The background of the slide features a textured, light brown or beige surface. Overlaid on this are various abstract, organic shapes in black, red, blue, and yellow. These include stylized cells with internal organelles, some with cilia-like appendages, and larger, more complex forms resembling microorganisms or small plants. Small, five-pointed starburst shapes are scattered throughout the space.

a
beginner's
guide

Biology.

First things first

This is **NOT** a comprehensive guide.

This is an **overview** of concepts and buzzwords that come up frequently in talks and courses, assembled arbitrarily, based on my experience alone.

If you find yourself interested in the details, **Wikipedia**.

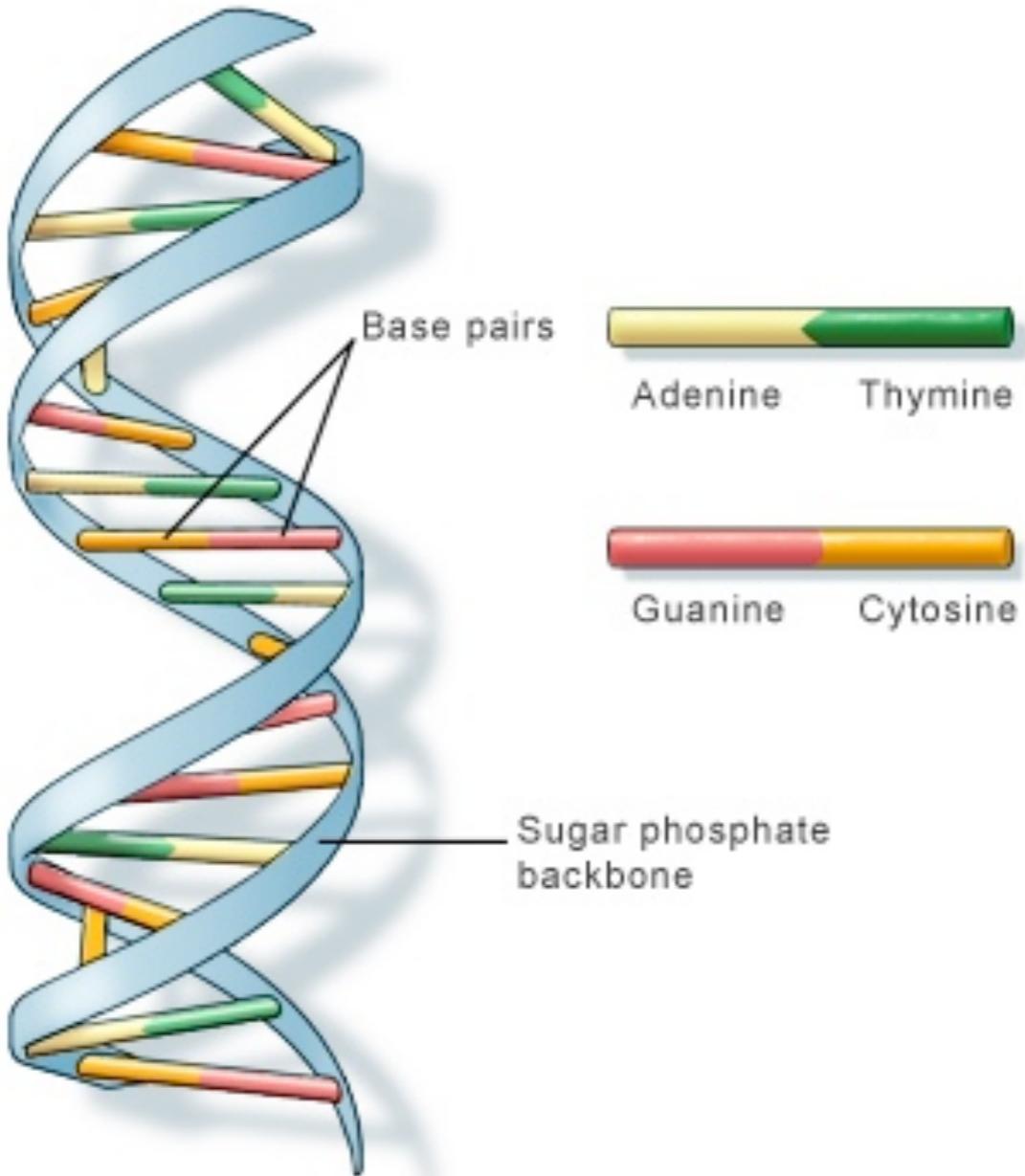
If you find yourself lost and confused, consider taking a **basic biology course**:

BILD 1, BICD 100, BICD 110, BICD 130, BIMM 100;
BGGN 220, BGGN 222, BGGN 223

In the
beginning,



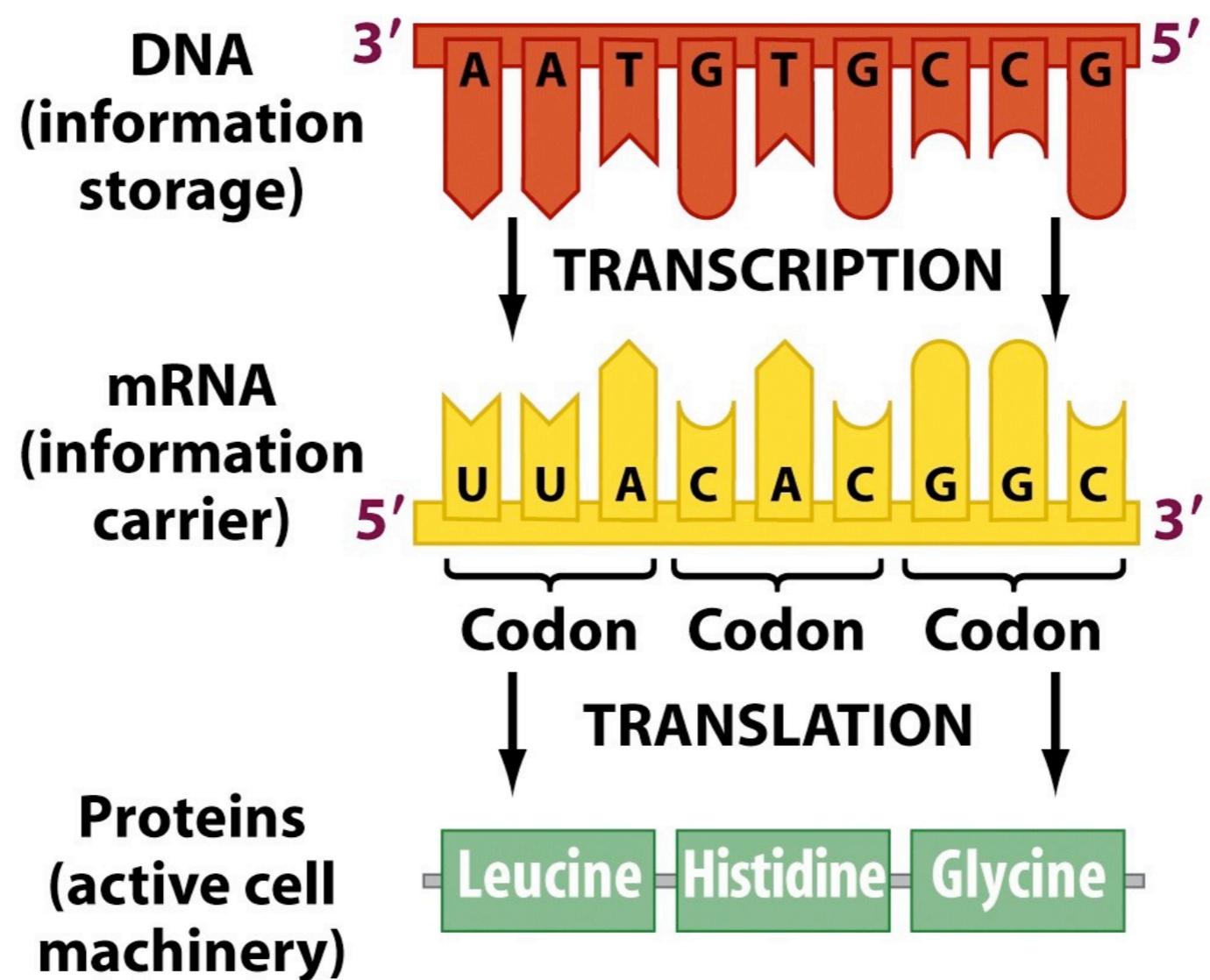
DNA



- Paired nucleotides on a sugar and phosphate backbone
- Base-4 biological bits: {A, C, G, T}
- Purines: A and G
- Pyrimidines: C and T
- Read from 3' → 5' during transcription
- 3 billion bases in the human genome, encoding 20K genes
- 150 billion bases in the *Paris japonica* genome, 10K in HIV
- Fun fact: George Church just stored his book (53K words) in DNA, in 96-bit data blocks each with a 19-bit locator indicating position in the fully assembled text.

Central Dogma

Information flows from DNA to RNA to proteins.

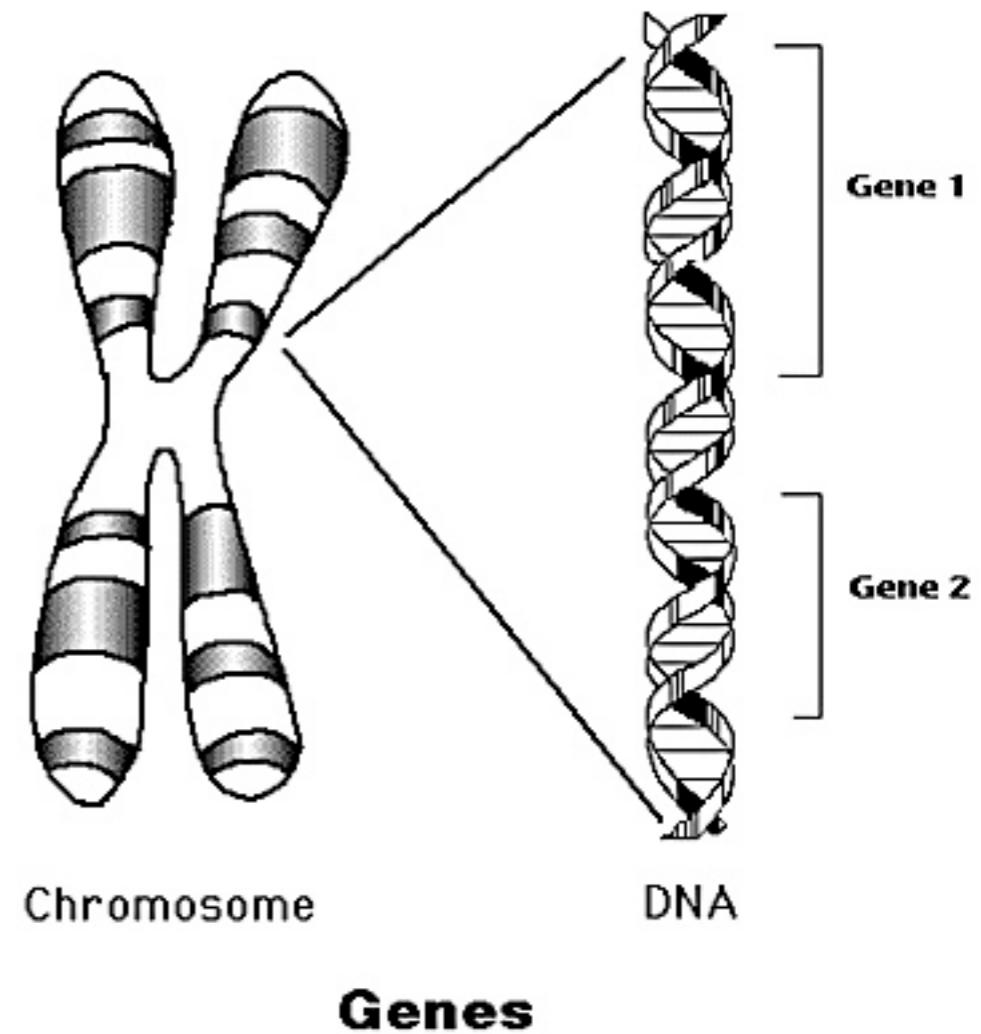


Central Dogma

DNA → RNA → Proteins.

What falls out from this?

- Genes are the crucial unit of the genome
- Proteins are the final output to measure, and RNA is useful as a proxy for proteins
- There is some loss of information from DNA to RNA to proteins



But then...



It's complicated.

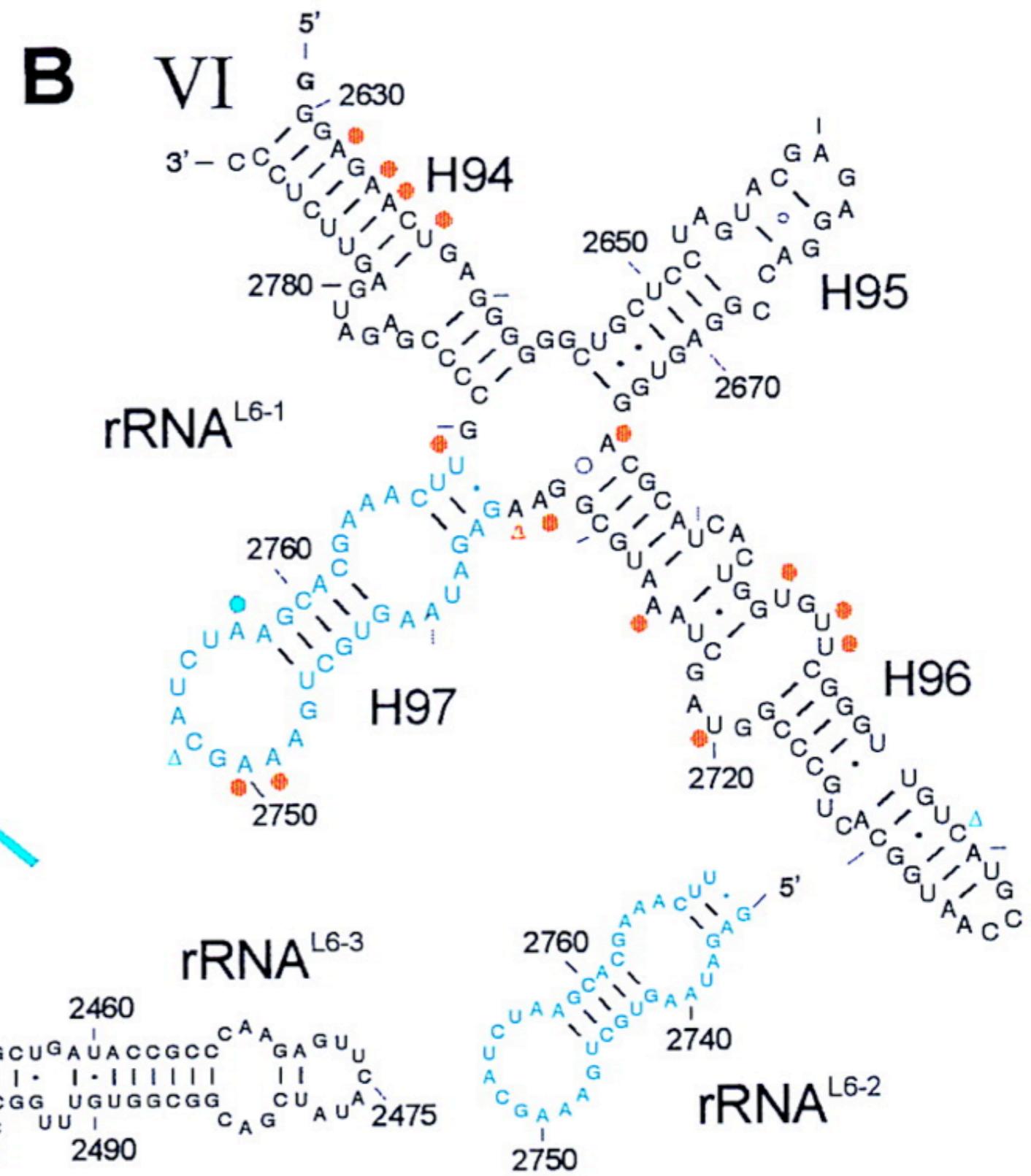
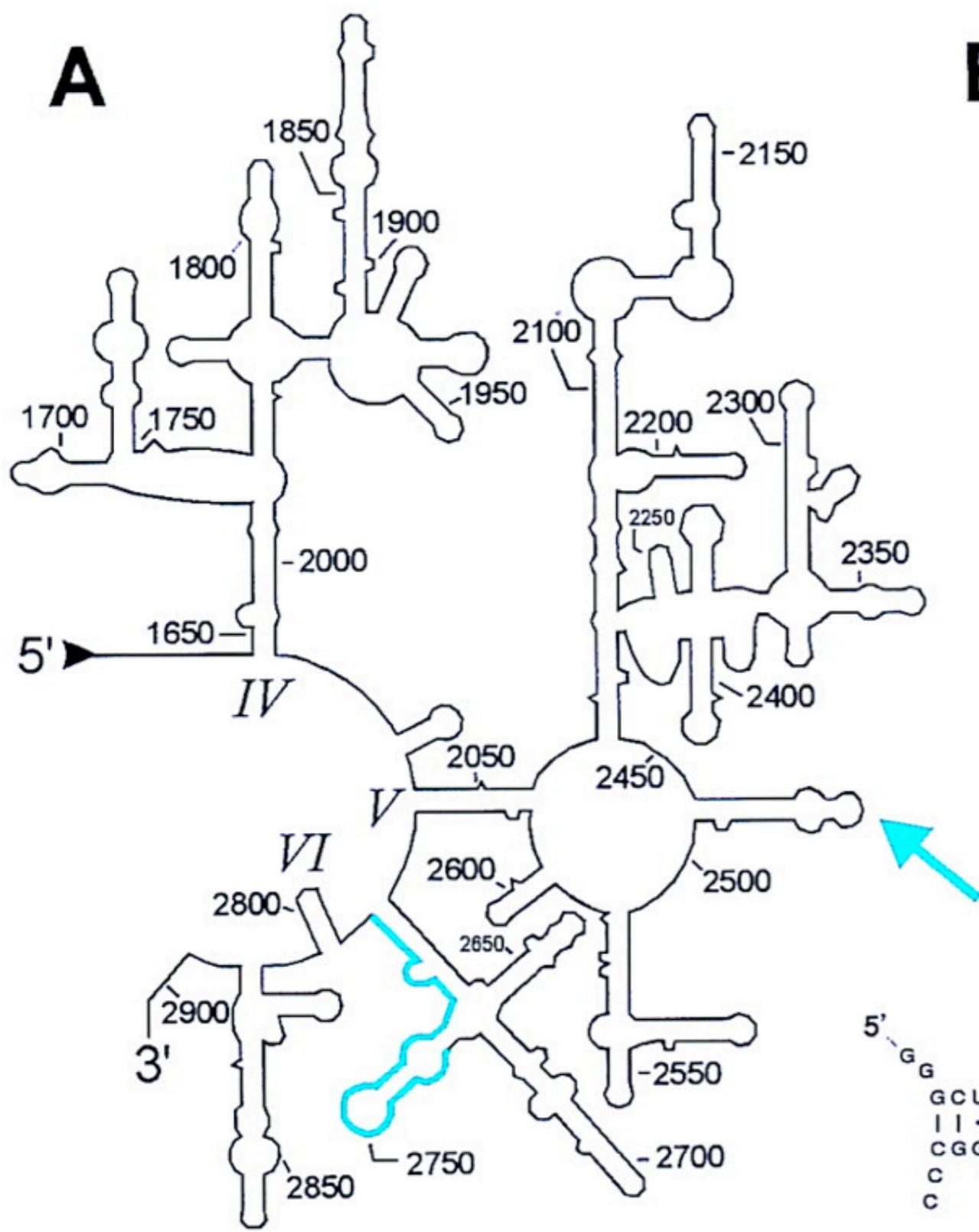
Recent ENCODE publications confirm several decades of findings:

“The vast majority (80.4%) of the human genome participates in at least one biochemical RNA- and/or chromatin-associated event in at least one cell type.”

“Despite intensive study, especially in identifying protein-coding genes, our understanding of the genome is far from complete, particularly with regard to non-coding RNAs, alternatively spliced transcripts and regulatory sequences. Systematic analyses of transcripts and regulatory information are essential for the identification of genes and regulatory regions, and are an important resource for the study of human biology and disease.”

The ENCODE Project Consortium. “An integrated encyclopedia of DNA elements in the human genome.” Nature 489, 57–74 (06 September 2012). doi:10.1038/nature11247

Non-coding RNA



Non-coding RNA

Structural

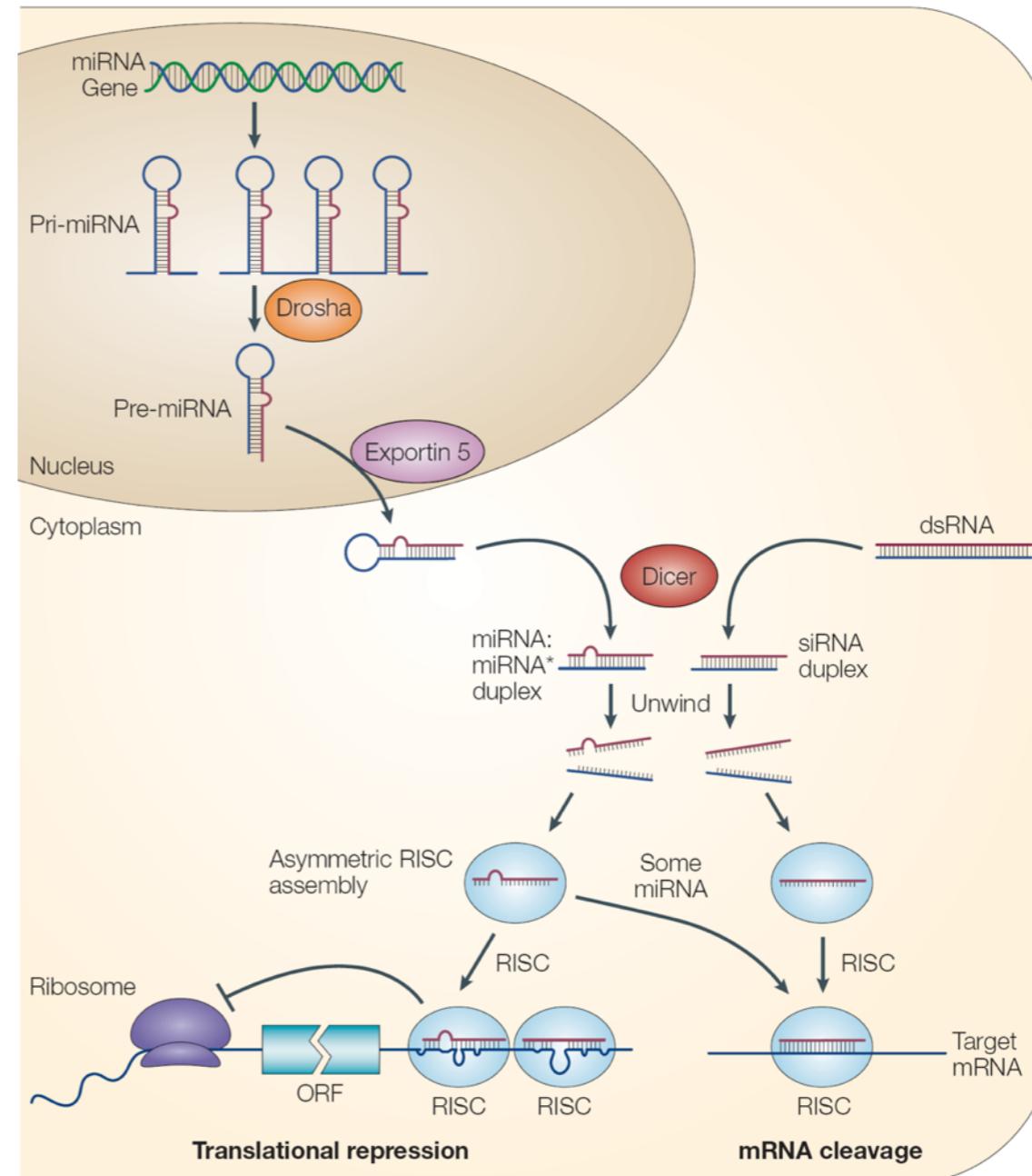
- rRNA
- tRNA
- snRNA, snoRNA

Regulatory

- miRNA, pre-miRNA
- lncRNA
- piRNA
- ceRNA

Manufactured

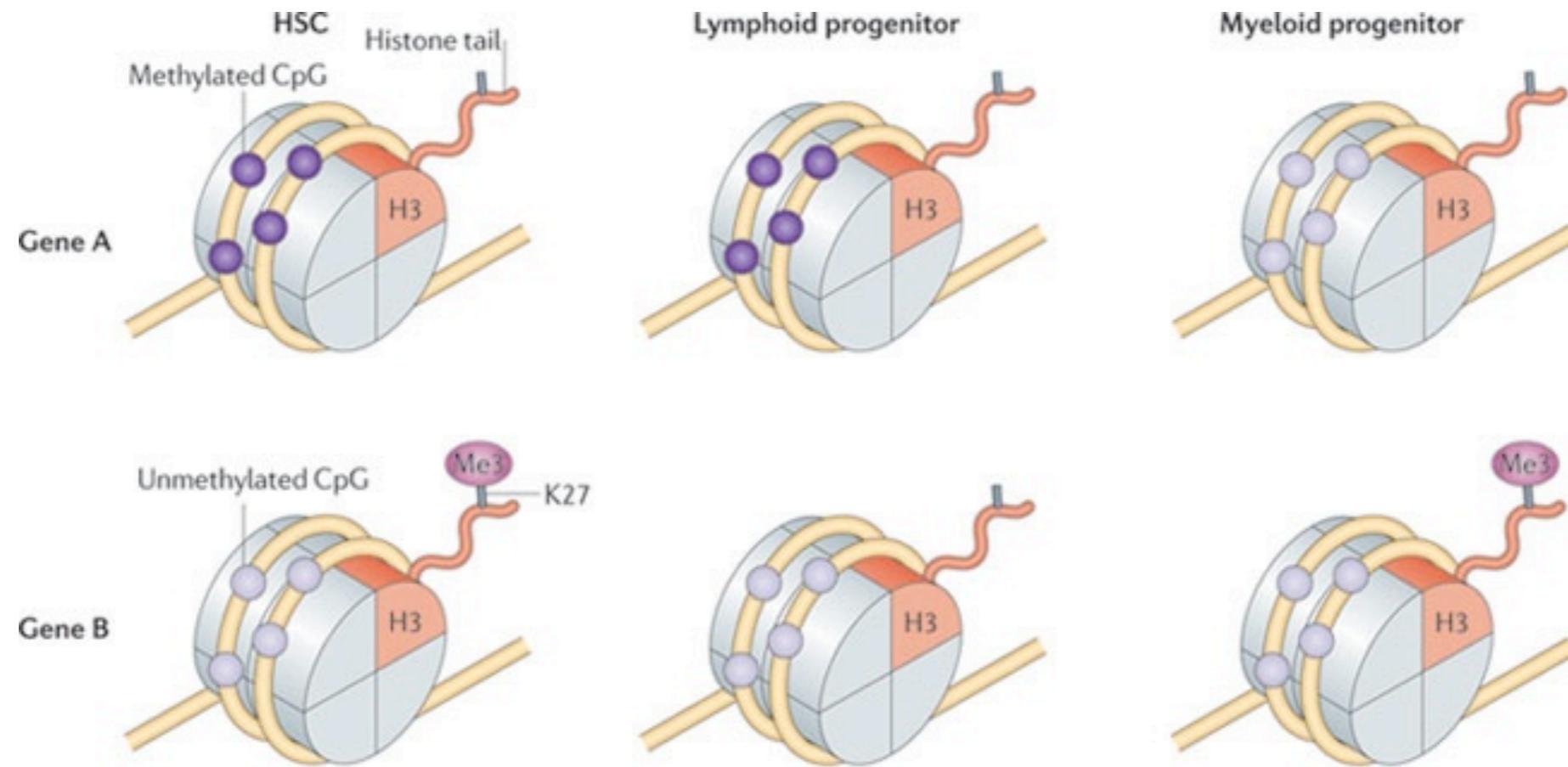
- siRNA
- shRNA



Anybody's guess

- eRNA
- promoter-associated RNA

Histones

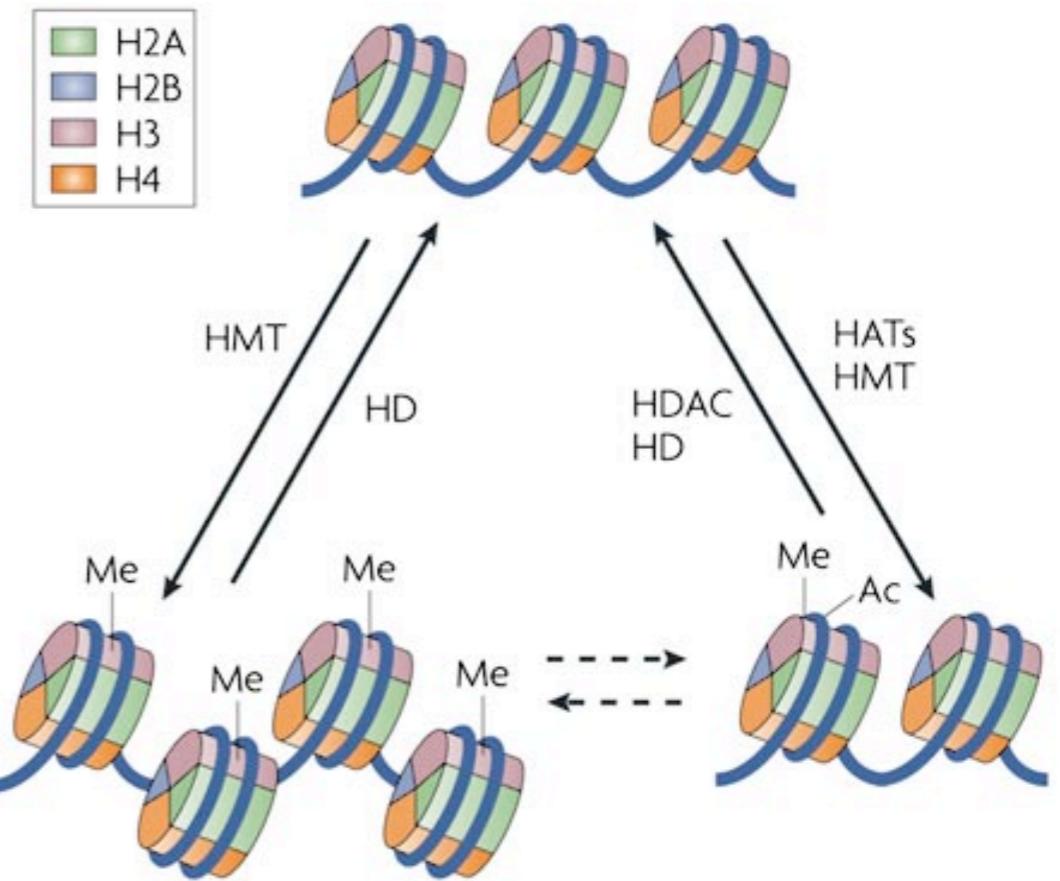


Nature Reviews | Immunology

The genes labelled A and B represent genes that are required for myeloid and lymphoid differentiation, respectively. In haematopoietic stem cells (HSCs), both are silenced, but by different mechanisms. The myeloid gene (A) is repressed by DNA methylation, whereas the lymphoid gene (B) is repressed by Polycomb-mediated trimethylation of histone H3 lysine 27 (H3K27me3). To enable differentiation to the lymphoid lineage, the Polycomb complex must be removed from gene B, whereas differentiation to the myeloid lineage requires demethylation of gene A. It should be noted that other repression mechanisms may also be involved in lineage determination.

Histones

- Spools around which DNA is wound
- Octameric proteins made up of pairs of H2A, H2B, H3, and H4
- About 147 bp of DNA wrap around to form a nucleosome
- Shift and unwind to allow transcription factors to bind and initiate transcription along unwound DNA
- Post-translational modifications called “histone modifications” change the way the surrounding DNA and DNA-binding proteins can interact
- H3K4me1 (“mono-methyl”) = promoter marker; activation
- H3K4me2/3 (“di/tri-methyl”) = enhancer marker
- H3K27me3 = repression



Examples of histone modifications

Heterochromatin

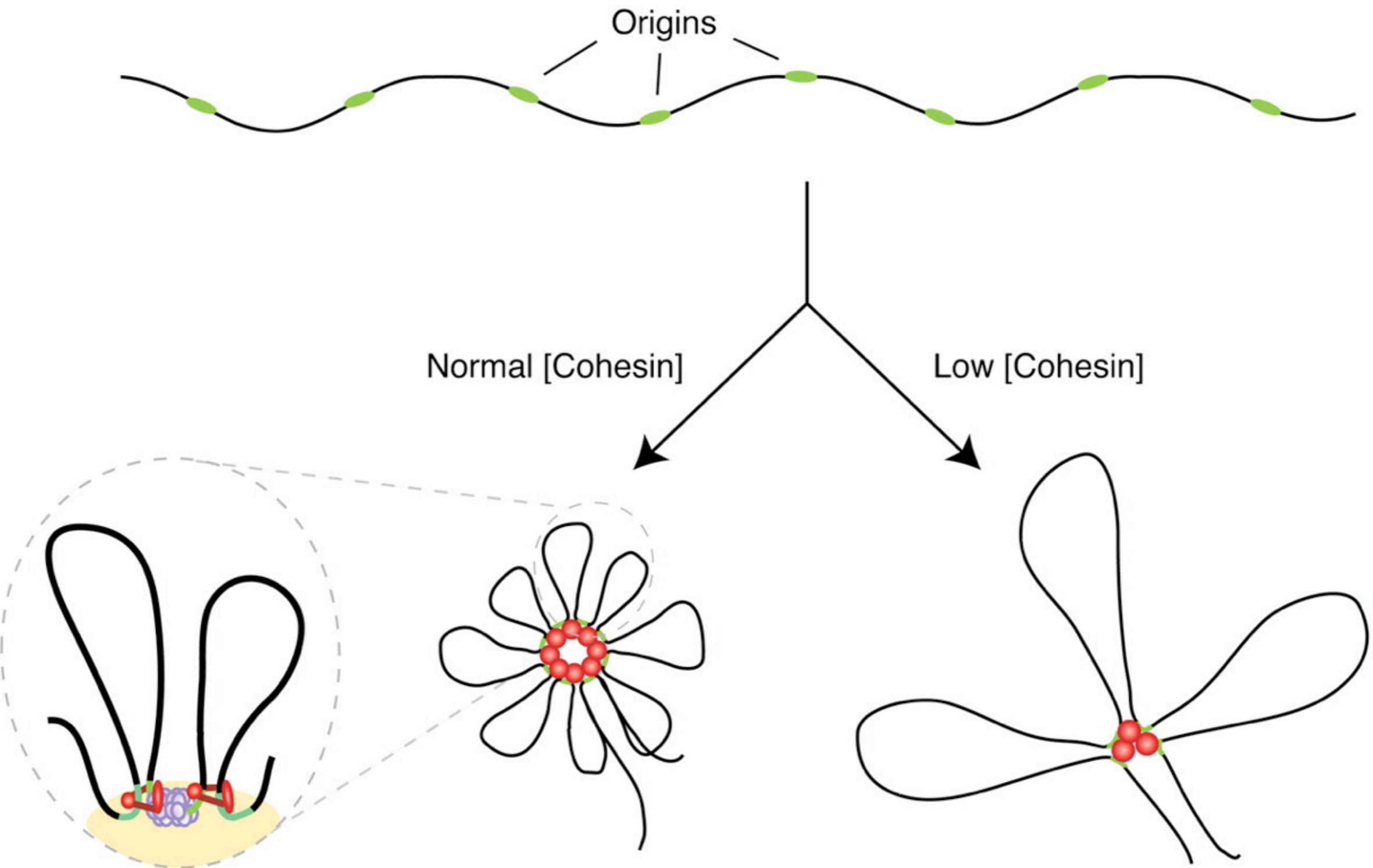
- H3K9me2,3
- H3K27me3
- H4K20me3

Euchromatin

- H3K4me2,3
- H4K5ac, H4K8ac
- H2AK5ac, H2BK12ac
- H2BK15ac

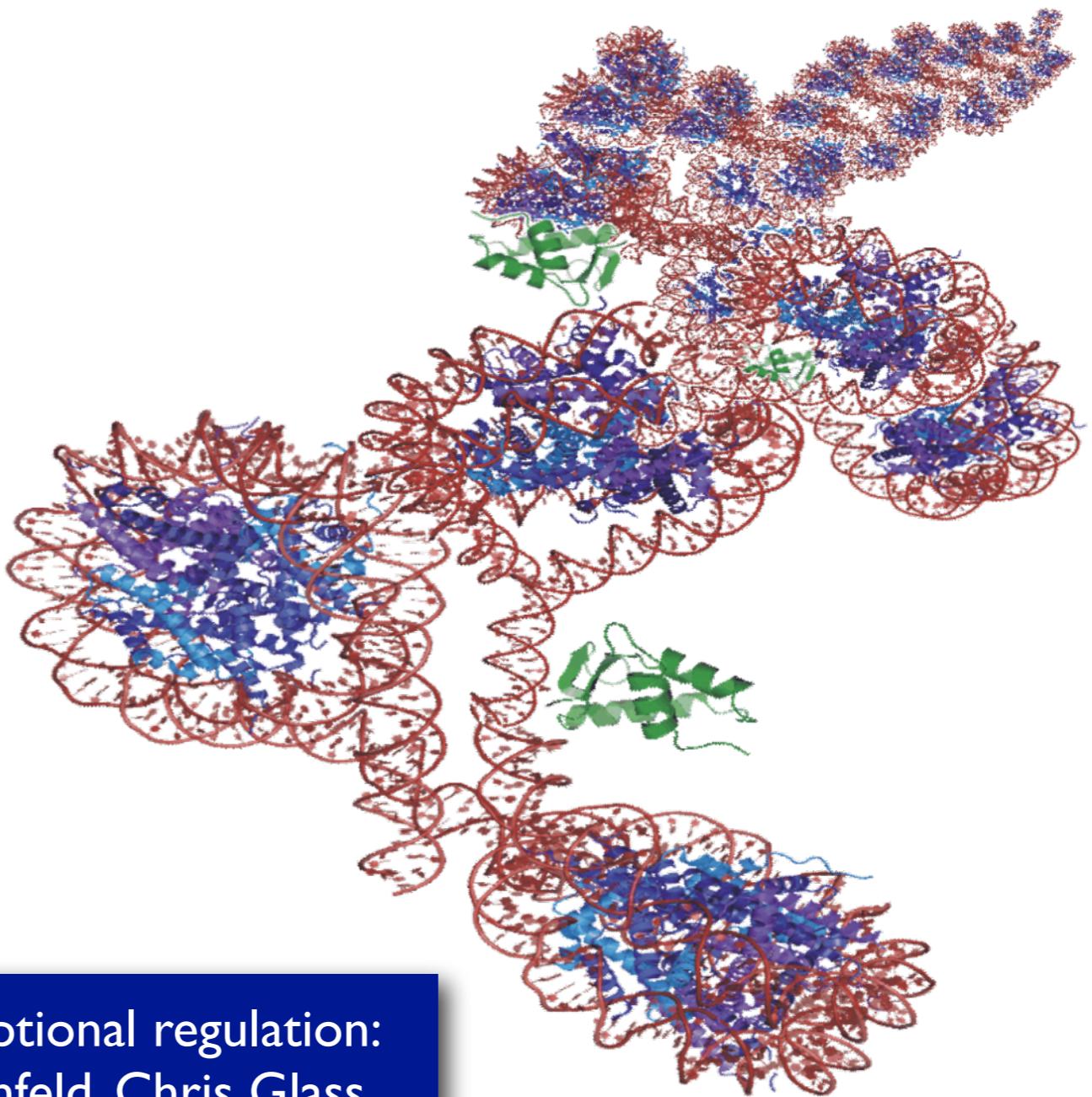
Nature Reviews | Microbiology

Chromatin structure



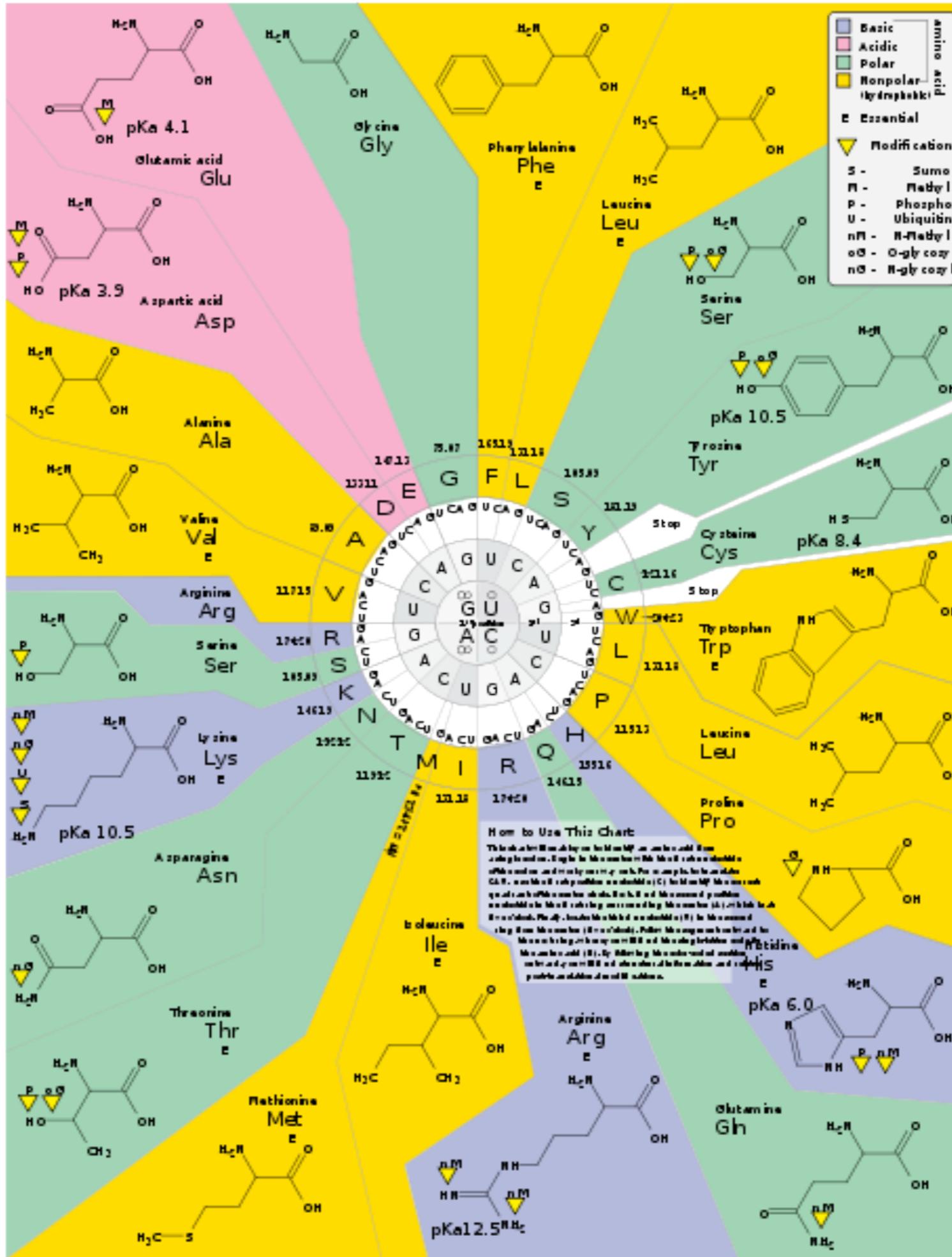
Chromatin structure

- All the DNA in chromosomes and associated proteins collectively is called chromatin
- DNA is a double helix, wound around histones, which in turn are organized into higher order structures and domains dependent on cell stage, type, and stimulation
- An open area of research
- “Active” and “inactive” regions
- Rosettes held together by CTCF, cohesin, allowing for long-range interactions



For more on transcriptional regulation:
Bing Ren, Geoff Rosenfeld, Chris Glass,
Wei Wang, Kun Zhang, Gene Yeo

PROTEINS!



Proteins!

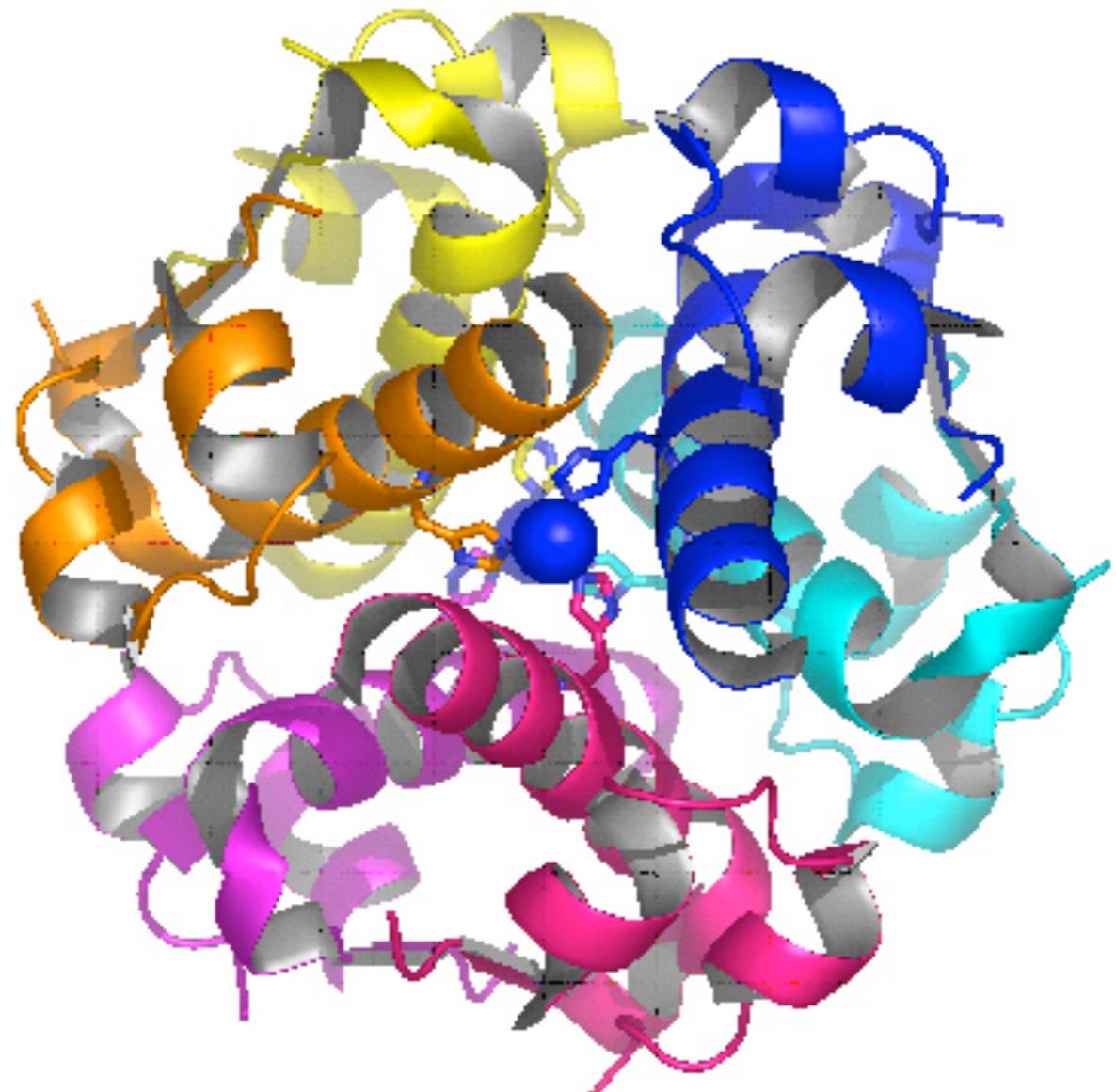
Many classes:

- Location: nuclear, cytoplasmic, cell-surface, secreted
- Binding activity: DNA-binding, RNA-binding, protein-binding, lipid-binding
- Structure: alpha-helices, beta-sheets, globin-like, carbohydrate-binding domain
- Function: receptor, ligand, enzyme, substrate, kinase, transporter, cytokine

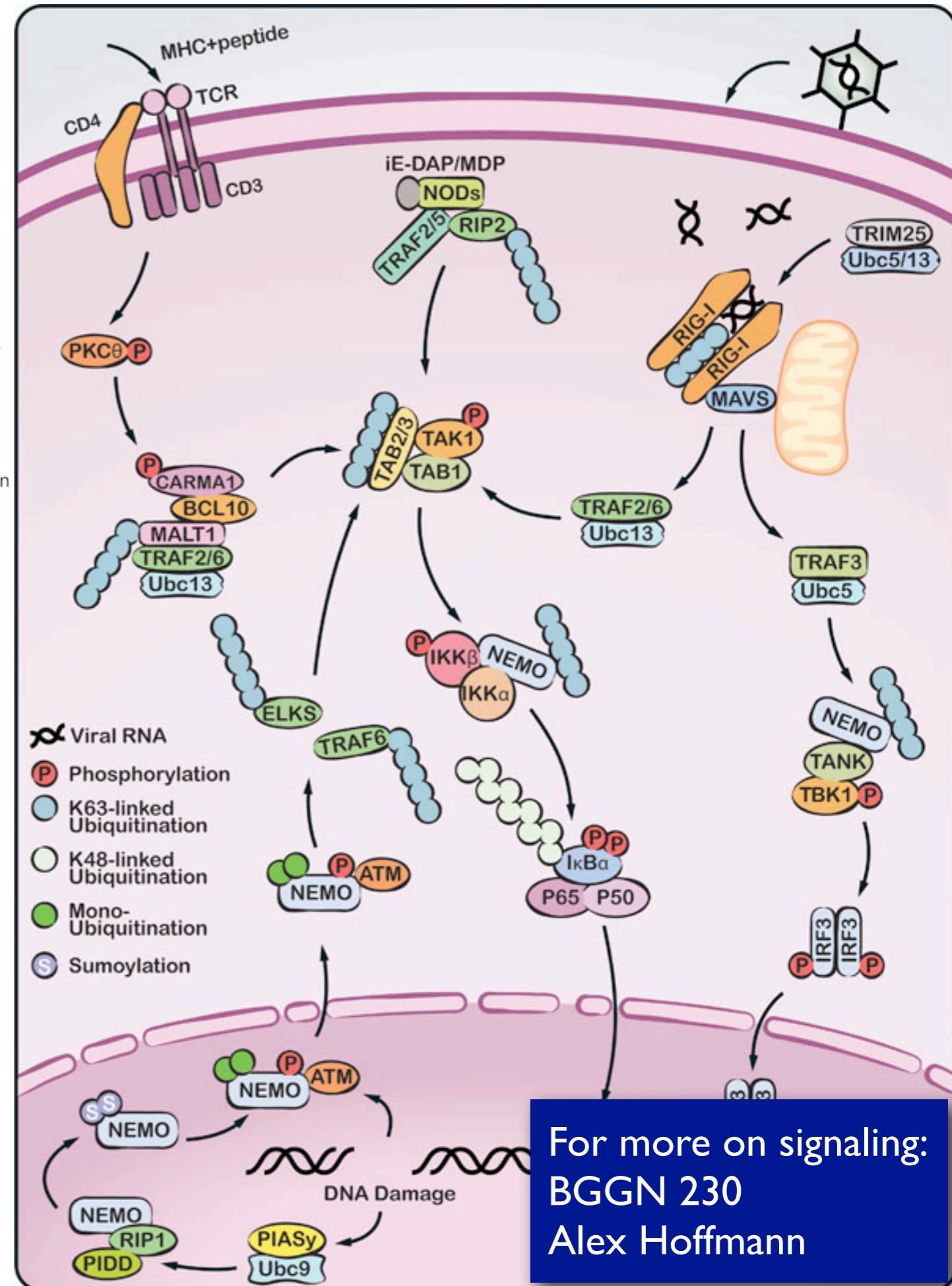
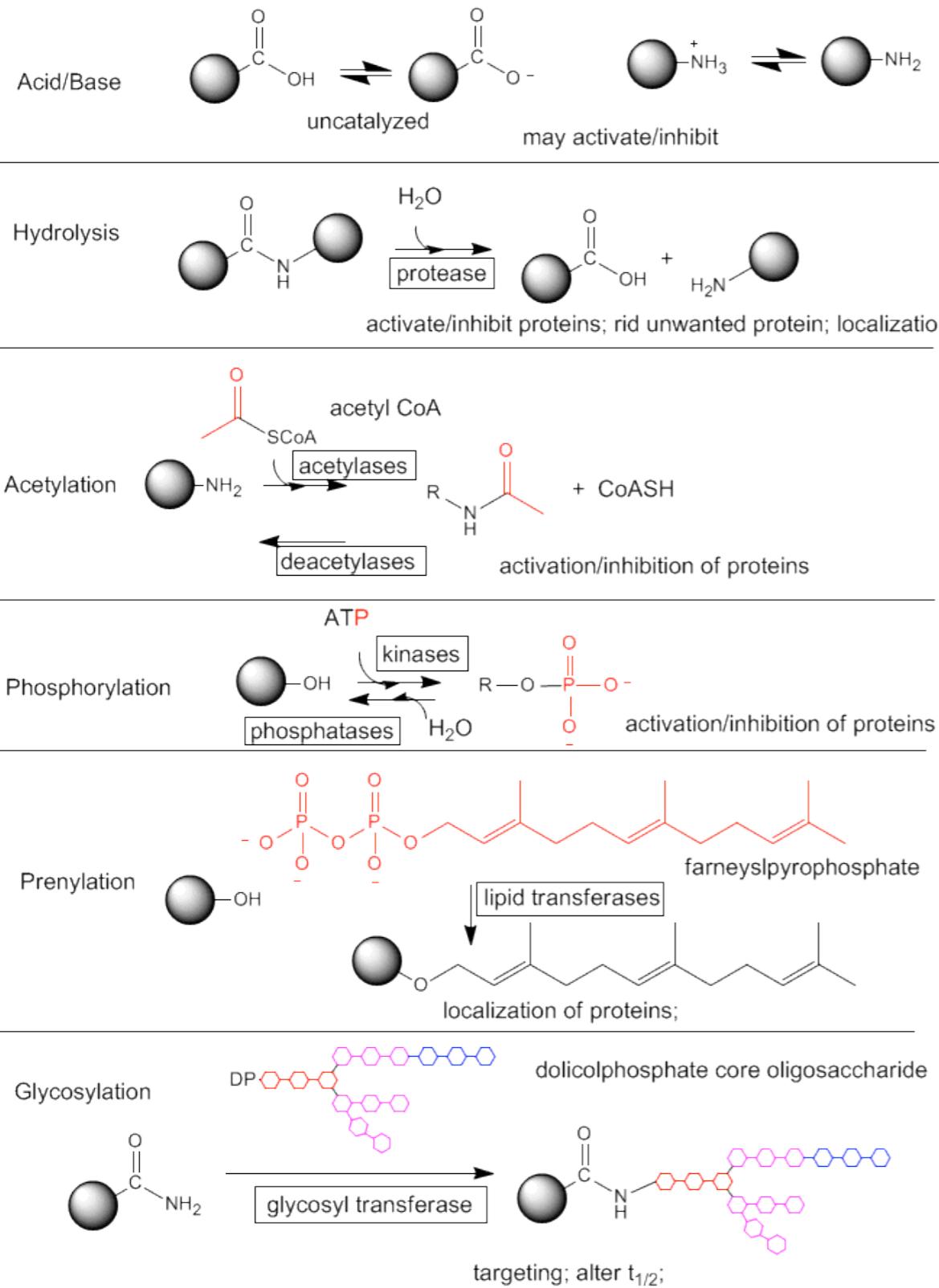
And post-translational modifications (PTMs):

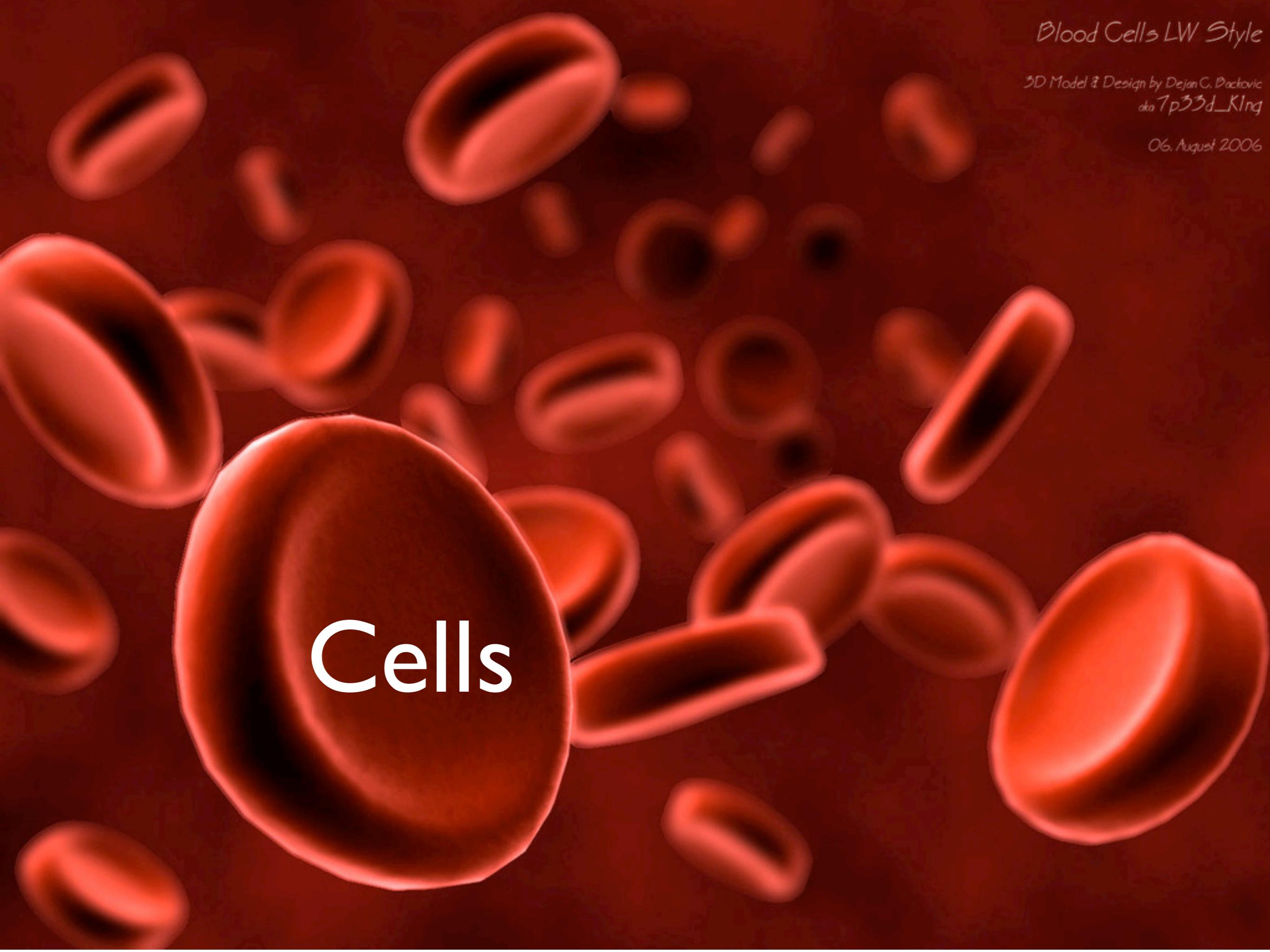
- Chemical groups: phosphorylation, acetylation, methylation
- Peptides: ubiquitination, SUMOylation
- Chemical localization signals: myristoylation, palmitoylation

For more on proteins:
Bioinformatics I
Phil Bourne, Vineet Bafna,
Nuno Bandeira



Proteins!



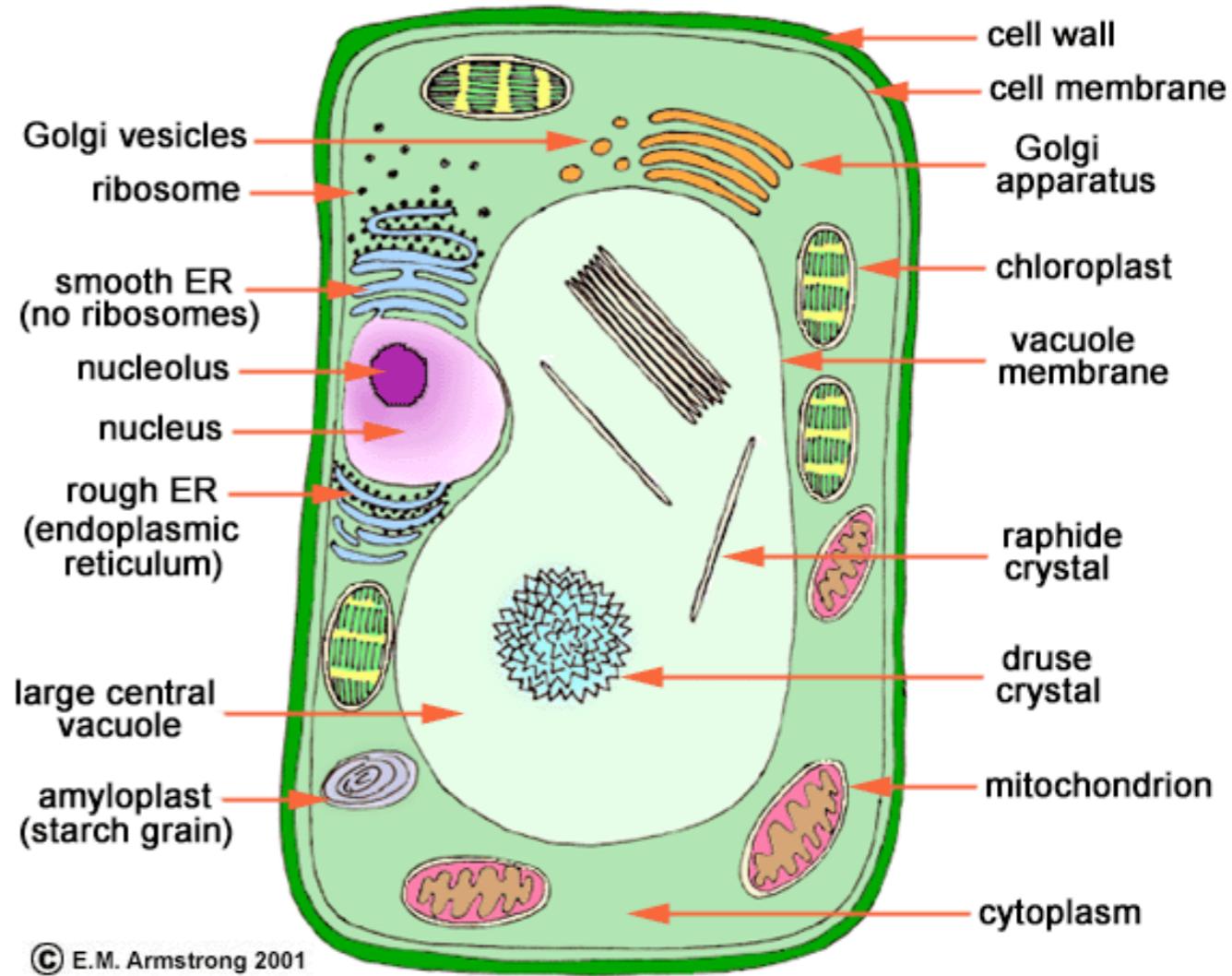
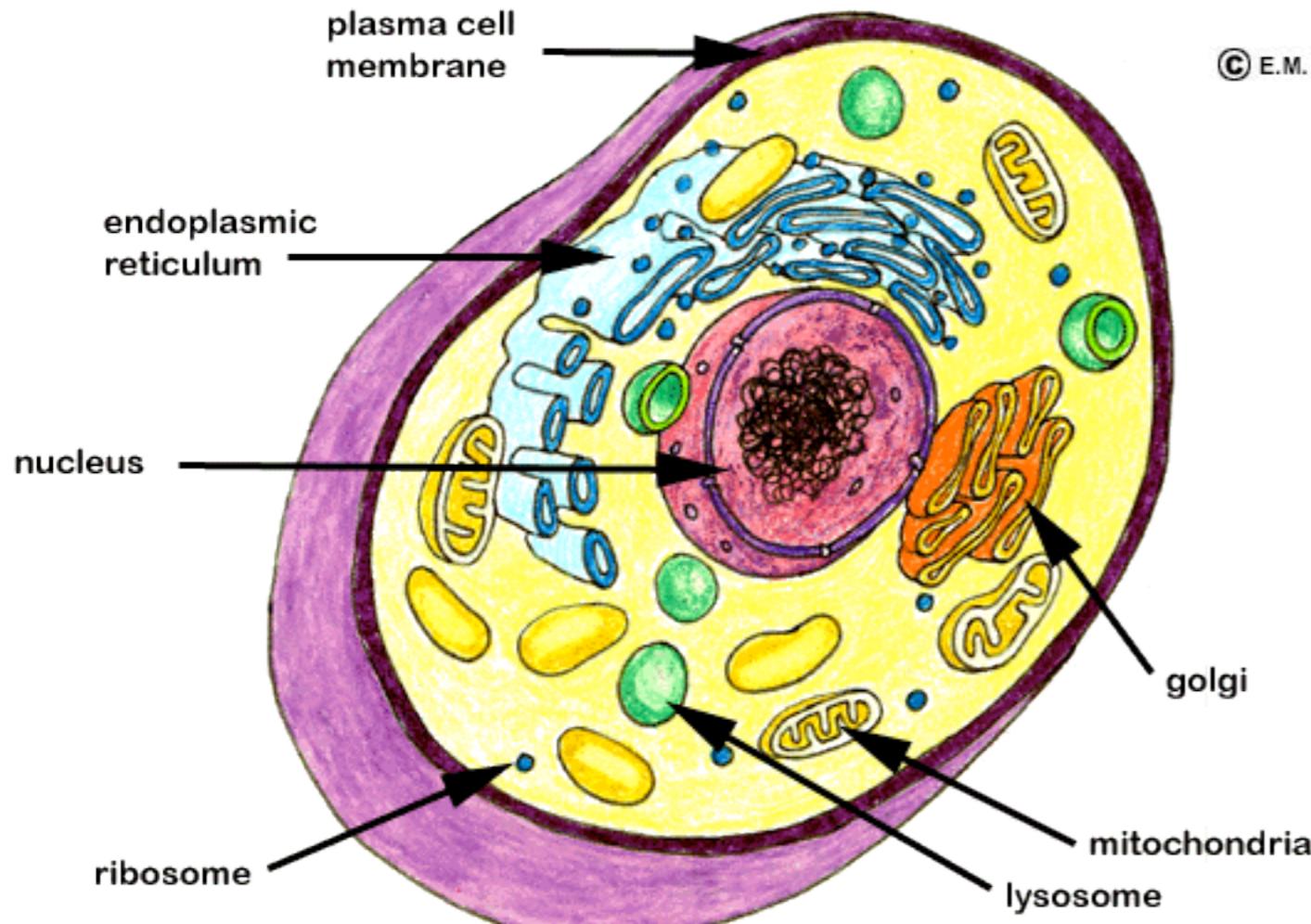
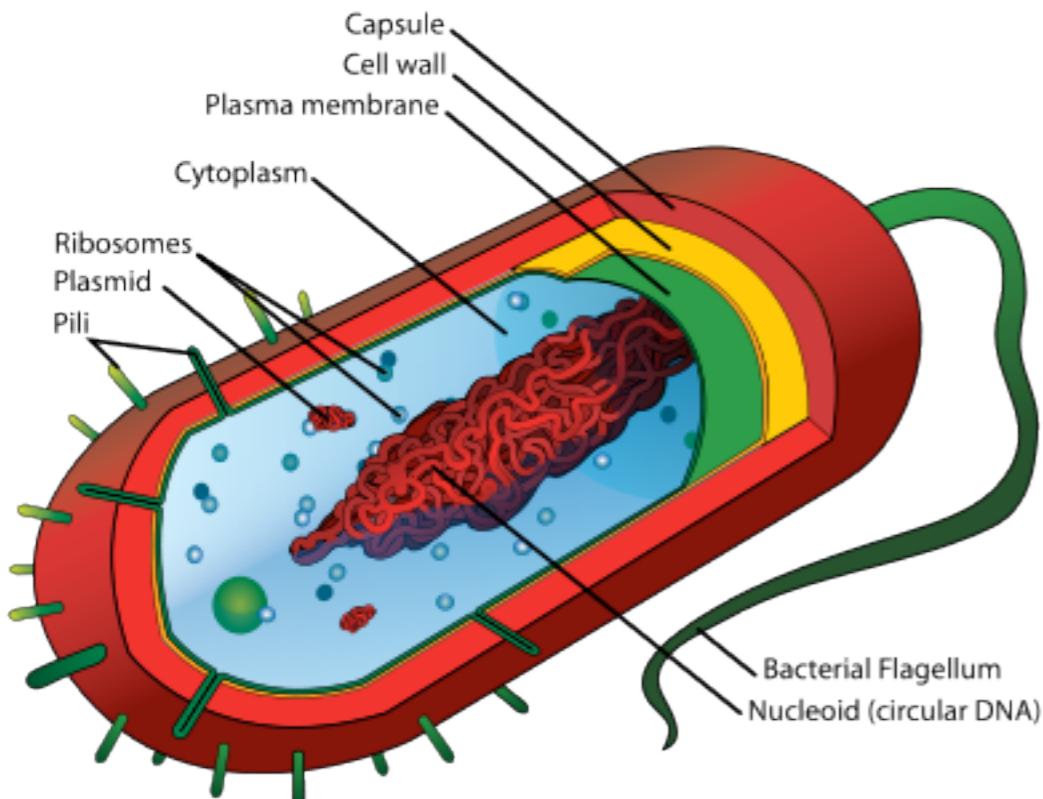
The background of the image is a dark, reddish-brown color, representing a dense concentration of red blood cells. The cells are rendered with a slight transparency and a soft lighting effect, giving them a three-dimensional appearance as if they are floating in a liquid medium.

Blood Cells LW Style

3D Model & Design by Dejan C. Backovic
aka 7p33d_King

06. August 2006

Cells

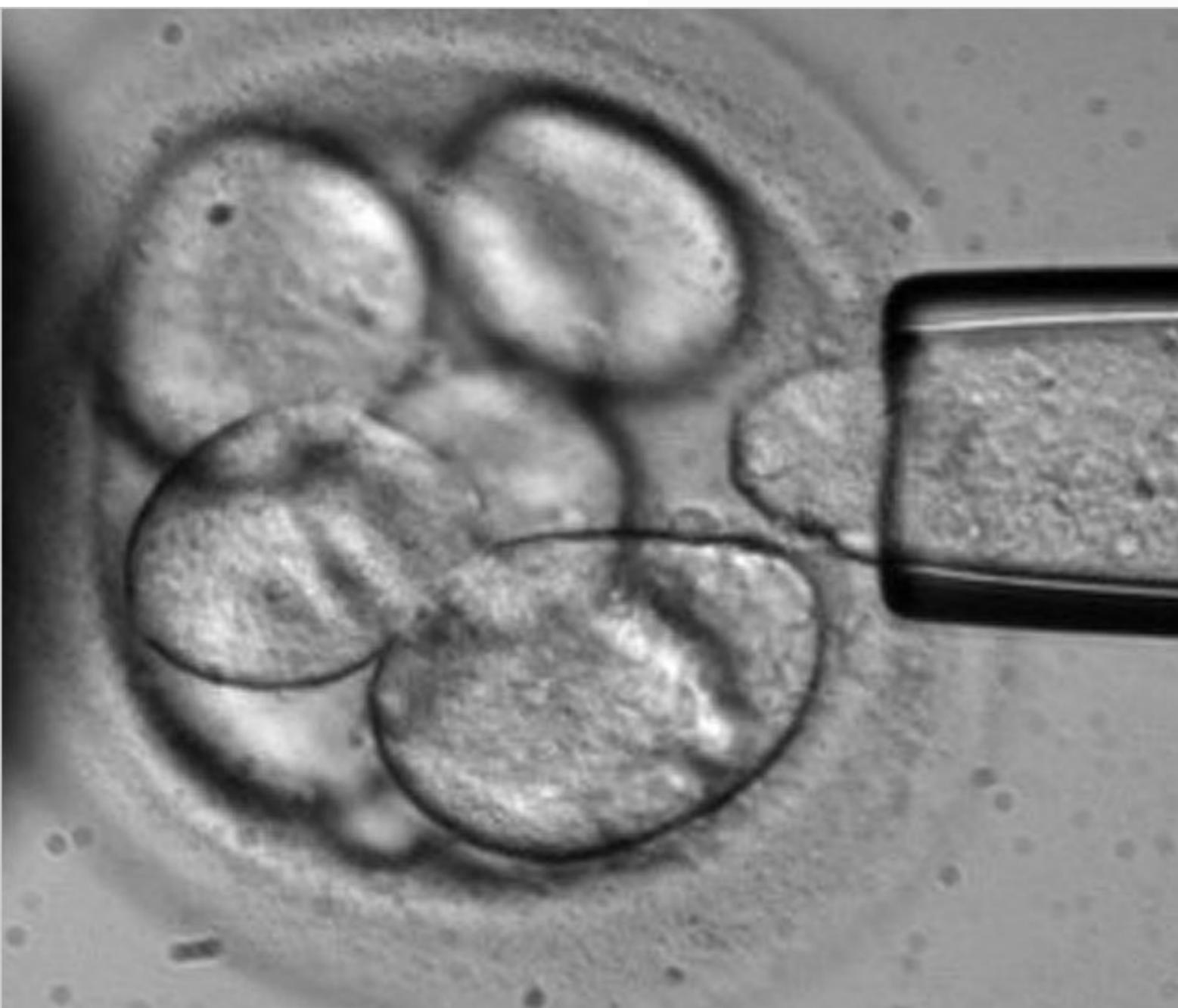
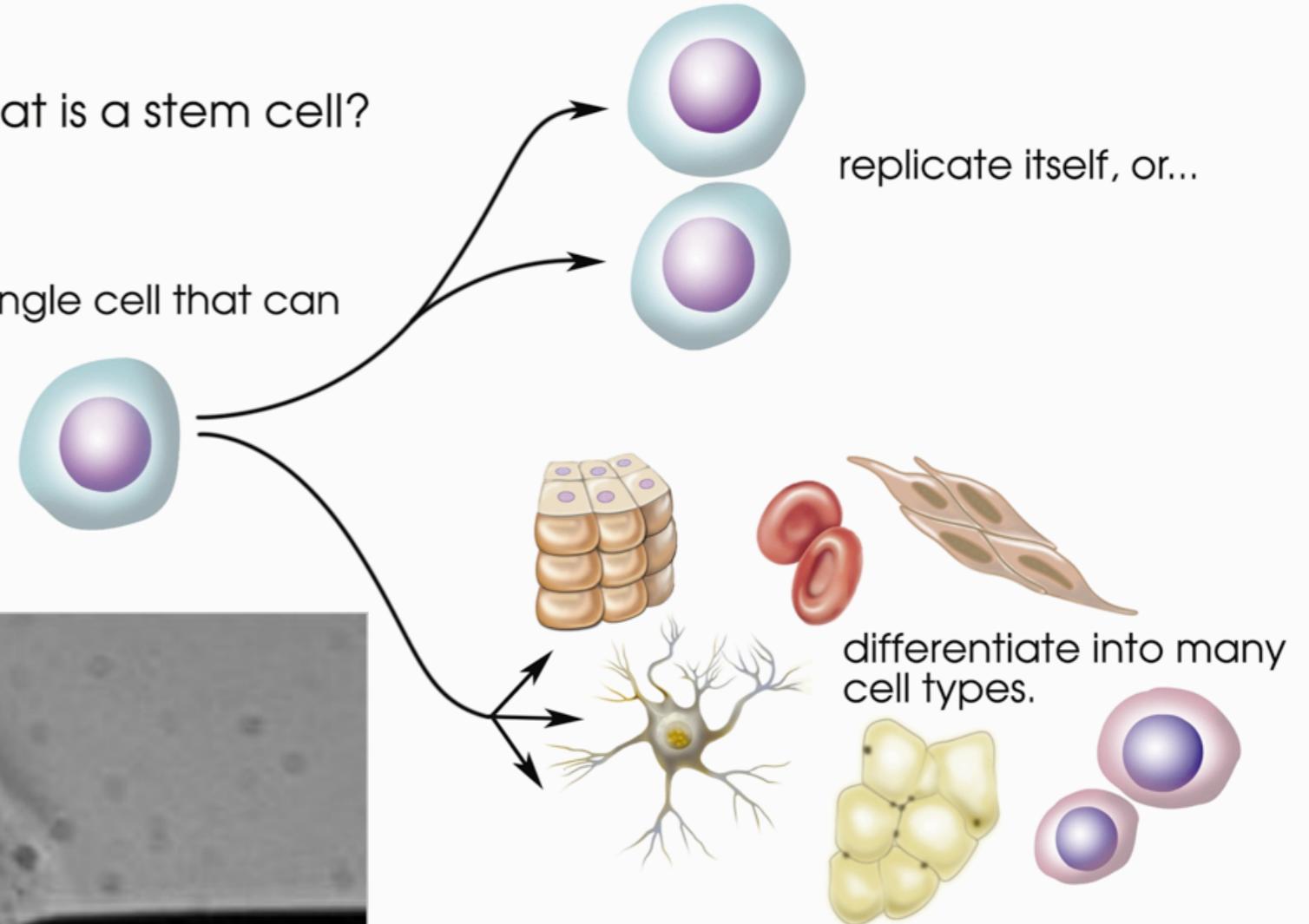


- Prokaryote vs. eukaryote
- Bacterial, animal, plant
- Nucleus, cytoplasm, ER, ribosome, vacuole, golgi, mitochondria, plasma membrane
- Cytokines, hormones, cilia, extracellular receptors, pores, channels

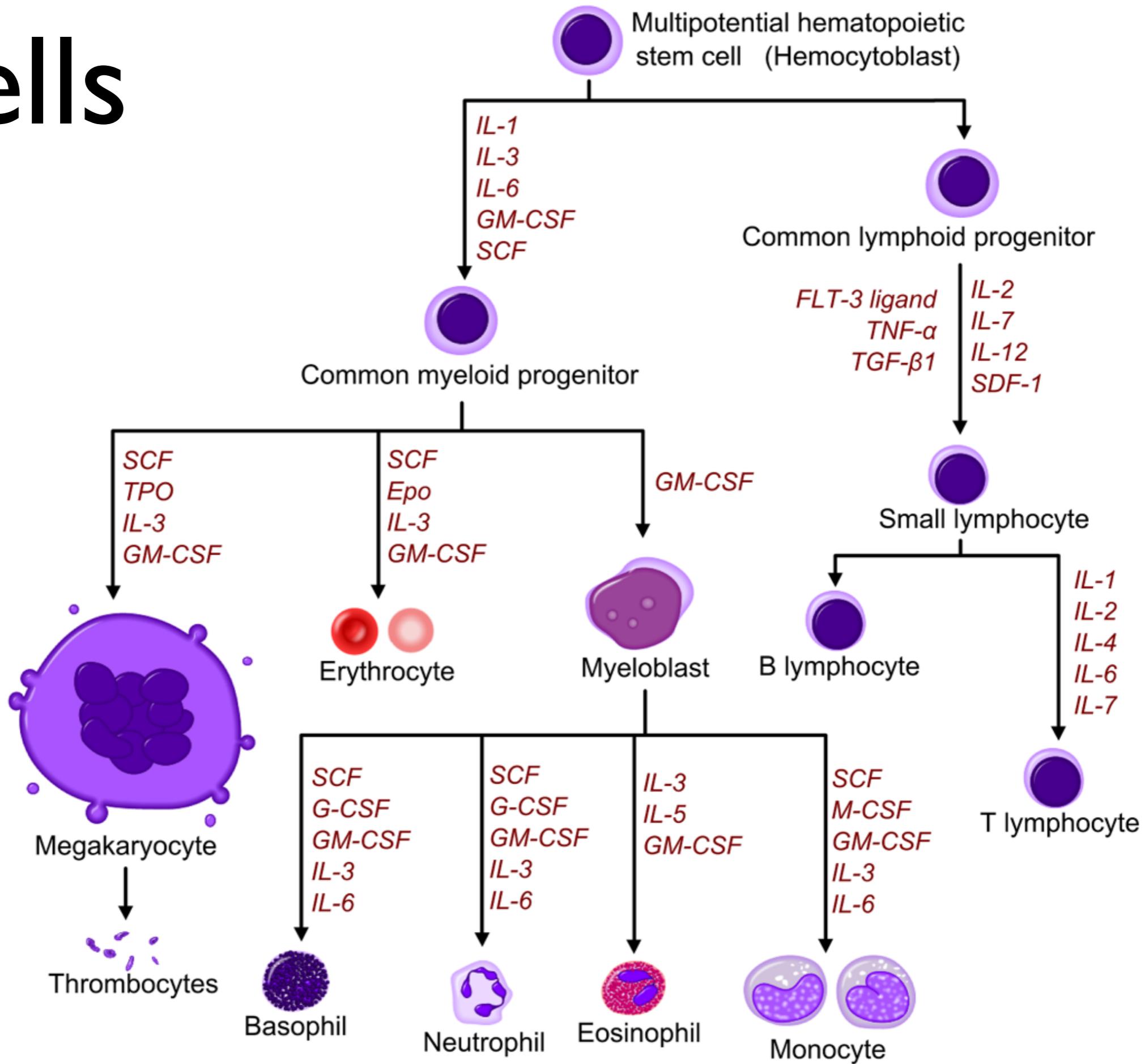
Cells

What is a stem cell?

A single cell that can

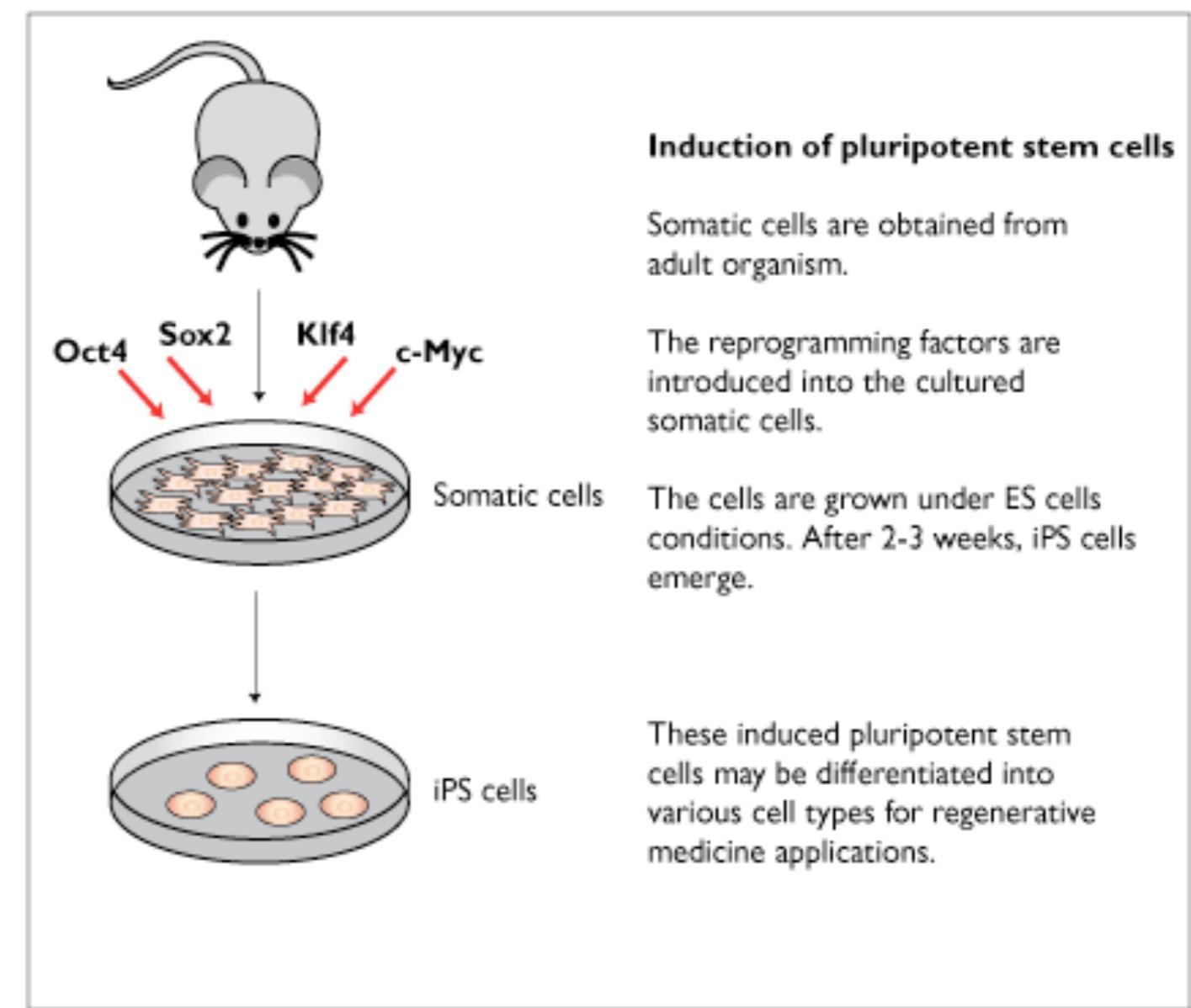
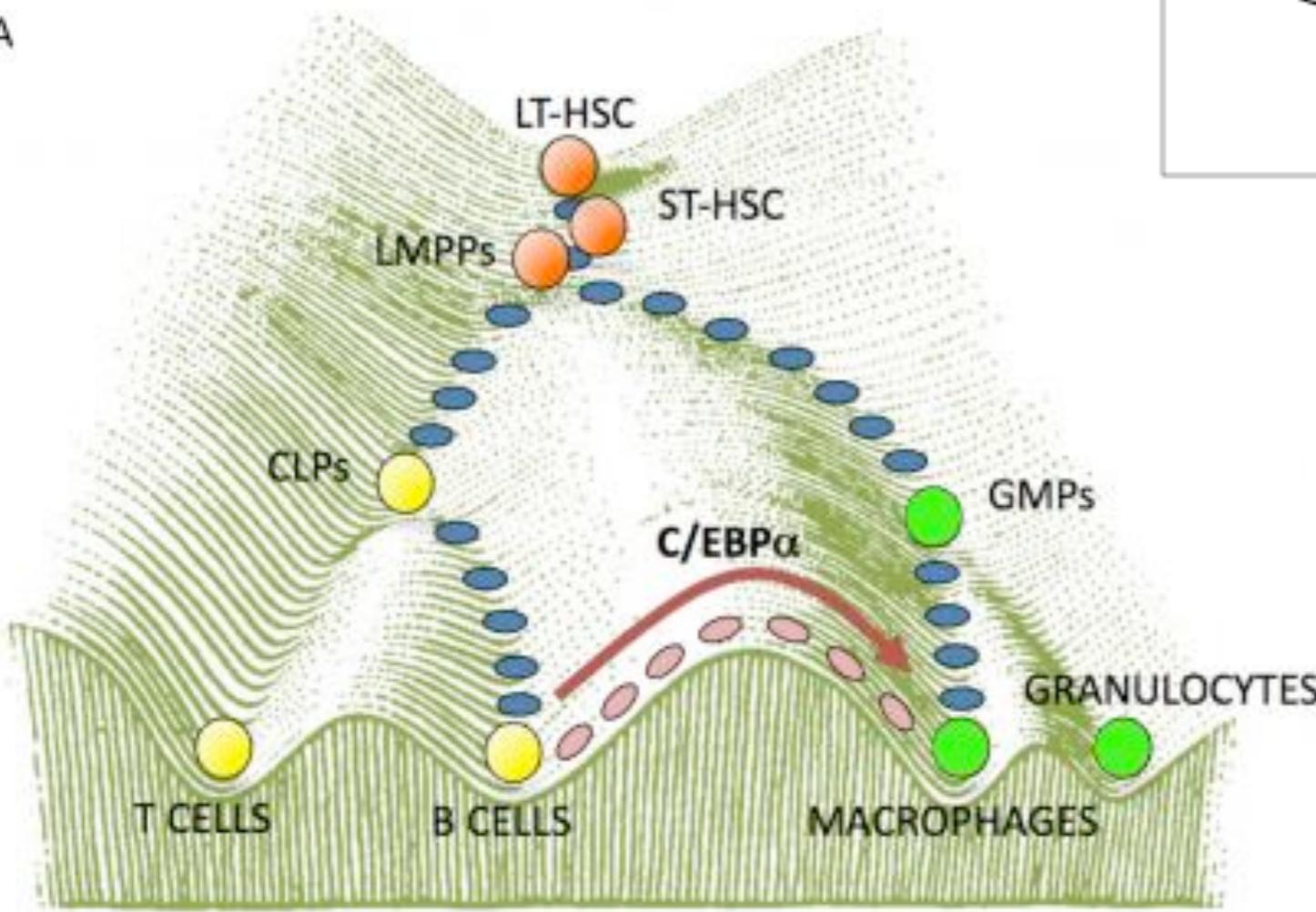


Cells



Cells

A



Induction of pluripotent stem cells

Somatic cells are obtained from adult organism.

The reprogramming factors are introduced into the cultured somatic cells.

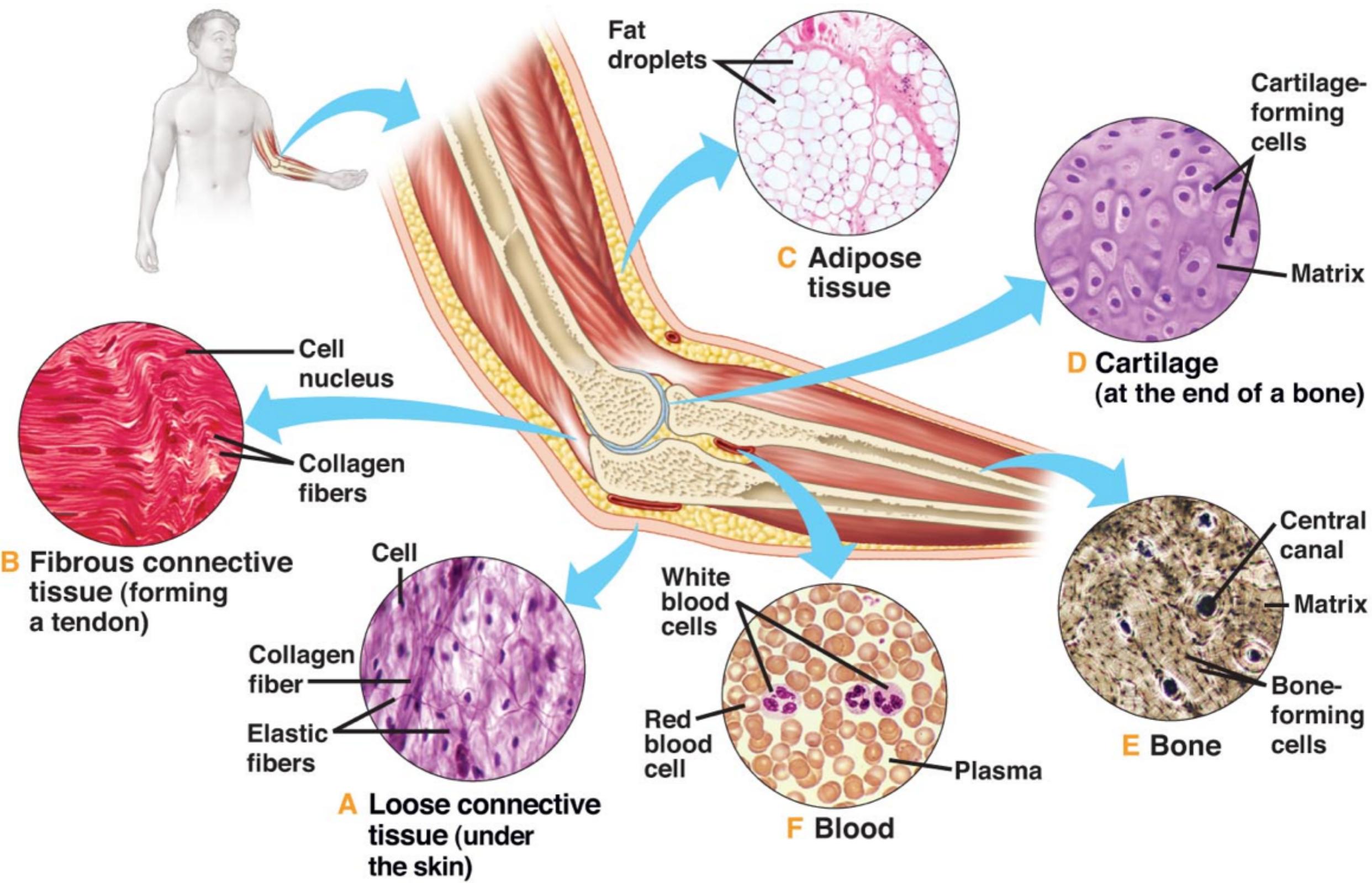
The cells are grown under ES cells conditions. After 2-3 weeks, iPS cells emerge.

These induced pluripotent stem cells may be differentiated into various cell types for regenerative medicine applications.

For more on stem cells:
BGGN 231
Steve Briggs, Bing Ren,
Gene Yeo, Kun Zhang

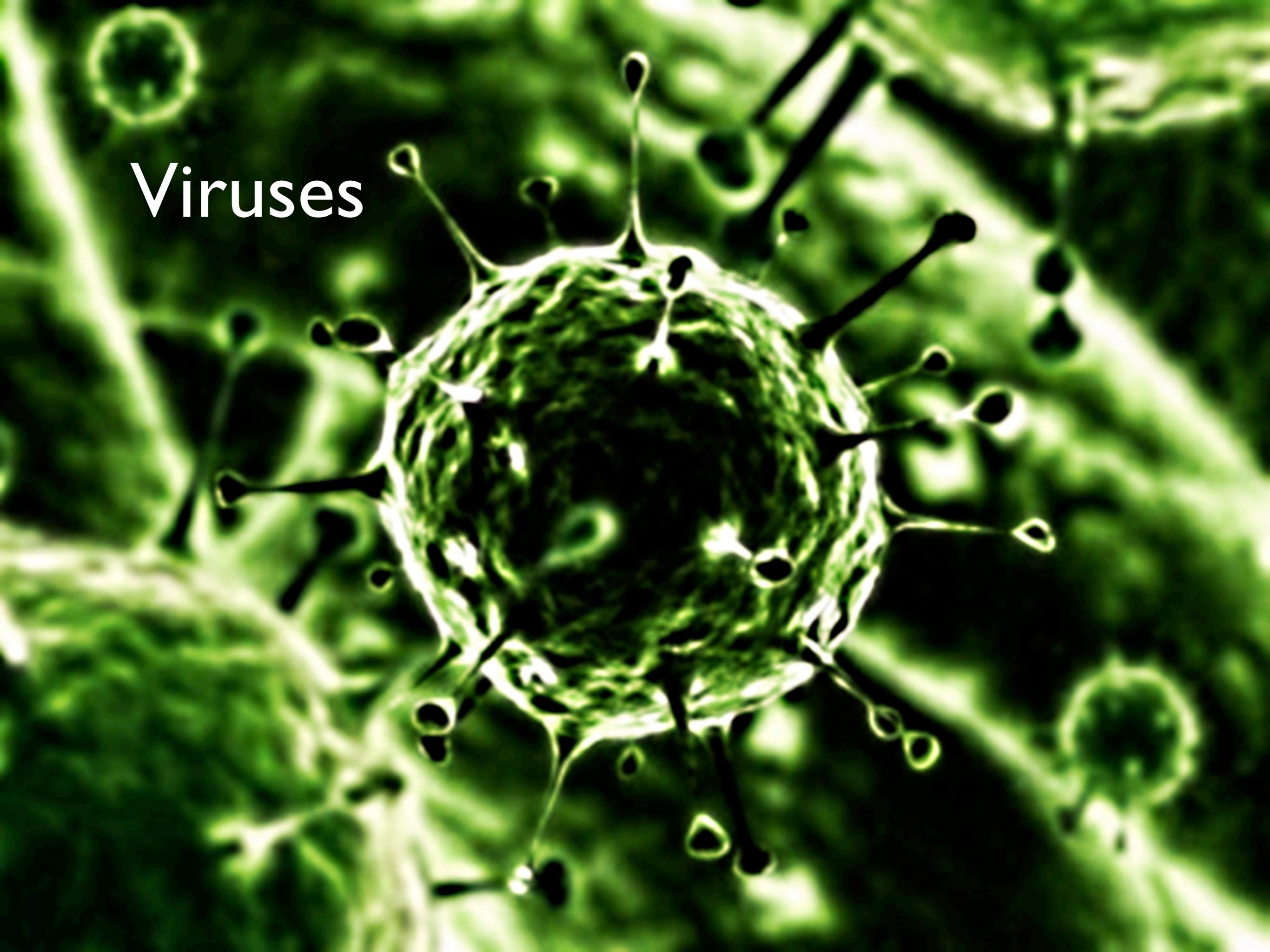
Waddington Landscape

Tissues



Organisms

EF 6691 • 5.0 kV × 15.0k 2.00 μ m

A high-magnification electron micrograph showing several viruses. The viruses are spherical particles with a distinct outer layer and internal structural details. Some have long, thin, tail-like projections extending from their surfaces. The background is dark, making the viral particles stand out.

Viruses

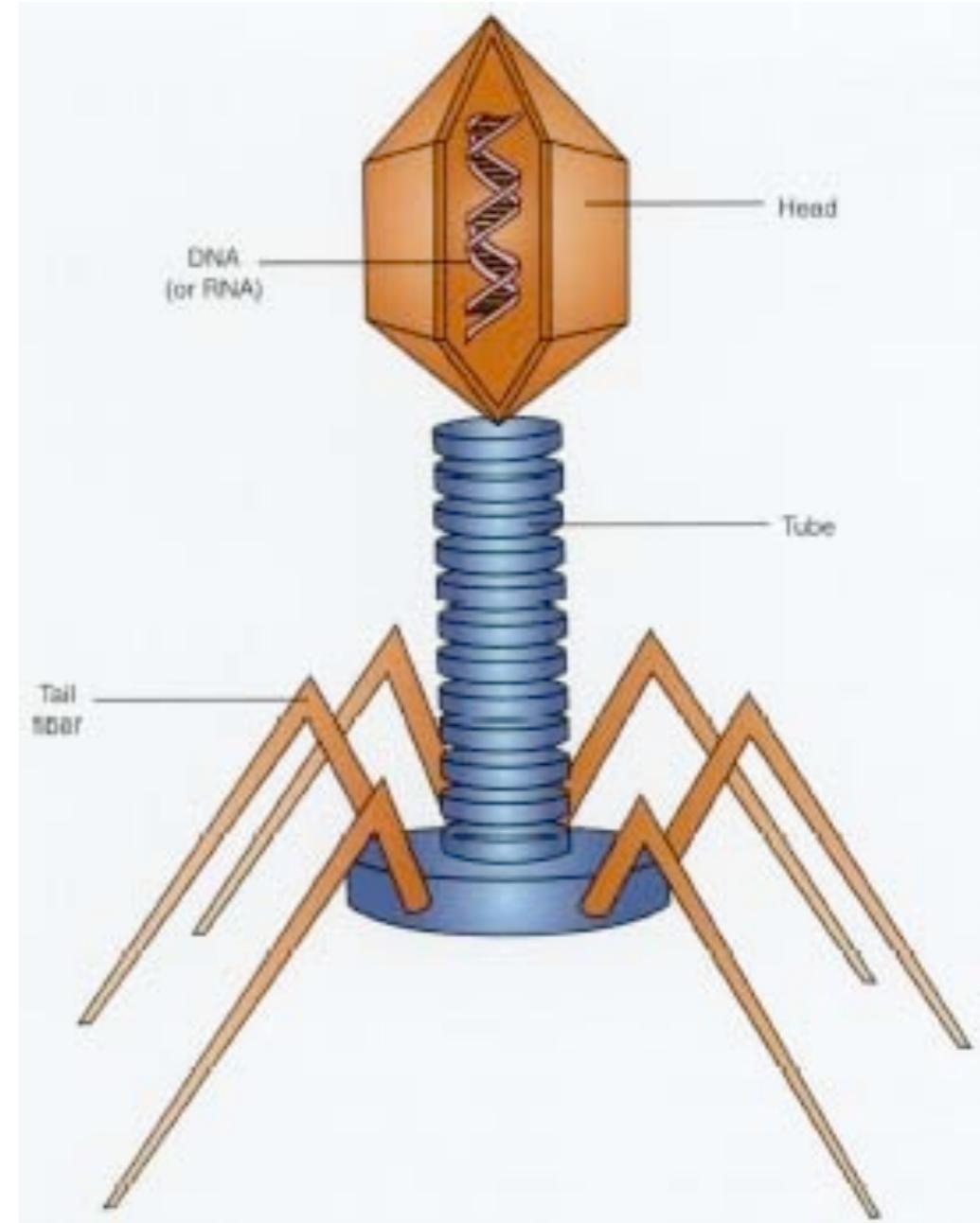
Viruses

Basics

- Nucleic acids surrounded by a protein or lipid envelope
- Genomes vary widely in size, from 2 kb to 1.2 Mb
- Co-opt host cellular machinery to replicate
- Many complex mechanisms of host-defense evasion have evolved

Examples

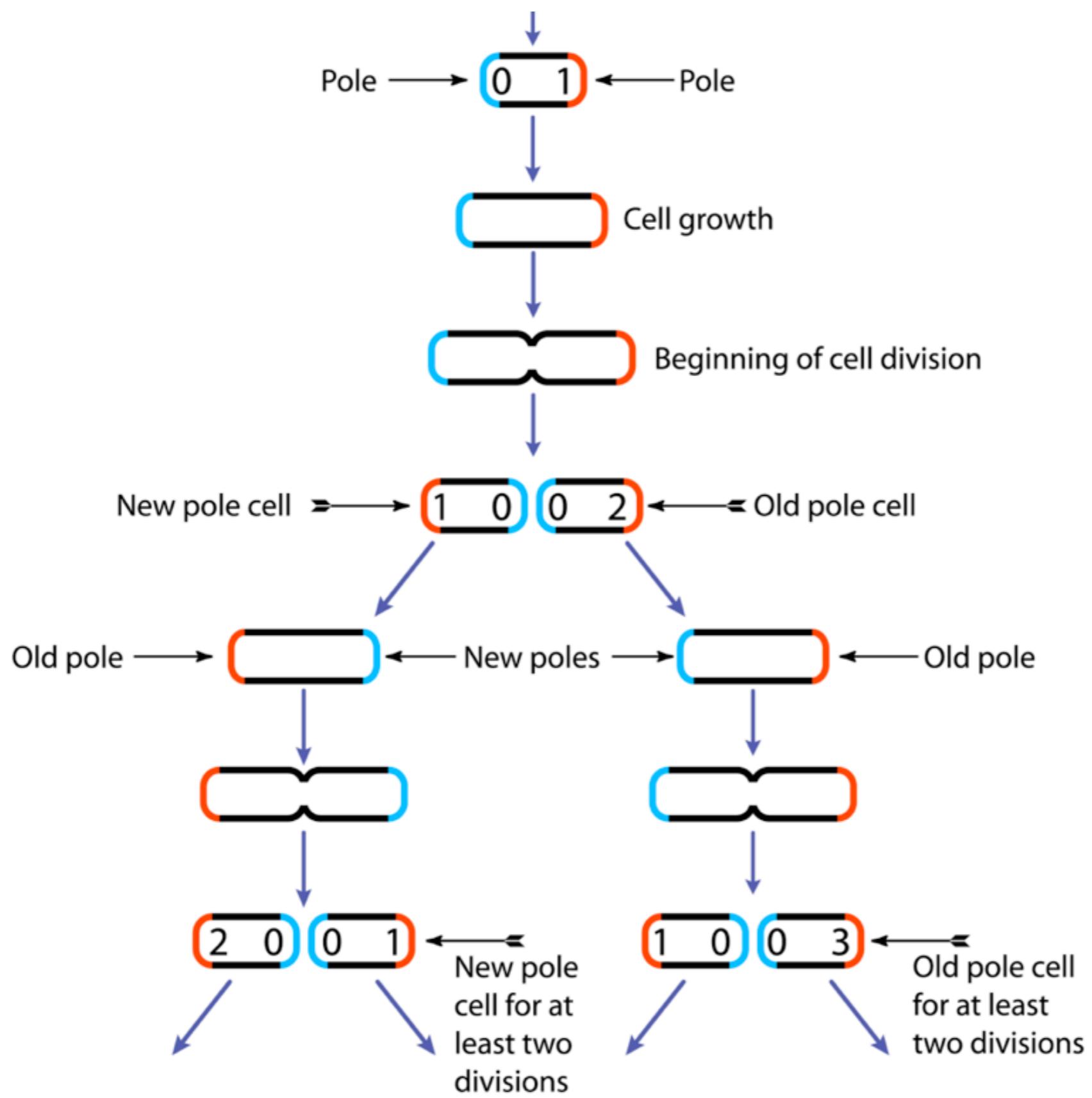
- HIV
- Herpesvirus
- Cytomegalovirus (CMV)
- Influenza, dengue, smallpox, SARS, hepatitis
- Viral vectors (lentivirus, adenovirus)
- Endogenous retroviruses



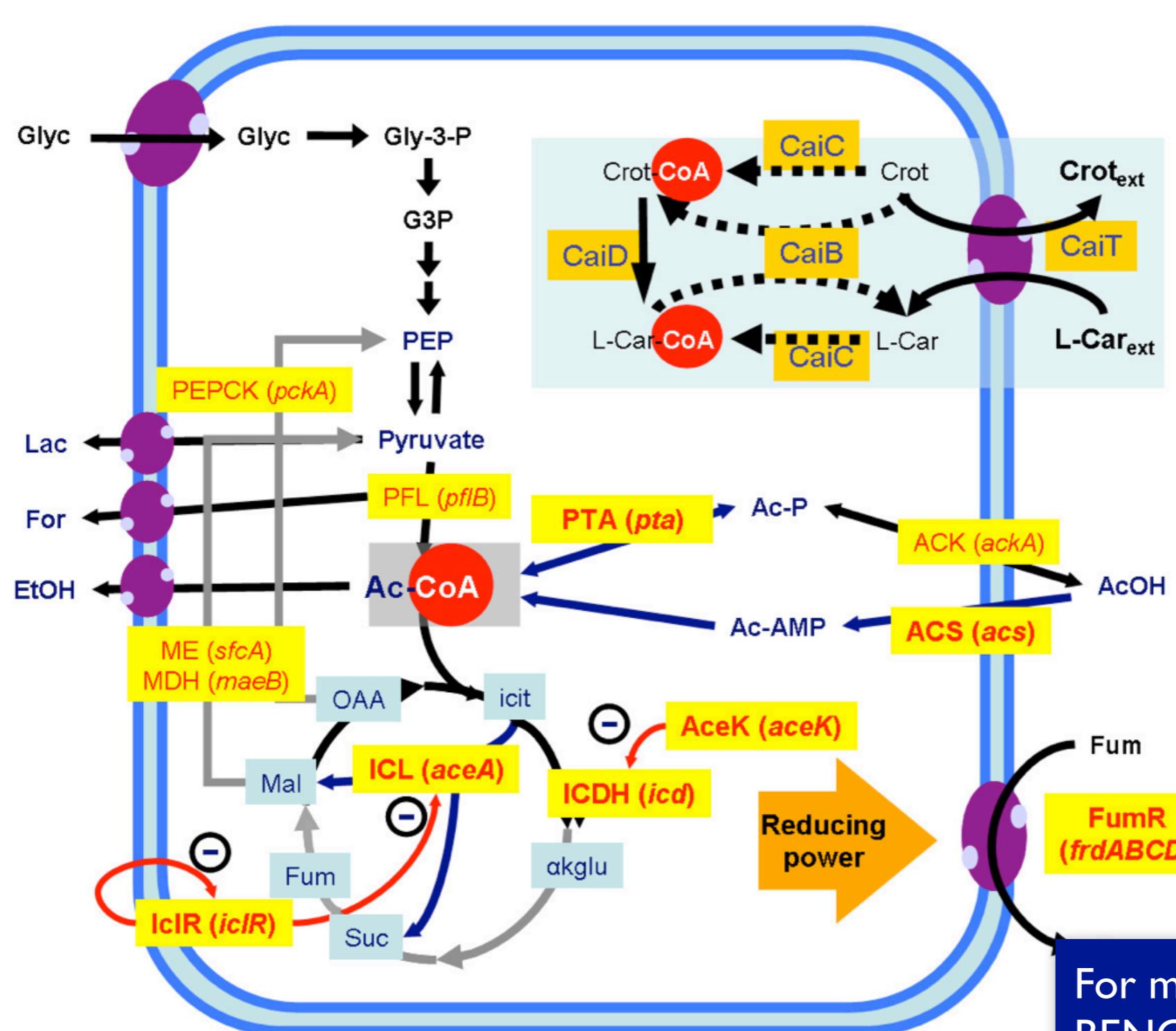
For more on viruses:
BIMM 114, BIMM 164, BGGN 226
Sergei Pond, Doug Richman, Chris Woelk

Escherichia coli (*E. coli*)

B

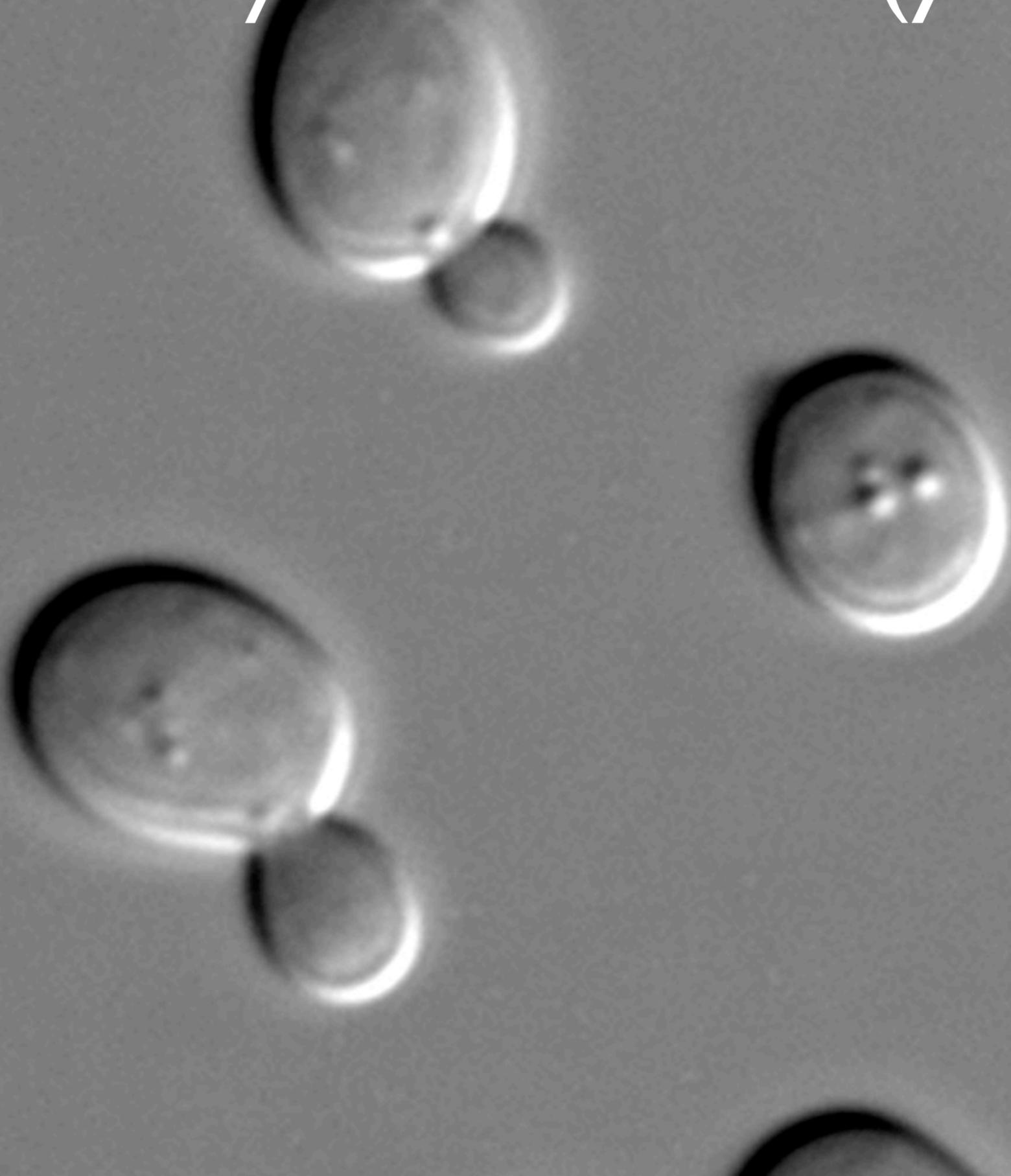


Escherichia coli (*E. coli*)

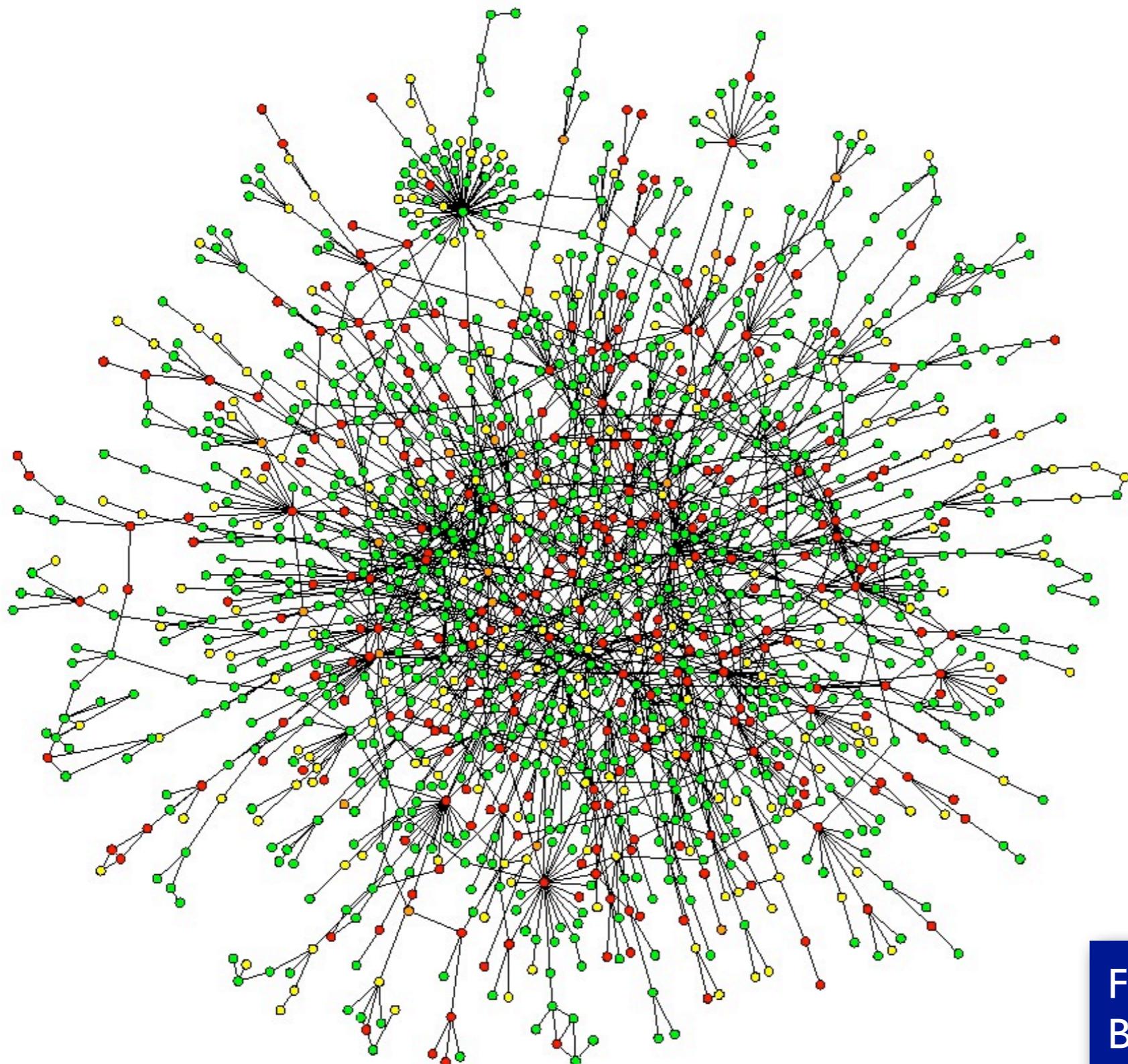


For more on metabolic models:
BENG 211 - 213
Bernard Palsson

Saccharomyces cerevisiae (yeast)

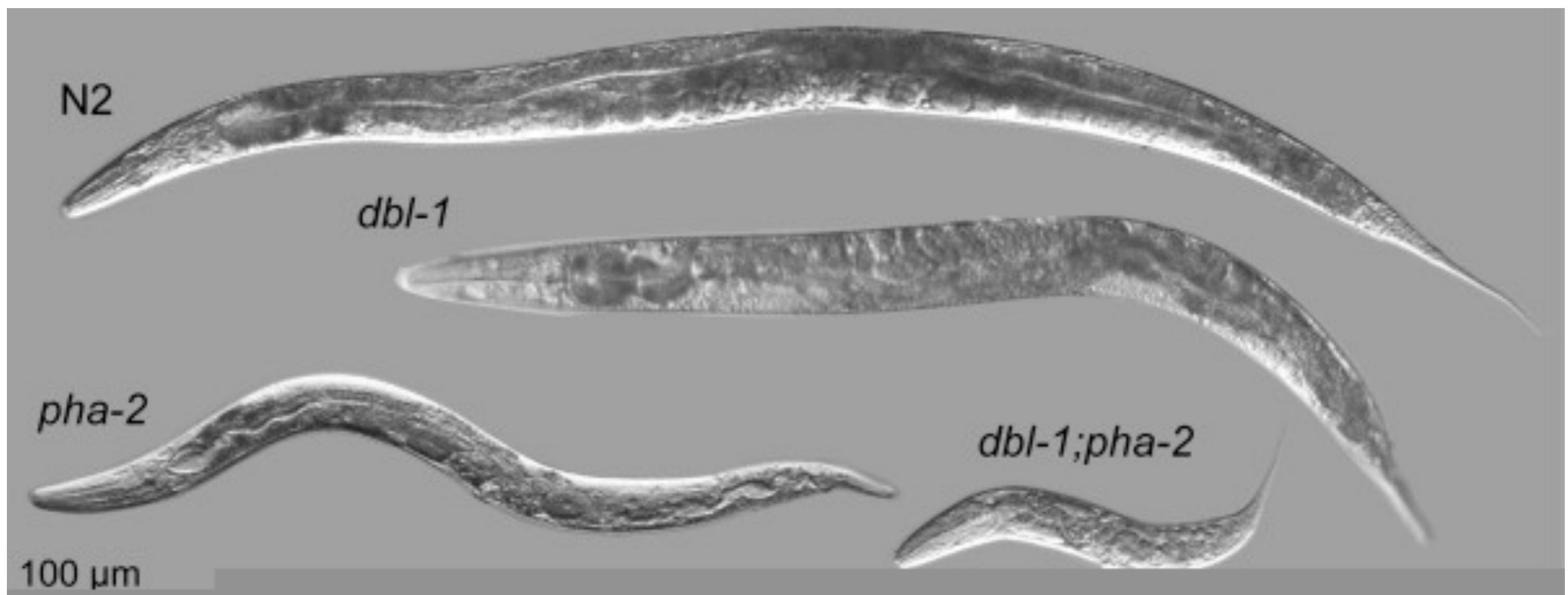
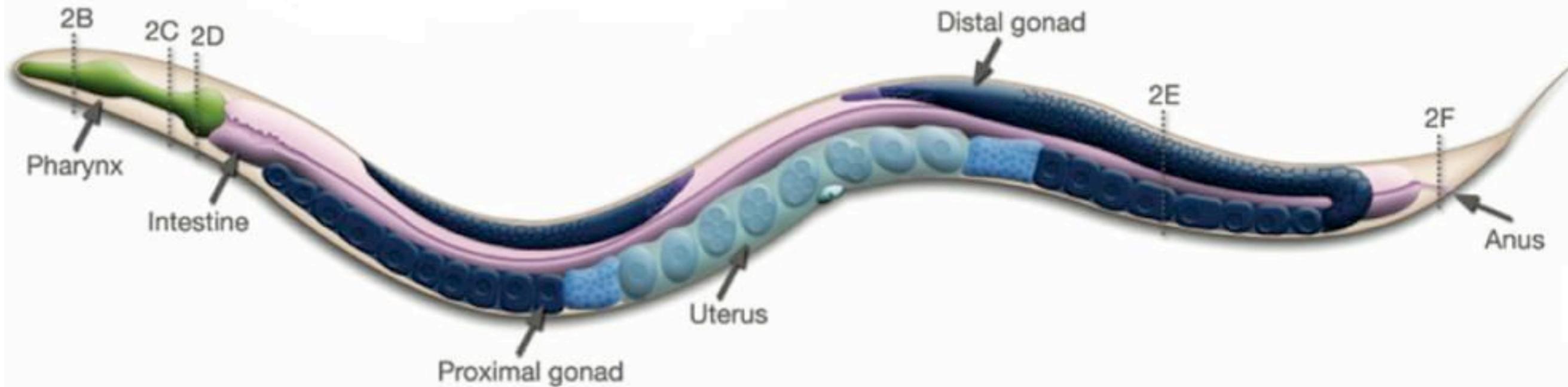


Saccharomyces cerevisiae (yeast)



For more network modeling:
Bioinformatics III, BENG 212
Trey Ideker

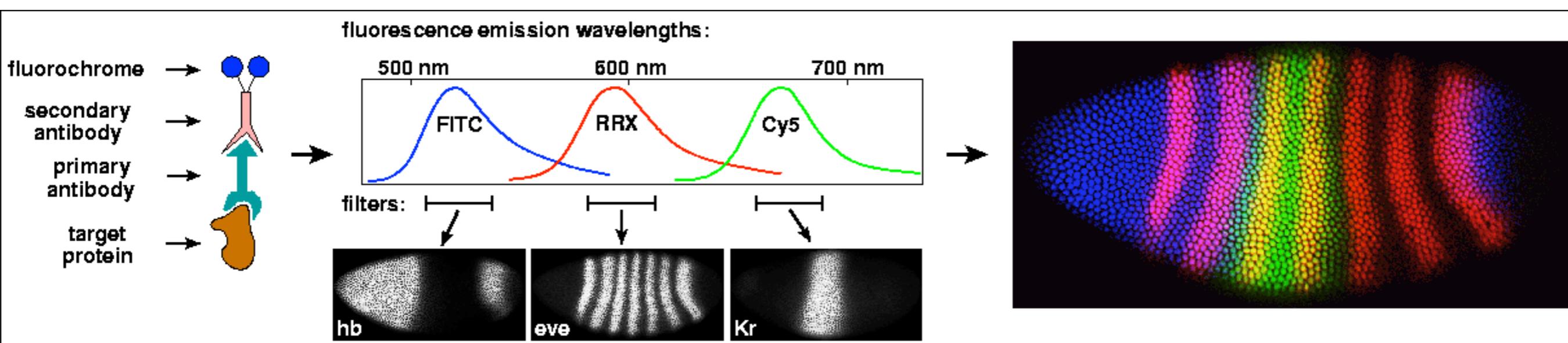
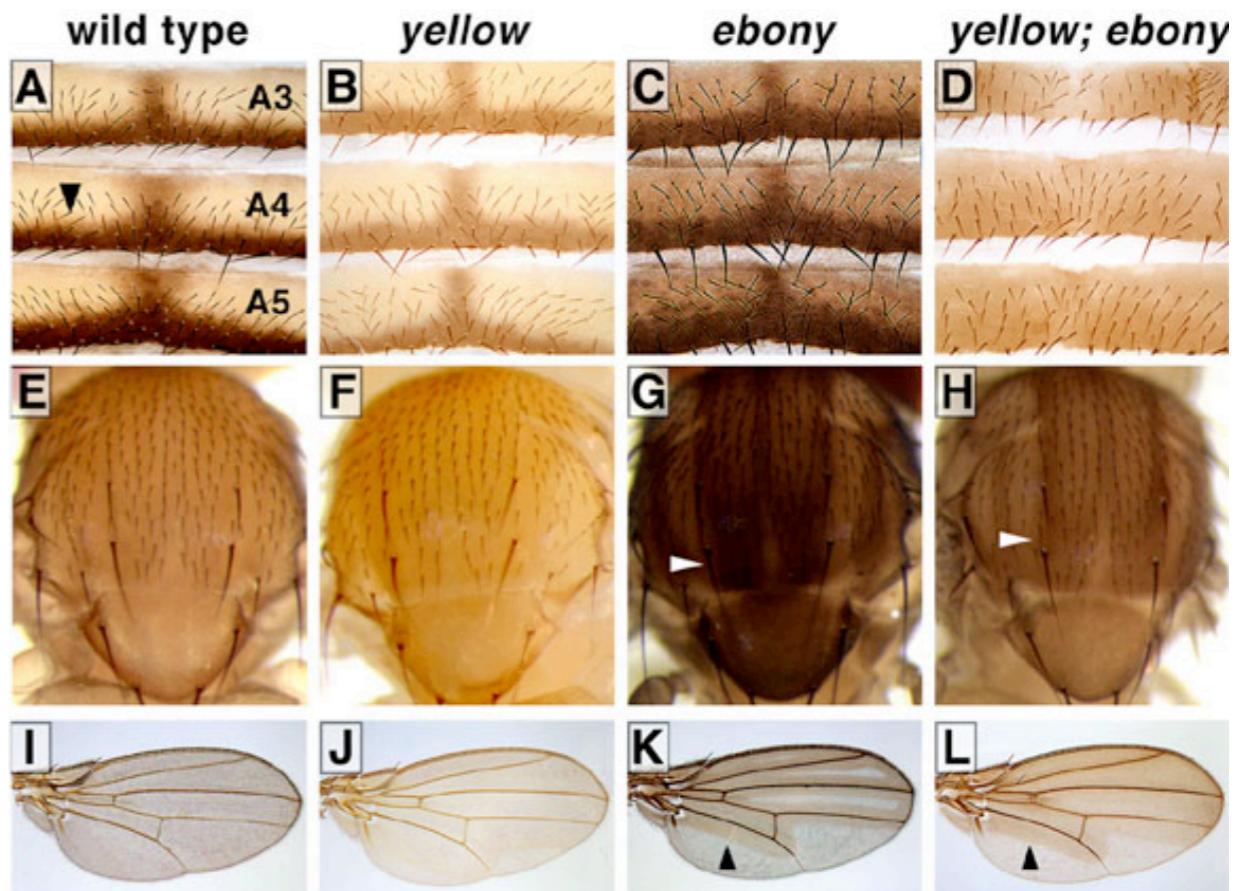
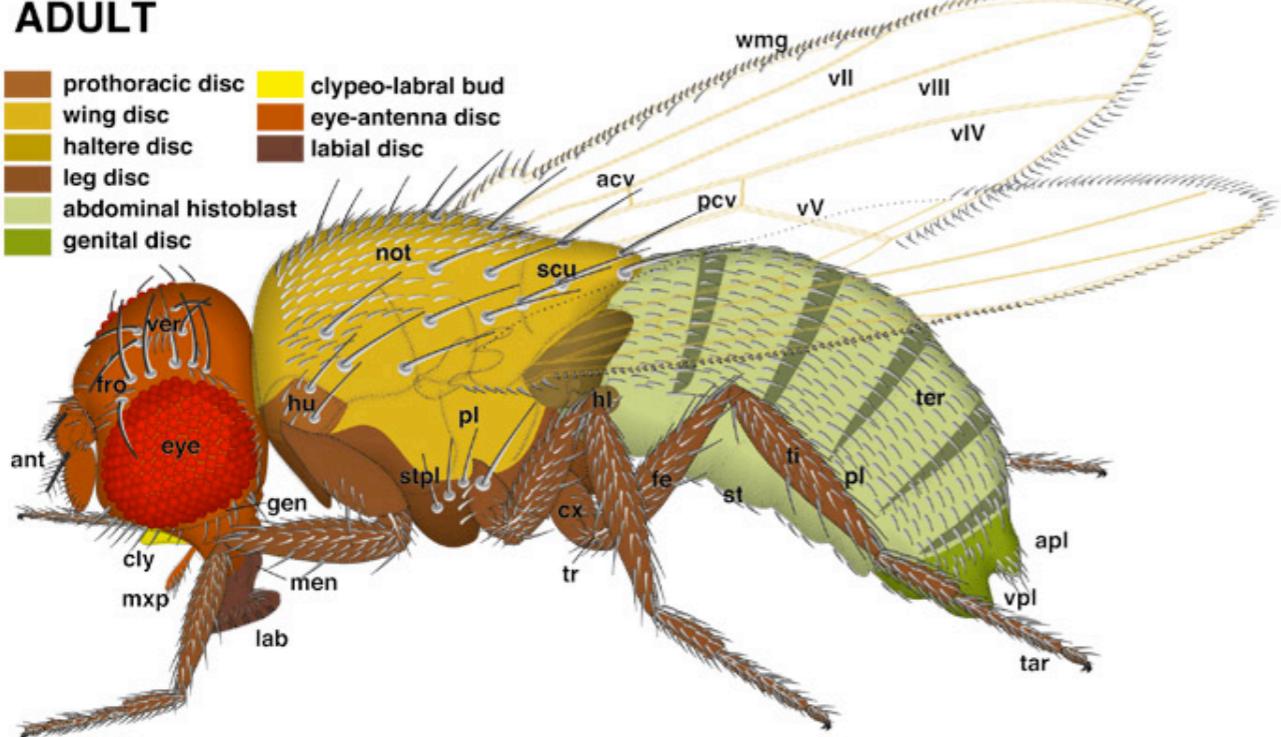
Caenorhabditis elegans (*C. elegans*)



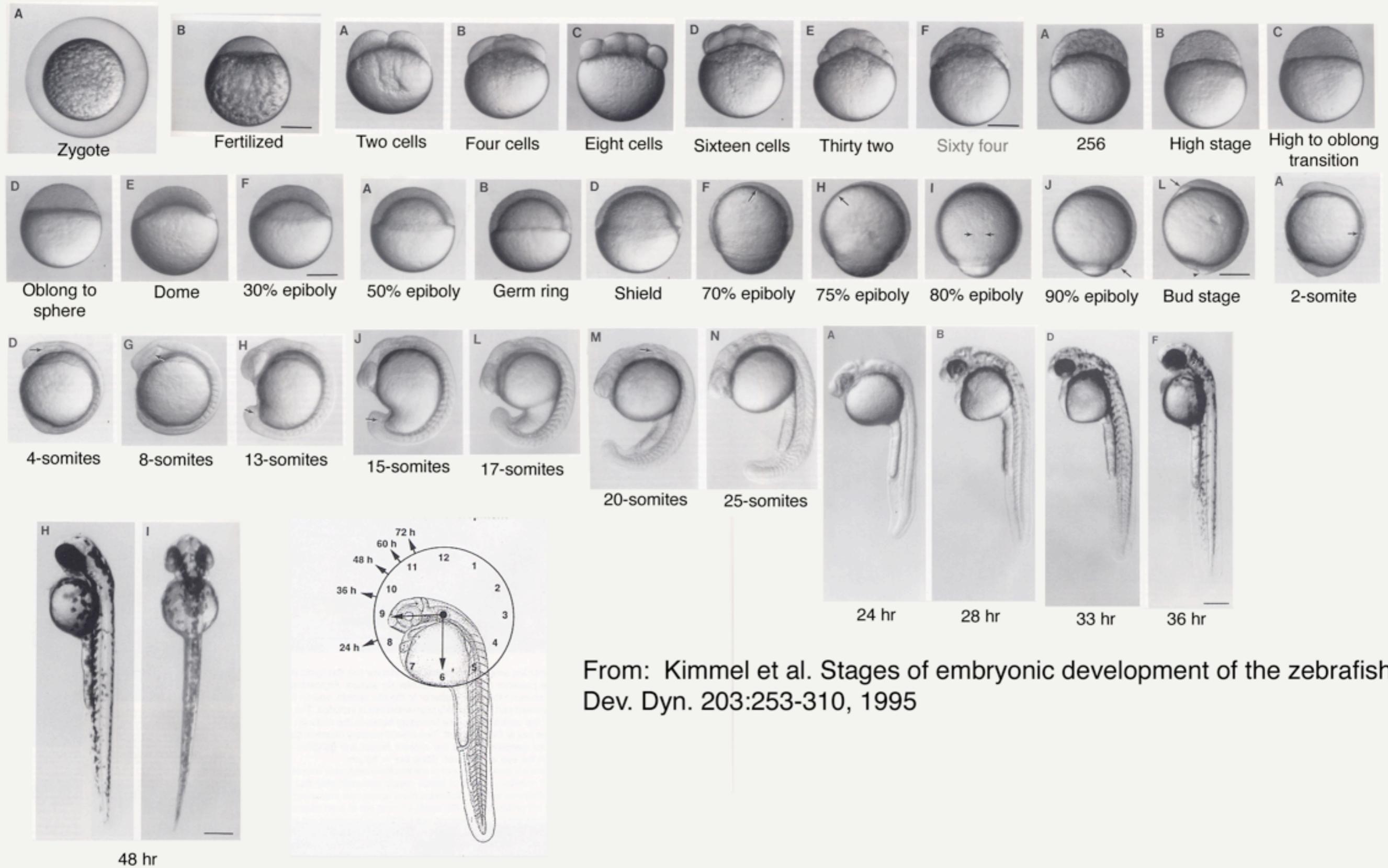
Drosophila

ADULT

prothoracic disc
 wing disc
 haltere disc
 leg disc
 abdominal histoblast
 genital disc



Zebrafish



From: Kimmel et al. Stages of embryonic development of the zebrafish
Dev. Dyn. 203:253-310, 1995

Mus musculus



Mus musculus

- C57Bl6, BALBc
- Transgenic
- Inbred
- Normal chow vs. High Fat Diet (HFD)
- Immuno-compromised, SCID
- Knock-out, knock-in, flox



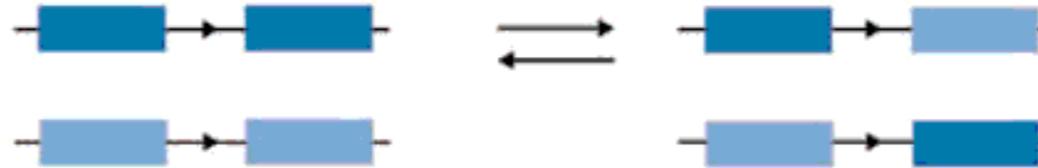
(A) **Excision-*cis*** place-
ment of *loxP* sites in
same directional
orientation.



(B) **Inversion-*cis*** place-
ment of *loxP* sites in
opposite directional
orientation.

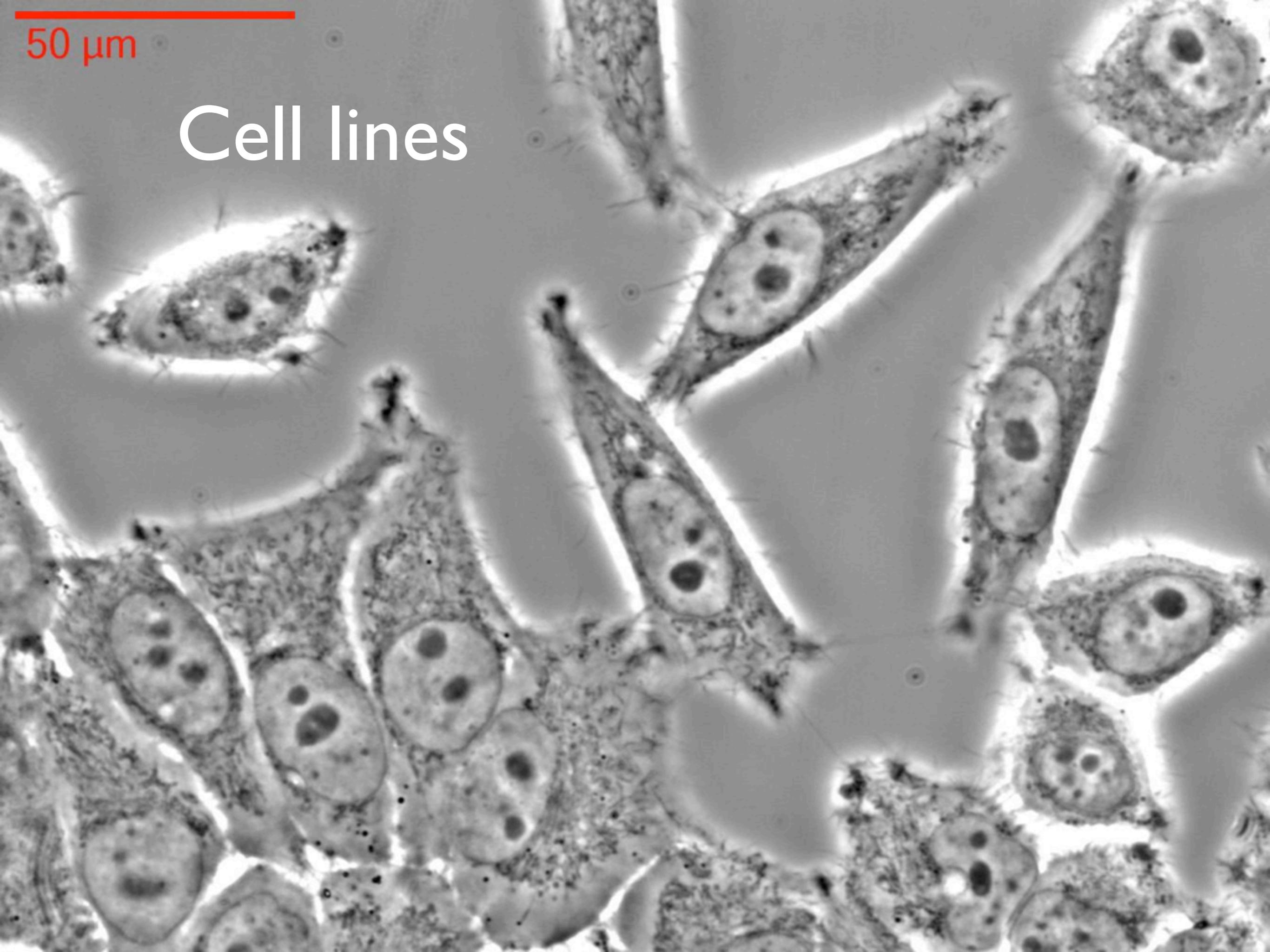


(C) **Translocation-*trans***
placement of *loxP* sites



50 µm

Cell lines



Cell lines

Human cell lines

- HeLa (cervical cancer)
- LnCap (prostate cancer)
- MCF-7 (breast cancer)
- THP-1 (acute myeloid leukemia)
- U87 (glioblastoma)
- GM12878 (lymphoblasts)
- K562 (chronic myelogenous leukemia)
- HepG2 (liver carcinoma)
- HUVEC (human umbilical vein endothelial cells)

Rat cell lines

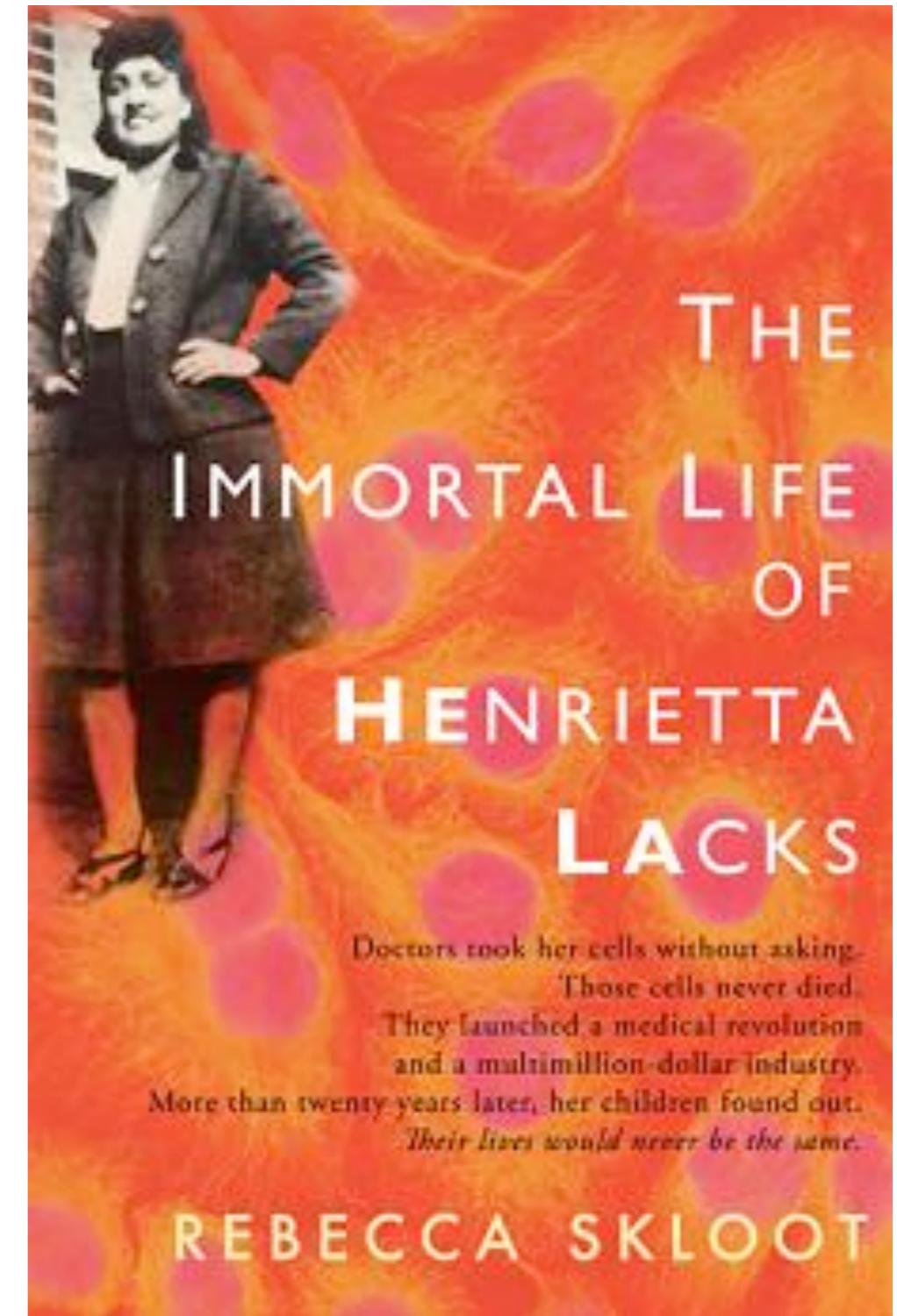
- GH3 (pituitary tumor)
- PC12 (pheochromocytoma)

Mouse cell lines

- MC3T3 (embryonic calvarium)
- 3T3 (fibroblasts)

Plant cell lines

- Tobacco BY-2 cells



Homo sapiens





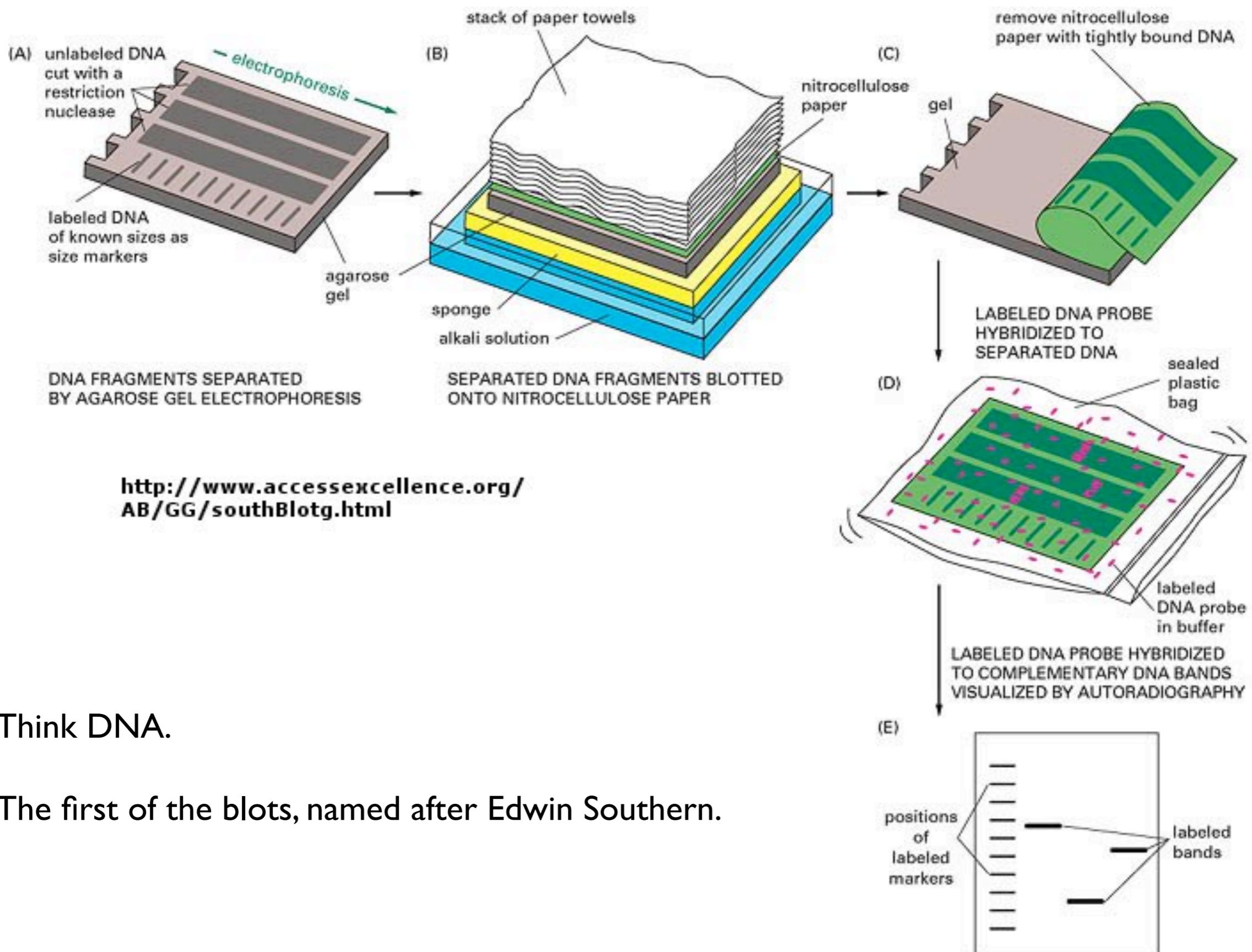
ManBearPig



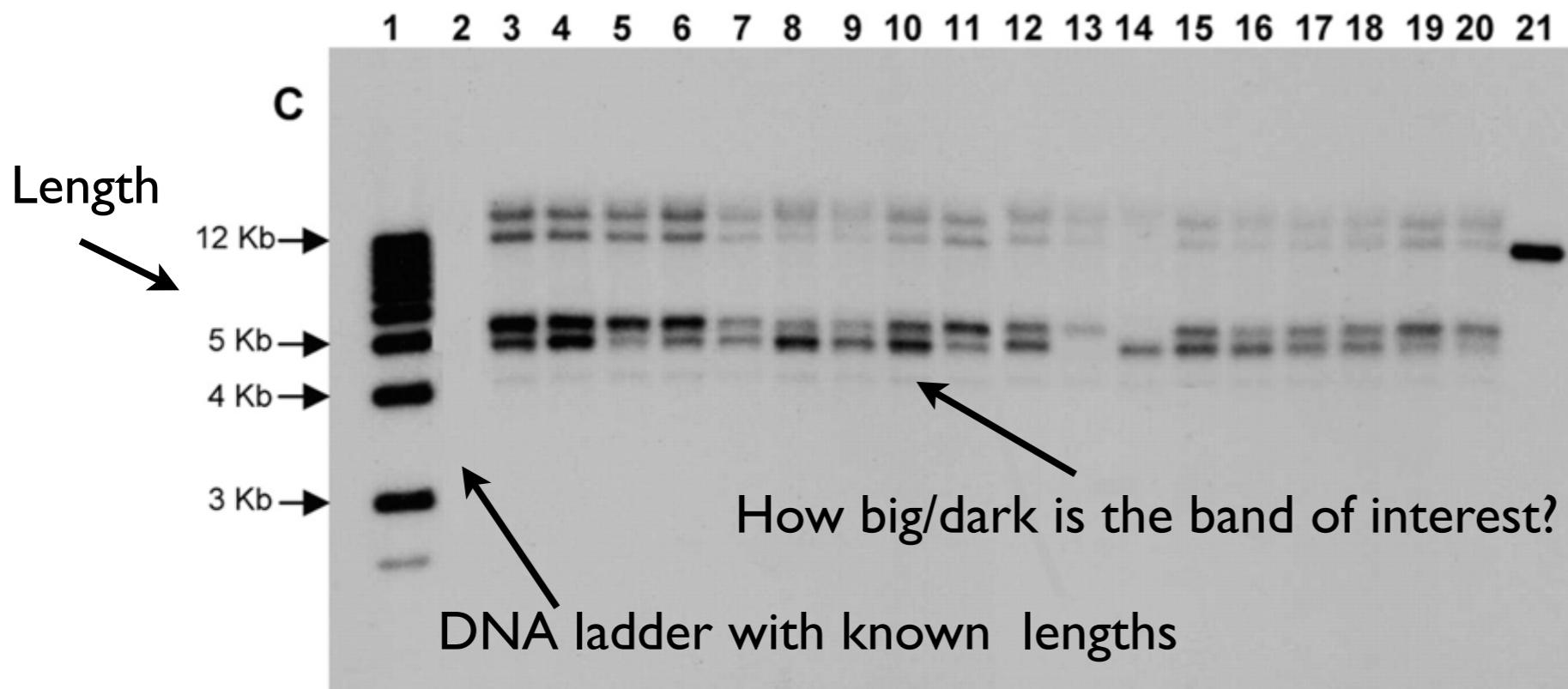
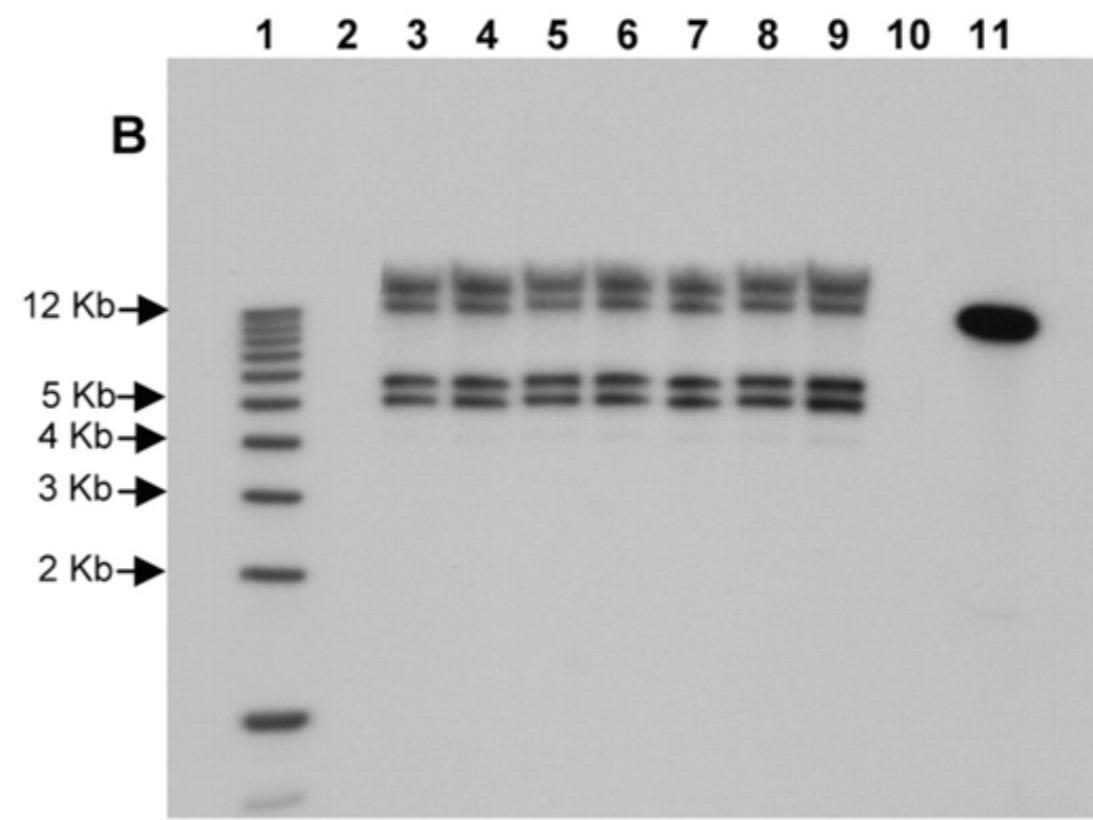
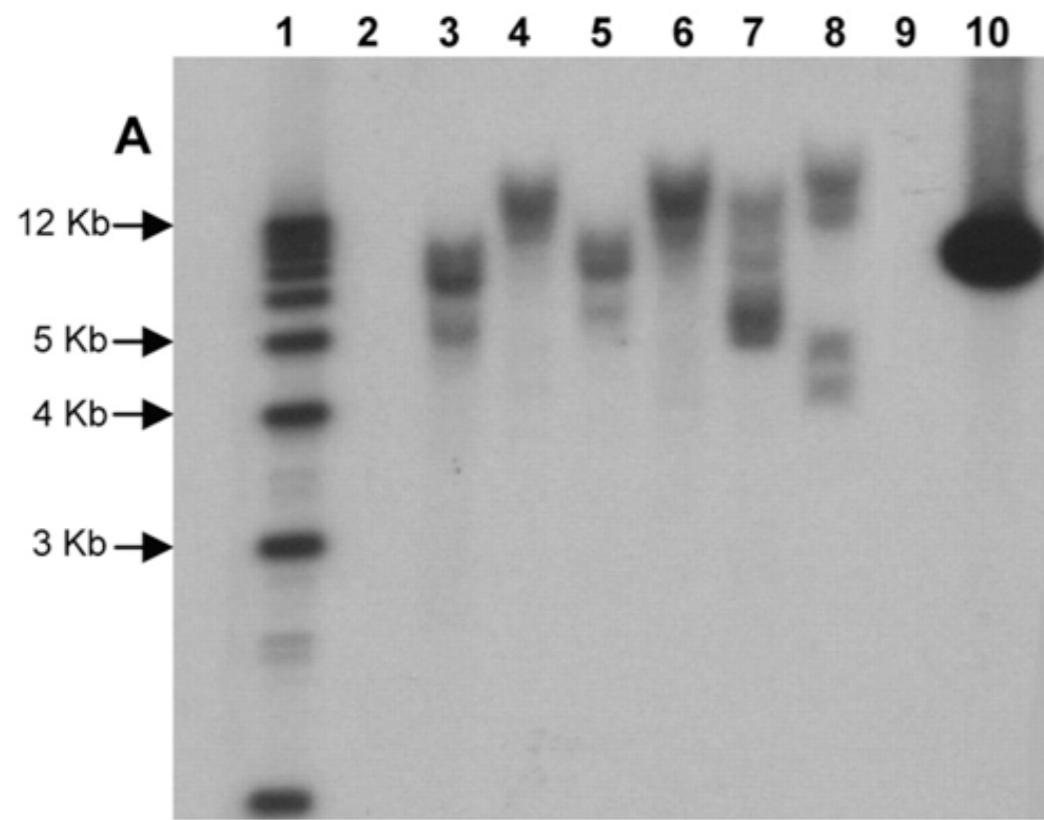
“To measure is to know.”

– Lord Kelvin (Sir William Thomson)

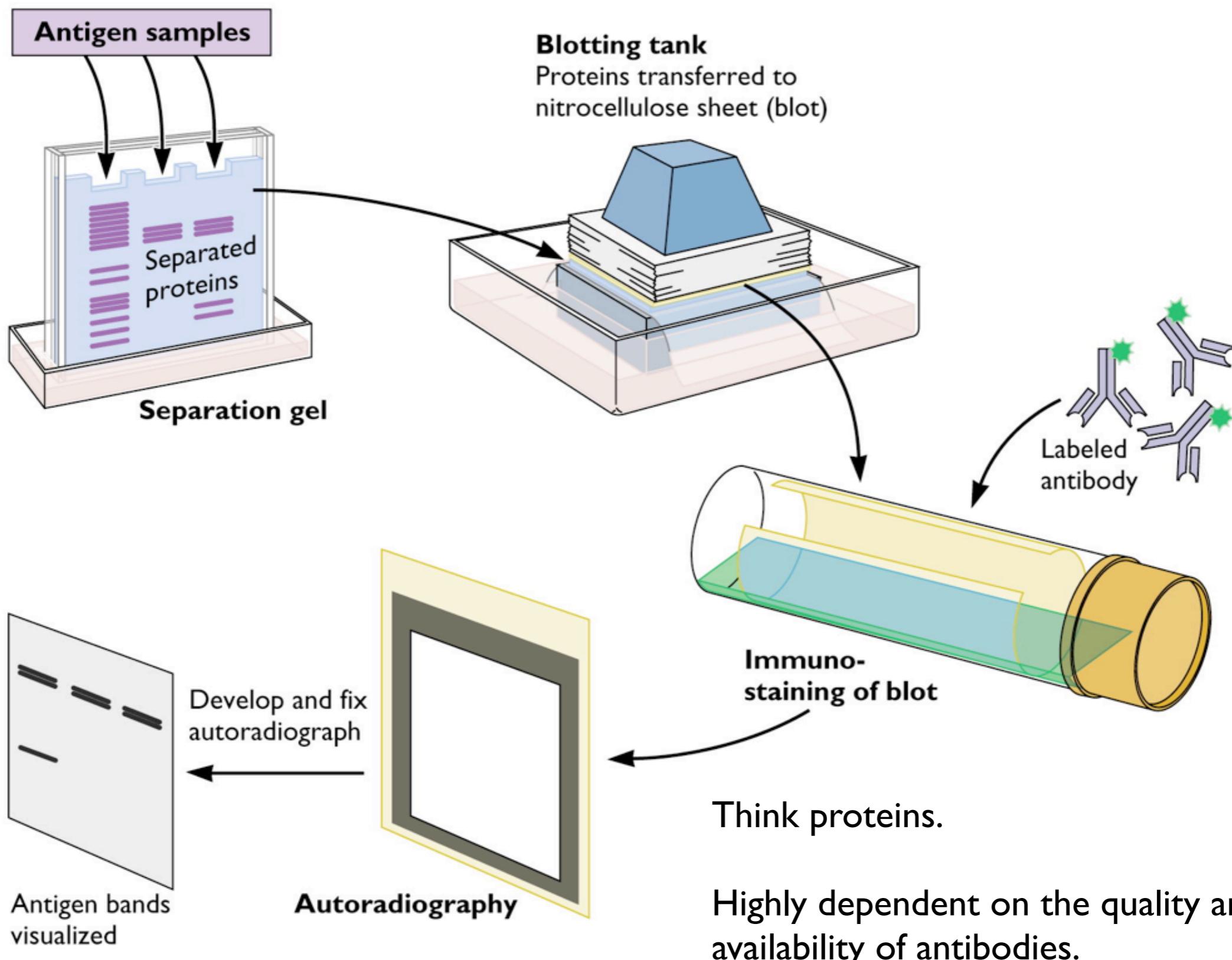
Southern blot



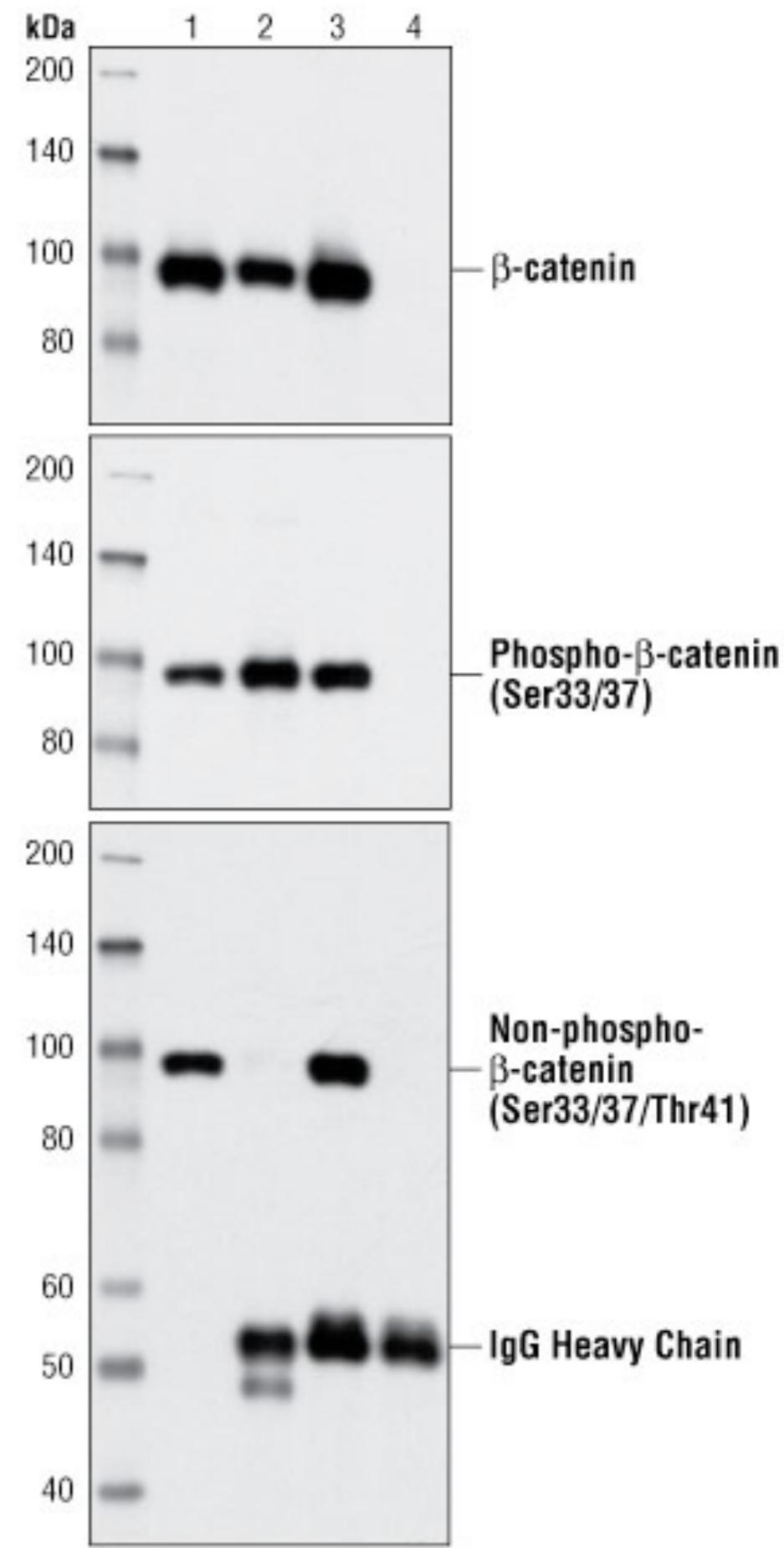
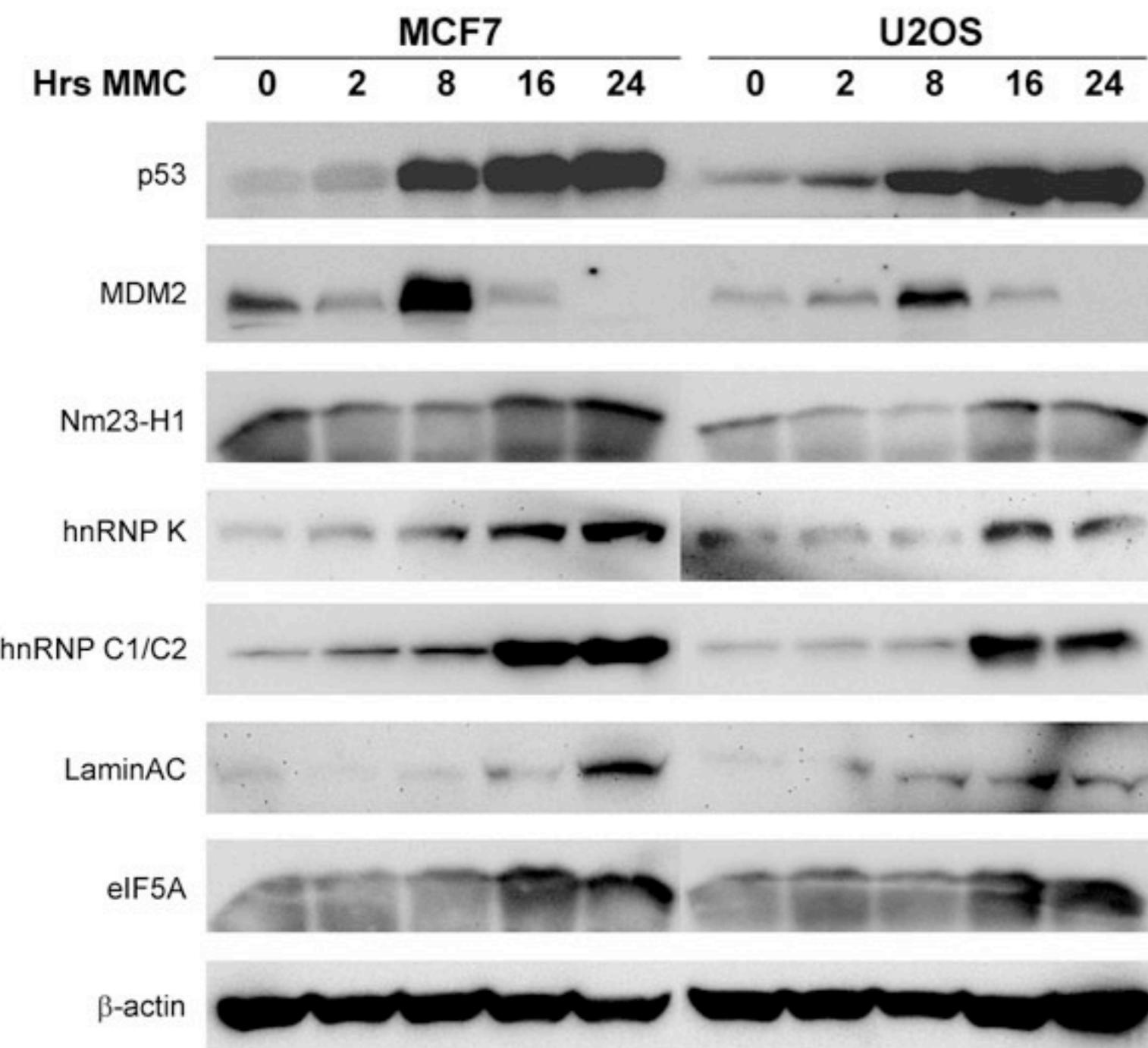
Southern blot



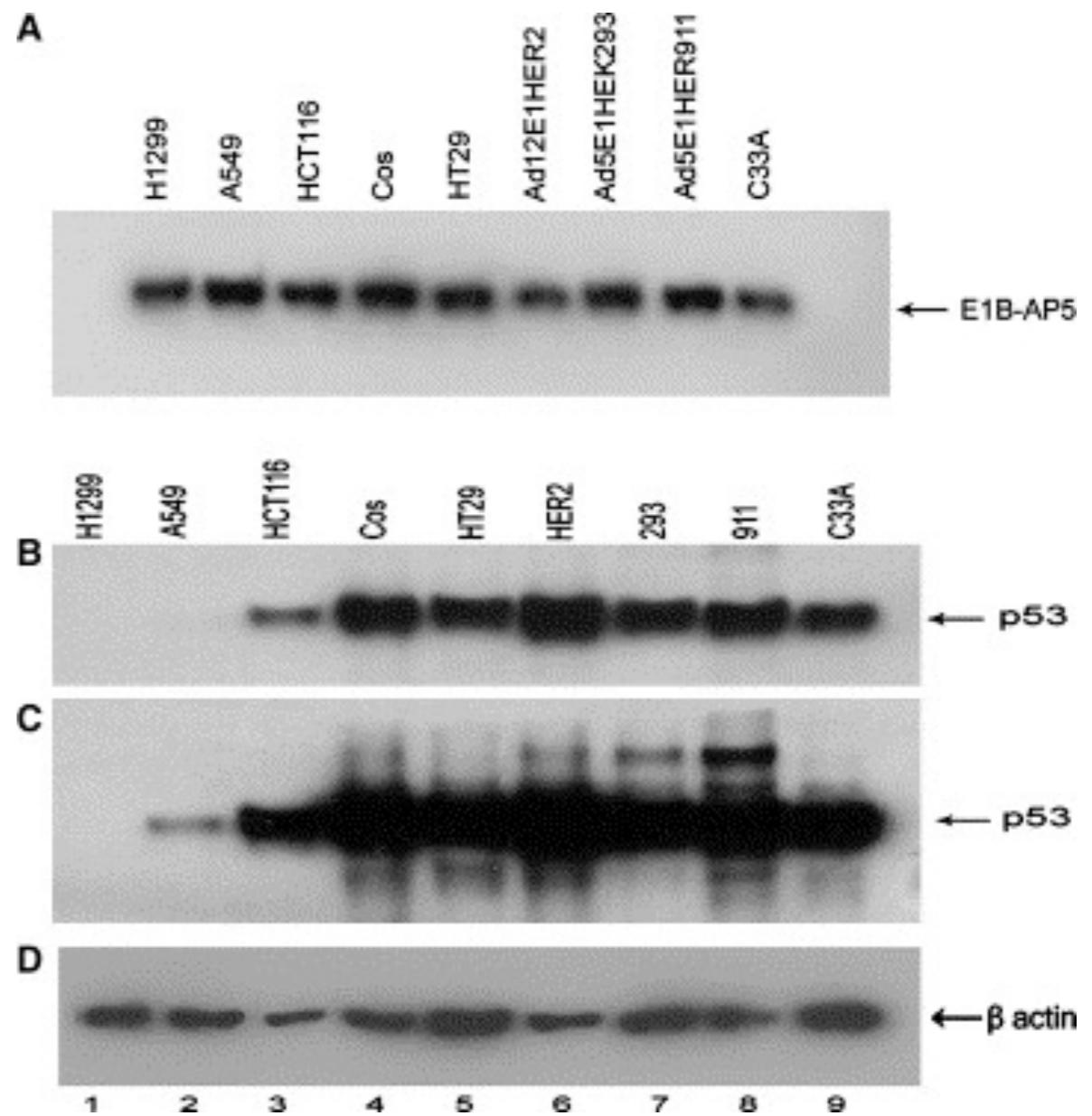
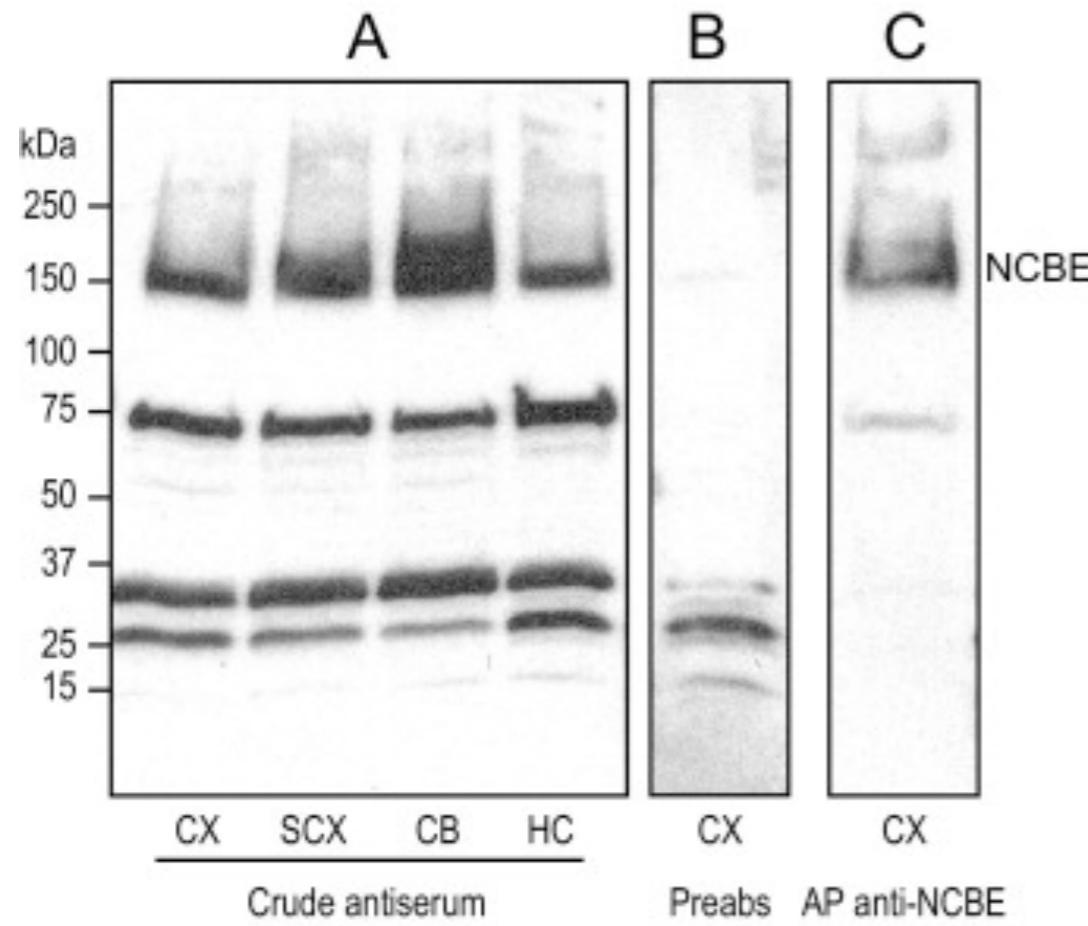
Western blot



Western blot



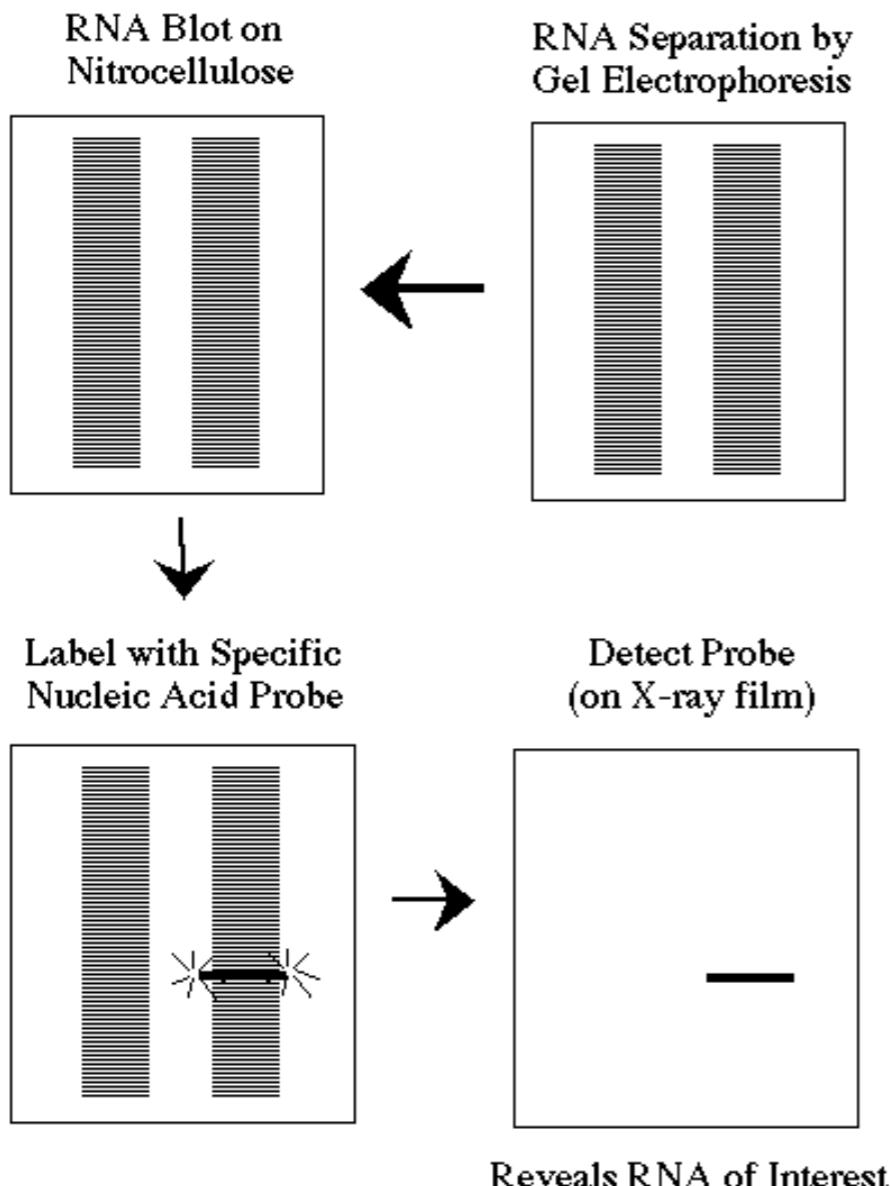
Western blot



You should be able to read text through the bands on the film.

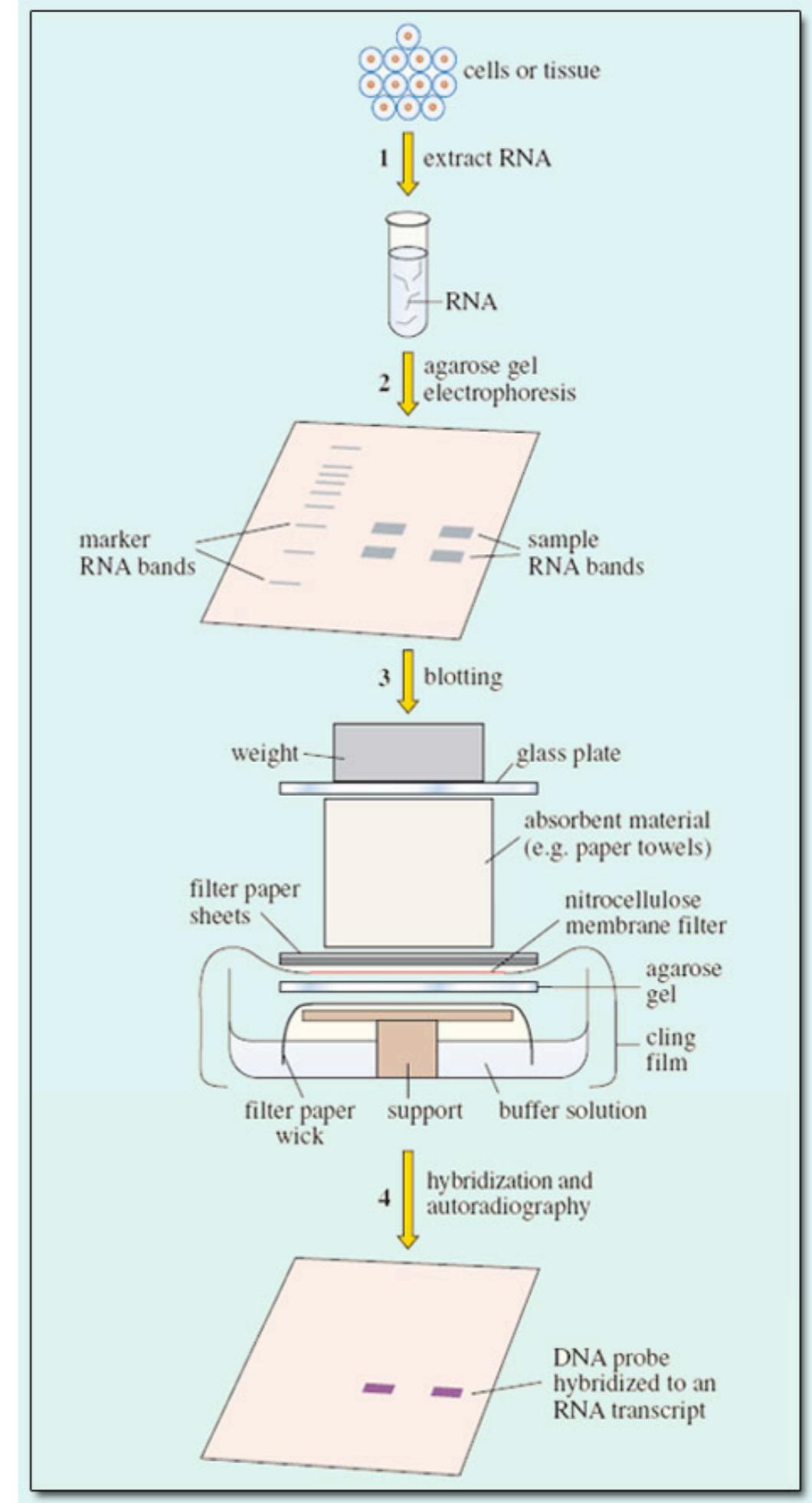
You should be able to read text through the bands on the film.

Northern blot

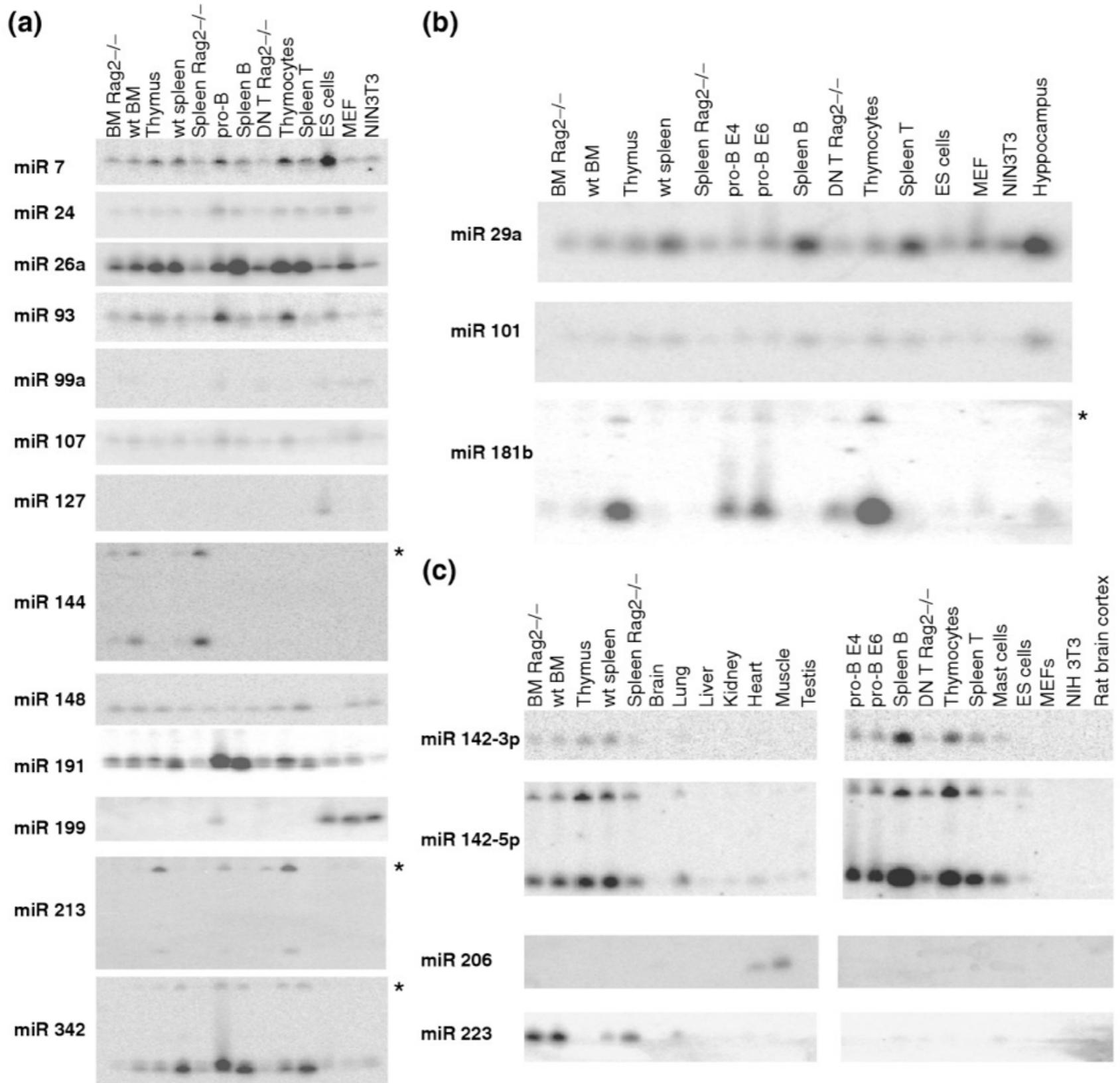


Think RNA.

Like Southern blotting, but closer to Santa.



Northern blot



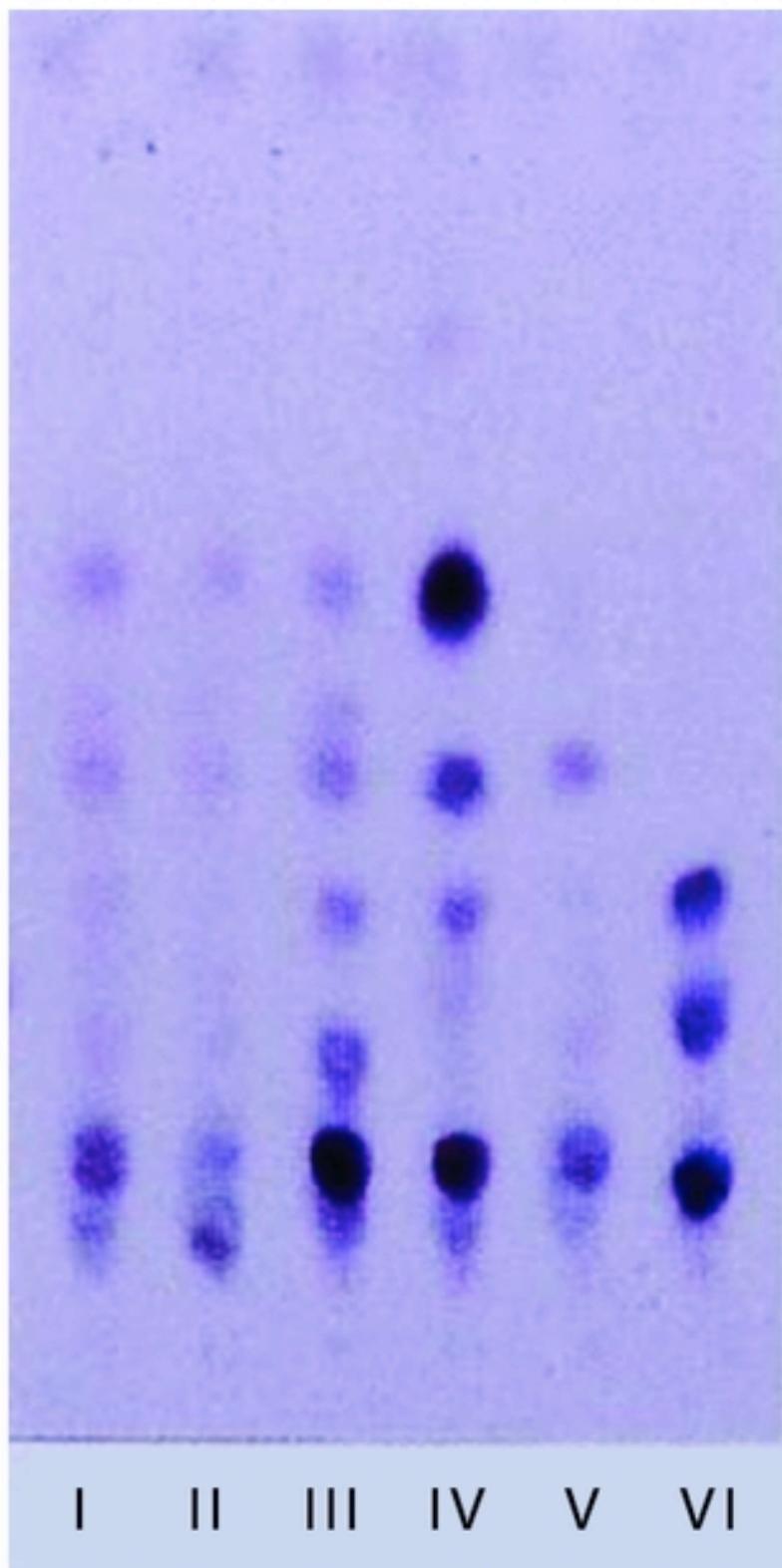
Eastern blot

Think PTMs.

Like Western blotting, but probe for lipids, glycoproteins, and other molecular modifiers of interest.

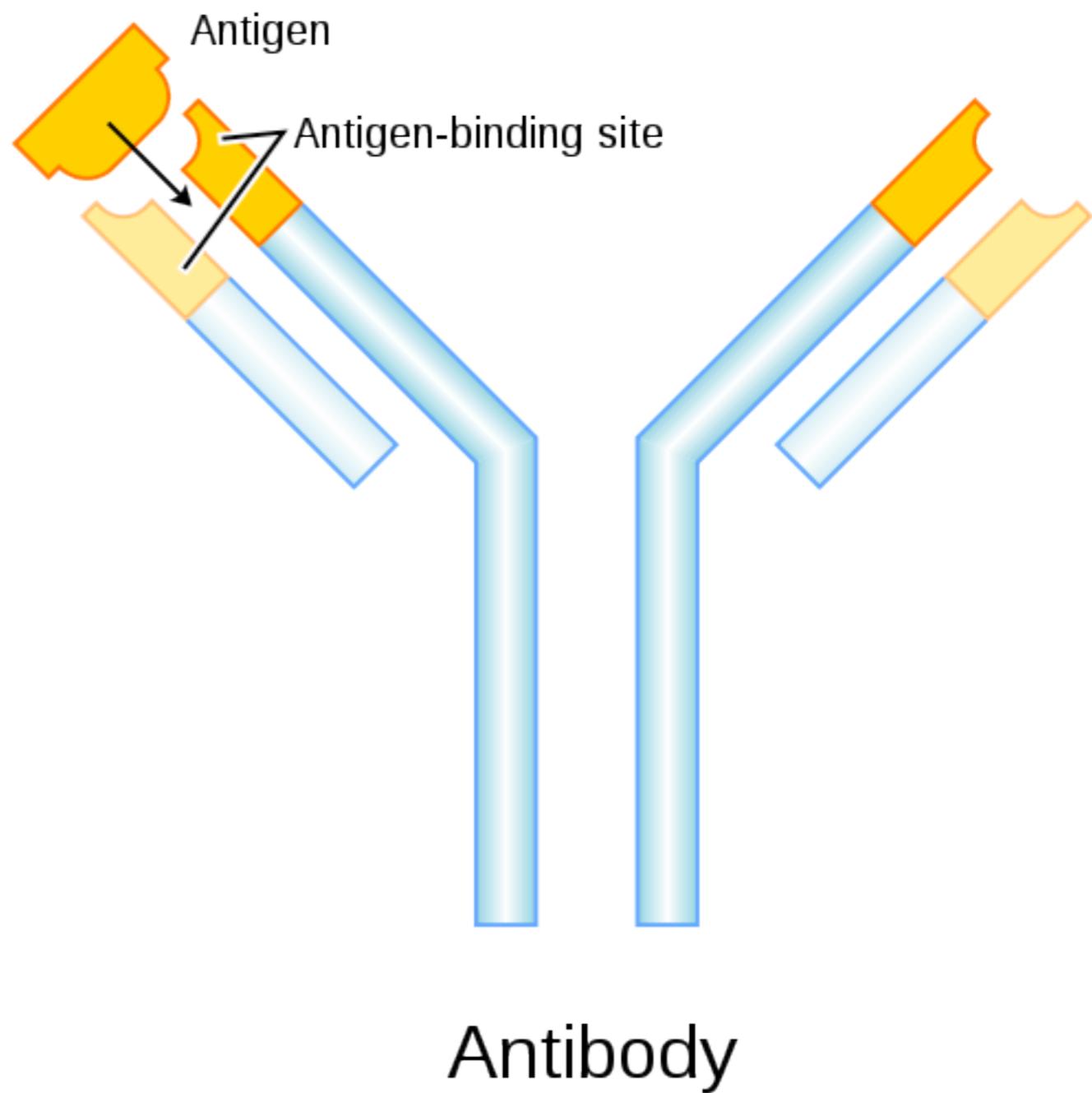
Not really substantially distinct from Western blotting, but we like the name.

Ginsenoside Rg₁ →
Ginsenoside Re →
Ginsenoside Rd →
Ginsenoside Rc →
Ginsenoside Rb₁ →



Antibodies (Abs)

- Made by B cells
- Bind to antigen, which can be proteins, peptides, viruses, molecular groups...
- In theory, specific for a particular epitope (the key to the Ab lock)
- In practice, can be cross-reactive
- Important in the natural immune system, but also in scientific assays
- Targeted Abs made by injecting the antigen-of-interest into rabbit, mouse, donkey, etc. Abs are then purified from serum (polyclonal), or clonal Ab-producing cells are isolated to produce monoclonal Abs.
- Can be conjugated to fluorescent molecules, tags, beads, etc.

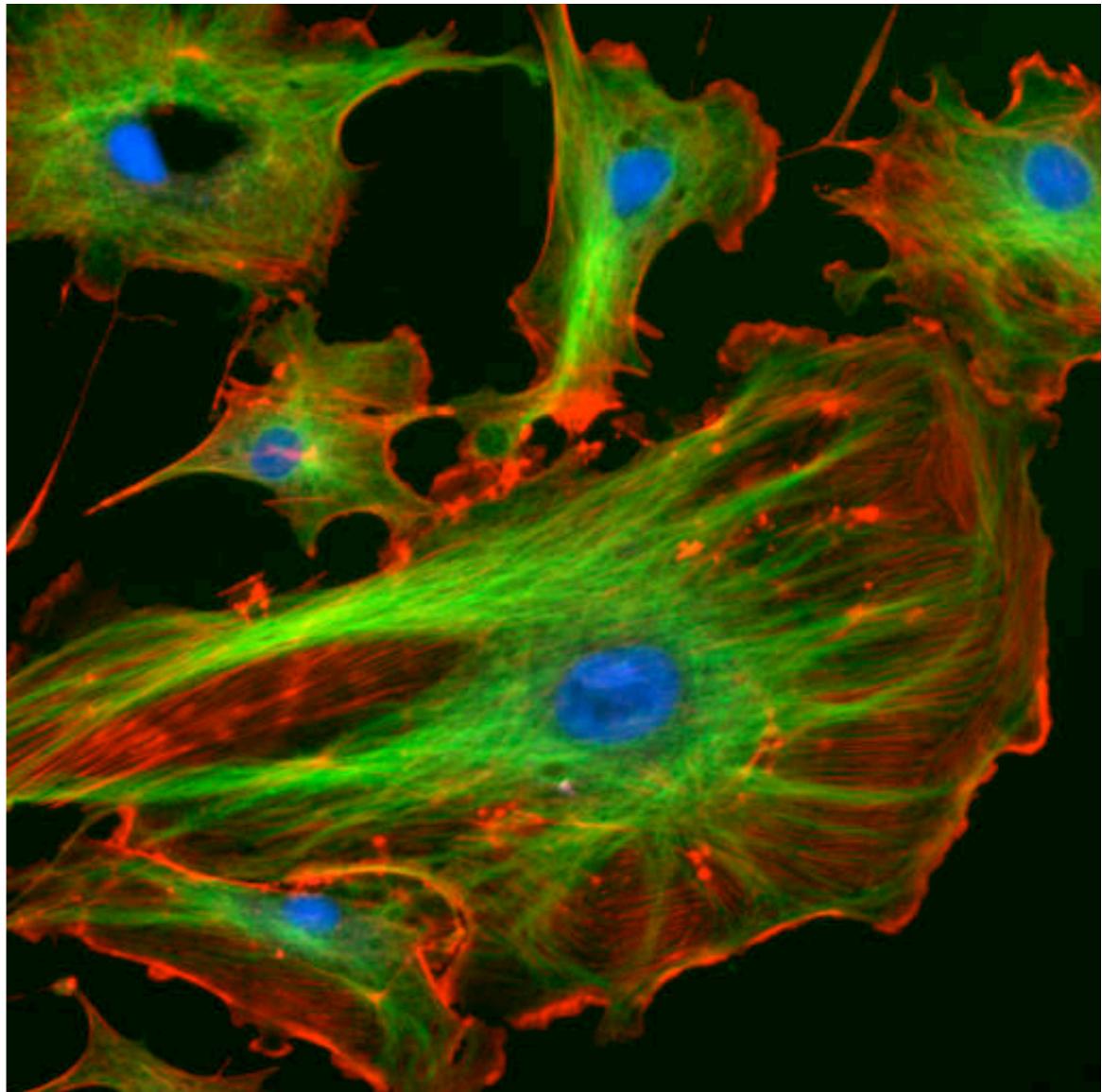


Antibodies (Abs)

- Five types; IgG (immunoglobulin G) is most common in serum, IgM is membrane-bound, IgE is involved in allergic reactions

Used in...

- Immunoprecipitation (IP; selection and isolation of proteins of interest using Abs)
- Chromatin IP (ChIP), Co-IP
- Western blotting
- Flow cytometry
- Immunofluorescent staining, microscopy
- Histology staining
- Diagnostics (for example, anti-GAD65 to diagnose type I diabetes)
- Therapies (for example, anti-TNF α for rheumatoid arthritis)
- Almost everything.



For more B cells and Abs:
BGGN 225
Kees Murre

Immunoprecipitation

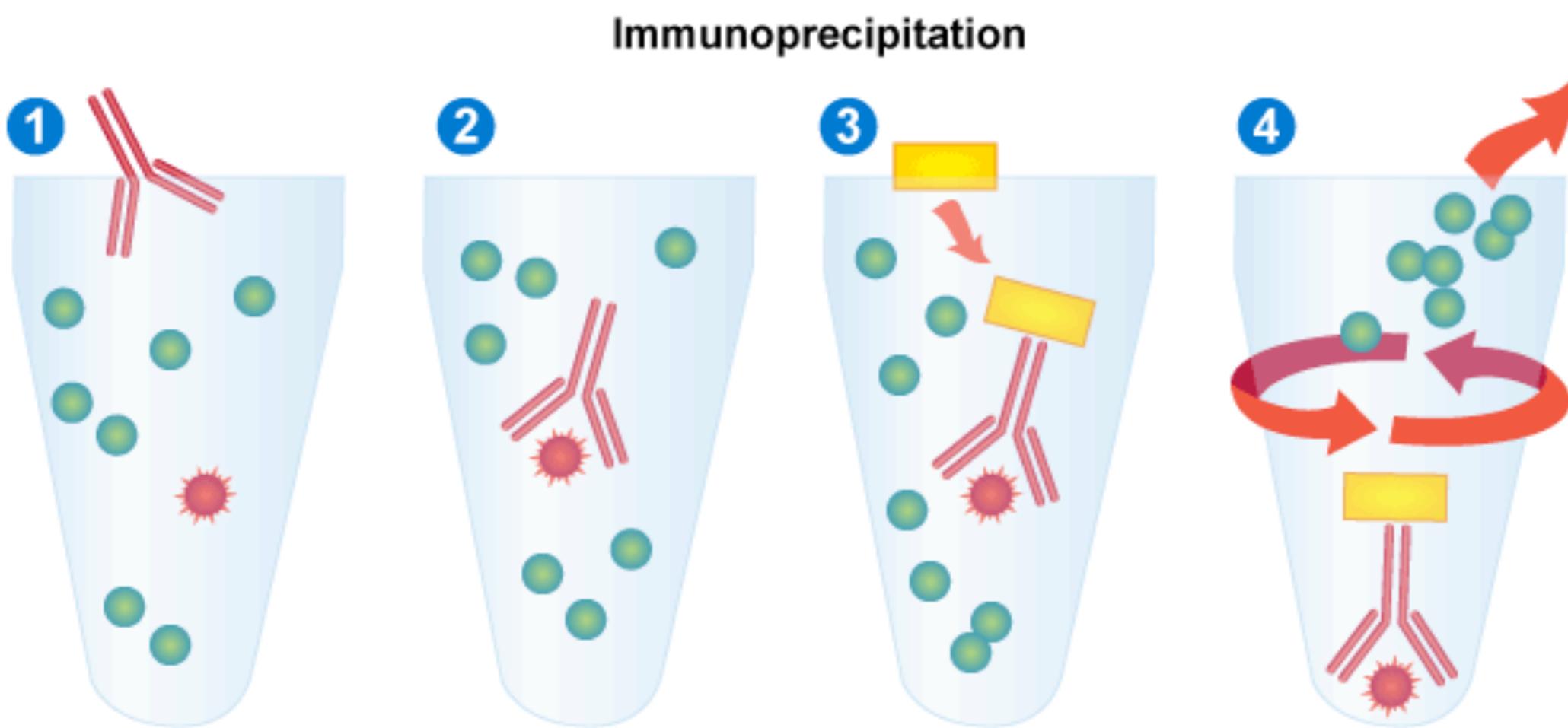
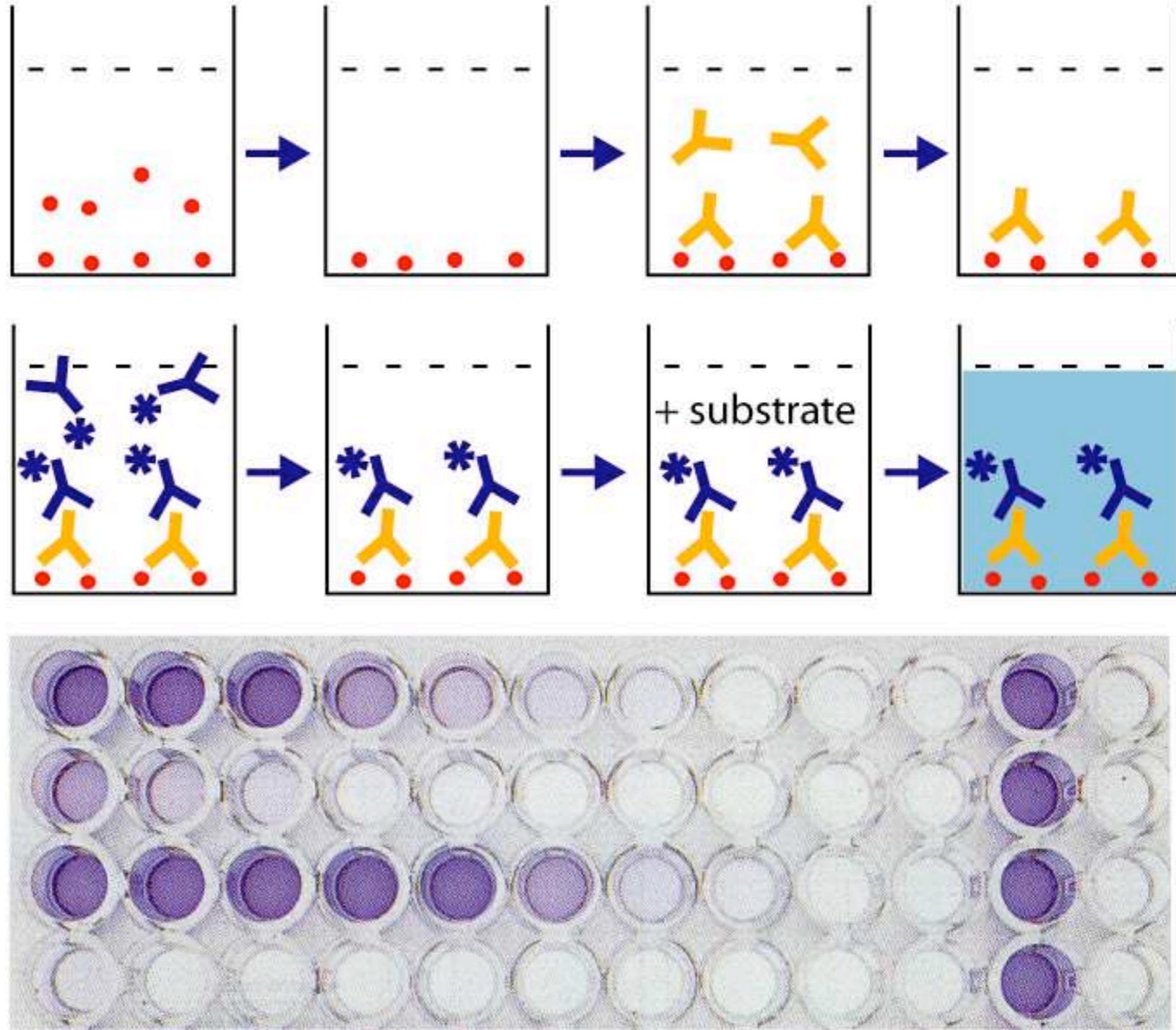


Diagram 1: Illustration of Immunoprecipitation process.

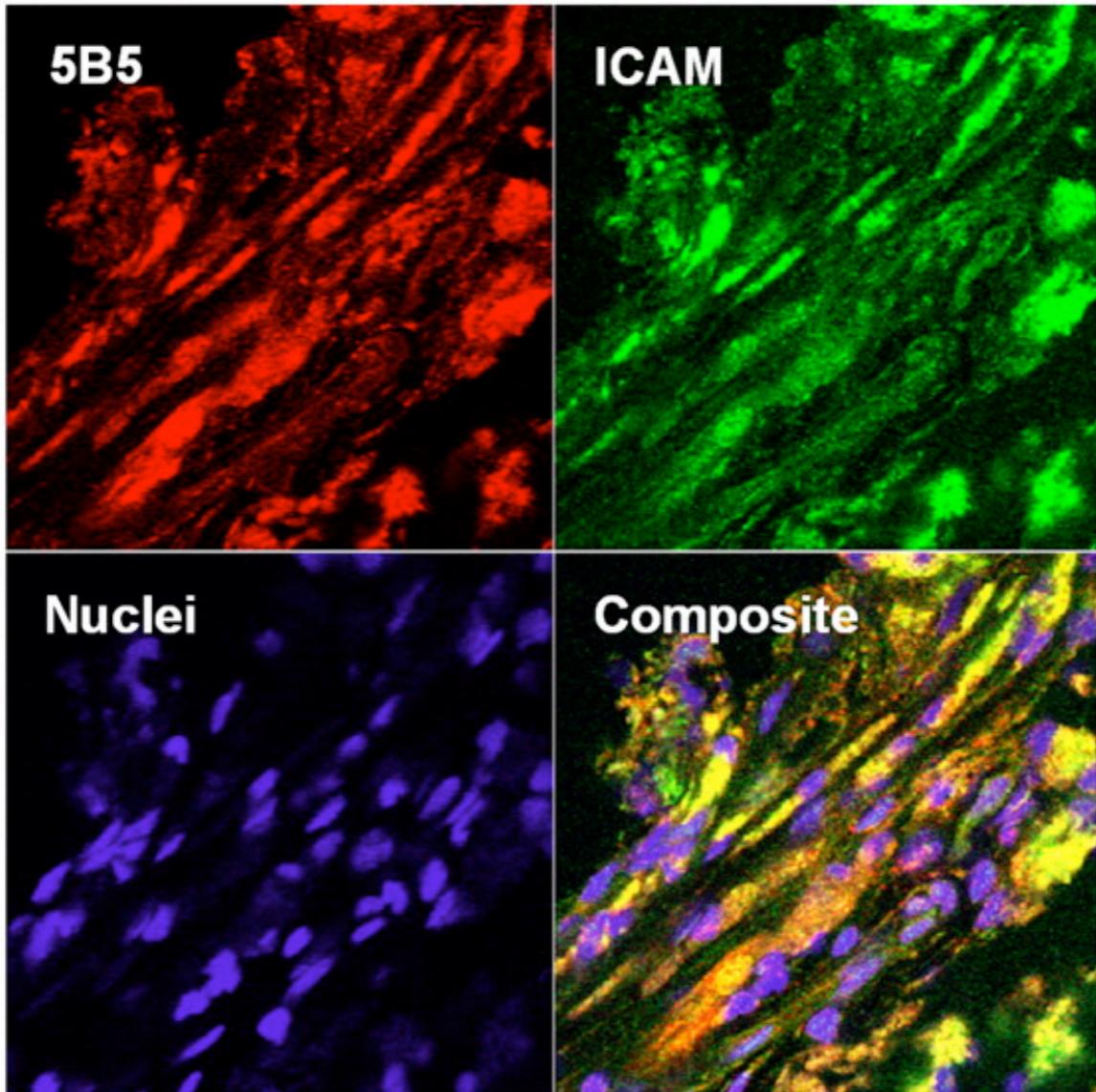
ELISA

- Enzyme-linked immunosorbant assay
- Sample with suspected antigen is immobilized on a plate
- Add antibody that binds to protein-of-interest
- Add an enzyme-linked secondary antibody that binds to primary antibody
- Add a substrate for the secondary antibody that emits signal (usually, turns a particular color) when in contact with the enzyme

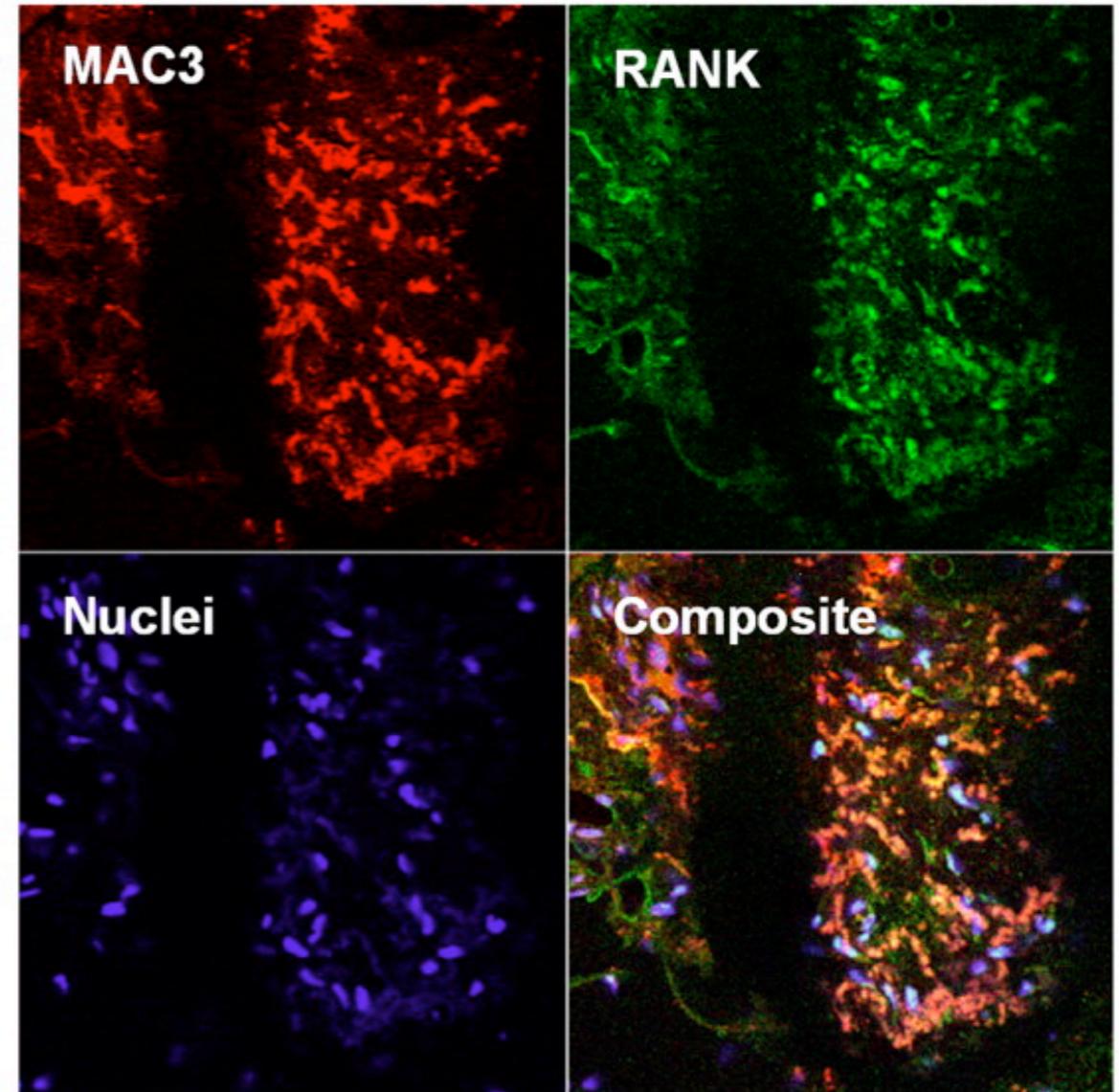


Fluorescence microscopy

1A.

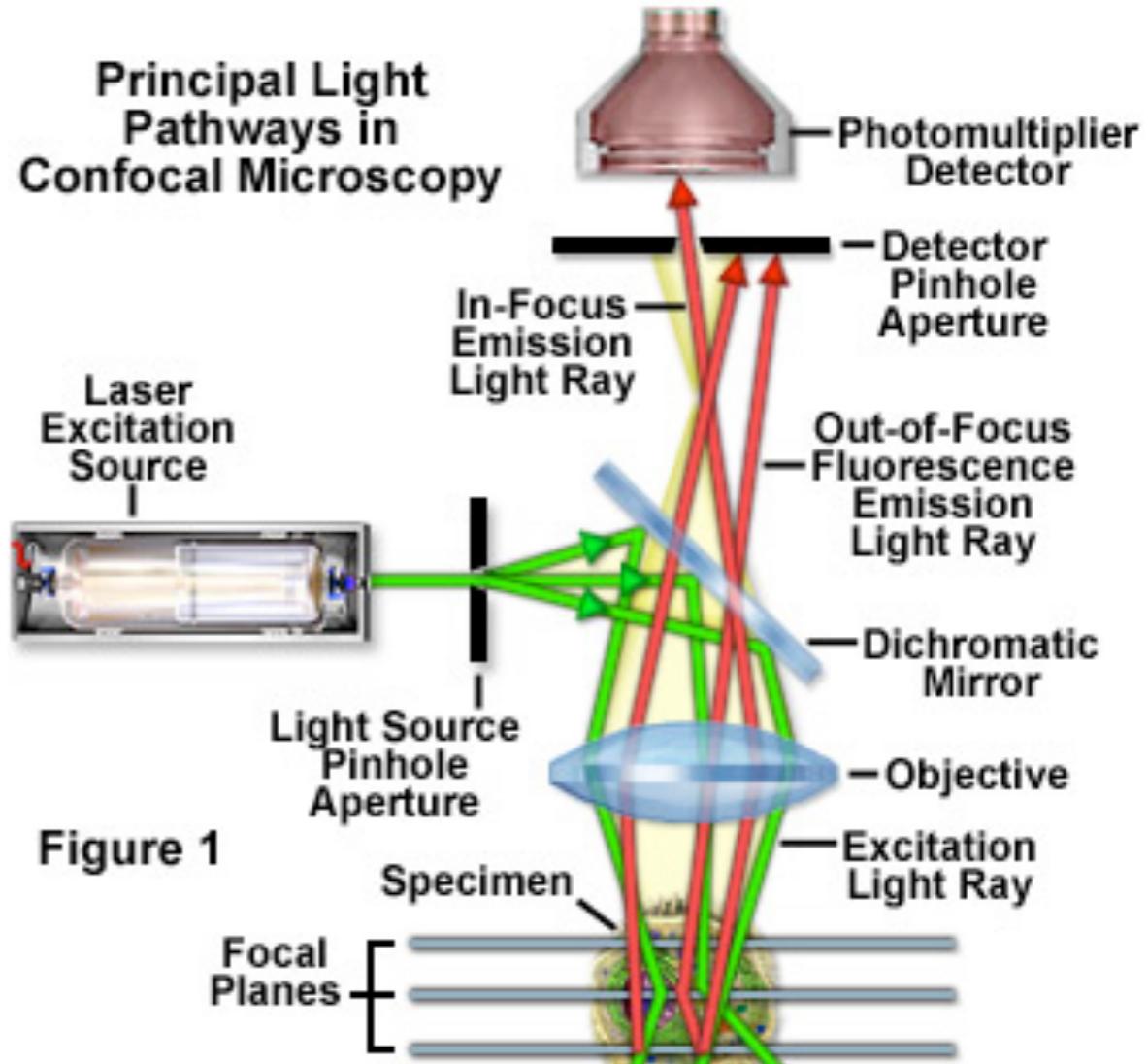


1B.



Fluorescence microscopy

- Immunofluorescence: stain samples with Abs for proteins of interest attached to fluorescent markers
- Additionally, more general reporter molecules are used: DAPI and Hoescht stains bind to DNA, so is used to label nuclei
- Or, green fluorescent protein (GFP) and similar constructs can be incorporated into genetically-modified cells. Resulting proteins fluoresce, and can be imaged
- Images can be separated by fluorescent marker and merged to show localization of entities of interest



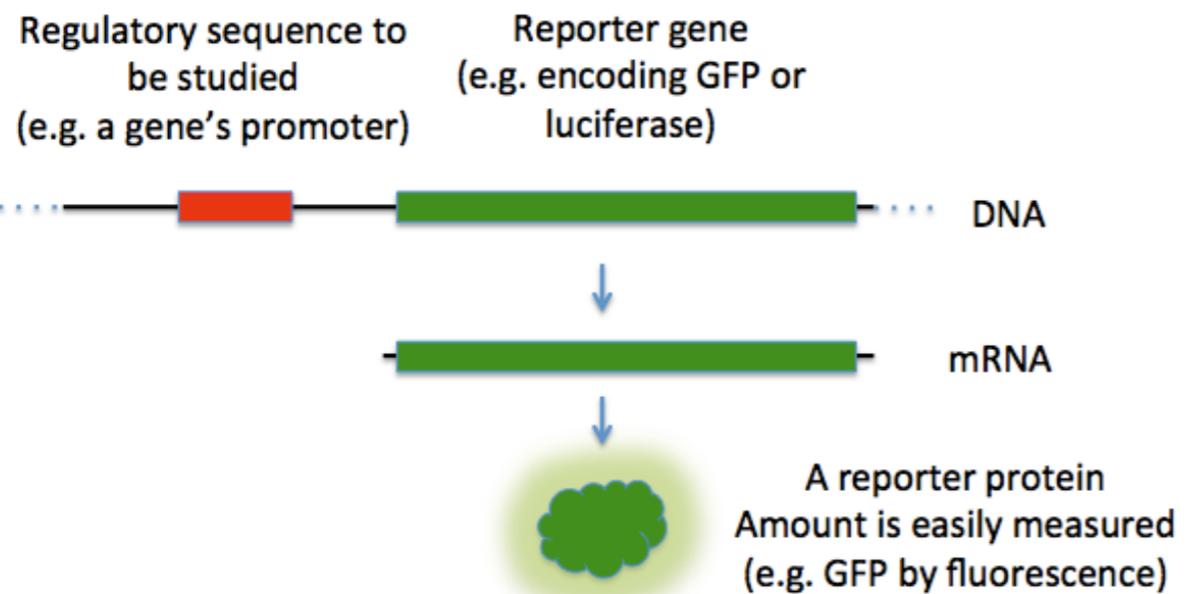
For more fluorescence:
BGGN 206
Roger Tsien, Scott Rifkin, Jeff Hasty

Cell culture

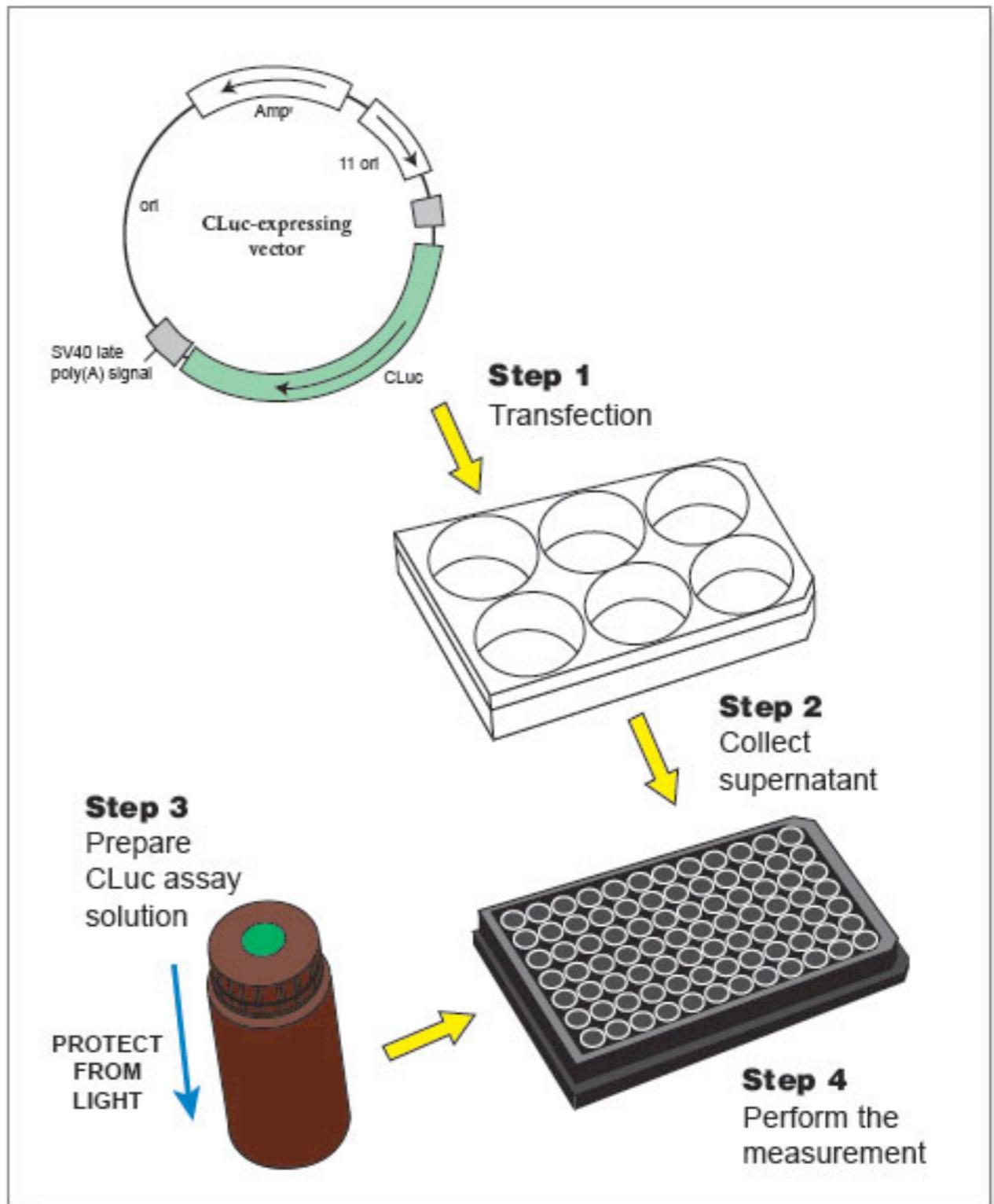
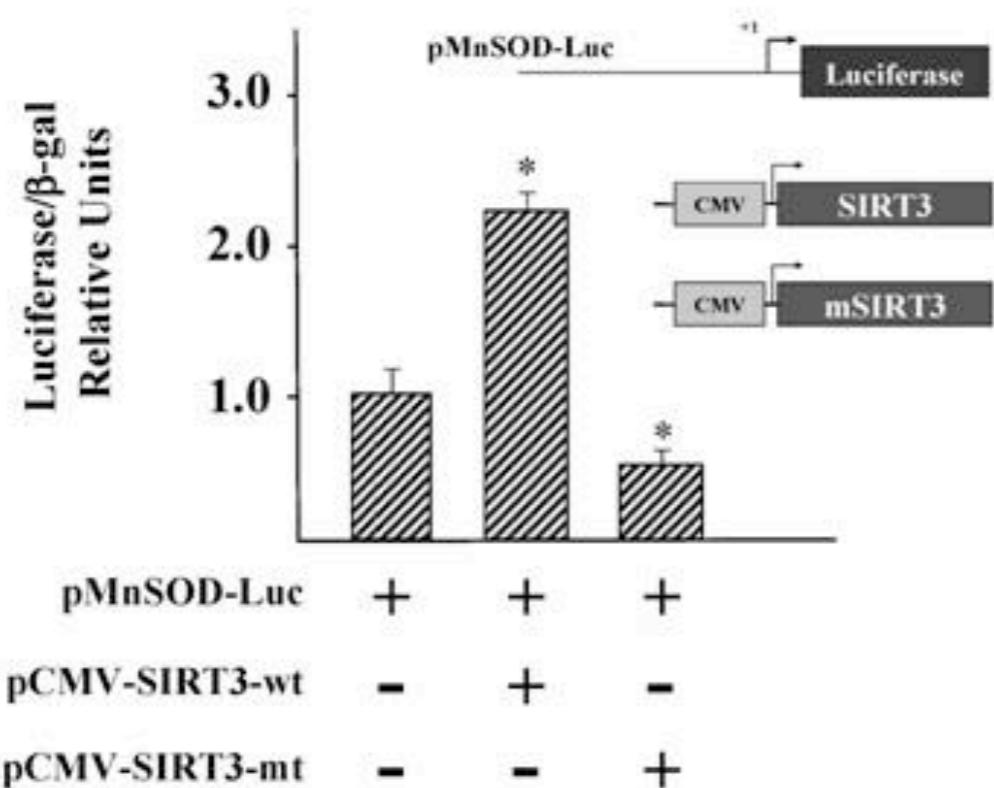
- Maintaining cells in a dish
- Can be direct from tissue (“primary cells”) or from cell lines
- Single cell types or a mix
- In suspension or adherent
- Incubators maintain temperature and air content for the cells (mammalian cells like to be at 37°C, 5% CO₂)
- Media types and components vary by cell type and context. Generally, you need a source of nutrients and a particular pH.
- Some common types of media: LB, agar, fetal bovine serum (FBS), RPMI
- Cells can be split and passaged (separated out into new containers)
- Watch out for contamination!



Reporter gene



B.



Transfection

Lipid-Mediated Transfection in Mammalian Cells

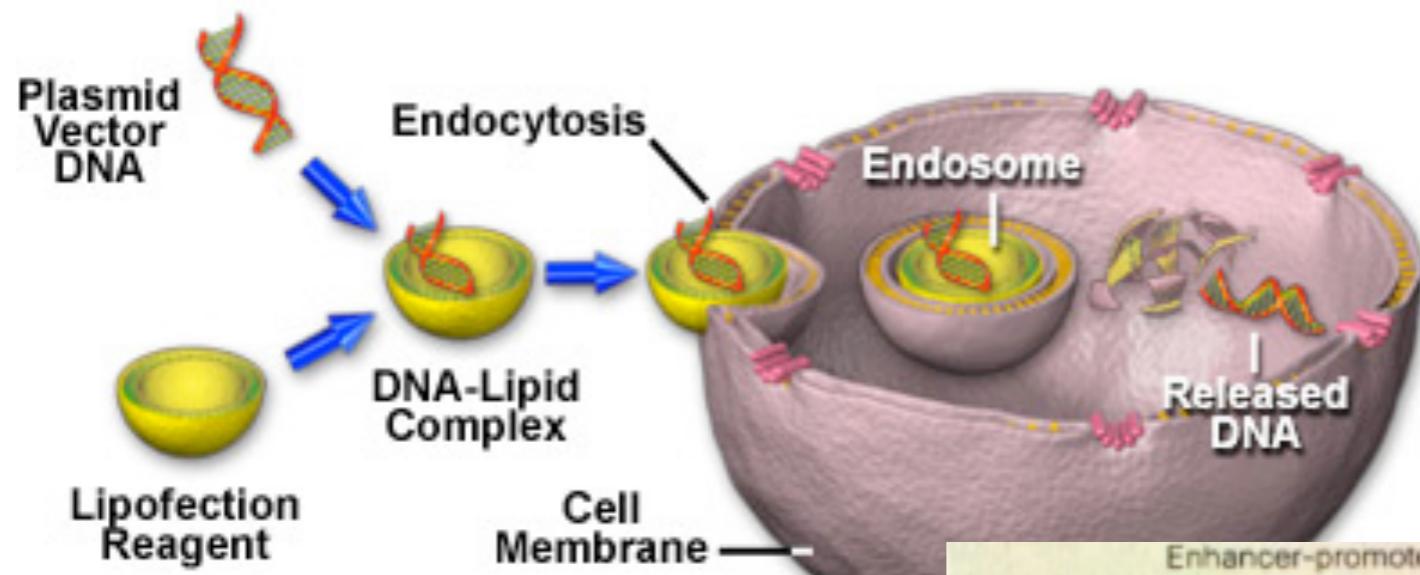
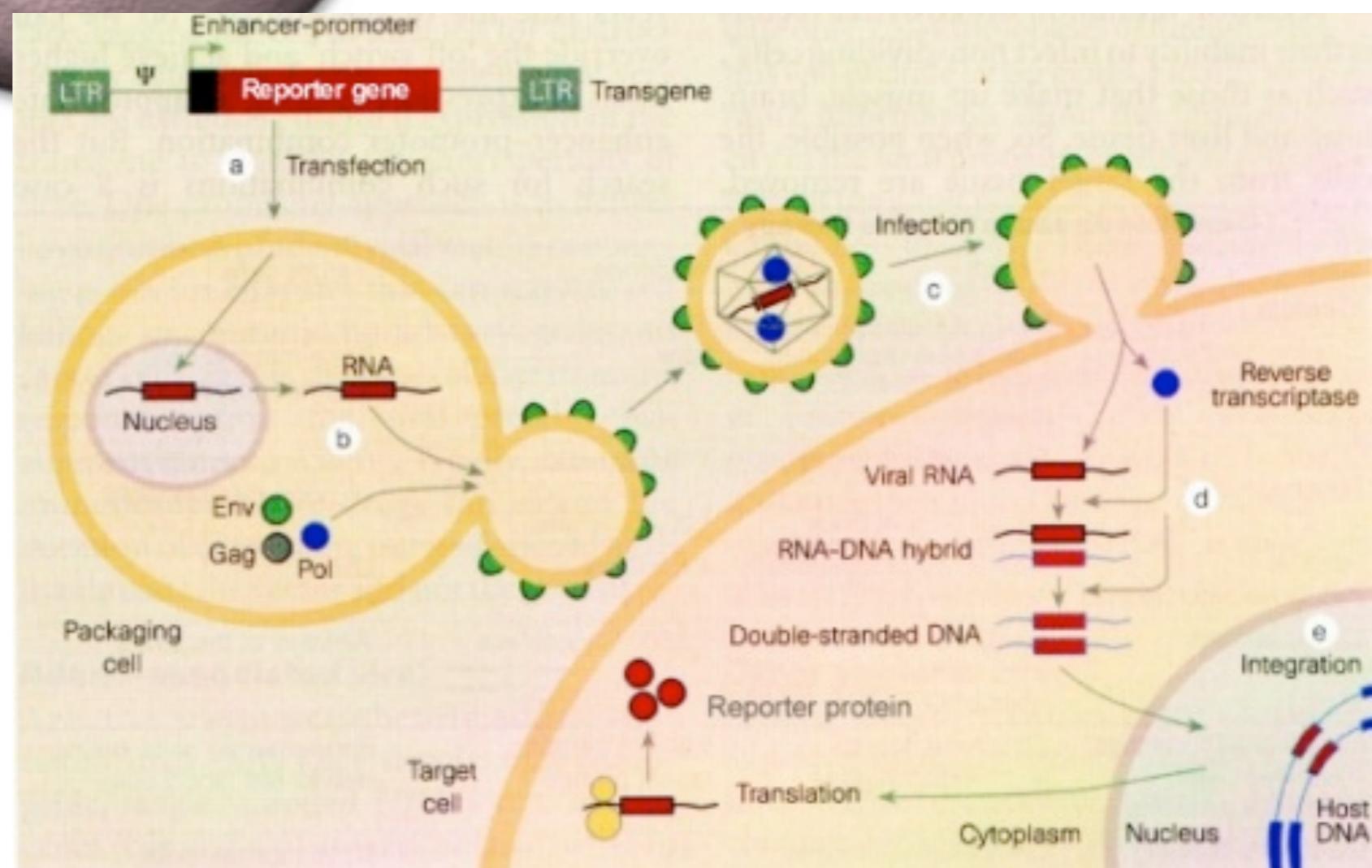
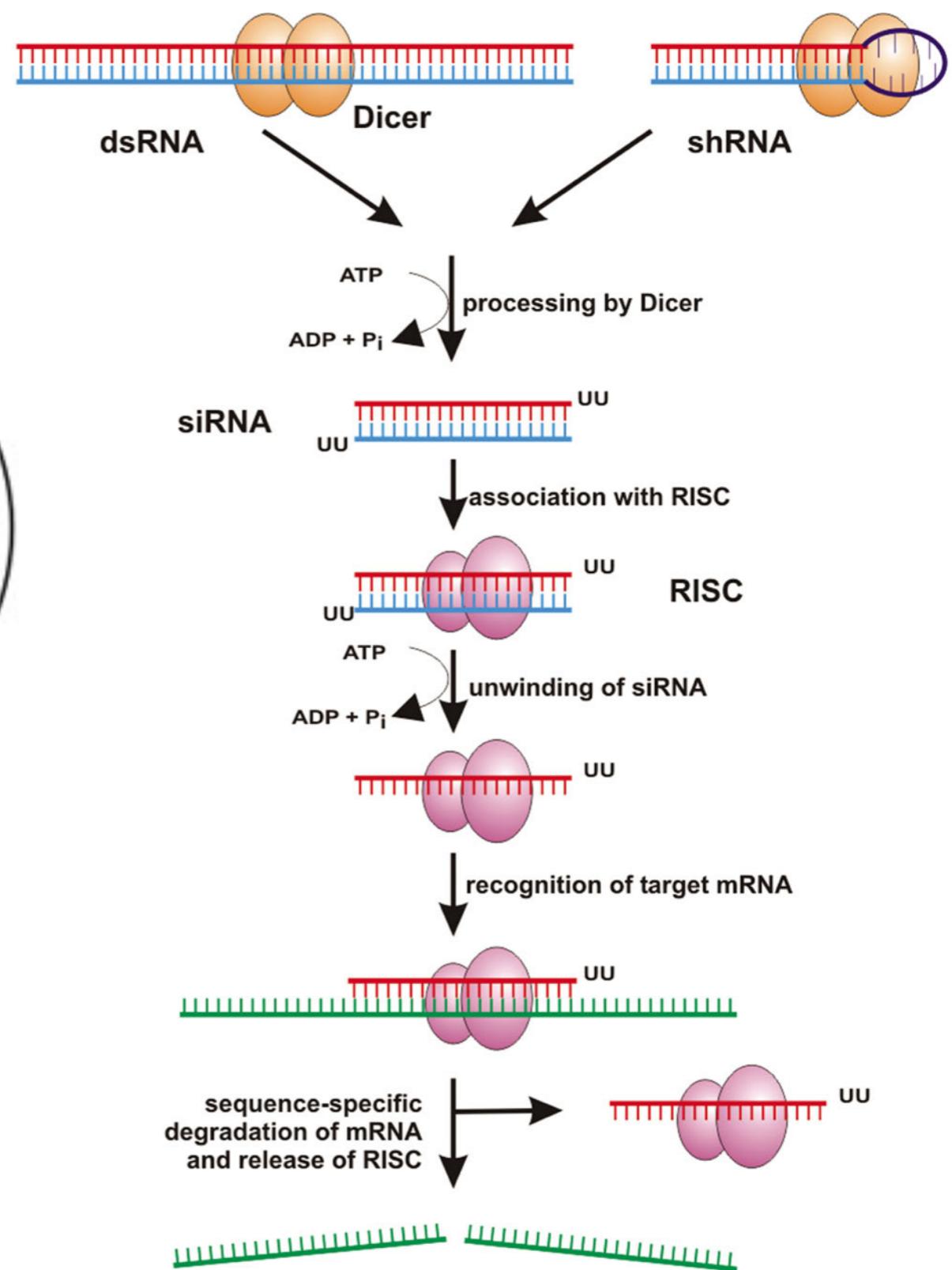
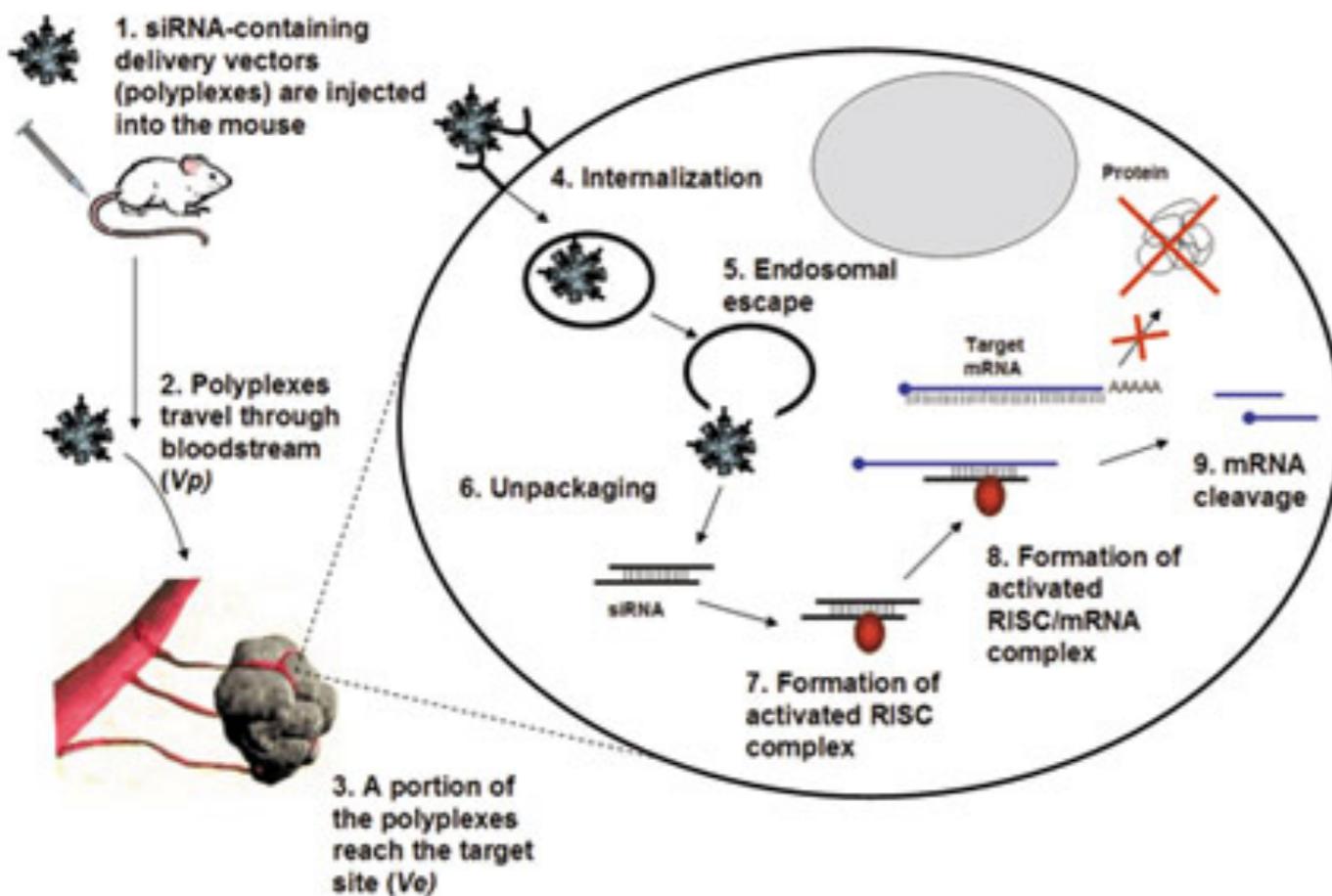


Figure 8



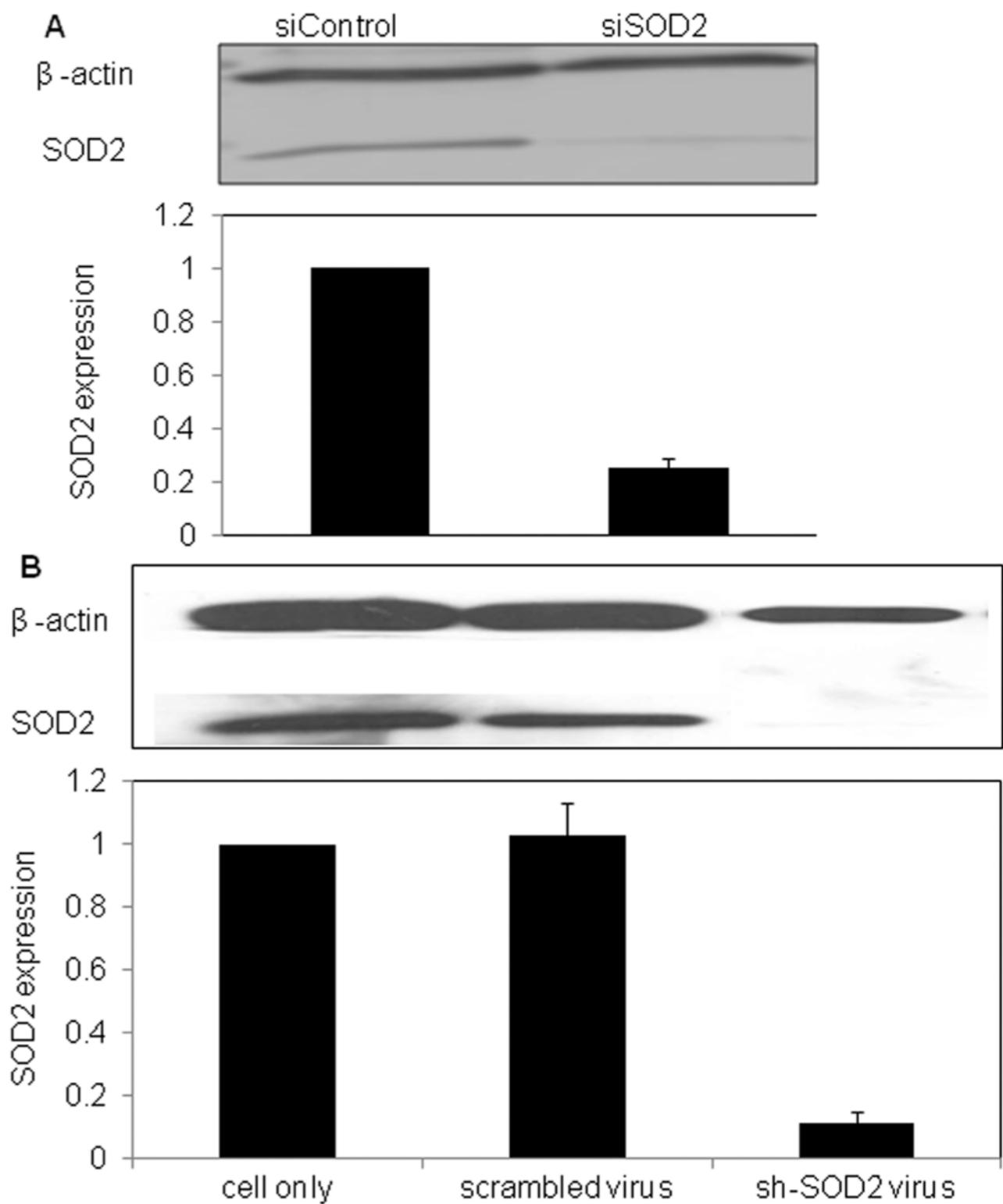
siRNA/shRNA



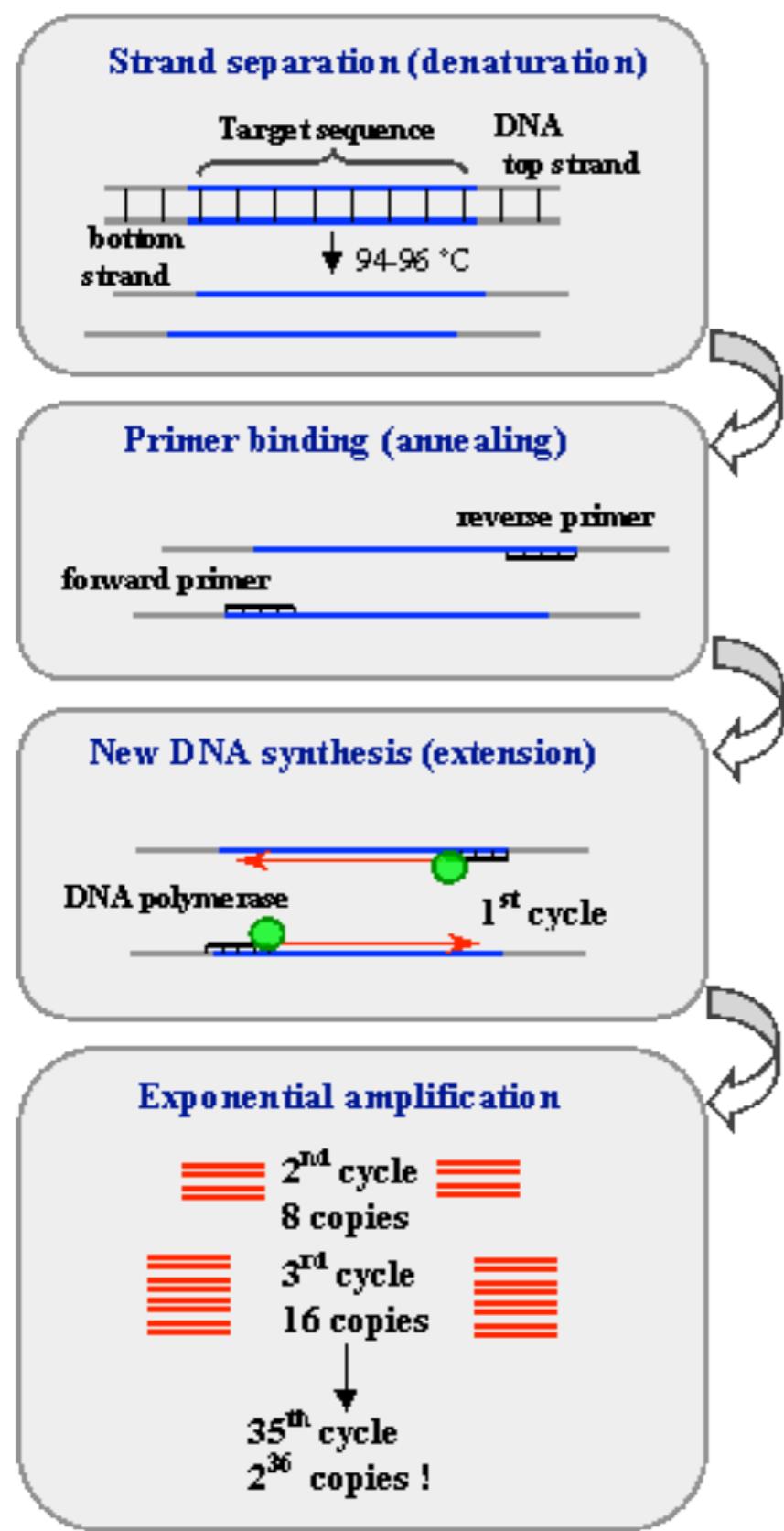
- Antisense RNA targeted at proteins of interest takes advantage of existing cellular machinery to knock down target mRNA
- Transfected into cells *in vitro* usually, though more recently *in vivo* and even in people

siRNA/shRNA

- Important to measure efficiency of knockdown
- And to watch for off-target effects
- Scrambled siRNA sequences are frequently used as controls for the transfection process itself
- Can be a single siRNA, or a pool that targets different parts of the mRNA sequence
- In theory, does not enter the nucleus, but acts on mRNA in the cytoplasm.
In theory.

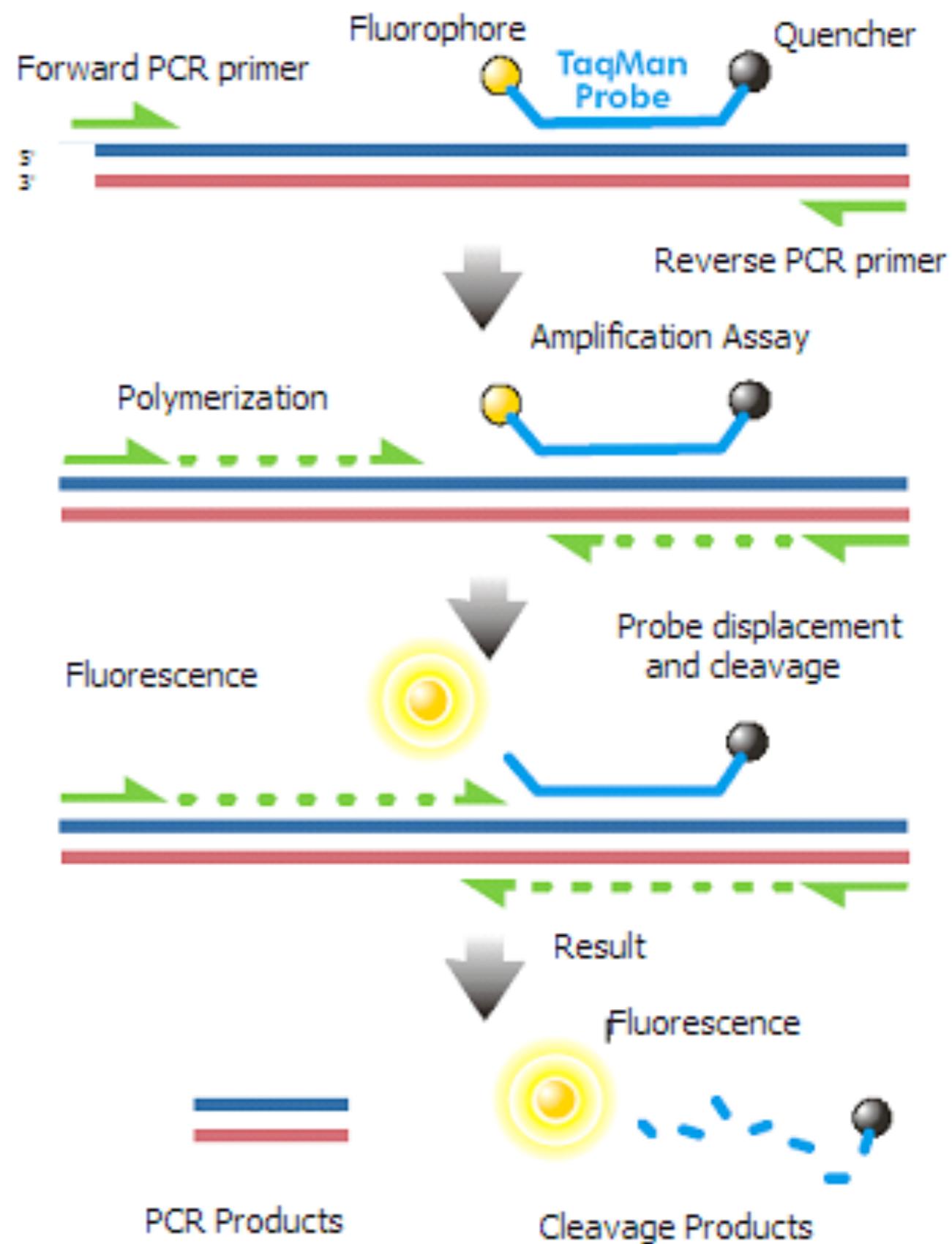


PCR

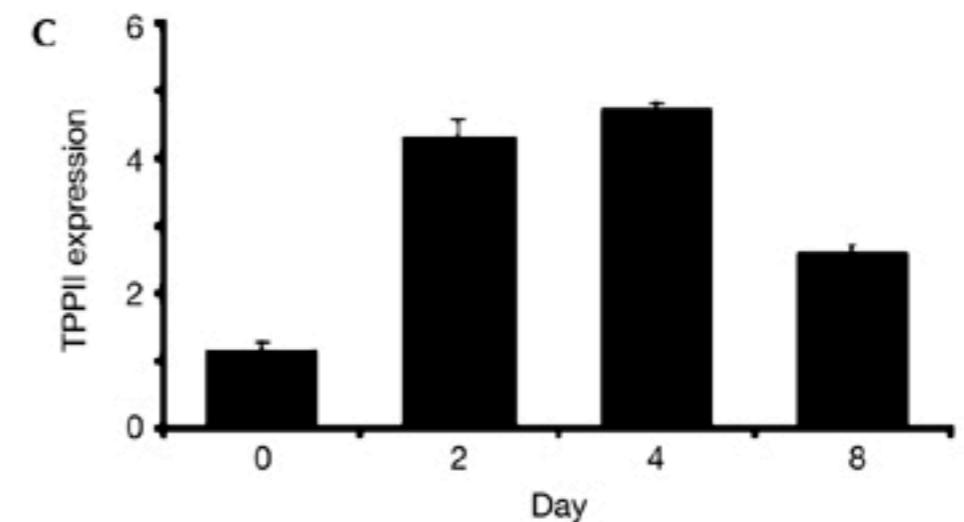
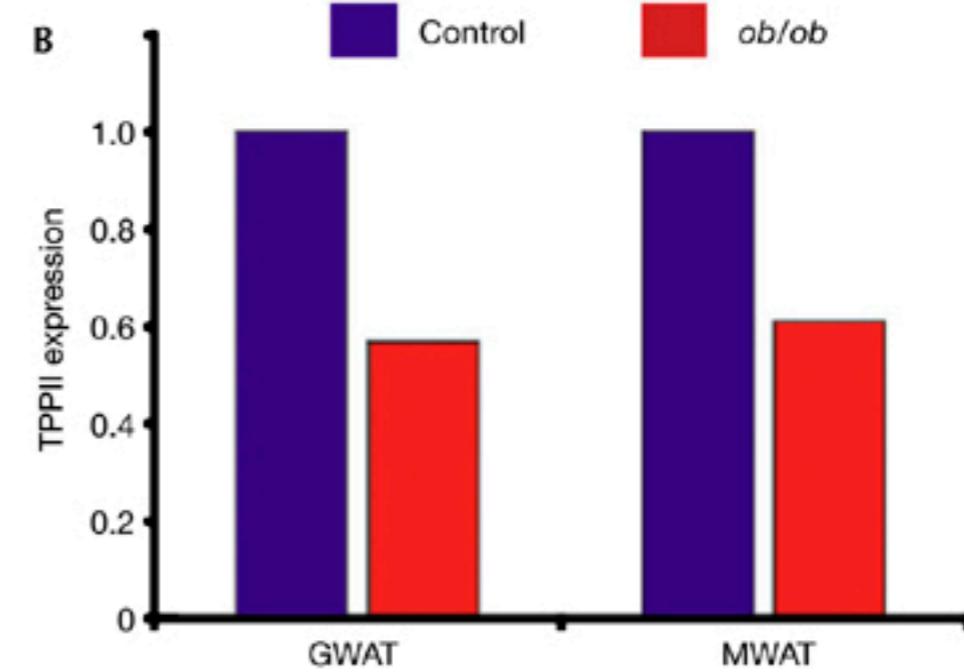
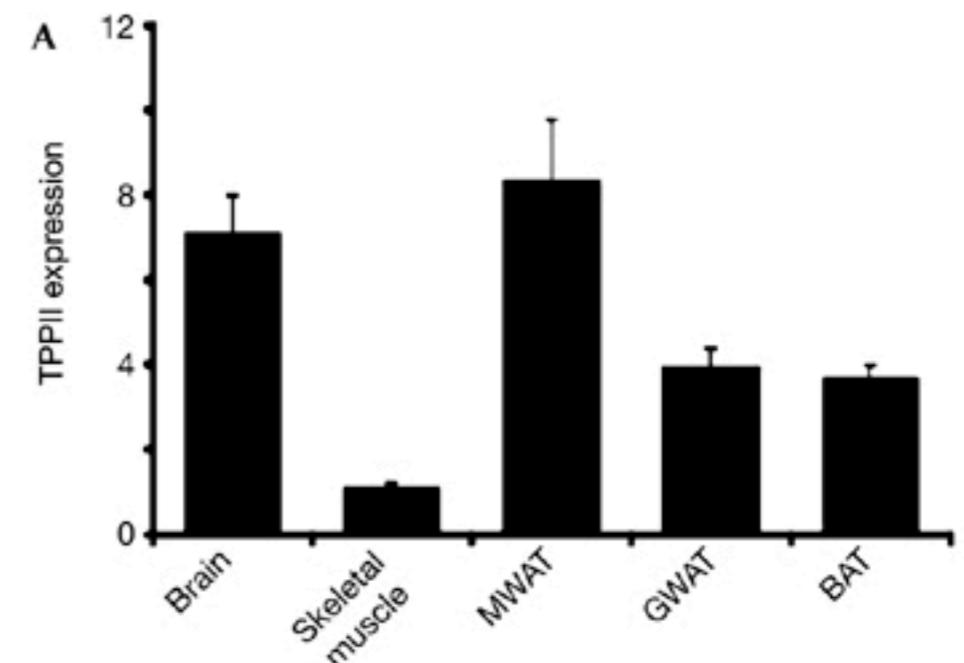
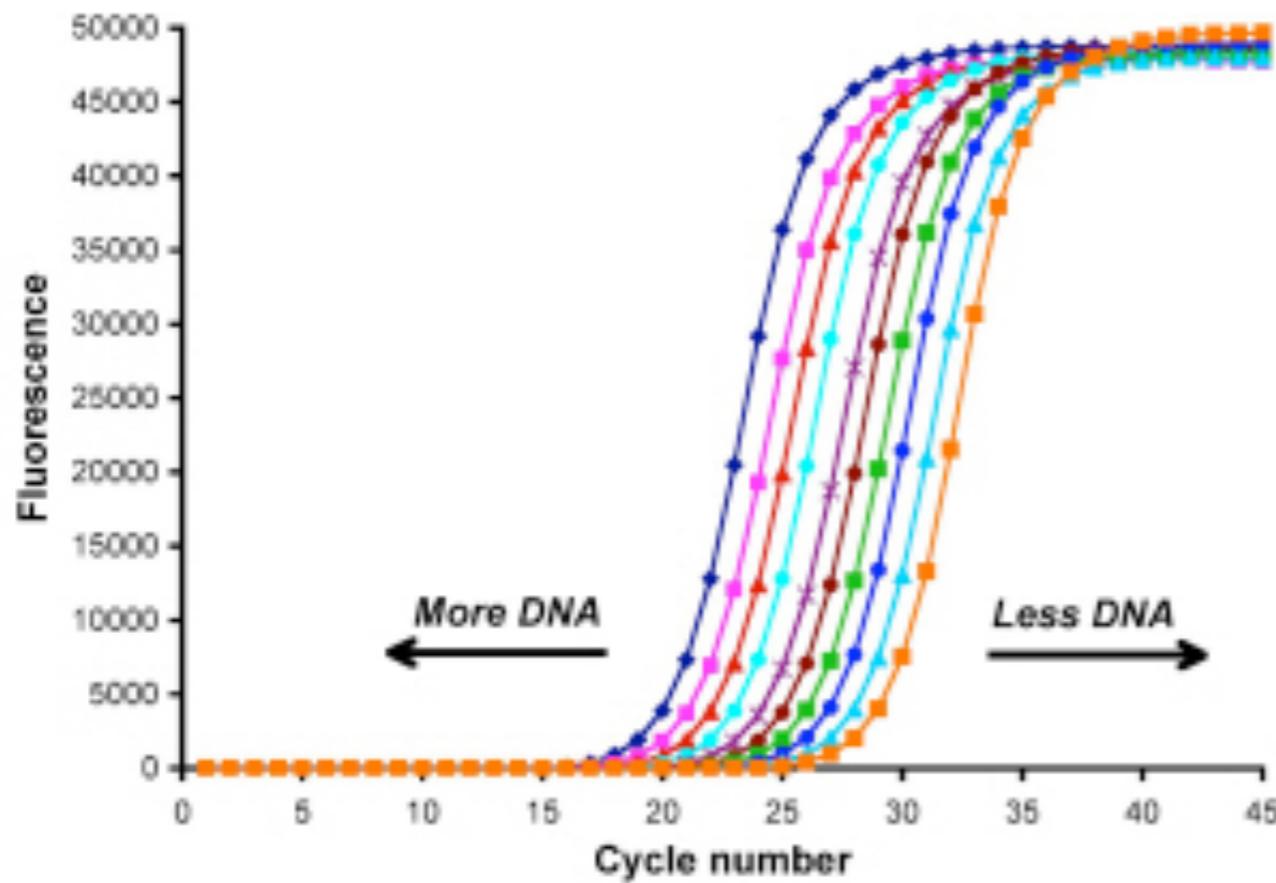


qPCR

- Measure levels of target mRNA, but more accurately than with a Northern blot
- First convert RNA into cDNA (complementary DNA) with reverse transcriptase
- Mix DNA, primer for target sequence, and dye that labels dsDNA
- As with PCR, ssDNA is amplified exponentially, but as dsDNA is generated, dye fluoresces and can be measured
- Because the sample has been amplified, final quantities are given as relative to some consistent control (i.e., a housekeeping gene)

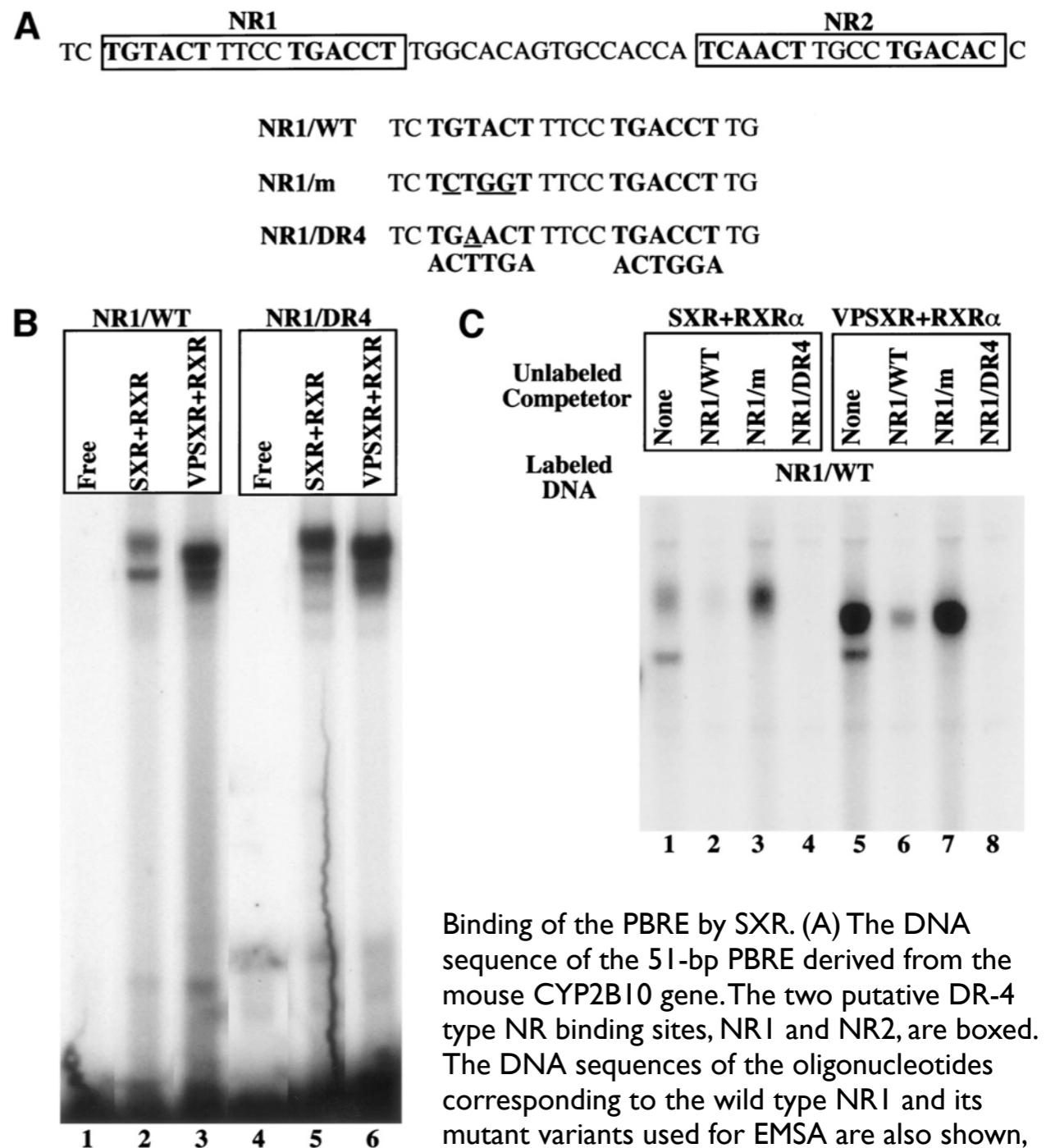


qPCR



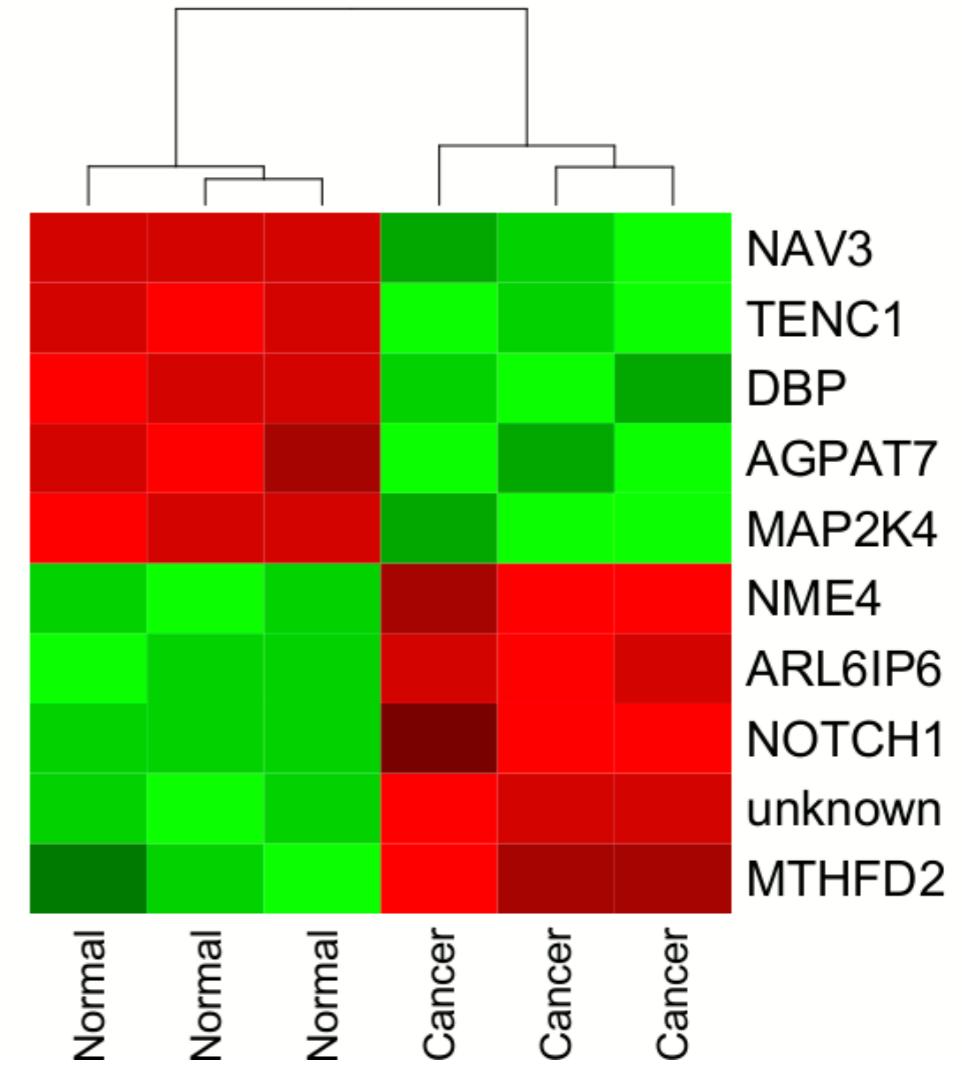
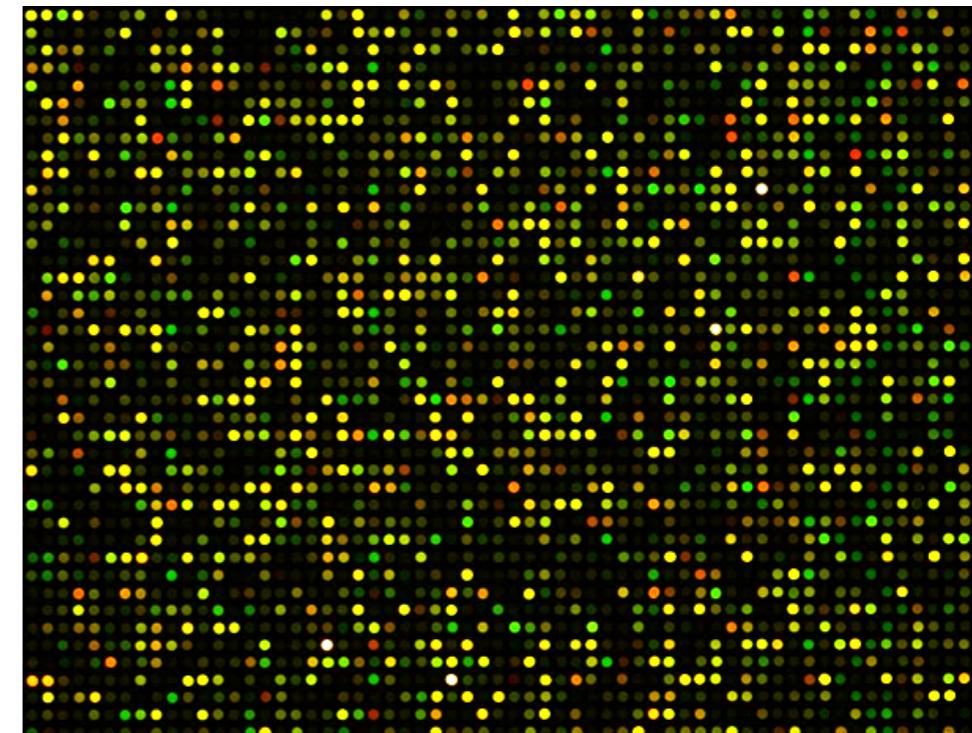
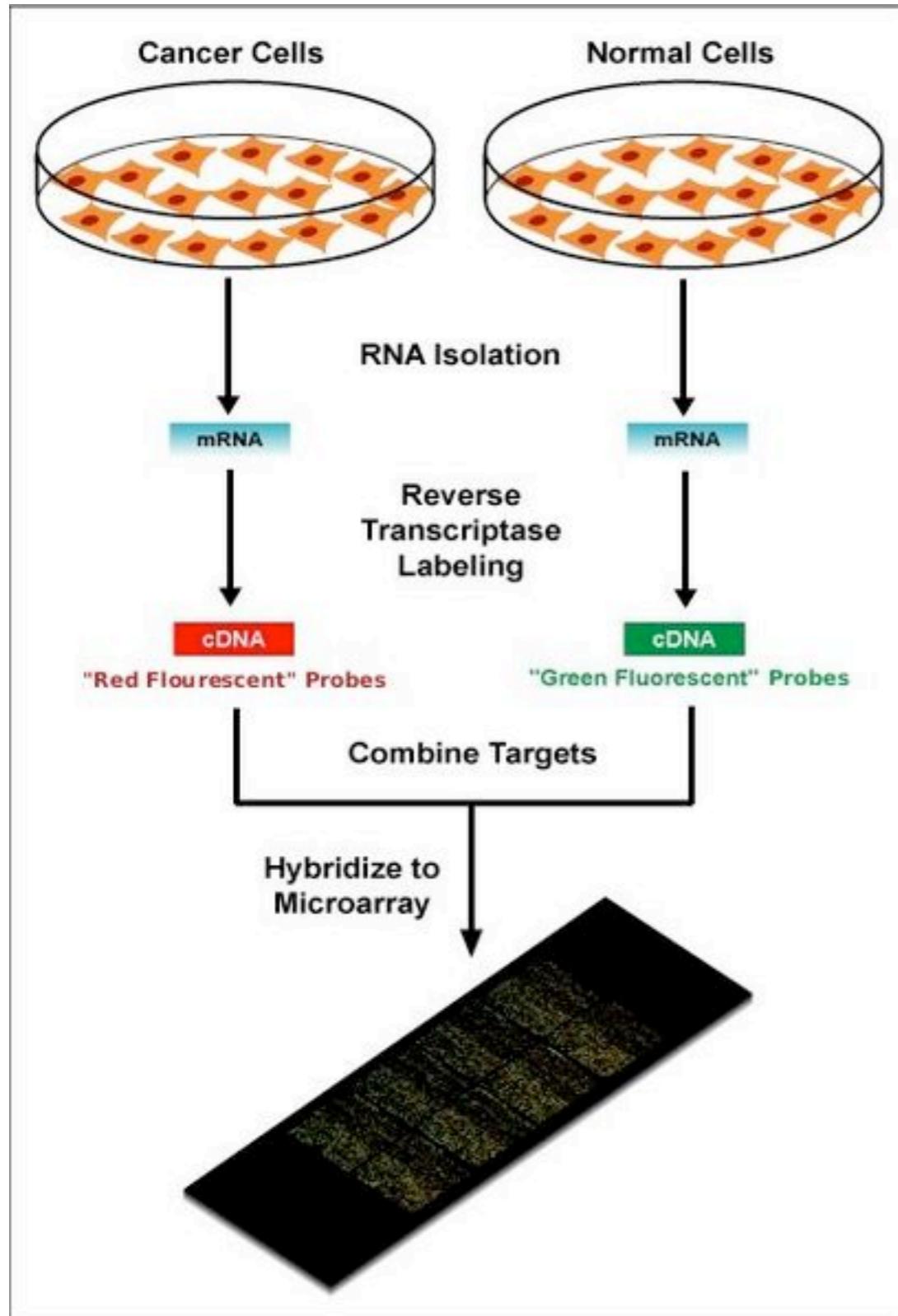
EMSA

- Electrophoretic mobility shift assay, or gel shift
- Measures protein-nucleotide interactions
- Based on the principle that complexes of different sizes and charges move through the gel at differing rates
- Control: unbound DNA/RNA
- If sample has protein bound, the band will travel less far down the gel in the allotted running time
- Supershift: add Ab of interest, which slows down the complex further and confirms the identity of the binding protein complex

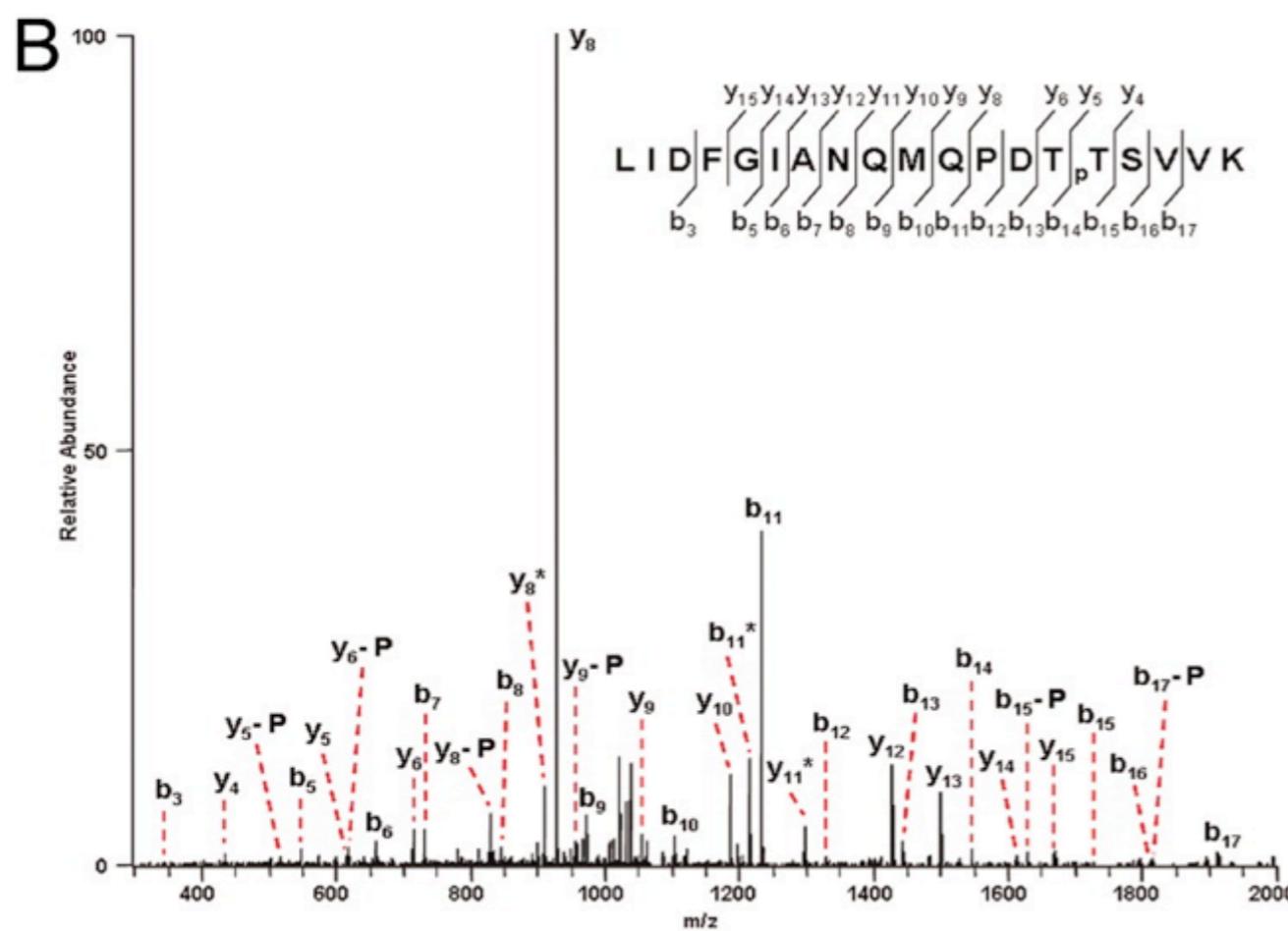
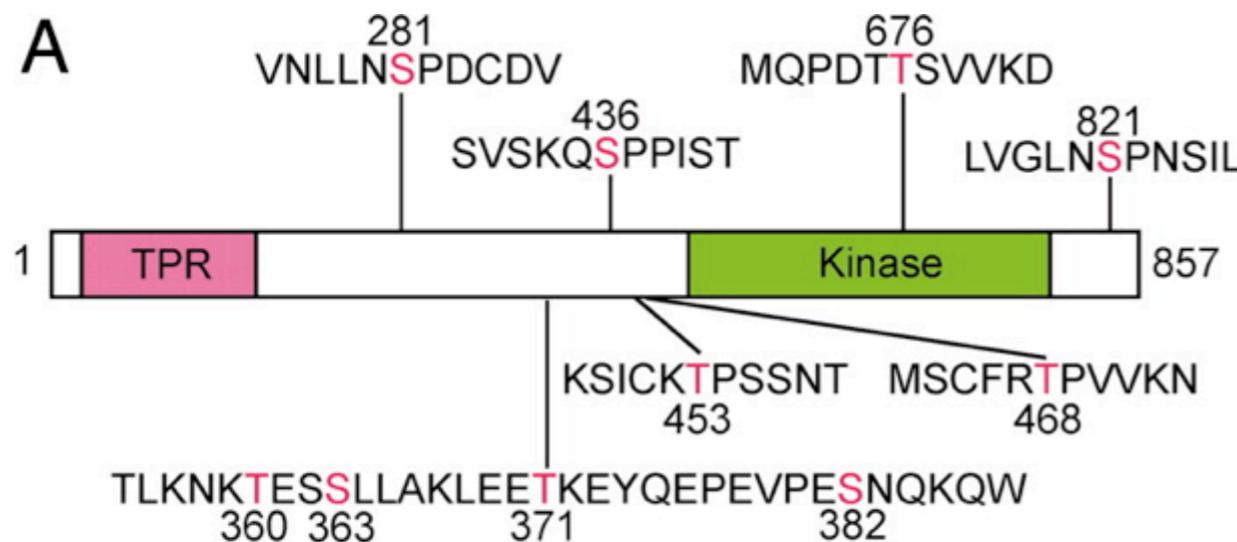


Binding of the PBRE by SXR. (A) The DNA sequence of the 51-bp PBRE derived from the mouse CYP2B10 gene. The two putative DR-4 type NR binding sites, NR1 and NR2, are boxed. The DNA sequences of the oligonucleotides corresponding to the wild type NR1 and its mutant variants used for EMSA are also shown, with the mutated nucleotides underlined. (B) SXR:RXR heterodimers bind to the PBRE. EMSA was performed using in vitro synthesized SXR, VPSXR, and RXR α proteins and radiolabeled oligonucleotides of NR1 (lanes 1–3) and its mutant NR1/DR4 (lanes 4–6) in which the imperfect DR-4 of NR1 was mutated to an AG(G/T)TCA type of DR-4. (C) The binding of NR1 by SXR/RXR α or VPSXR/RXR α can be efficiently competed away by excessive unlabeled NR1/WT (lanes 2 and 6), NR1/DR4 (lanes 4 and 8), but not by NR1/m (lanes 3 and 7). The free probes ran off the gel.

Microarrays

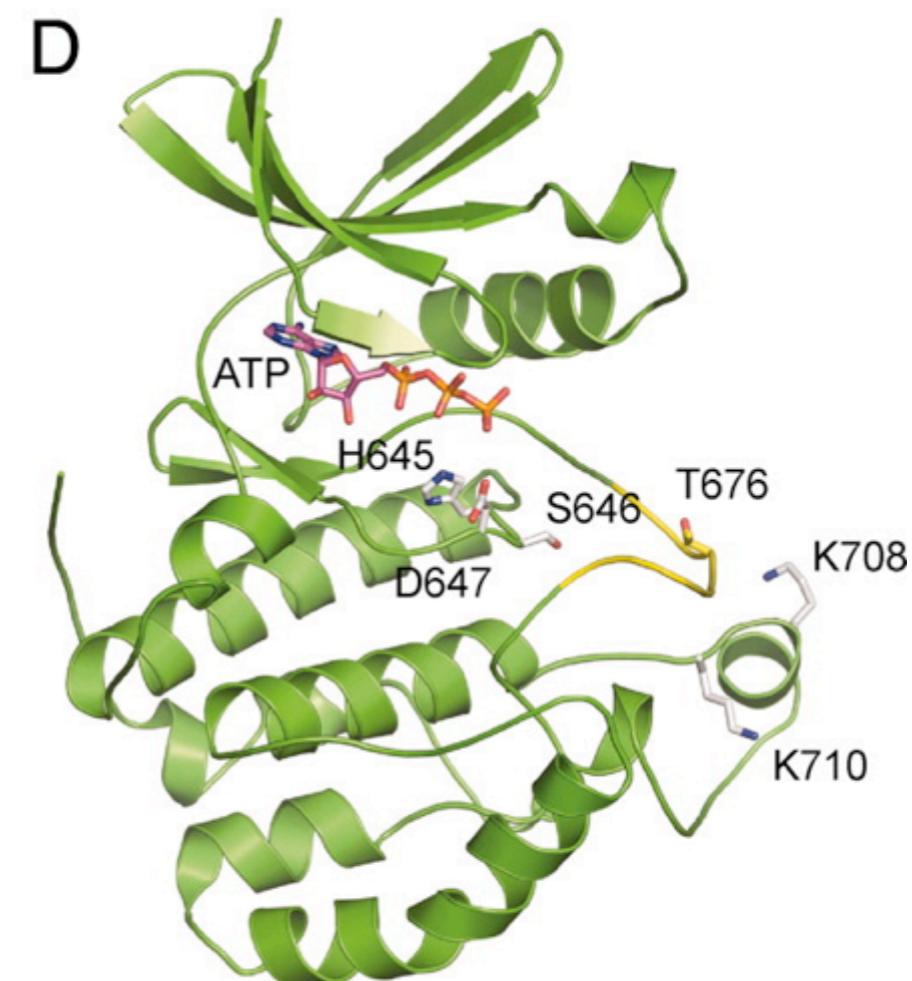


Mass spectrometry



C

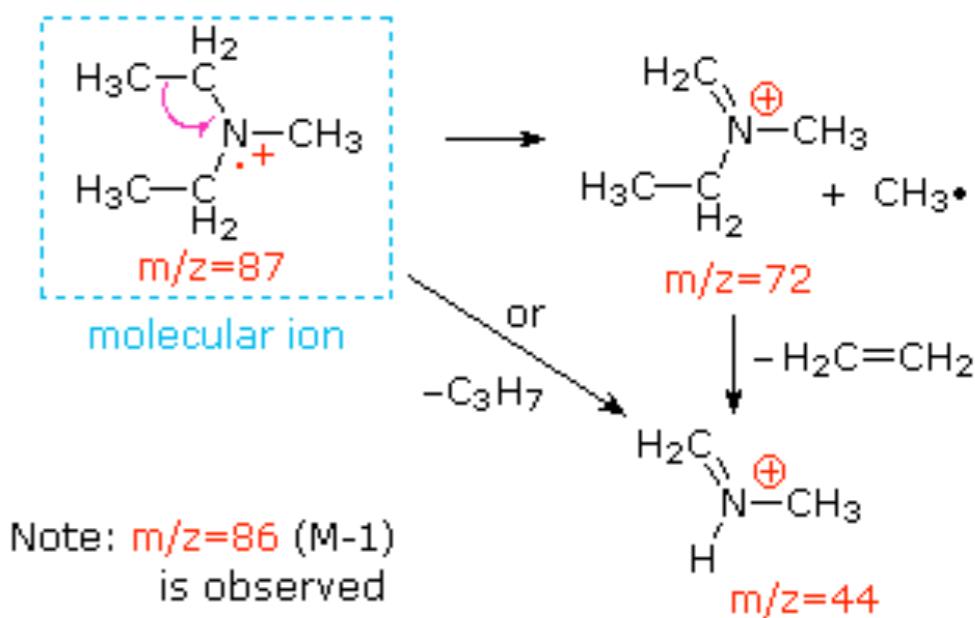
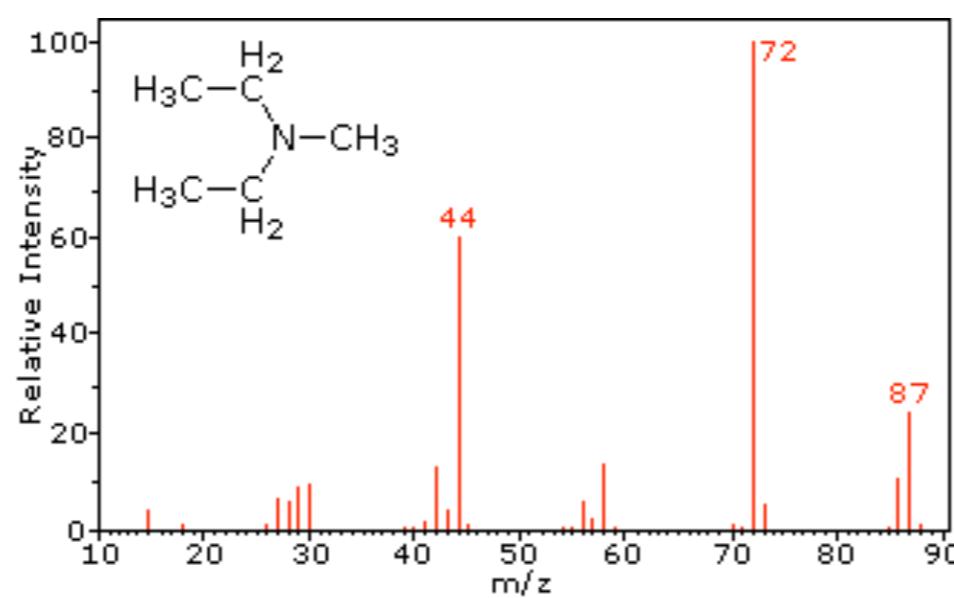
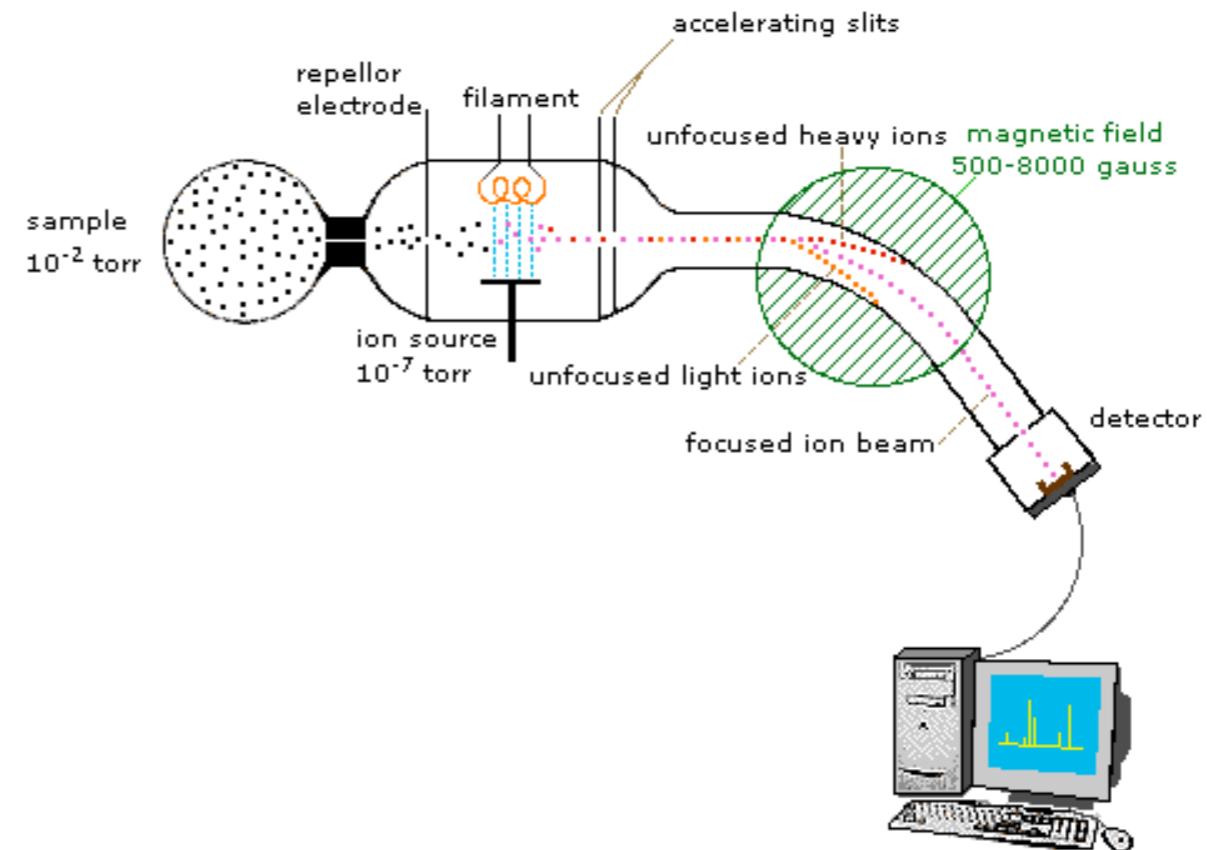
Hs	Mps1	664	DFGIANQM Q PDT T SVVK
Mm	Mps1	676	DFGIANQM Q PDT T SIVK
Xl	Mps1	680	DFGIANQ I QPDVT S IVK
Dm	Mps1		DFGIASNIAVD S TSIIK
Sp	Mps1		DFGIAKAIGND T NIHR
Sc	Mps1		DFGIANAVPEHTVNIYR



Identification of phosphorylation sites in human Mps1 protein

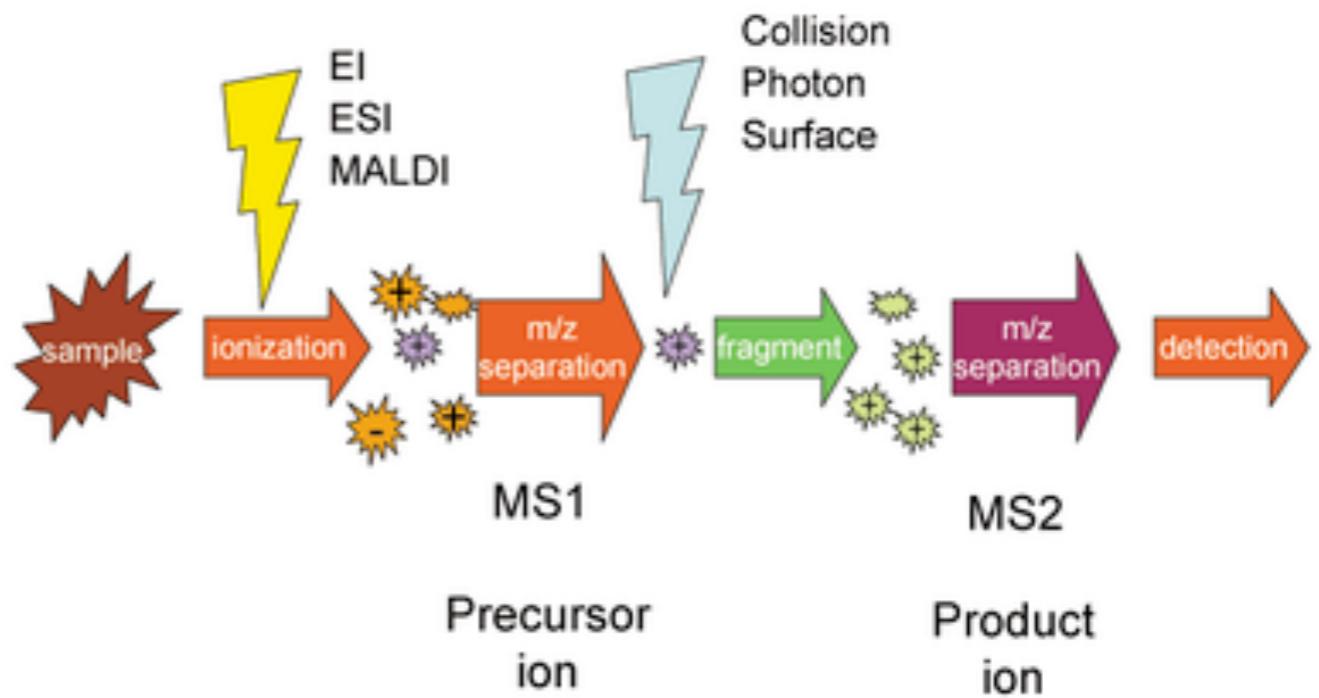
Mass spectrometry

- Used to determine the mass of particles, which can help us determine the identity of peptides in a sample
- The idea: chop up protein sample; ionize the pieces; throw them against a detector; measure mass-to-charge ratio by determining how long particles take to hit the detector in relation to the charge induced when they hit; map mass-to-charge (m/z) ratios back to known amino acid sequences.



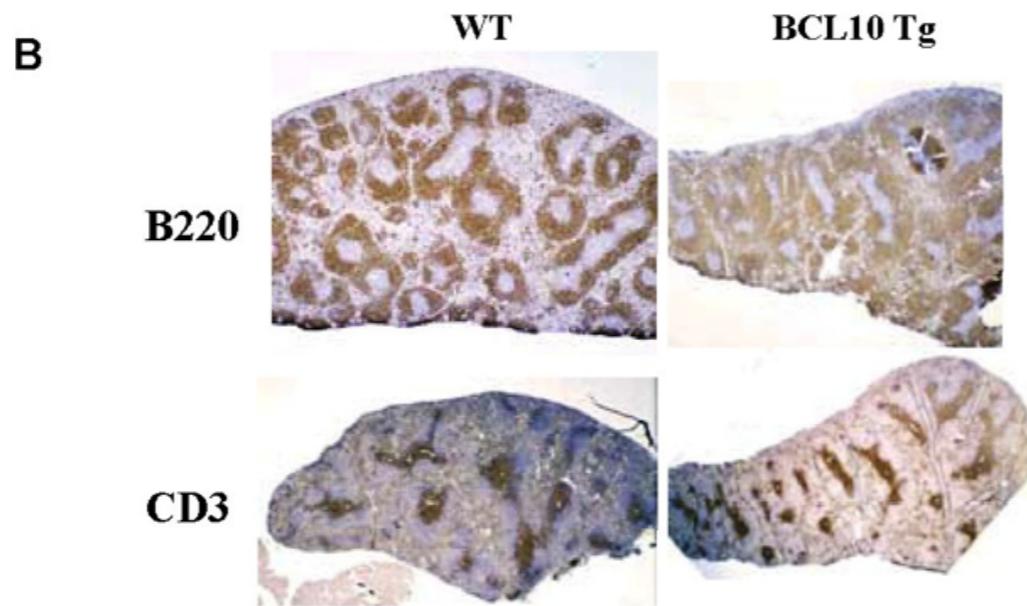
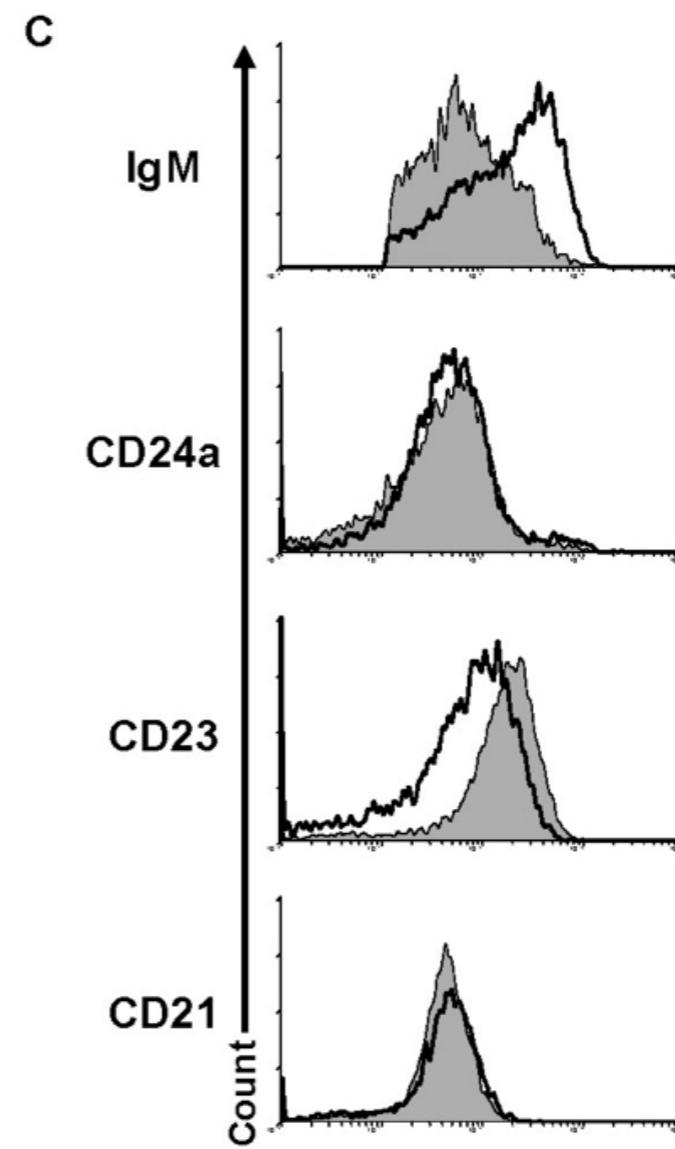
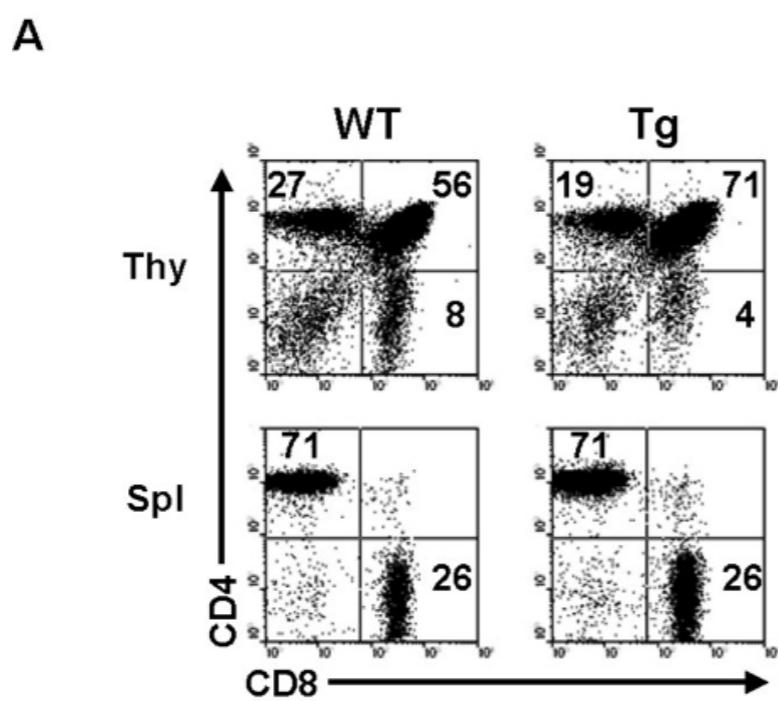
Mass spectrometry

- Tandem mass spectrometry is mass spectrometry run several times in sequence to separate out proteins of interest
- A big bioinformatic challenge: bin-packing problem, gone high-throughput. Given the mass of a group of particles, what particles are represented? Given the existence of several groups of particles, what peptides are represented? Proteins? PTMs?

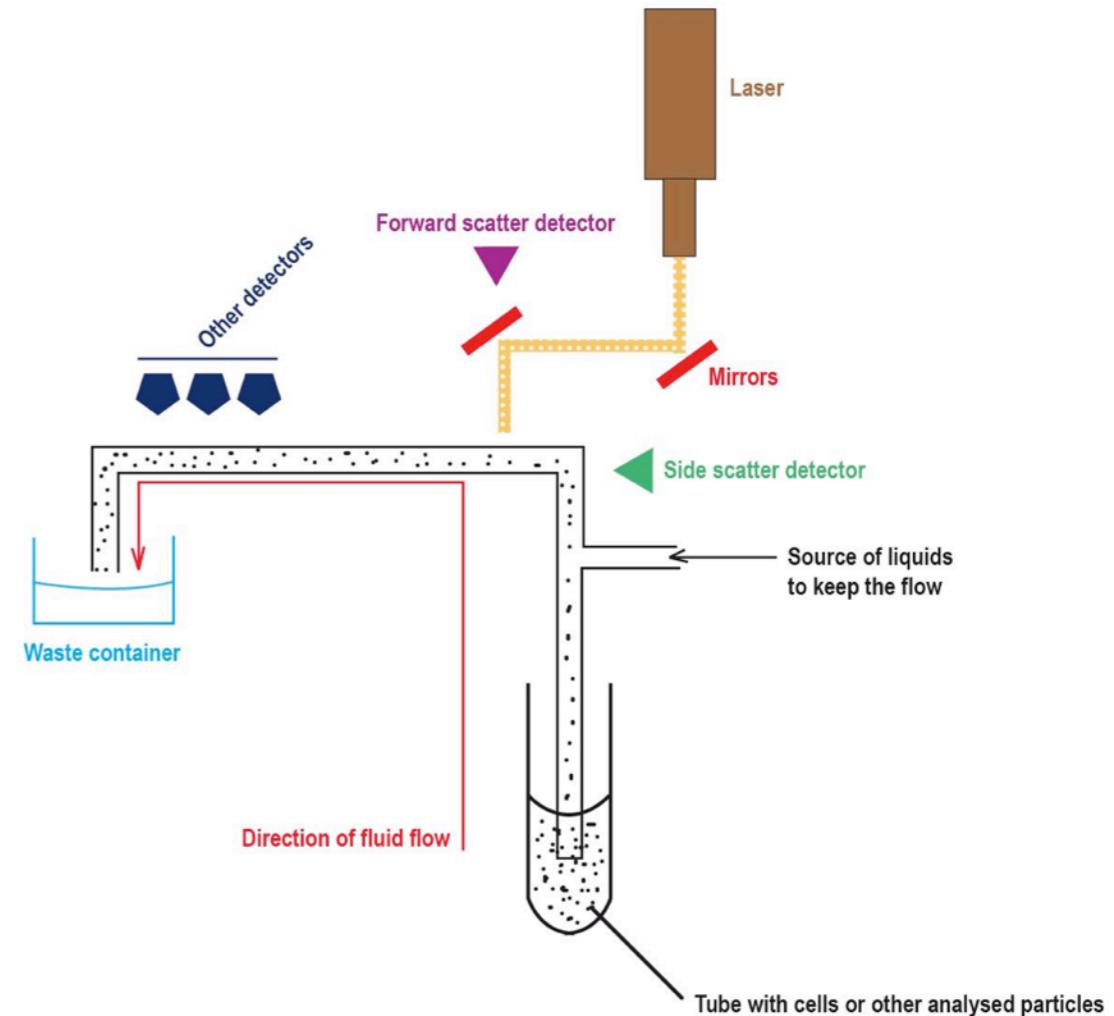


For more mass spec:
Bioinformatics III
Vineet Bafna, Nuno Bandeira, Steve Briggs,
Pieter Dorrestein, Vivian Hook

Flow cytometry

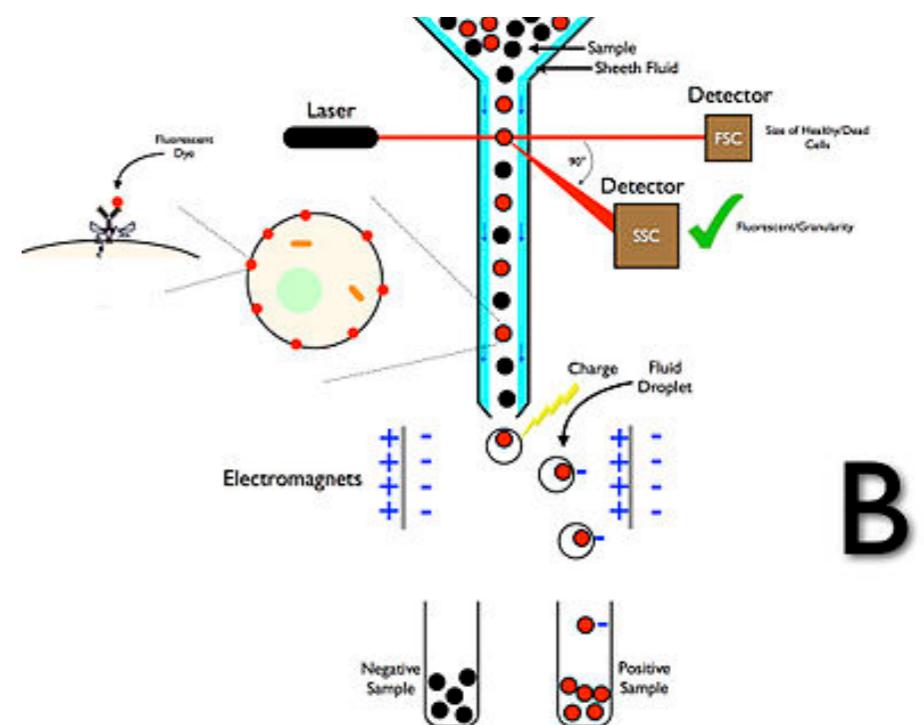


Flow Cytometer Scheme

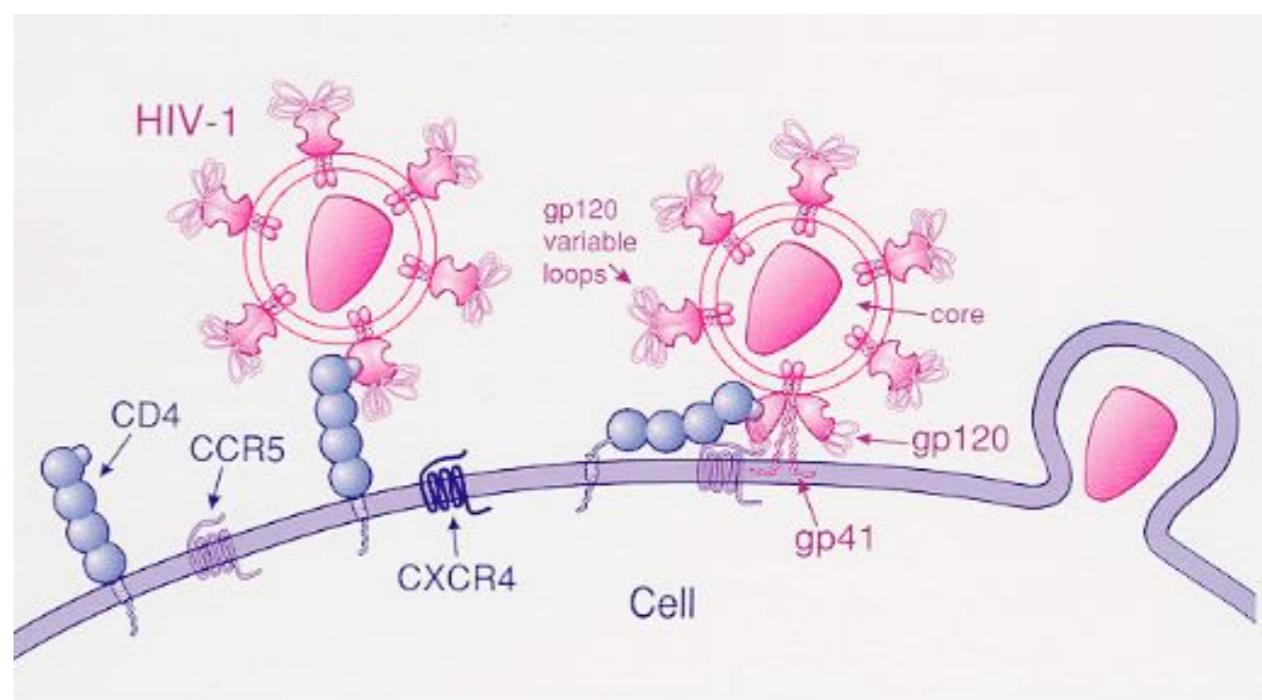
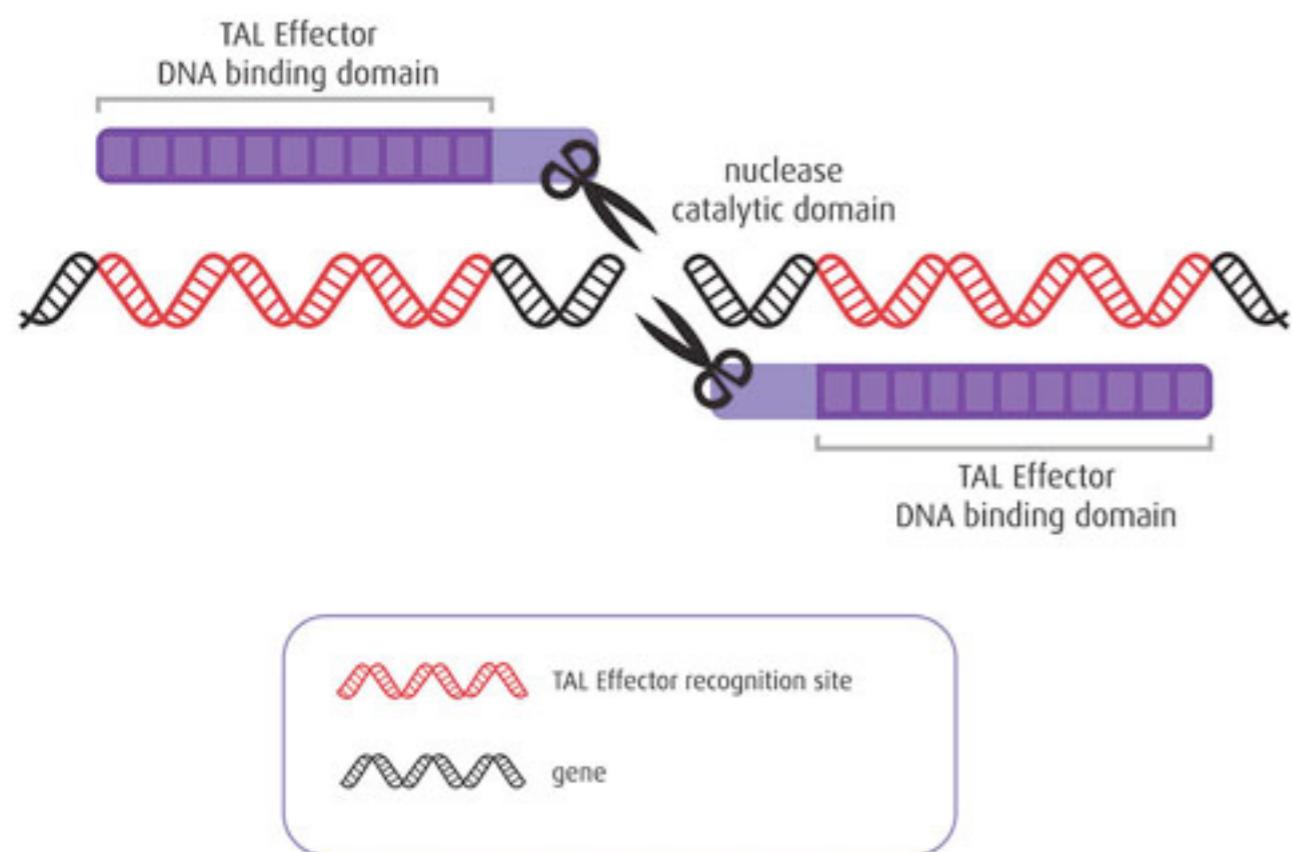


Flow cytometry

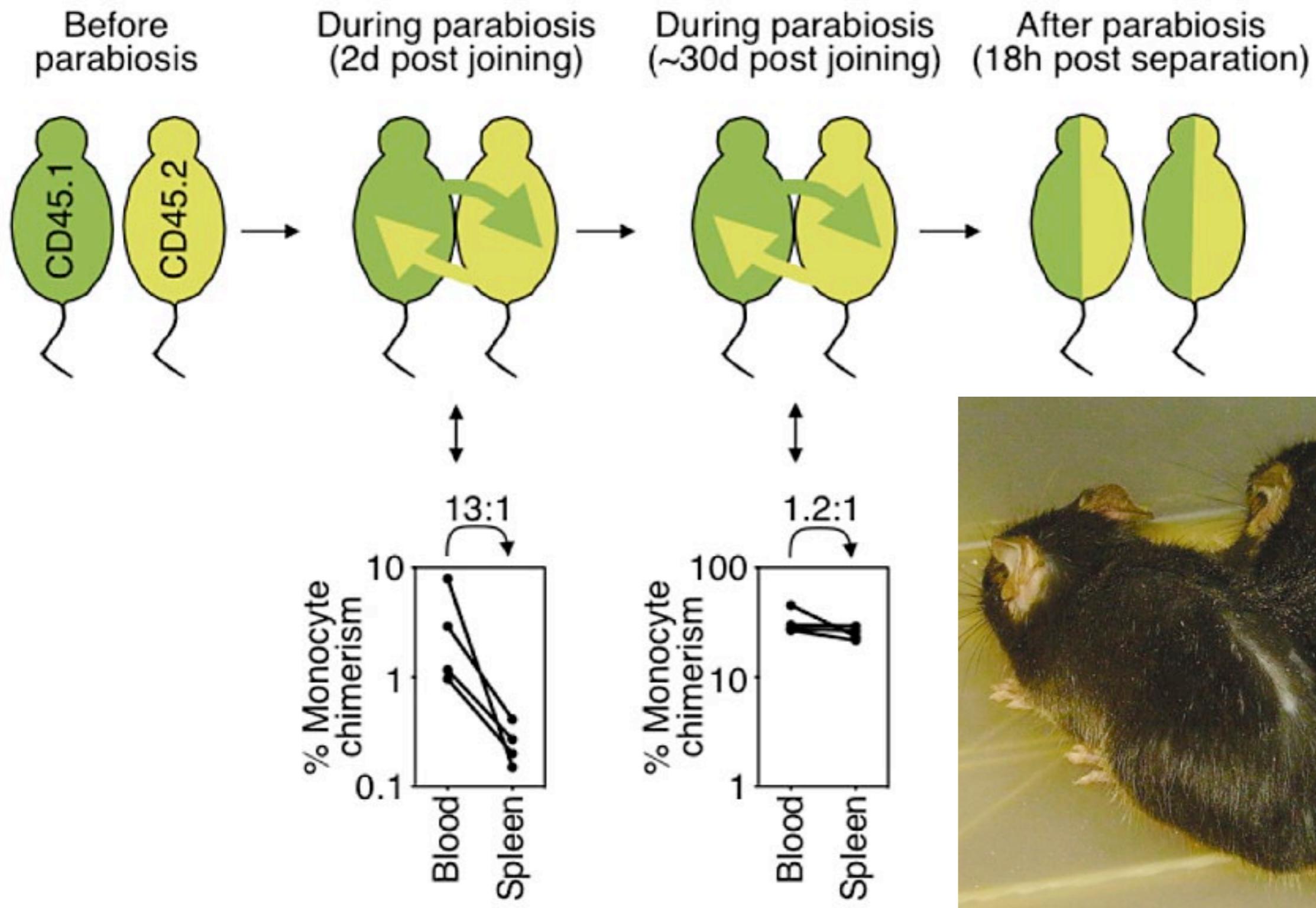
- Optics-based measurements of cell properties
- Highly used in immunology to separate and characterized cell populations by external markers (CD4 vs CD8, etc)
- FSC: size
- SSC: granularity
- Fluorescence along several channels: Abs of interest
- A bit of an art form— quadrants and gates often set manually by flow cytometers
- “What did they gate on?” The goal is to gate out noise and irrelevant cells... or just data that doesn’t conform to hypothesis...
- Can be used to sort cells: Fluorescence activated cell sorting (FACS) passes cells through the cytometer, then redirects cells one-by-one into separate containers based on Ab fluorescence.



TALENs and ZFNs



Parabiosis



Name that tool!



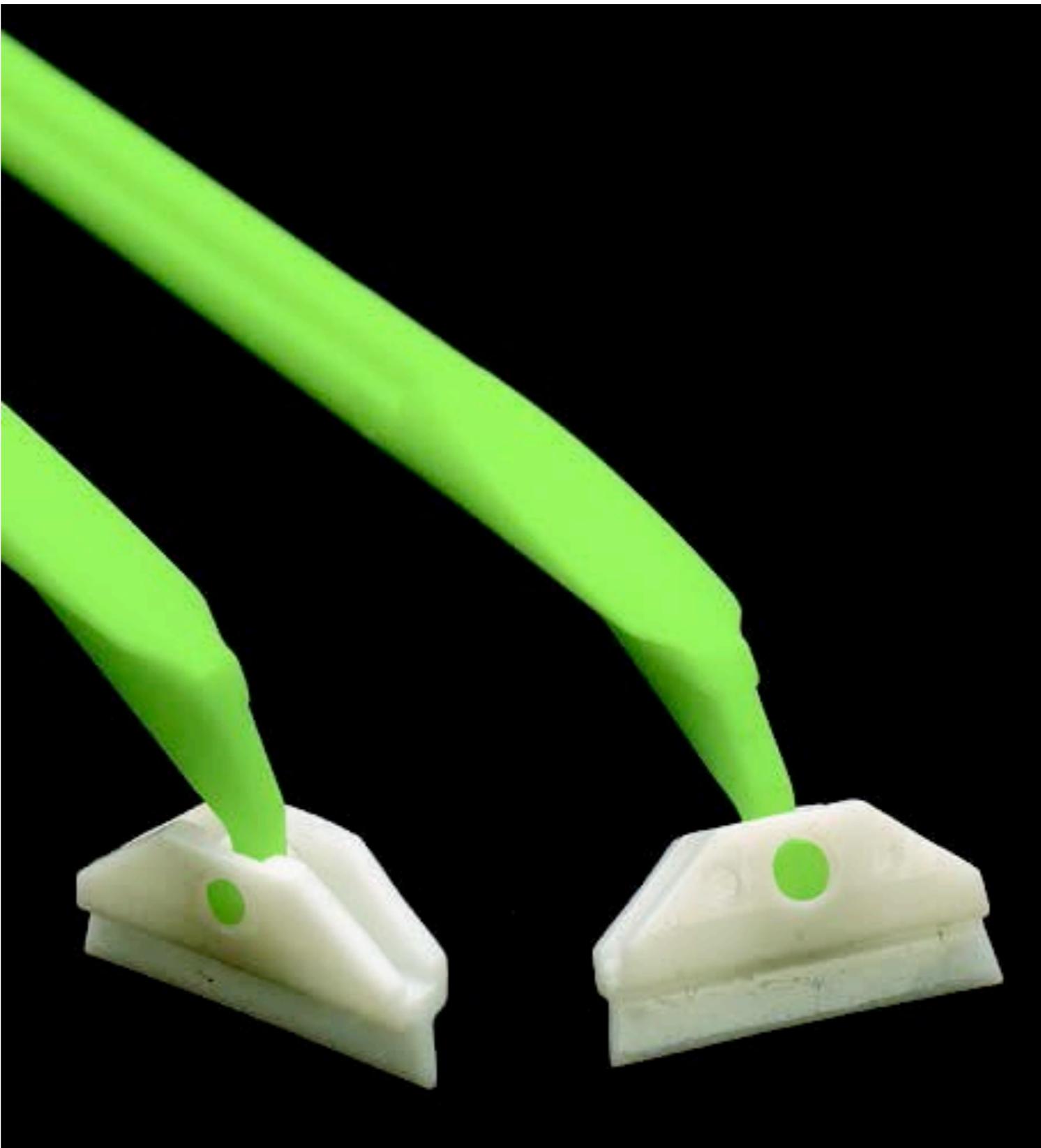


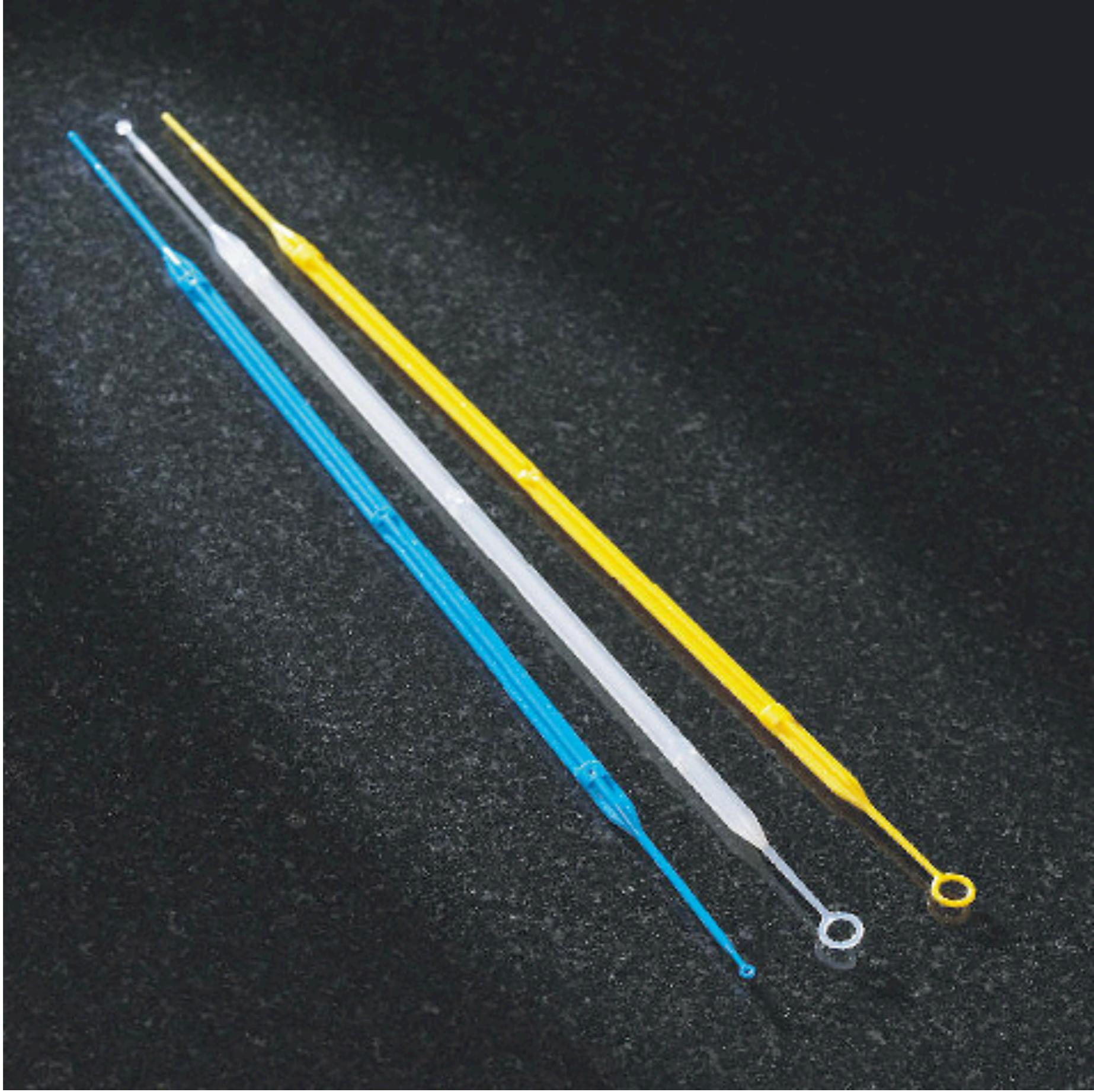
















In conclusion,
be curious.
Biologists don't bite.