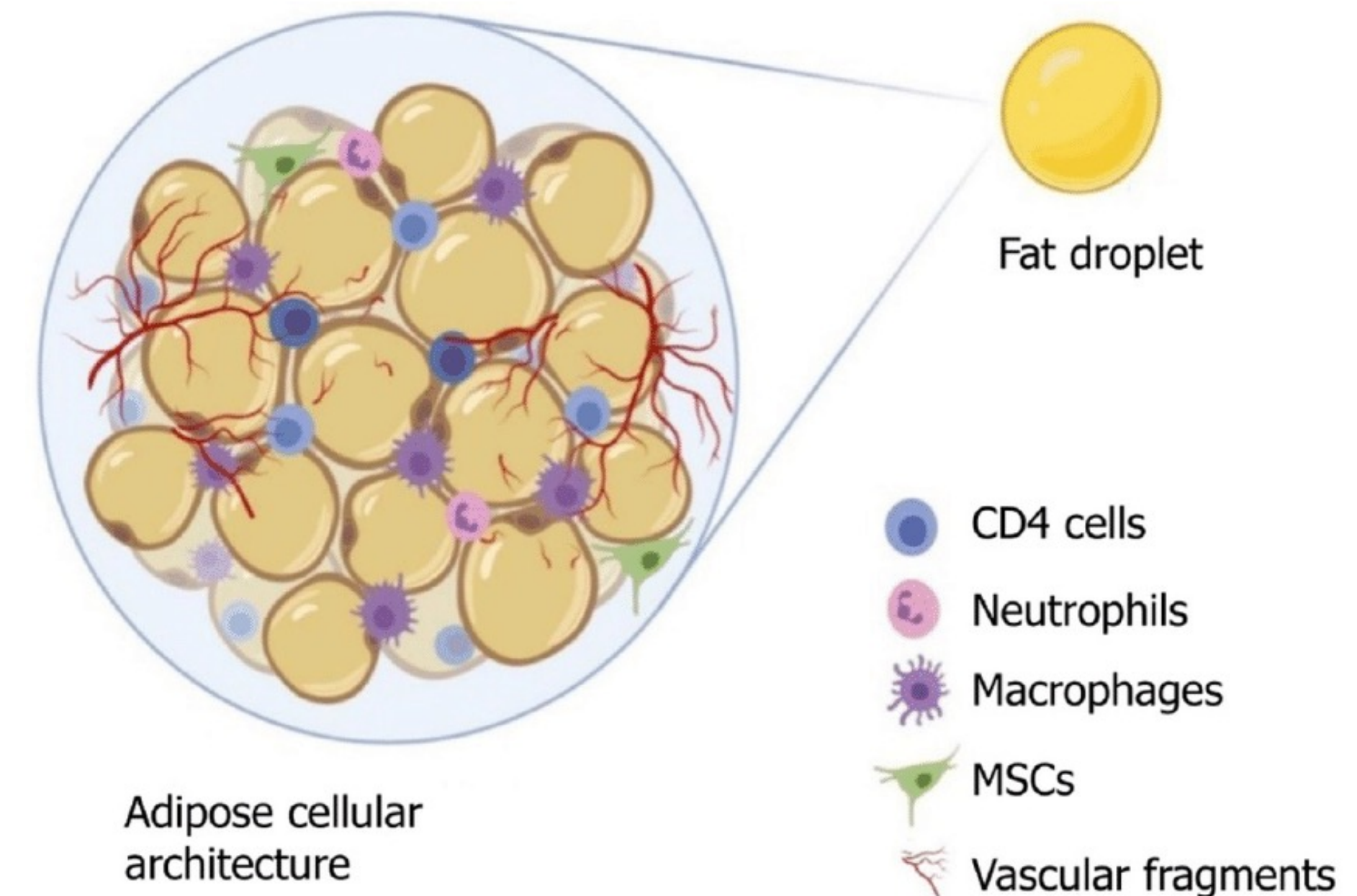


Read count matrix

	Sample 1
Gene A	5
Gene B	2
Gene C	6

Limits of Bulk RNA sequencing

- Requires usually **thousands or millions of cells**
- Useful for **tissues or cell lines**.
- **Rare cell types and states cannot** be analyzed
- Tissues are generally **heterogeneous**
- Most often **multiple cell types** in a tissue
- Bulk RNA-seq sample is an **average** transcriptome of many cells
- No information about expression levels or heterogeneity of **individual cells**

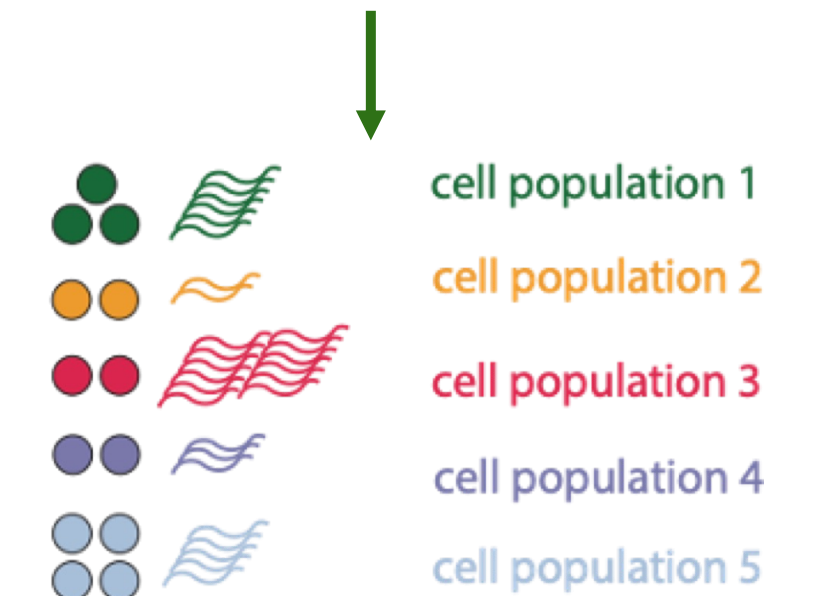
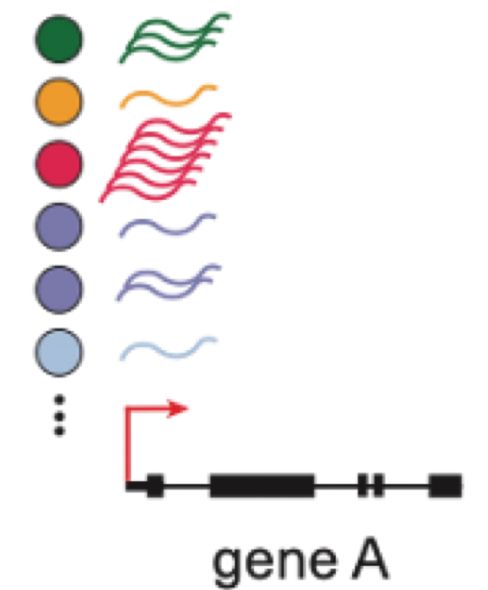


Single cell (sc) RNA sequencing

- **scRNA-seq enables:**
- Identification of **rare cell types** (e.g. early development, stem cells, circulating tumour cells)
- Elucidating **gene regulatory networks**
- Studying **heterogeneity** (e.g. tissue composition, cancer, temporal processes)
- Exploring changes in expression while **incorporating** other information e.g spatial, regulatory or protein information
- **Single cell phenomena** (gene expression stochasticity, mono-allelic expression)

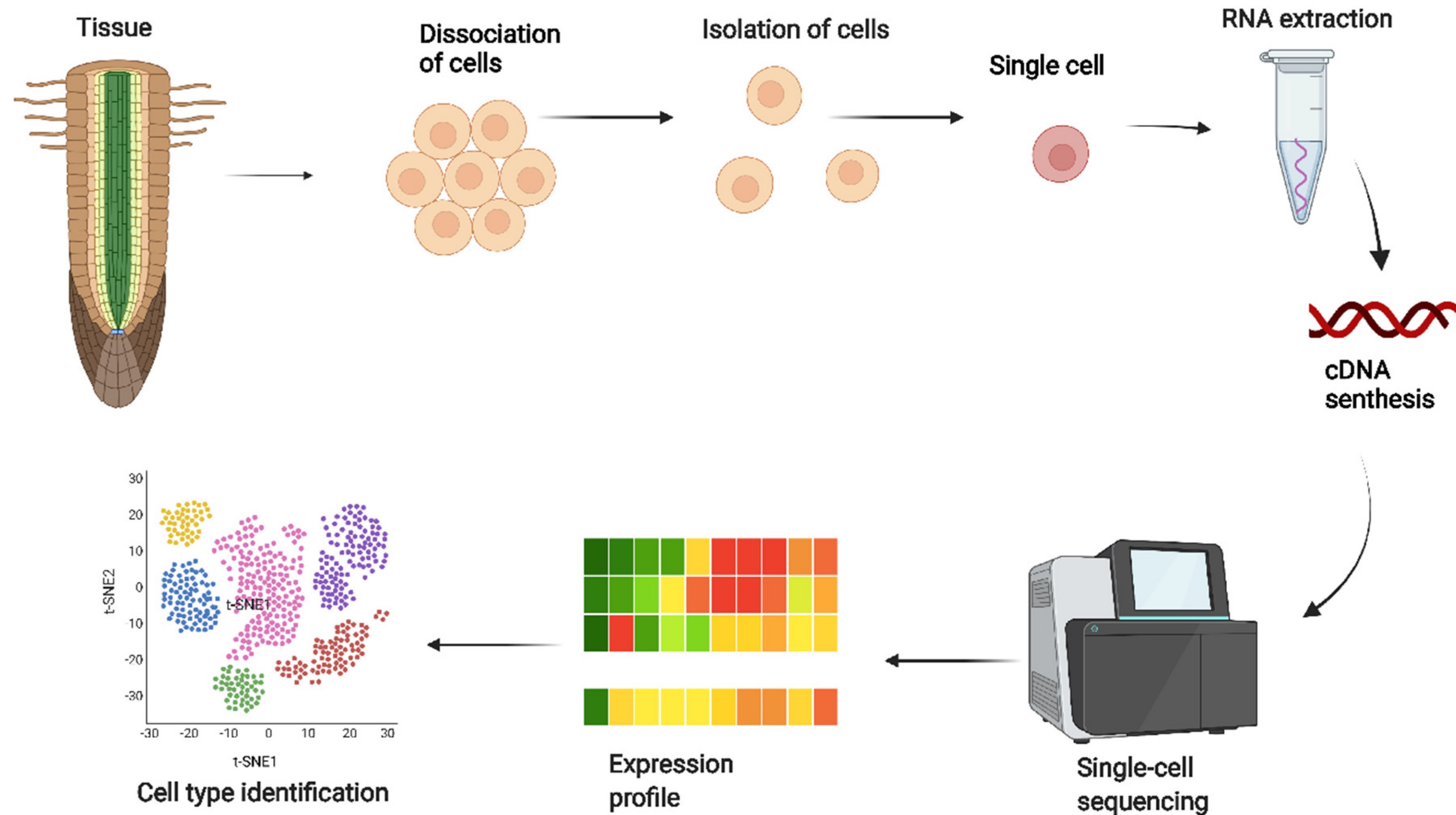


Sample



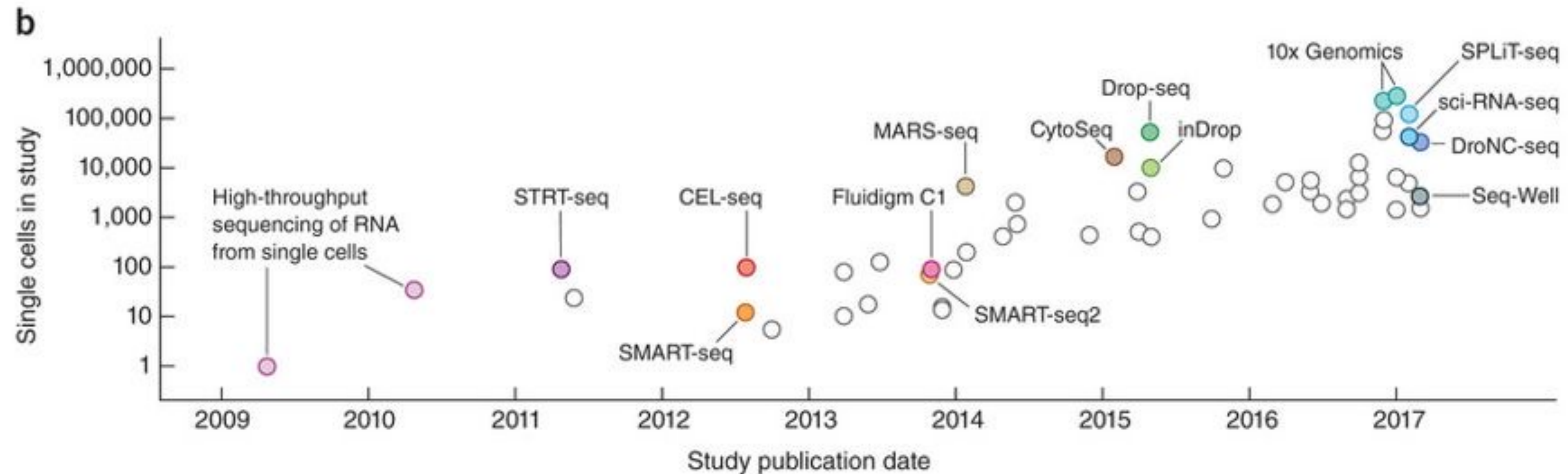
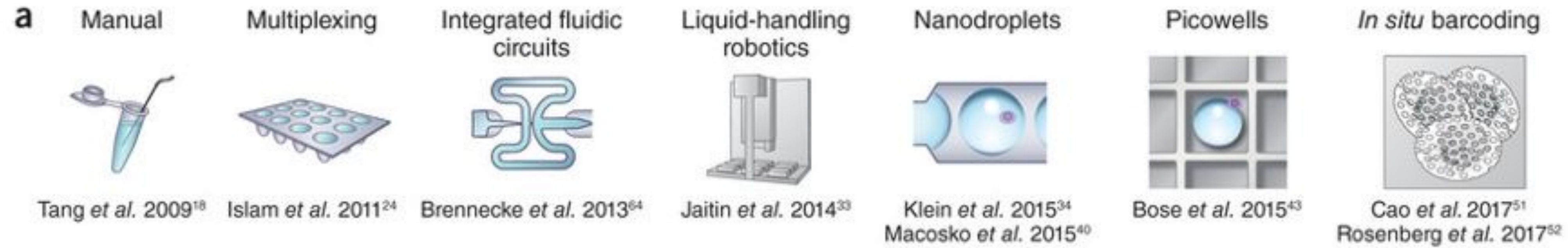
Distribution of gene expression levels in a cell population

scRNA sequencing overview



Bawa et al, IJMS, 2022

scRNA sequencing technology

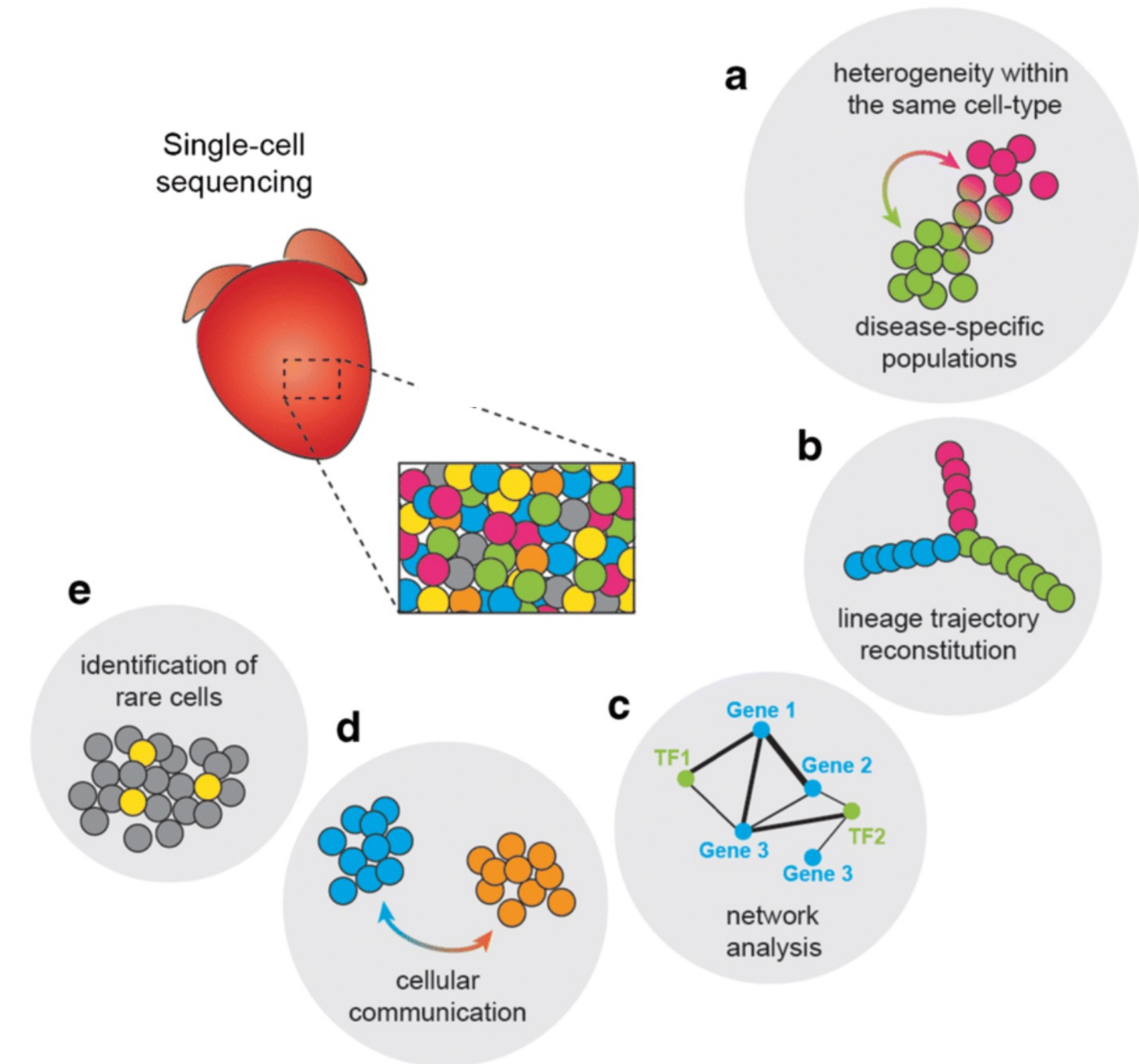


Drop sequencing

- Most commonly used current methods

10x genomics - commercial

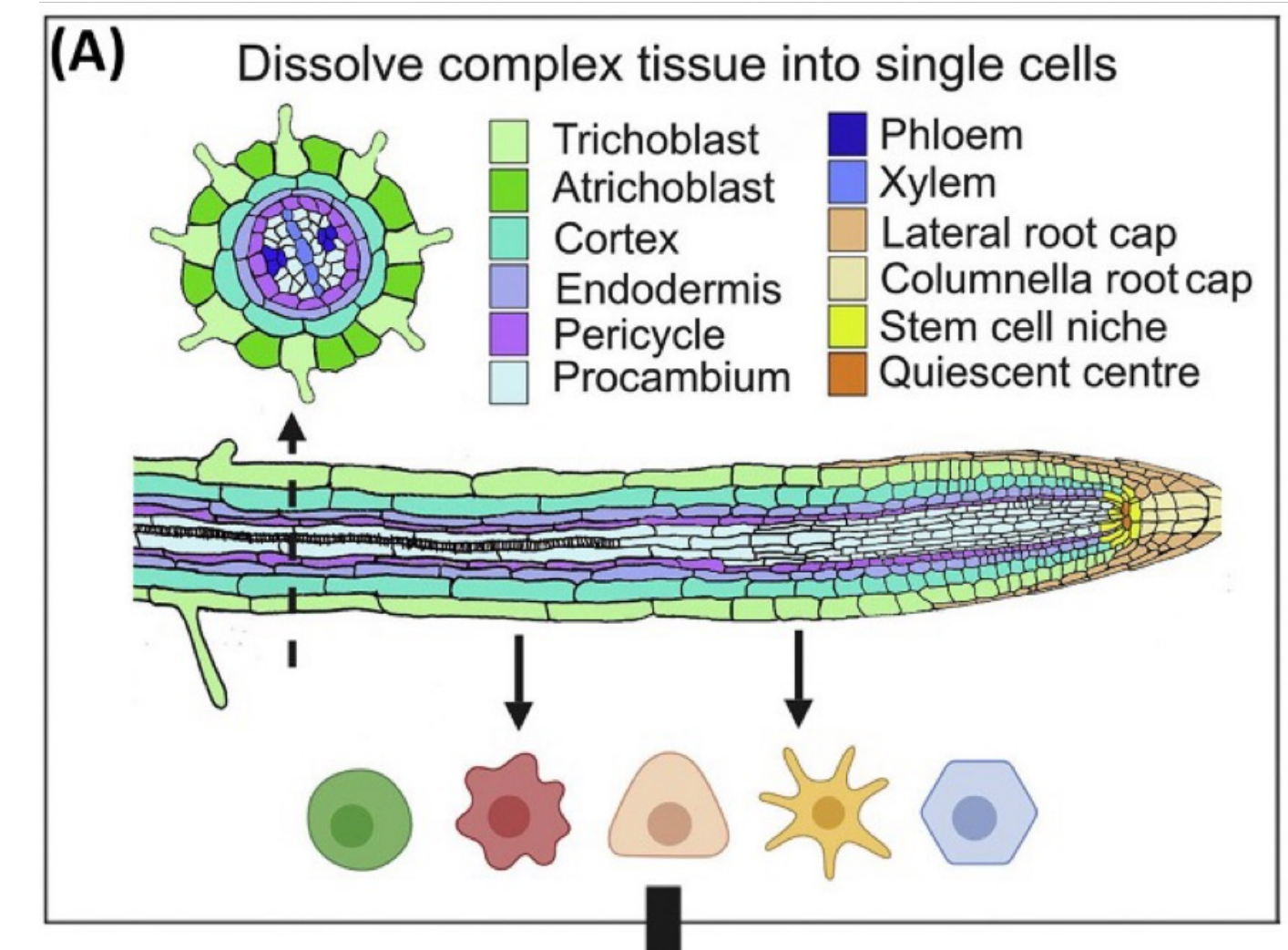
Drop-seq - cheaper, more complicated to set-up



scRNA sequencing

scRNA-seq procedure

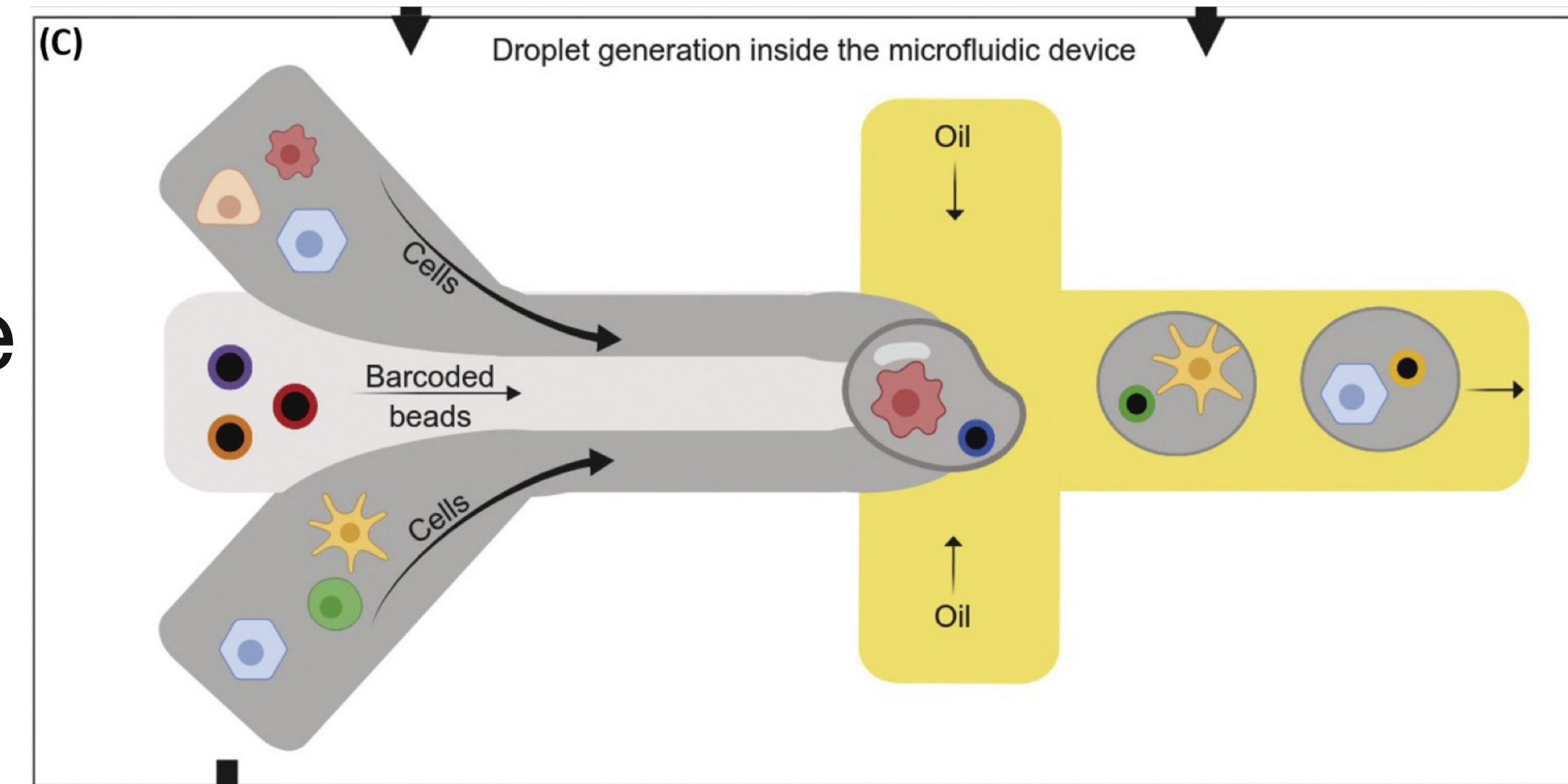
- Disassociate tissue into **single cells**



scRNA sequencing

scRNA-seq procedure

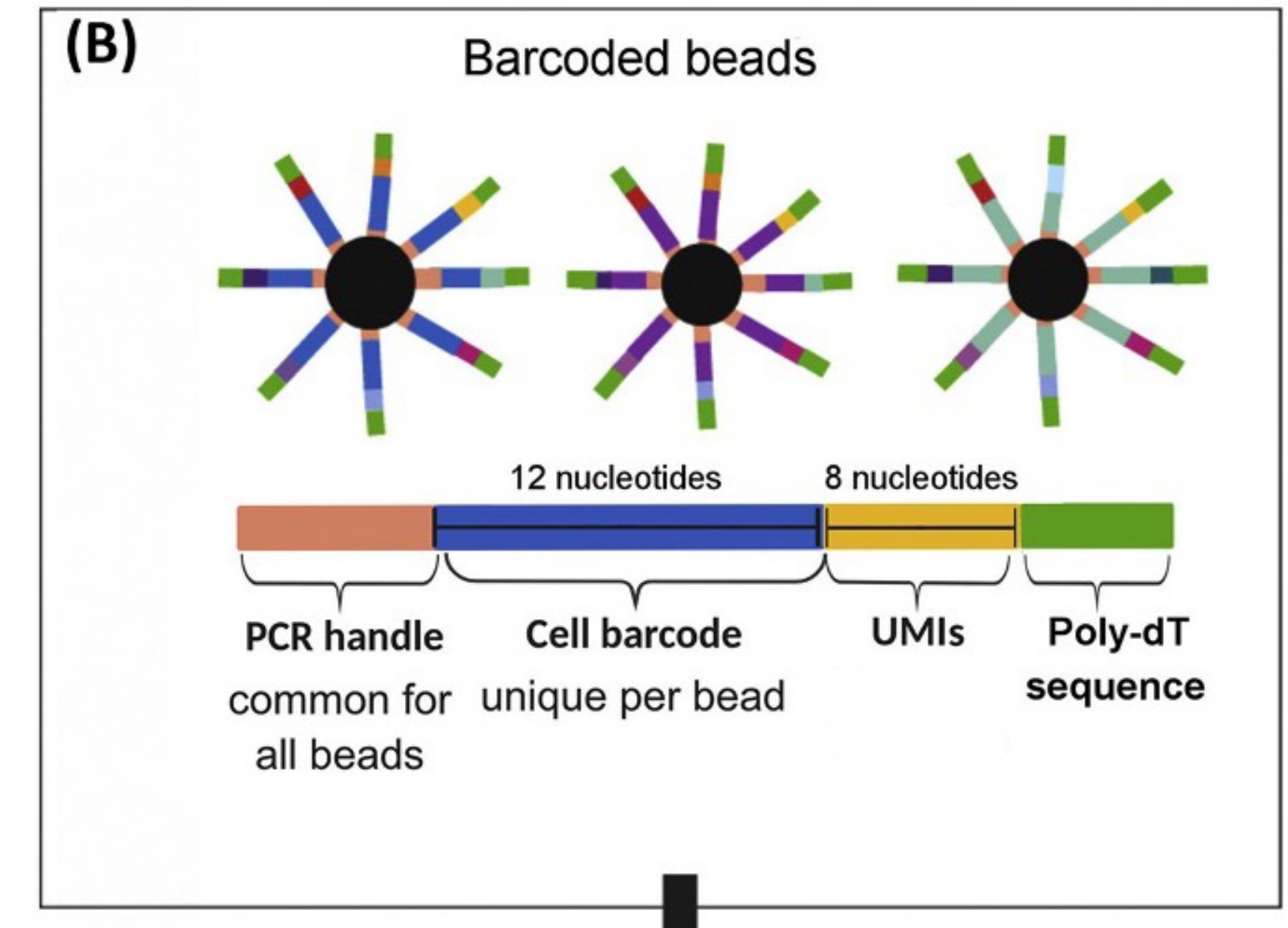
- Disassociate tissue into **single cells**
- Introduce cells into **microfluidic** device



scRNA sequencing

scRNA-seq procedure

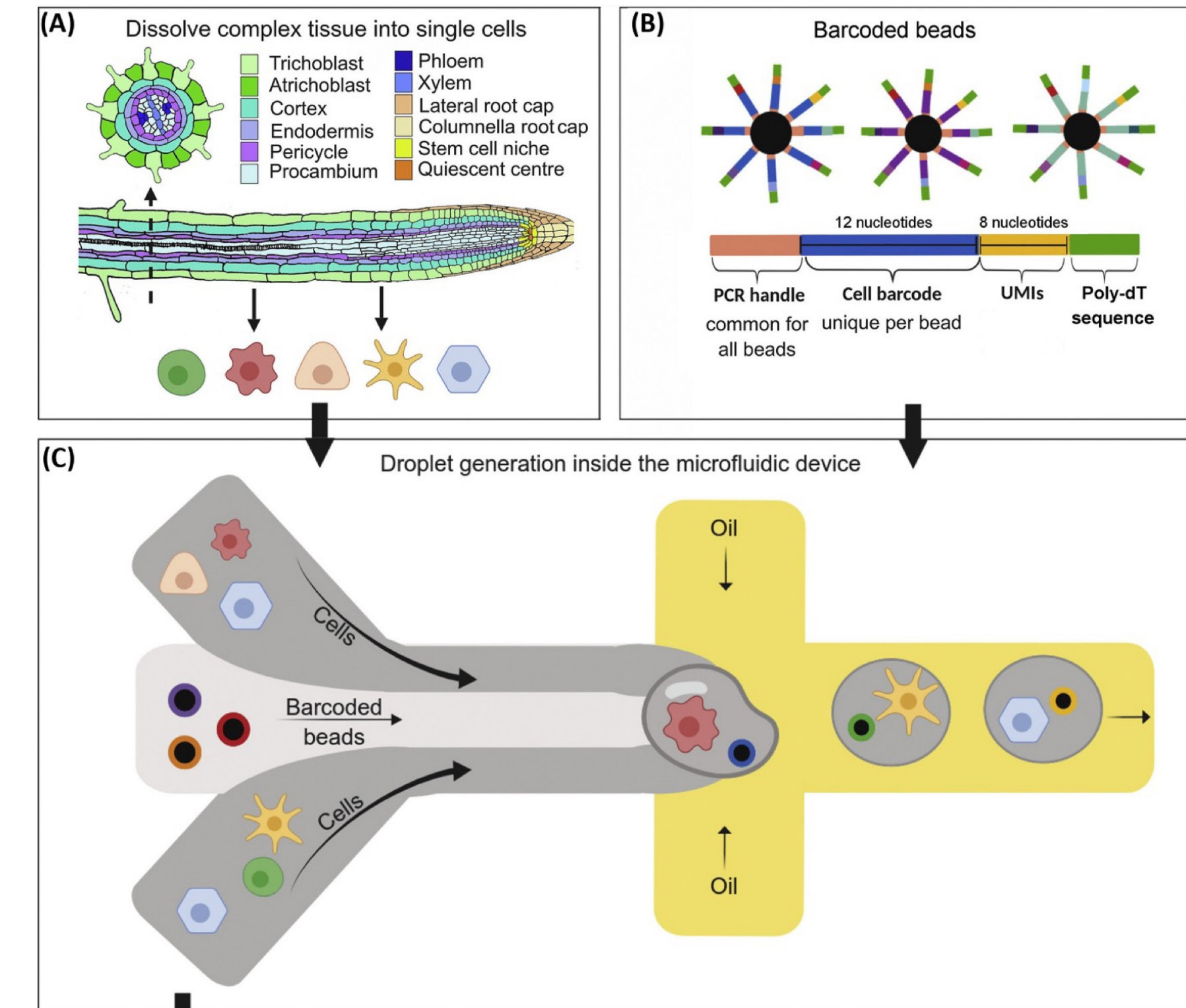
- Disassociate tissue into **single cells**
- Introduce cells into **microfluidic** device
- Combine cells with **barcoded beads** into a droplet



scRNA sequencing

scRNA-seq procedure

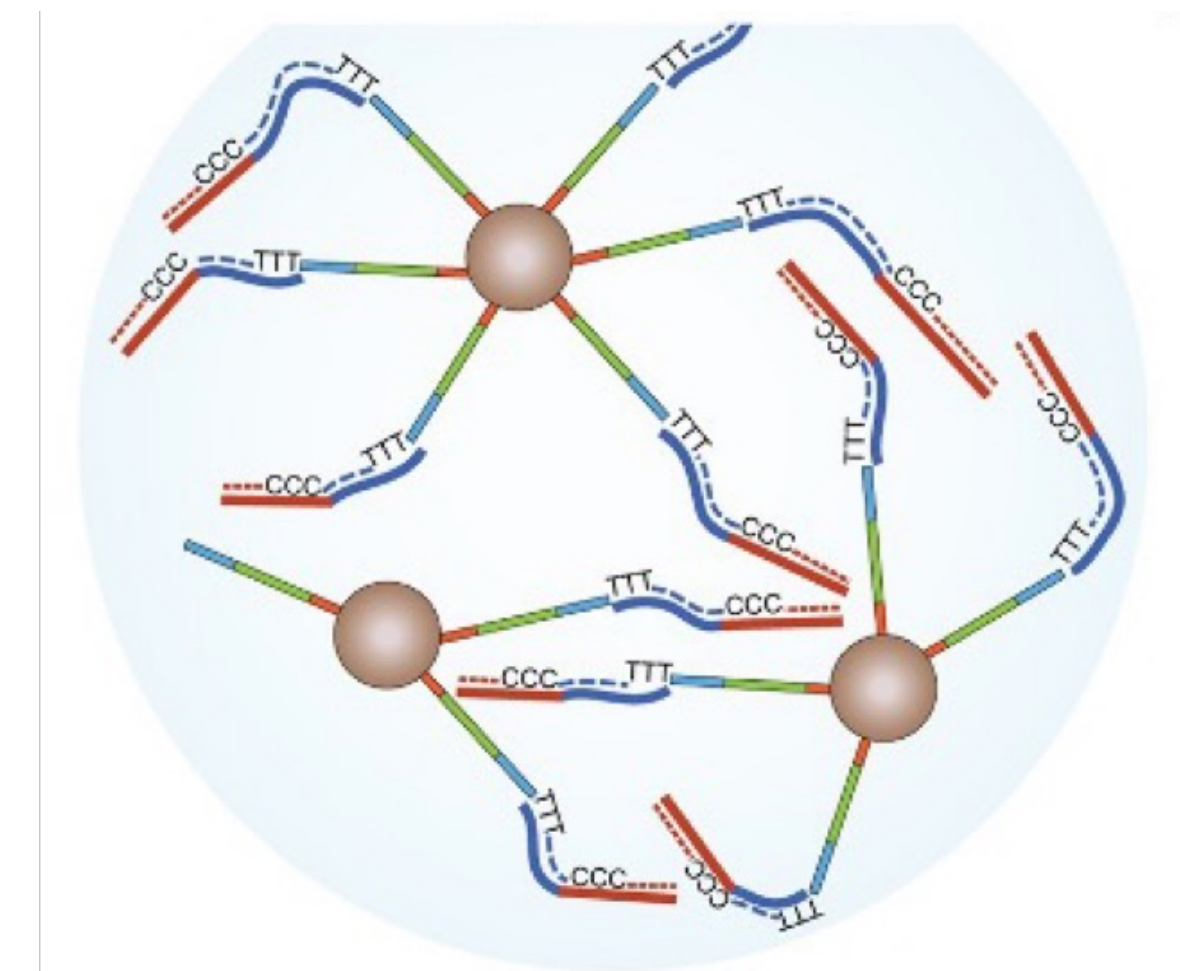
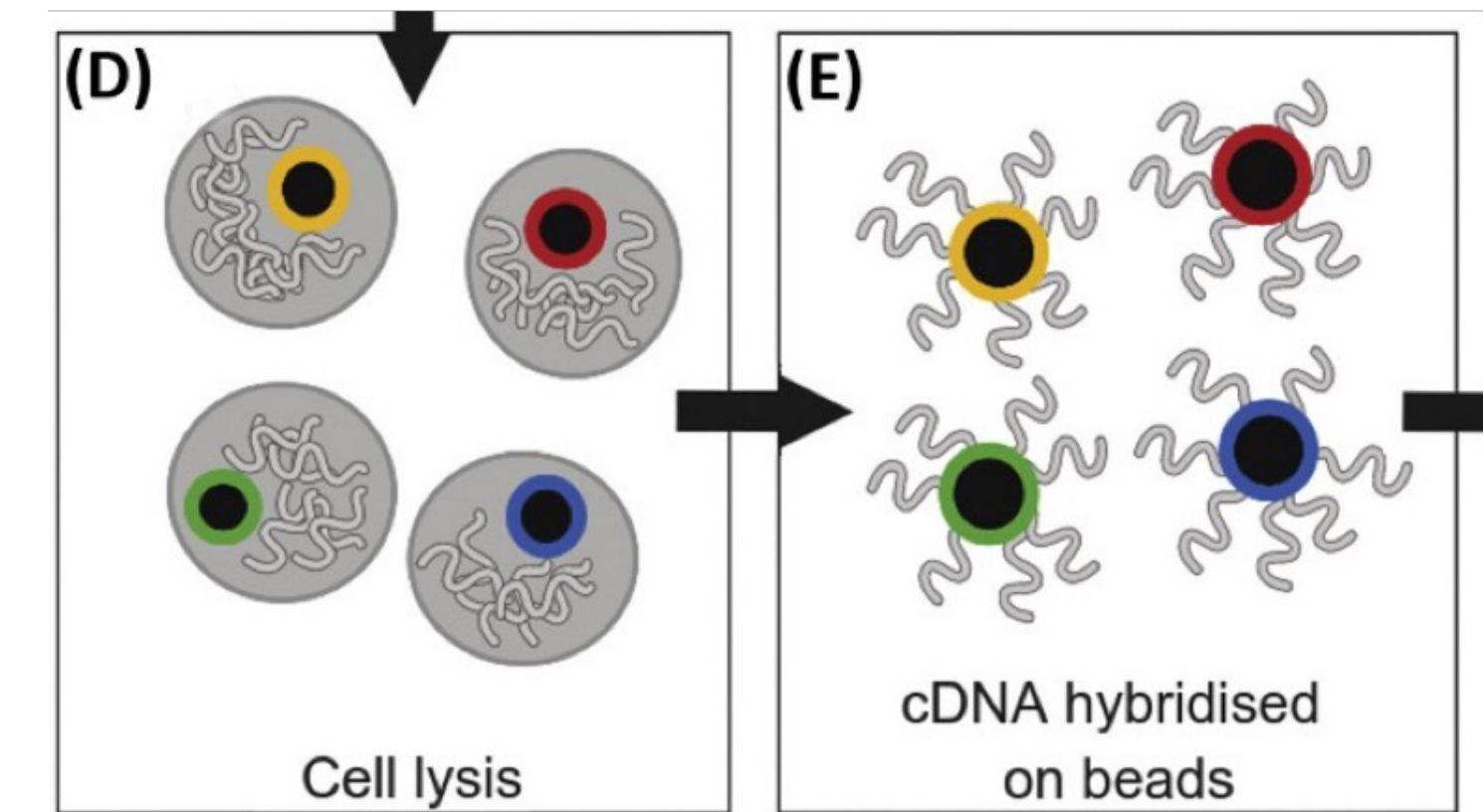
- Disassociate tissue into **single cells**
- Introduce cells into **microfluidic** device
- Combine cells with **barcoded beads** into a droplet



scRNA sequencing

scRNA-seq procedure

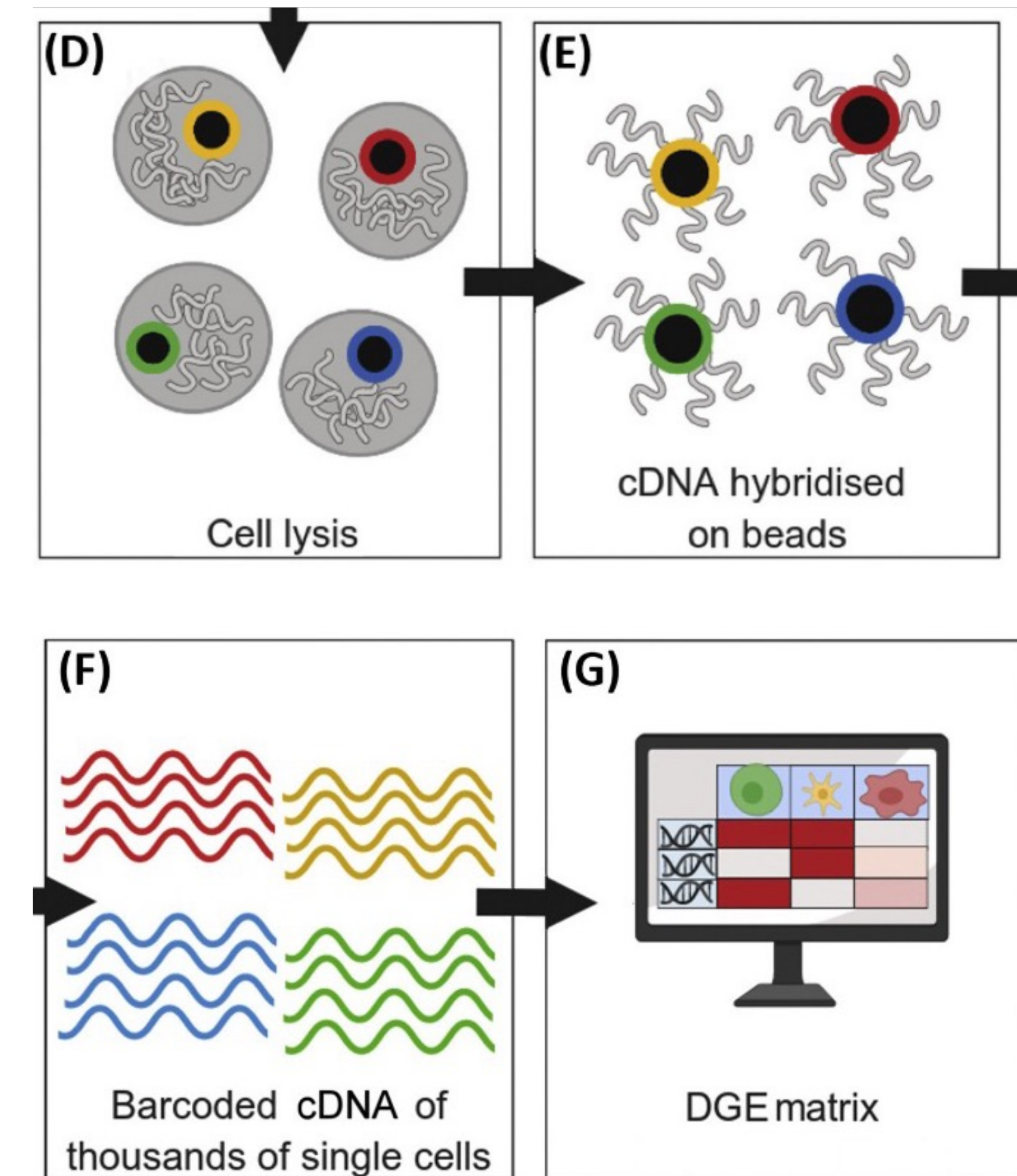
- Disassociate tissue into **single cells**
- Introduce cells into **microfluidic** device
- Combine cells with **barcoded beads** into a droplet
- **Lyse** cells to release RNA and **reverse transcribe** to cDNA
- cDNA derived from cell **hybridises** to beads



scRNA sequencing

scRNA-seq procedure

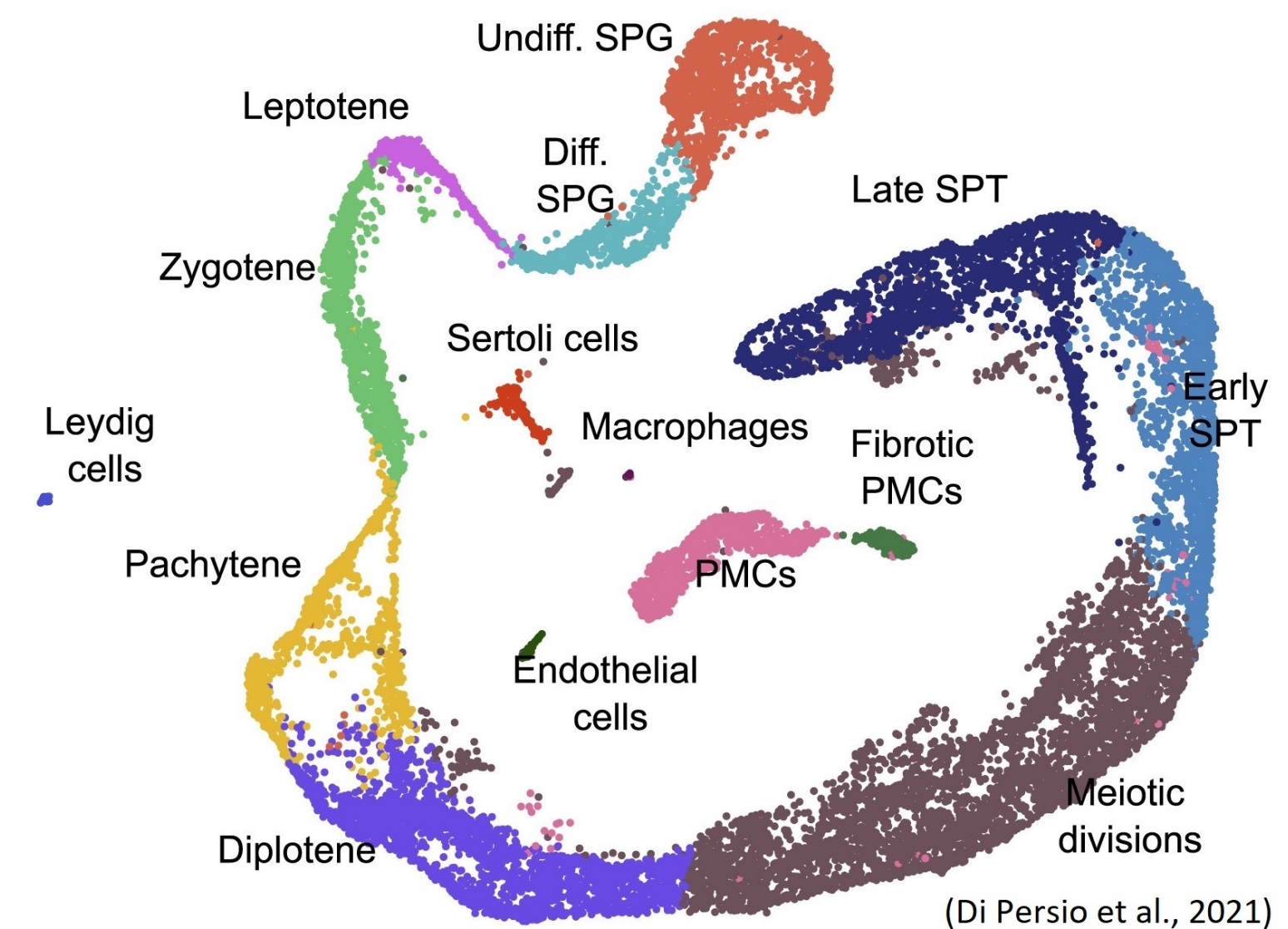
- Disassociate tissue into **single cells**
- Introduce cells into **microfluidic** device
- Combine cells with **barcoded beads** into a droplet
- **Lyse** cells to release RNA and **reverse transcribe** to cDNA
- cDNA derived from cell **hybridises** to beads
- **Barcoded cDNAs** then sequenced



scRNA sequencing

scRNA-seq procedure

- Disassociate tissue into **single cells**
- Introduce cells into **microfluidic** device
- Combine cells with **barcoded beads** into a droplet
- **Lyse** cells to release RNA and **reverse transcribe** to cDNA
- cDNA derived from cell **hybridises** to beads
- **Barcoded cDNAs** then sequenced
- **Analysis** of data is complex



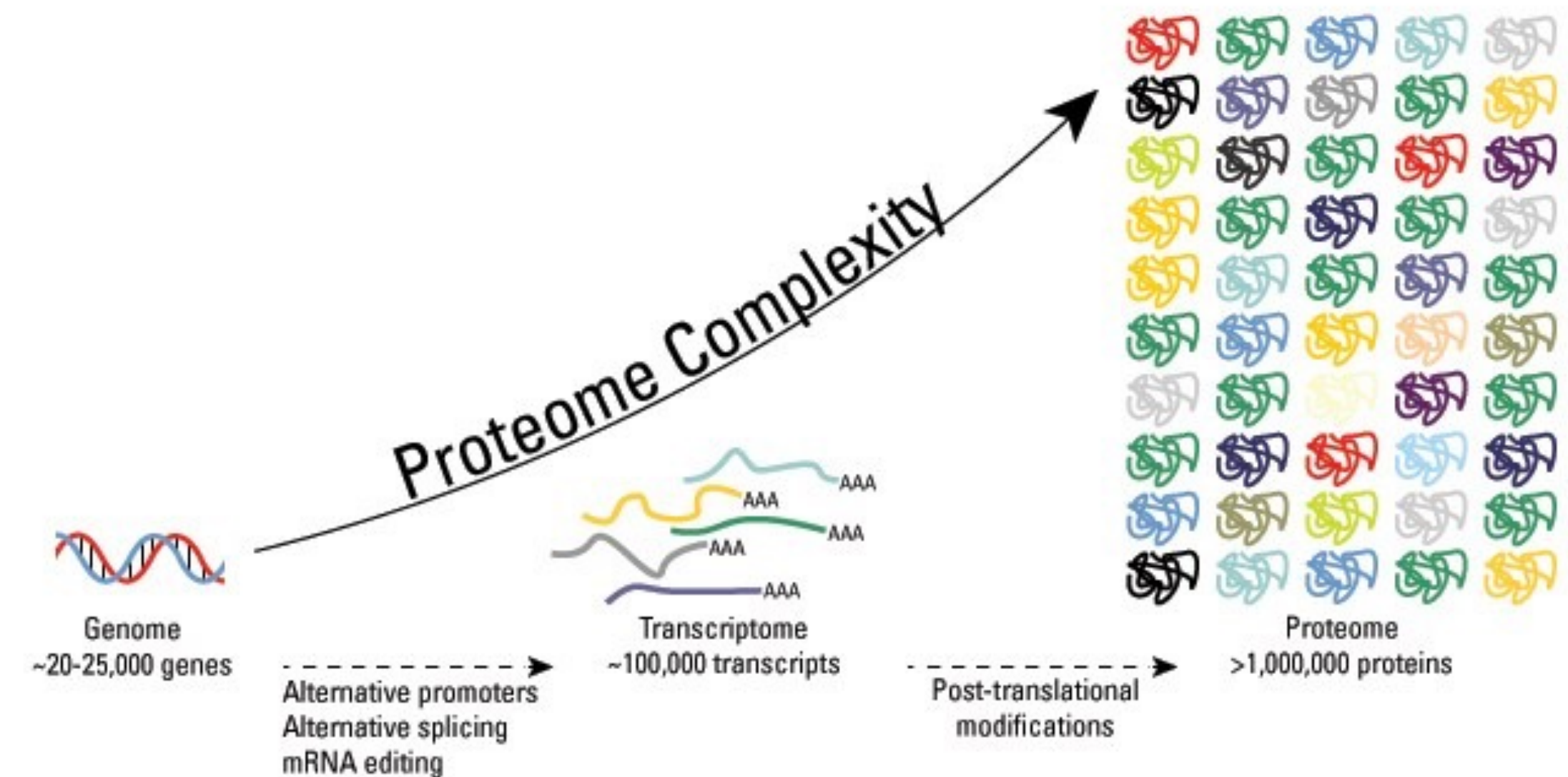


Manipulating Proteins

Proteome vs Genome

Human protein coding

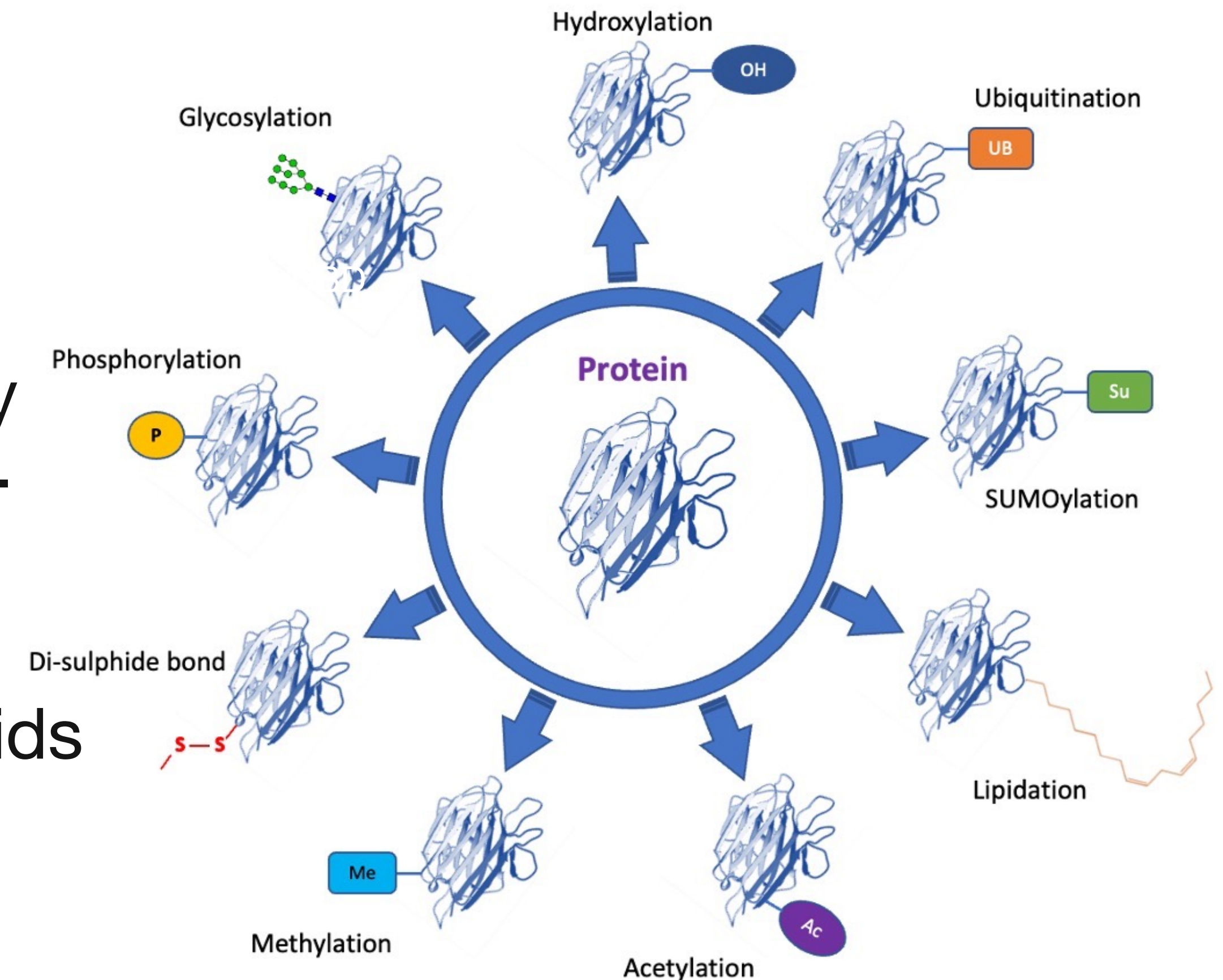
- ~20,500 human genes
- > 100,000s transcript variants
- Many proteins undergo **Post-translational modification (PTM)**
- **Millions of protein** variants per cell type



Post-translational modifications

PTMs

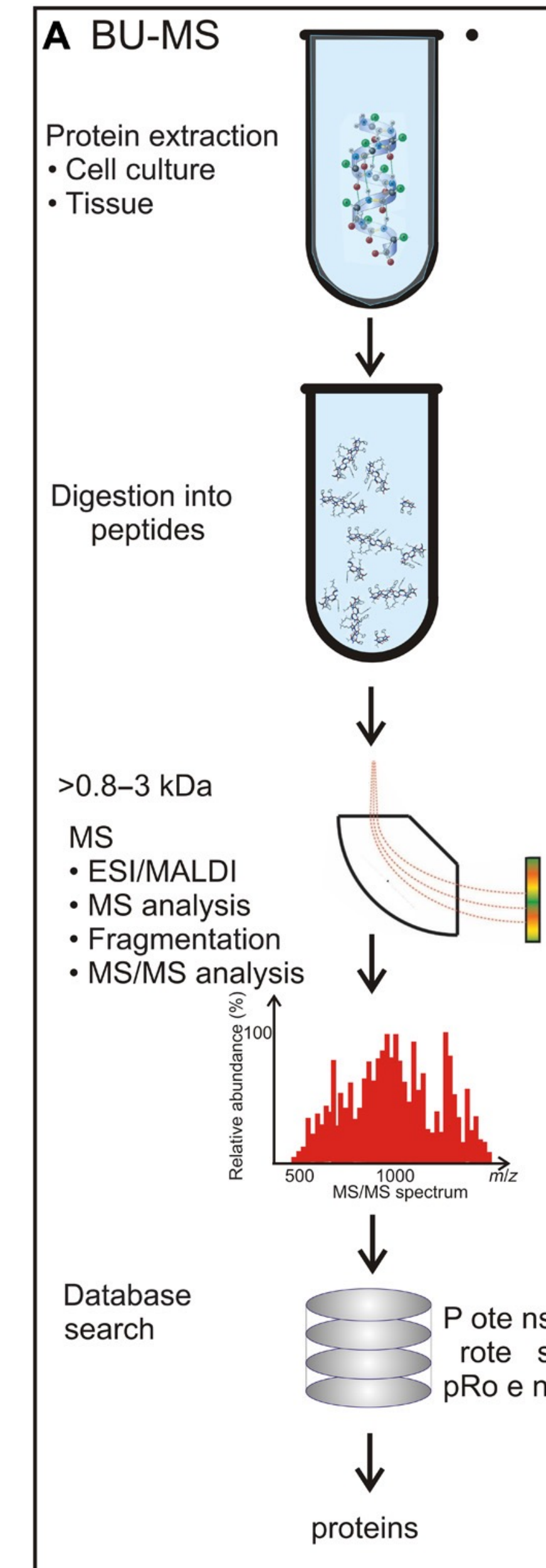
- PTMs increase the **functional diversity** of the proteome
- Include the covalent addition of **functional groups**, **proteolytic cleavage** of regulatory subunits, or **degradation** of entire proteins.
- PTMs can regulate **activity, localization, and interaction** with other molecules such as other proteins, nucleic acids, sugars, lipids and other cofactors
- 5% of the proteome are enzymes that perform **> 200 types of PTMs** to other proteins



Protein Identification

Mass Spectrometry (MS)

- **BU-MS** (Bottom UP MS)
- **Enzymatic digestion** of the proteins (Trypsin) into short peptides (5 -20 Aas)
- Ionized peptides are then analysed by **electrospray ionization** or matrix-assisted laser desorption/ionization (**MALDI**)
- Analyse their masses, **compare to database**
- Does capture **PTMs**
- **Short peptide sequences** can be similar complicating identification, low sensitivity, short reads means 'novel' peptides or PTMs can be hard to identify.

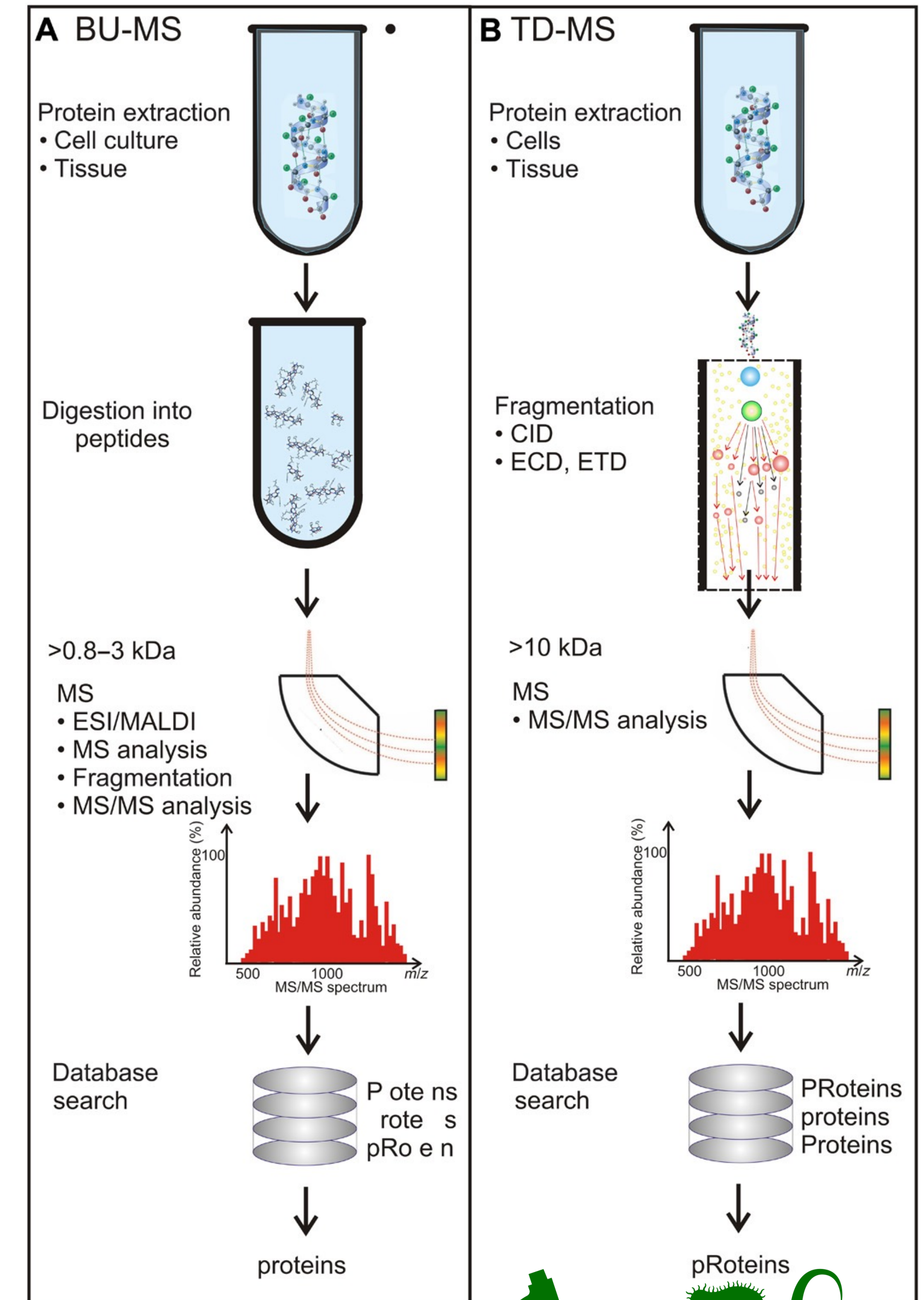


<https://doi.org/10.1126/sciadv.aax8978>

Protein Identification

Mass Spectrometry (MS)

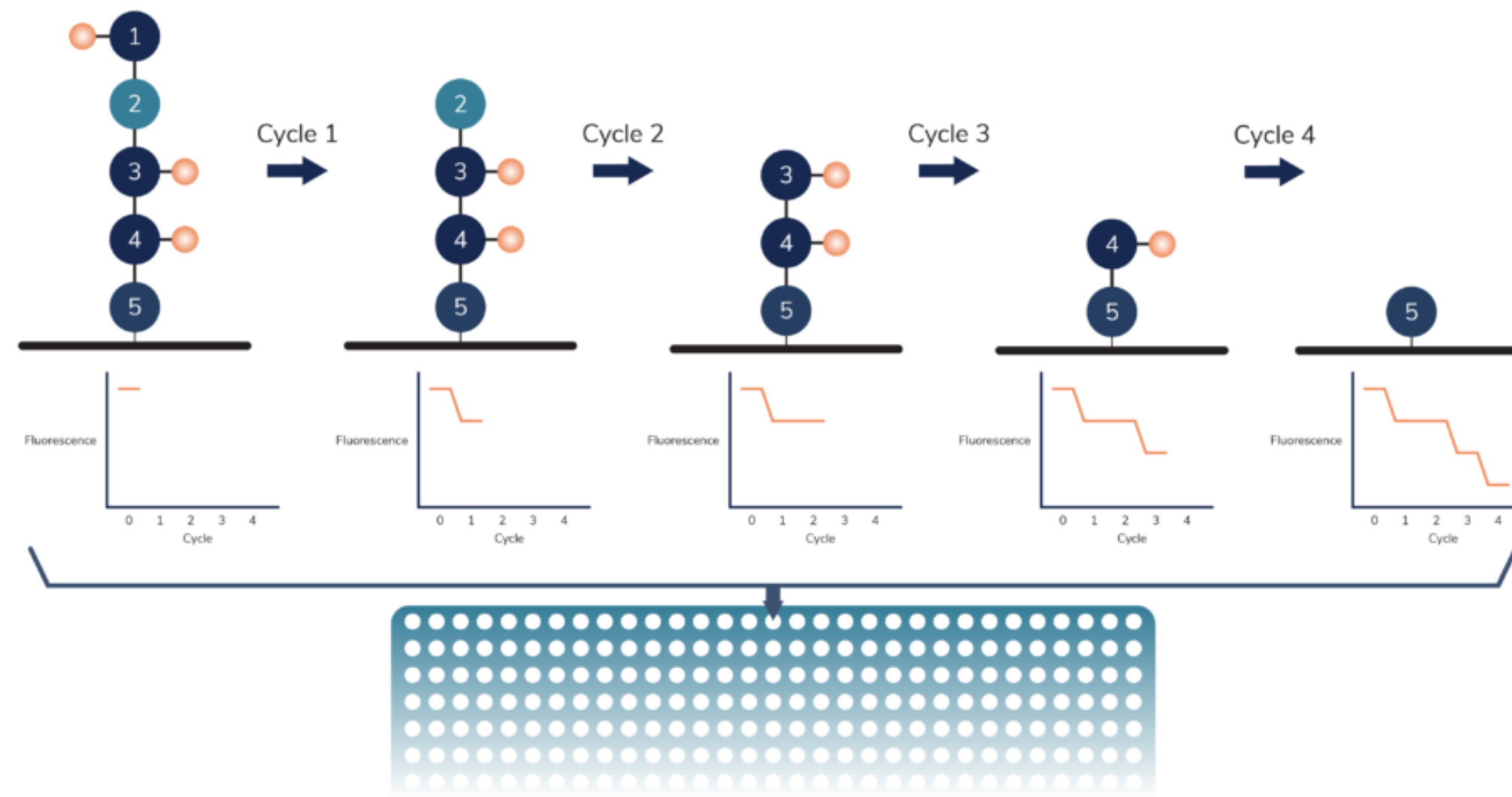
- **TD-MS** (Top Down UP MS)
- Introduce **intact proteins** into the gas phase by electrospray ionization that are **subsequently fragmented**
- Can reveal **primary structure** of the protein with PTMs
- **Limited in size** to 70kDa
- **Sensitivity is poor** vs short peptides in BU-MS
- Much **less frequently** used



<https://doi.org/10.1126/sciadv.aax8978>

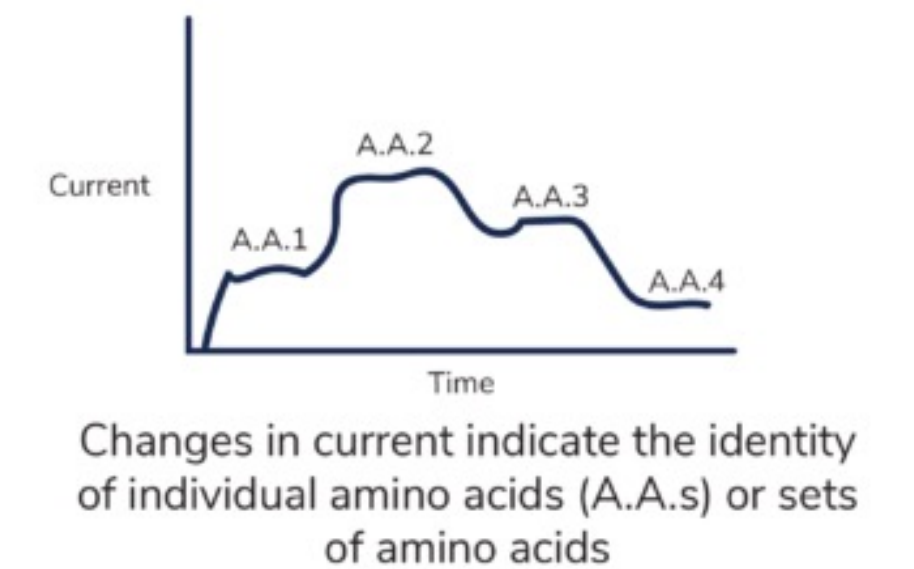
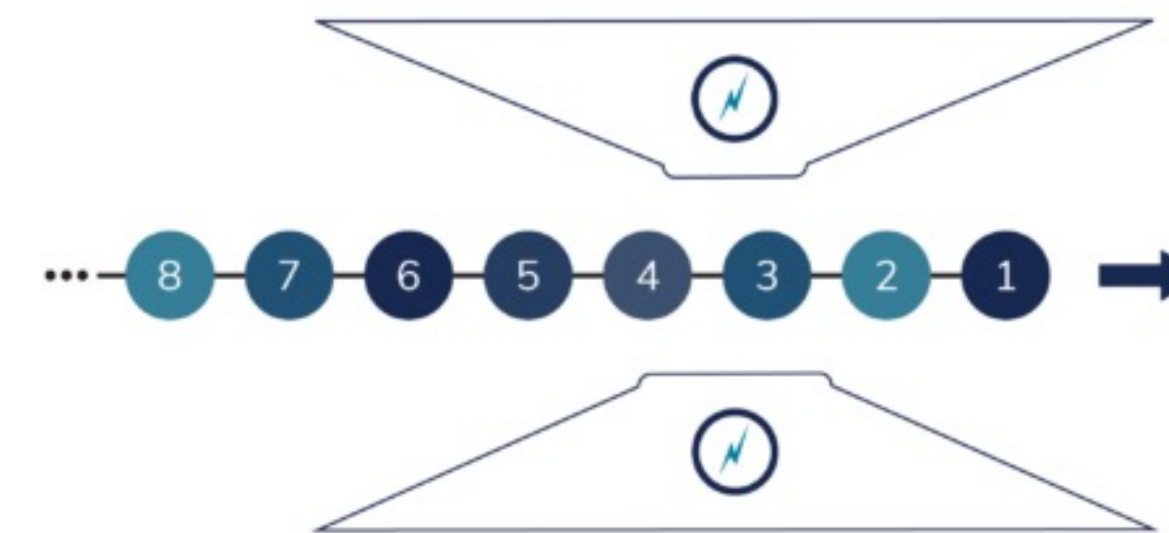
Next Generation Peptide sequencing

Edman Degradation based



Pore based

Denatured proteins

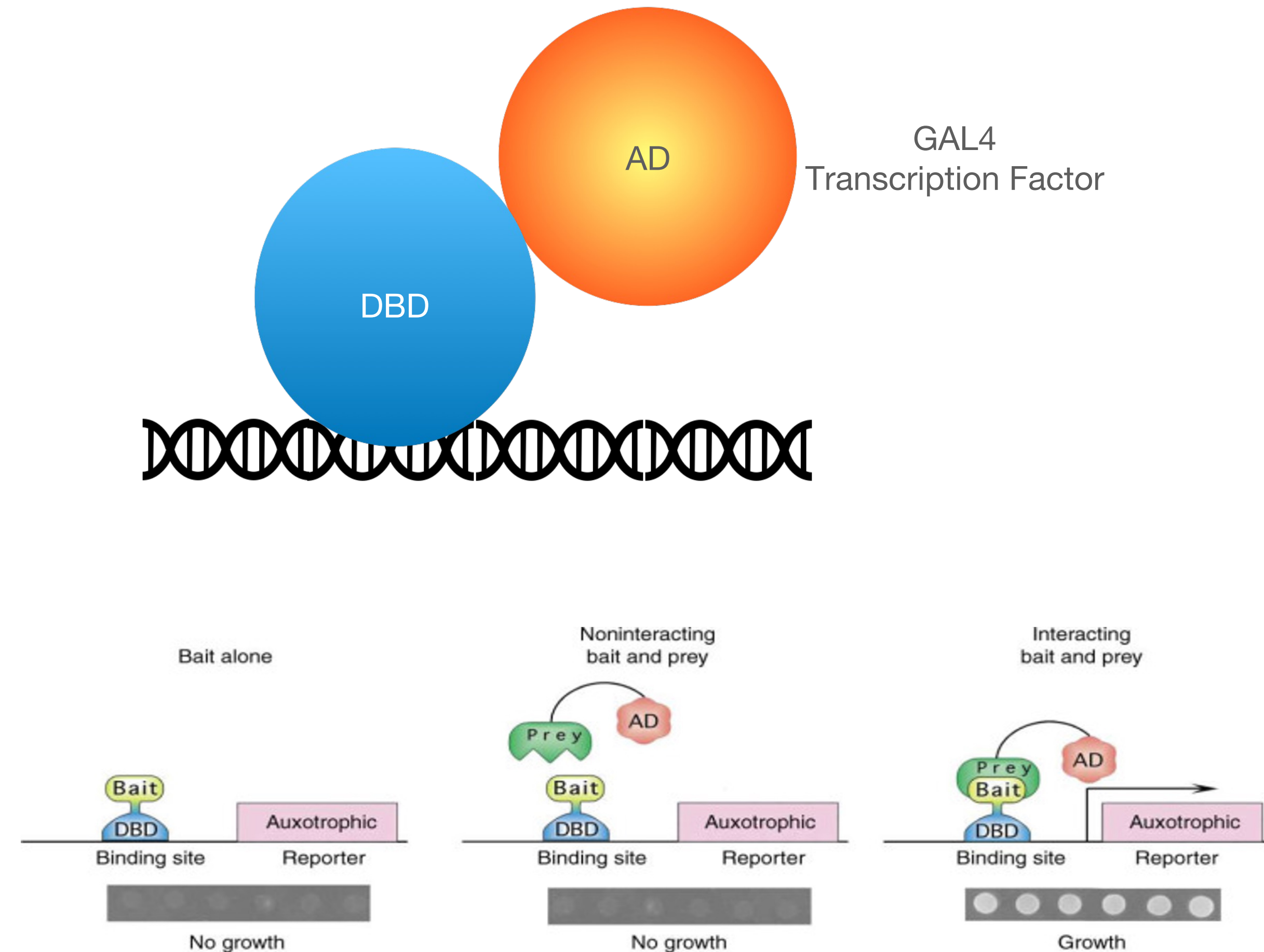


- **Not database** dependent
- **Single molecule** protein measurements possible
- **Slower**, must be **parallelised**
- **more expensive**, but technology is evolving rapidly

Finding Protein interactions

Yeast 2-hybrid system

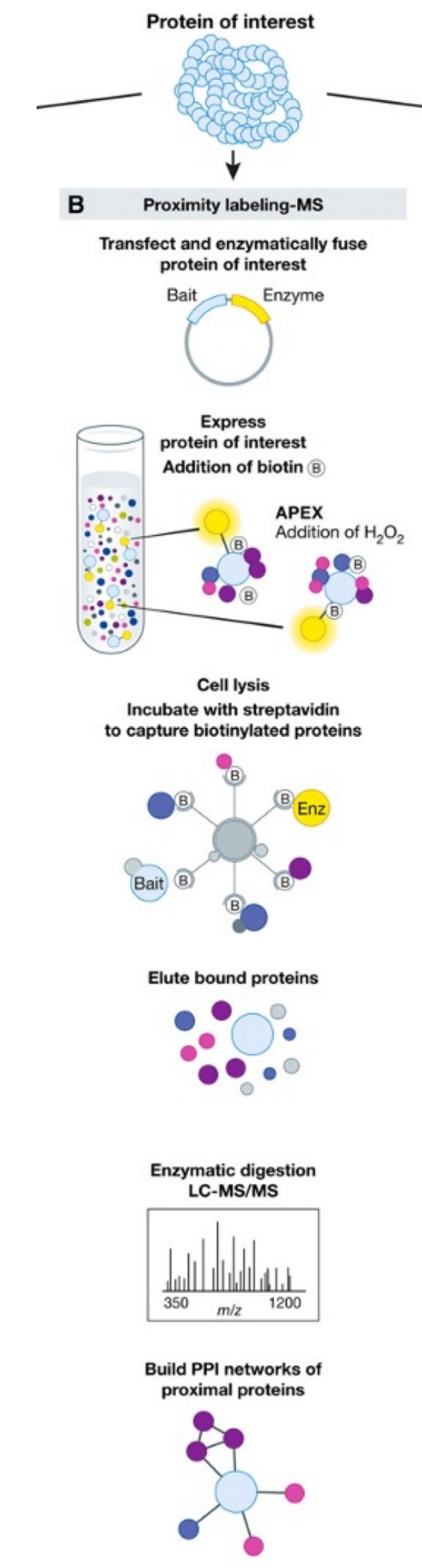
- Key concept creation of **hybrid fusion proteins**
- Transcription factors (e.g. GAL4) have **DNA binding domain** (DBD) and **Activation domain** (AD) protein regions (AD) that can be split apart
- DBD region can be **fused to your protein of interest** [POI] (bait)
- This can be screened against a **library** of other proteins (prey) that are fused to the AD of the transcription factor
- When the POI interacts with a prey protein, this brings the DBD and AD together **enabling transcription**.
- This is used to express a **reporter gene**



Finding Protein interactions

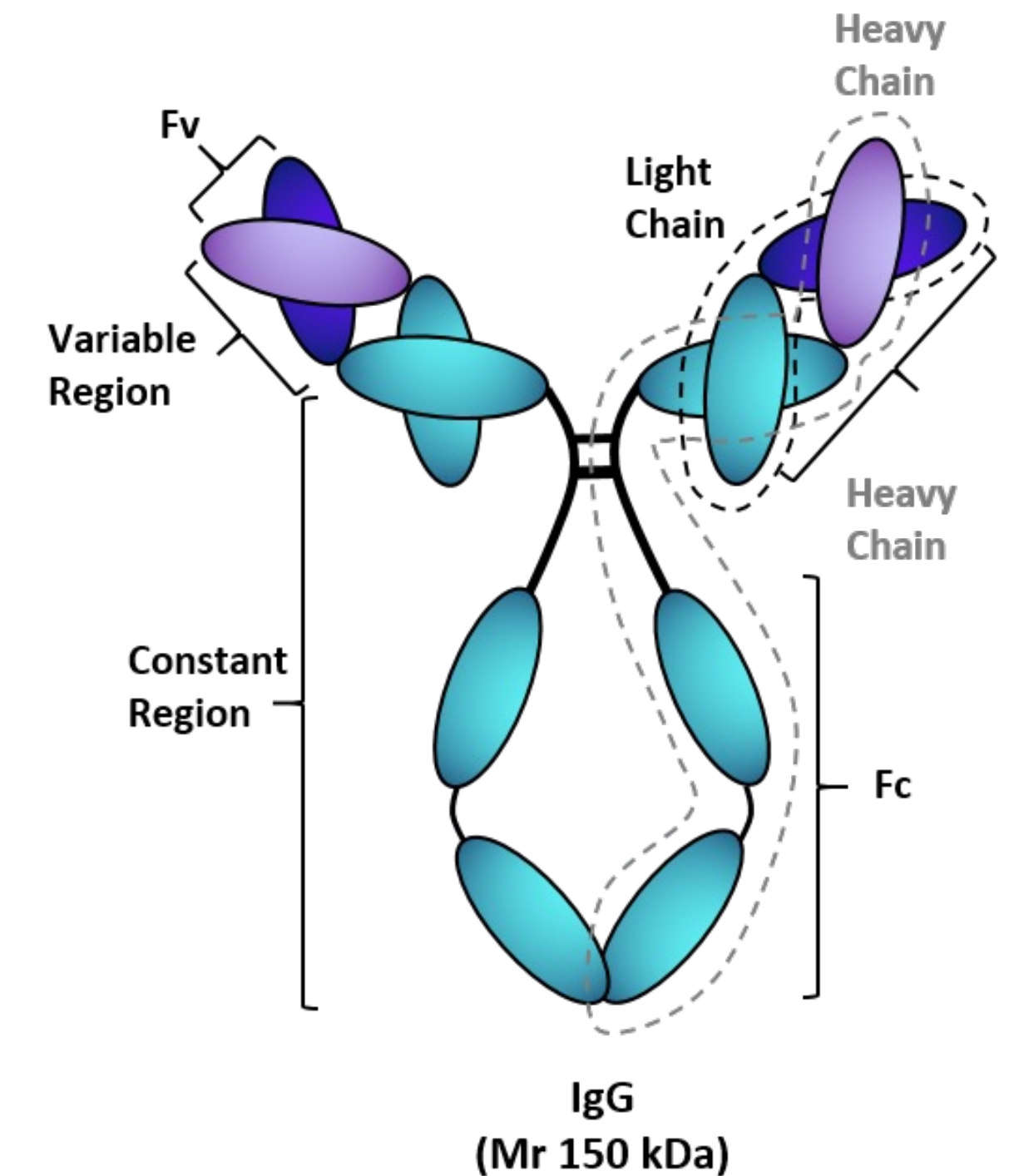
Proximity labelling methods

- **Protein of interest [POI]** is fused to an enzyme (Apex) that biotinylates other proteins
- This hybrid protein is expressed in cells and **biotin** is provided
- Proteins that interact with the POI are **biotinylated** by Apex activity
- These can be purified from all other proteins using **streptavidin** which binds to biotin with high affinity
- Interacting proteins are then **identified** e.g. by mass spectrometry

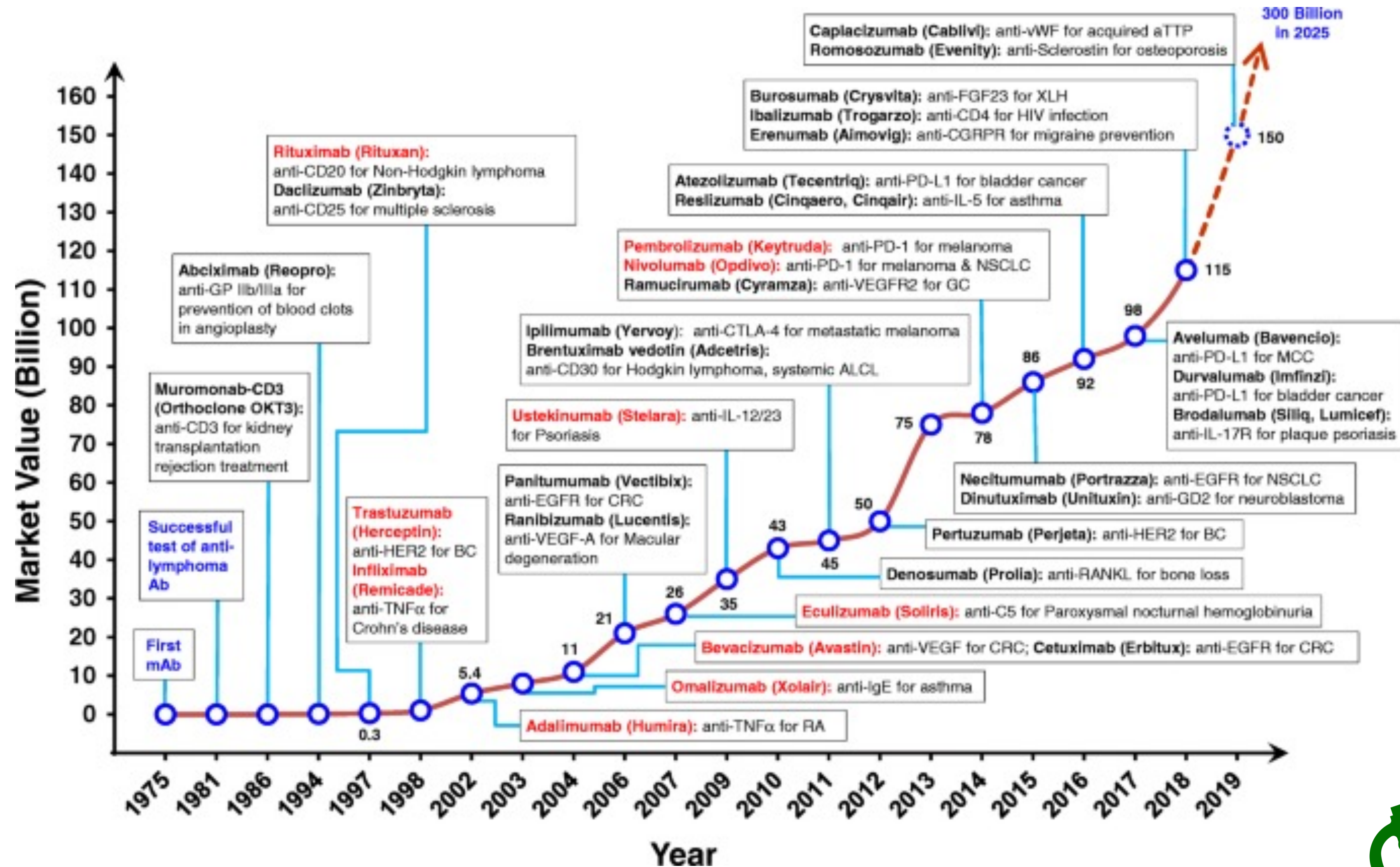


Changing Protein Activity

- **Drugs** small molecules
- **Antibodies** focus on monoclonal antibodies (mAbs) and nanobodies.

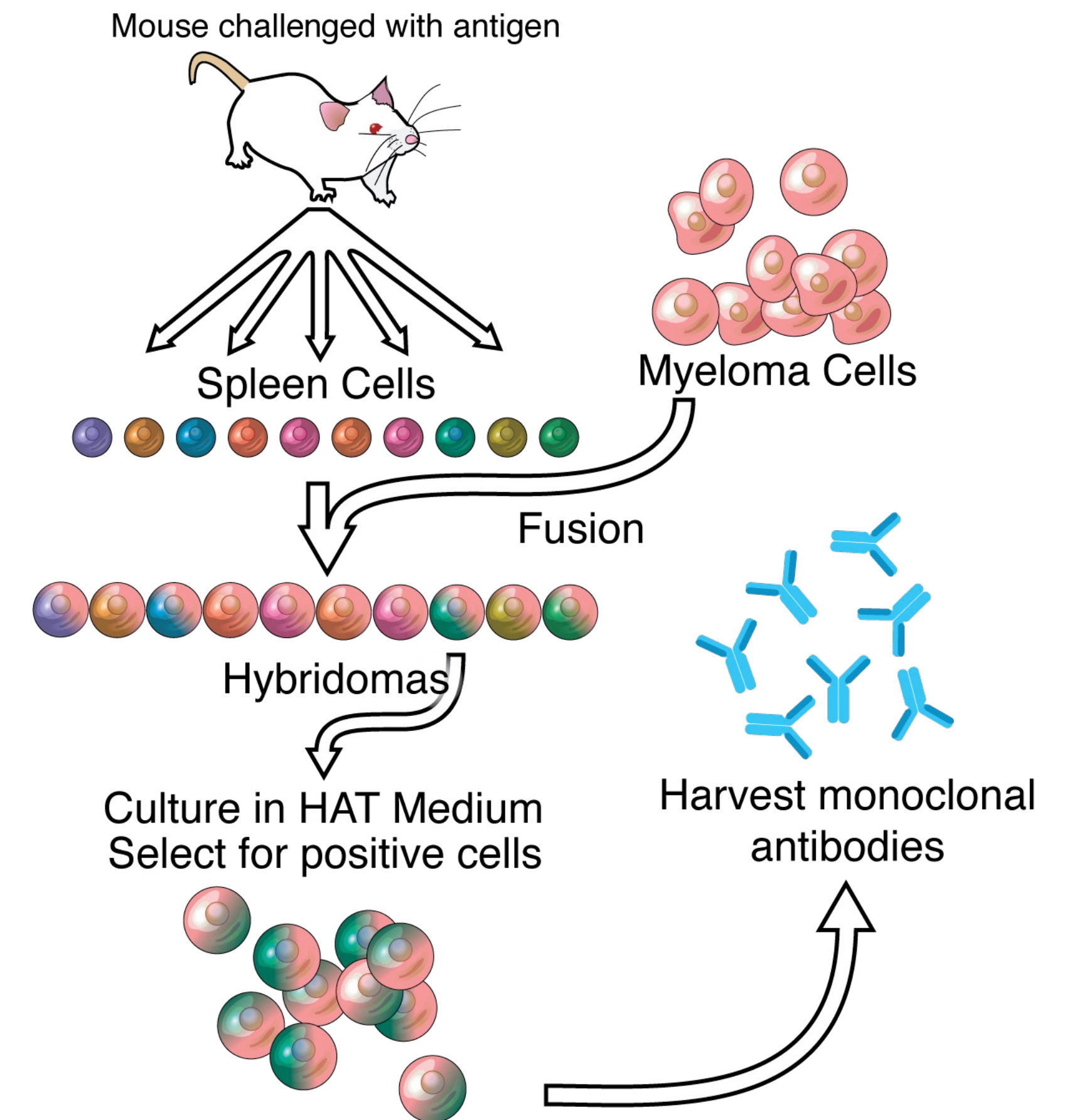


mAbs in medicine



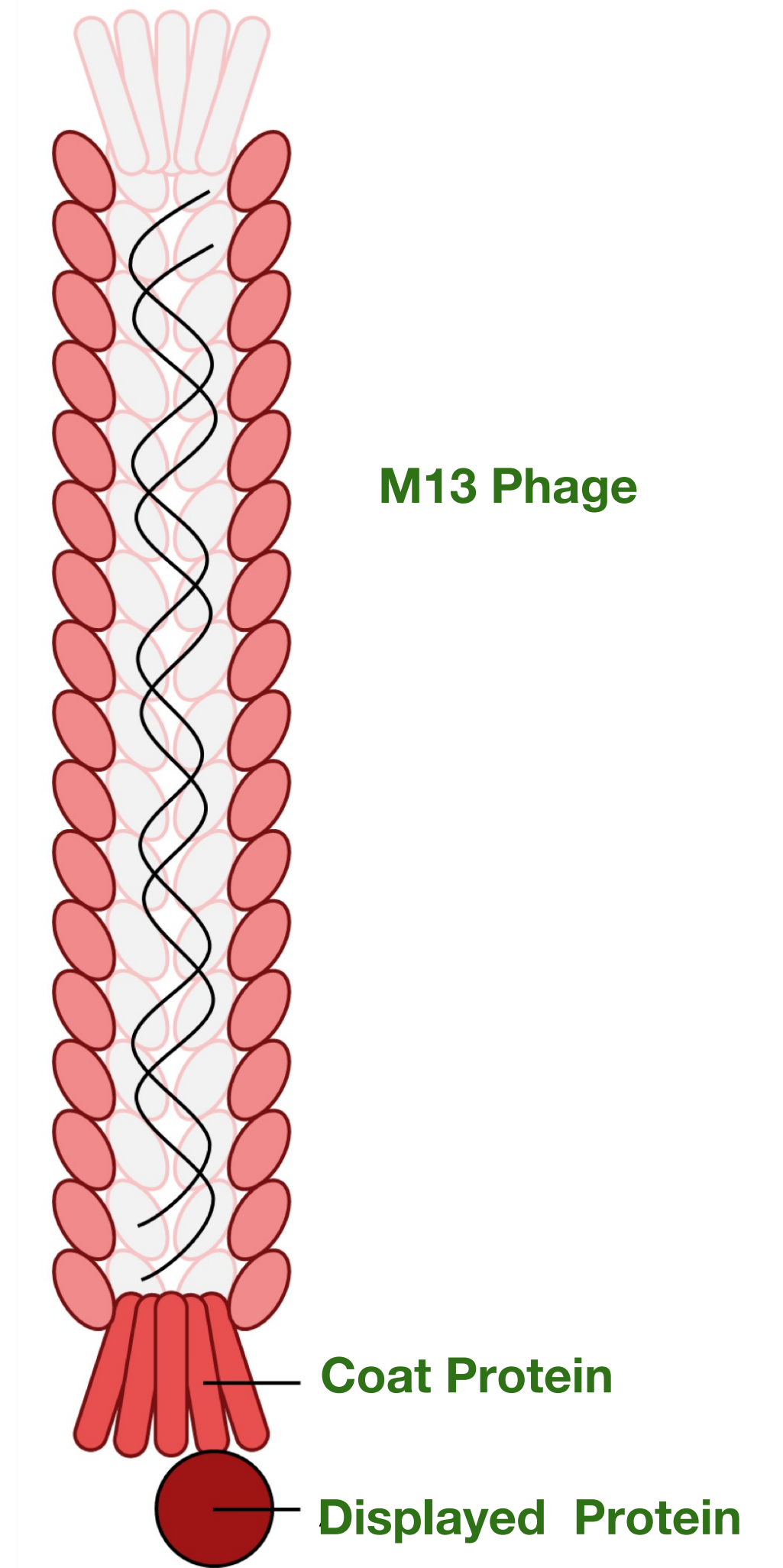
Monoclonal antibodies

- **Monoclonal antibodies (mAbs)** are produced from a hybridoma cell lines made from spleen derived B cells
- mAbs usually have **monovalent affinity** i.e. bind only to one epitope (different to polyclonal antibodies)
- Once a cell line is established Mabs can be produced in **unlimited quantities**
- Can be used to **detect and visualise proteins** in tissues or for biochemistry assays
- For human therapeutics, mAbs are often '**humanized**' to avoid immunogenicity against rodent sequences.



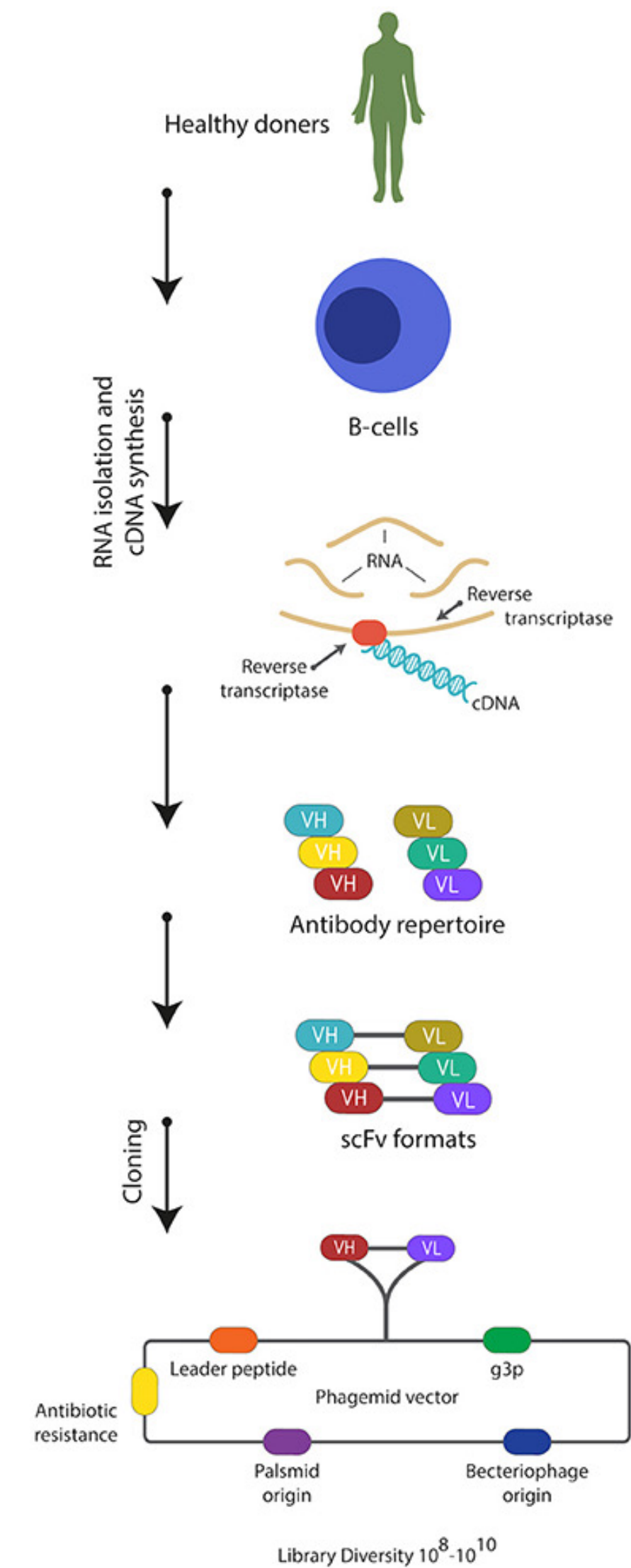
Phage Display

- **Genes** encoding protein sequences are inserted into a bacteria phage such as M13
- The inserted protein is then **displayed on the outside** of the phage.

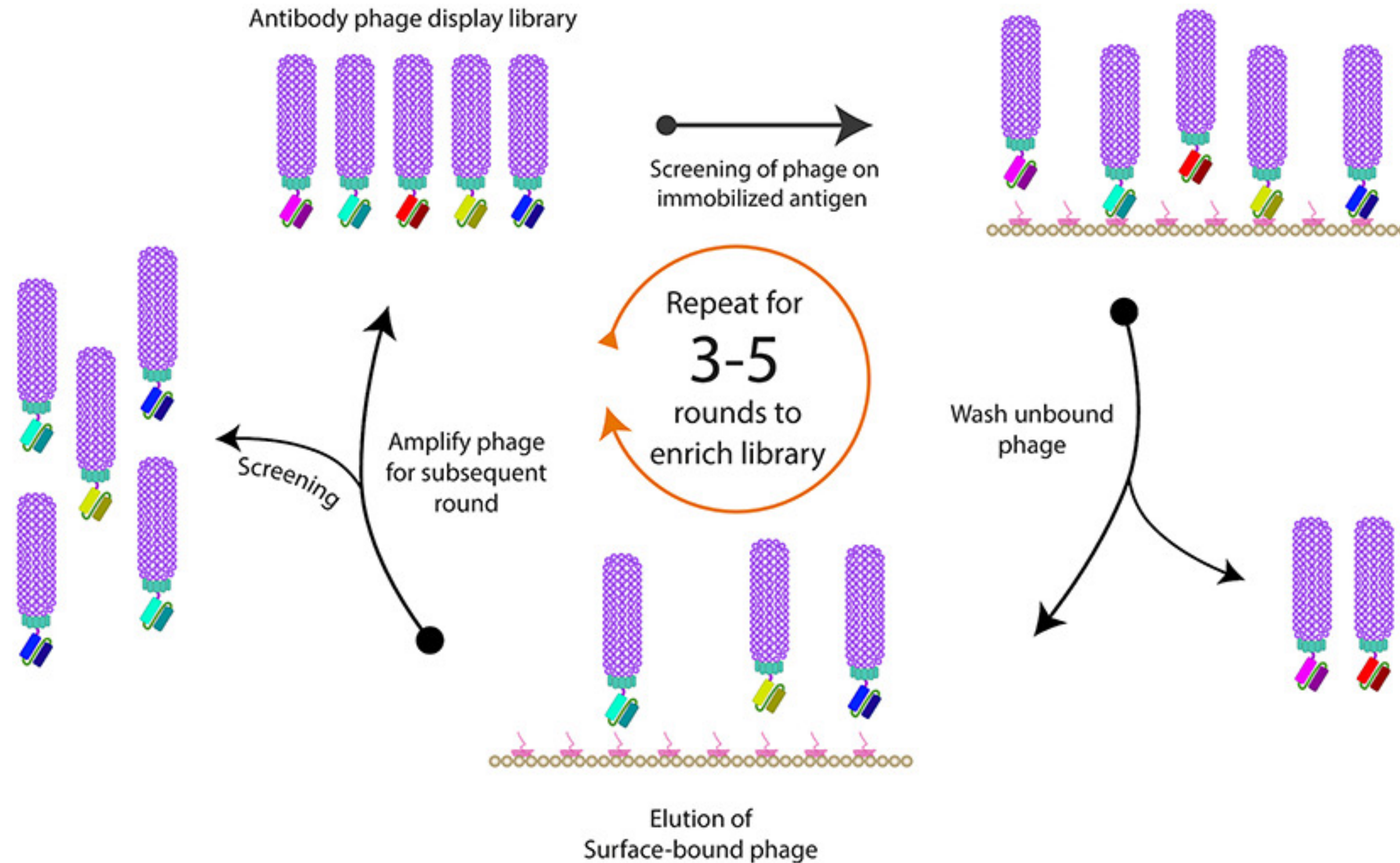


Phage Display

- **Genes** encoding protein sequences are inserted into a bacteria phage such as M13
- The inserted protein is then **displayed on the outside** of the phage.
- A **library of phages** can be made that express antigen binding sequences on their surface
- These can then be screened in vitro to **select antigen binding** sequences with high affinity
- Once identified they can be engineered to **make therapeutic antibodies**

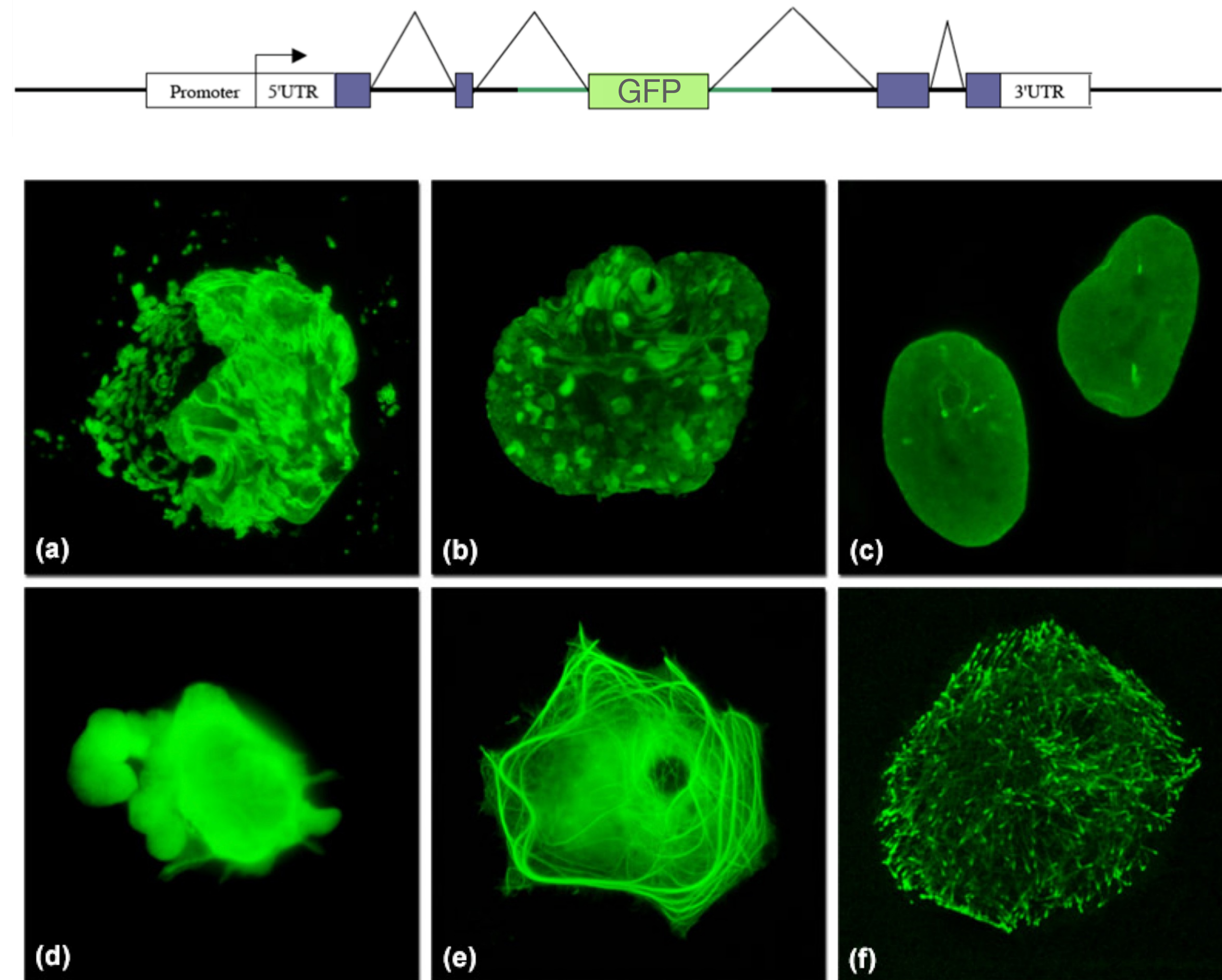


Phage Display



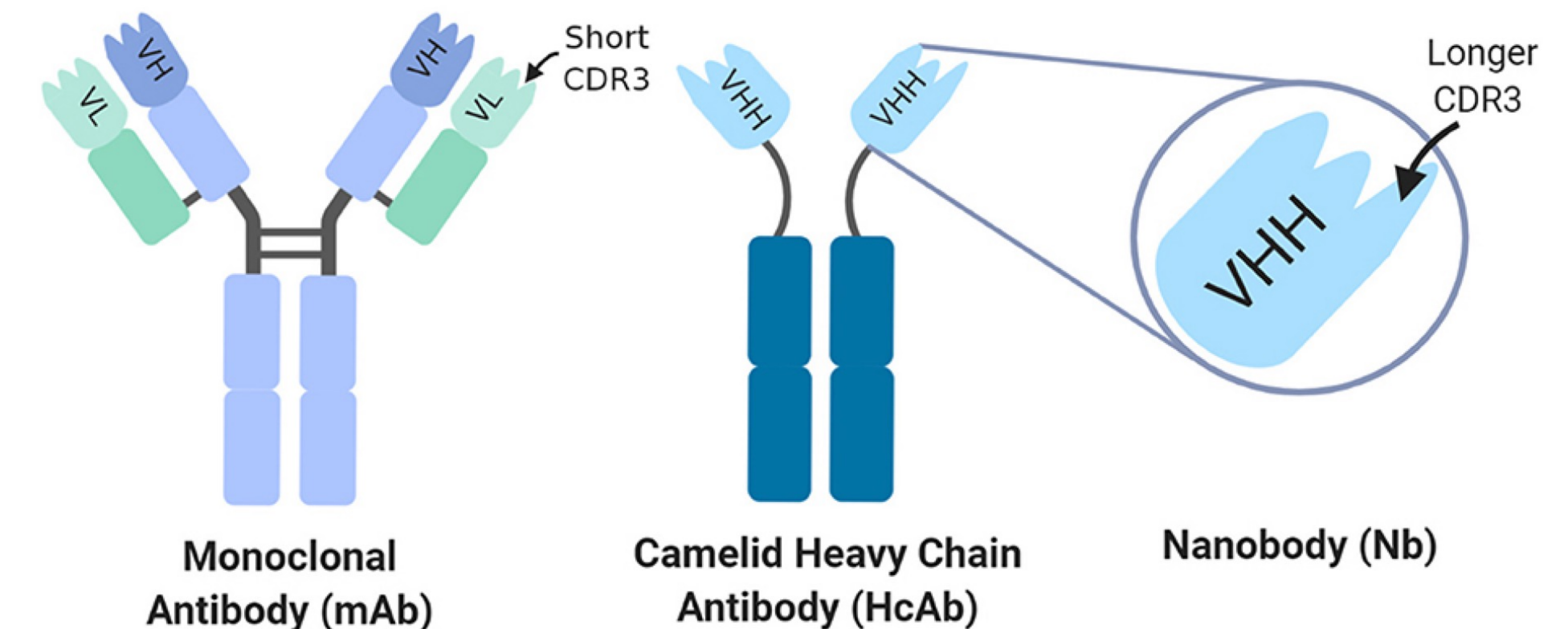
Tagging proteins

- **Protein tags** are peptide sequences genetically introduced into proteins
- Can be added to either **transgenic or endogenous** proteins
- Can be used to visualise or isolate proteins e.g. **epitope tags** (HA, V5 etc.) which are recognised by mAbs
- **Fluorescent protein tags** (e.g. GFP) allow protein dynamics to be observed in vivo

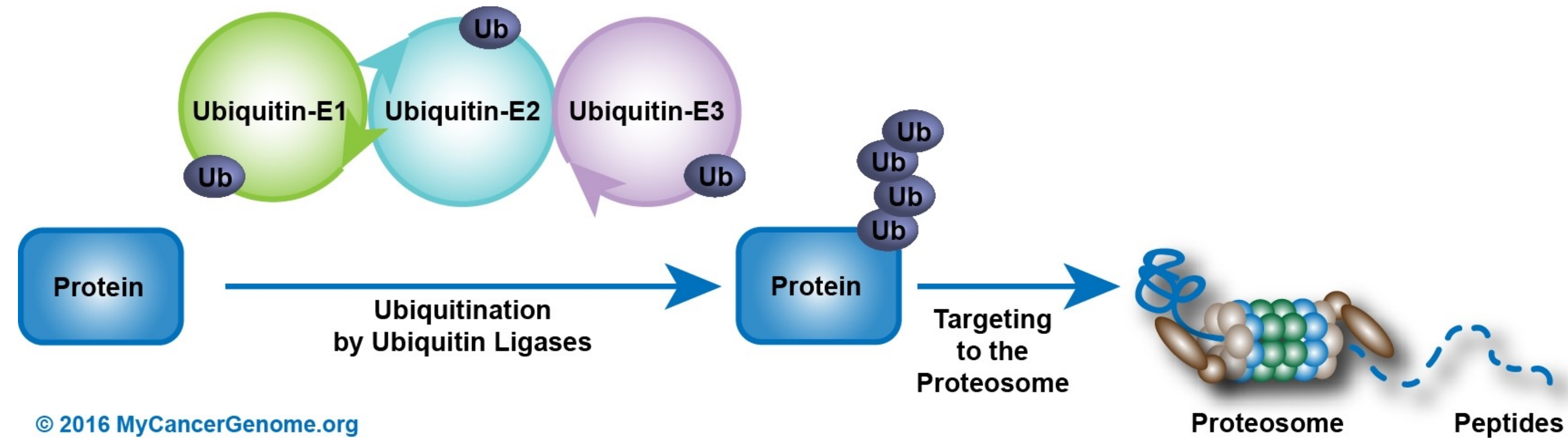


Nanobodies

- **Nanobodies** are the recombinant variable domains of heavy-chain-only antibodies
- Heavy-chain-only antibodies were first identified in **camels** but also found in cartilaginous fish.
- Nanobodies have a **small size**, excellent solubility, superior stability
- Can be cloned, engineered and **produced in bacteria**.

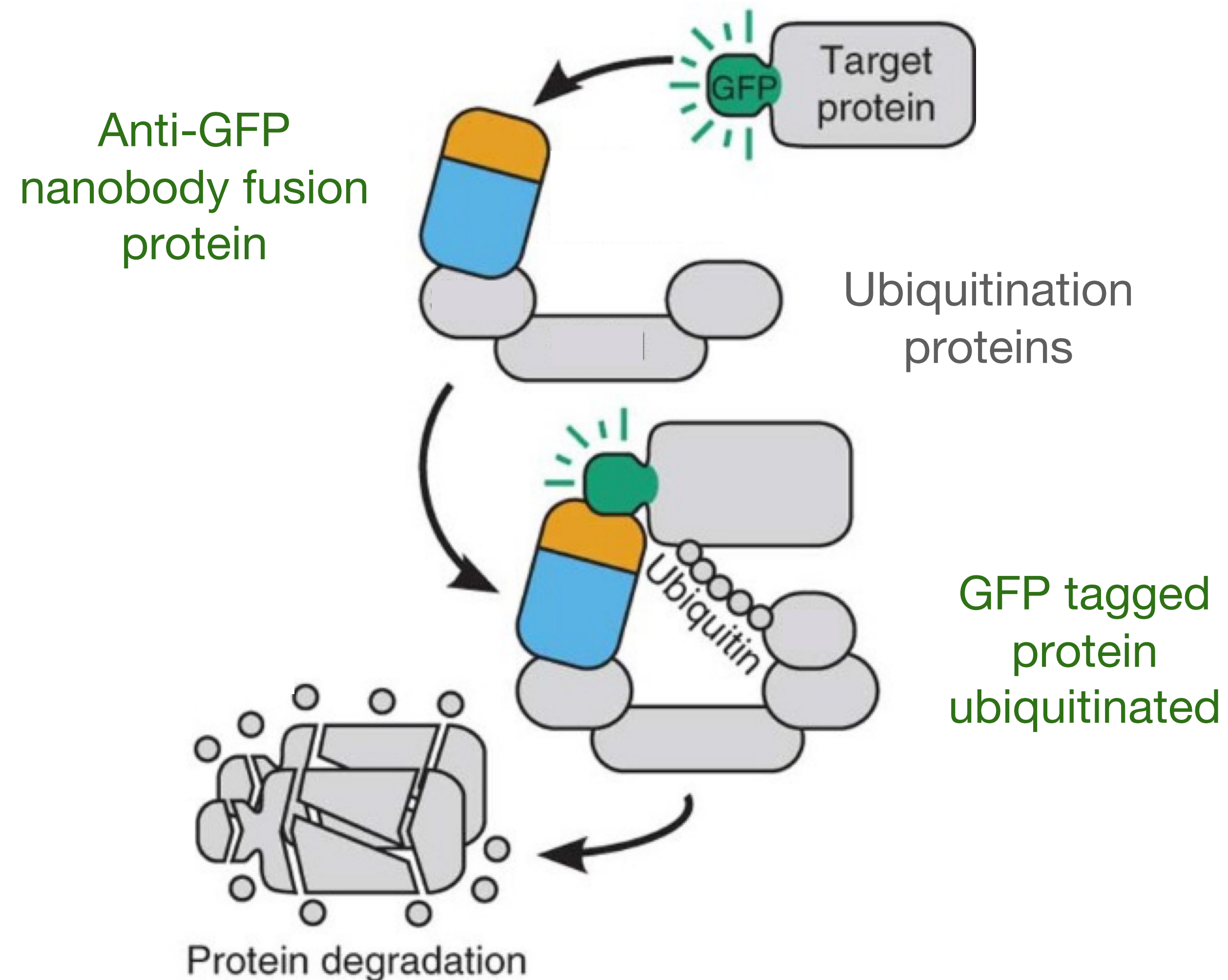


Ubiquitination degrades proteins



- **Ubiquitination** is the process in which proteins are labelled with ubiquitin in a post-translational modification.
- **Ubiquitin** is small protein which is attached in chains to Lysine residues in the ubiquitinated protein via the action of ubiquitin ligases
- Ubiquitination most often targets a protein for **degradation** by the proteasome.

Nanobody promoted Ubiquitination





How **Brian**

Why **Pierre**

What **Felix**