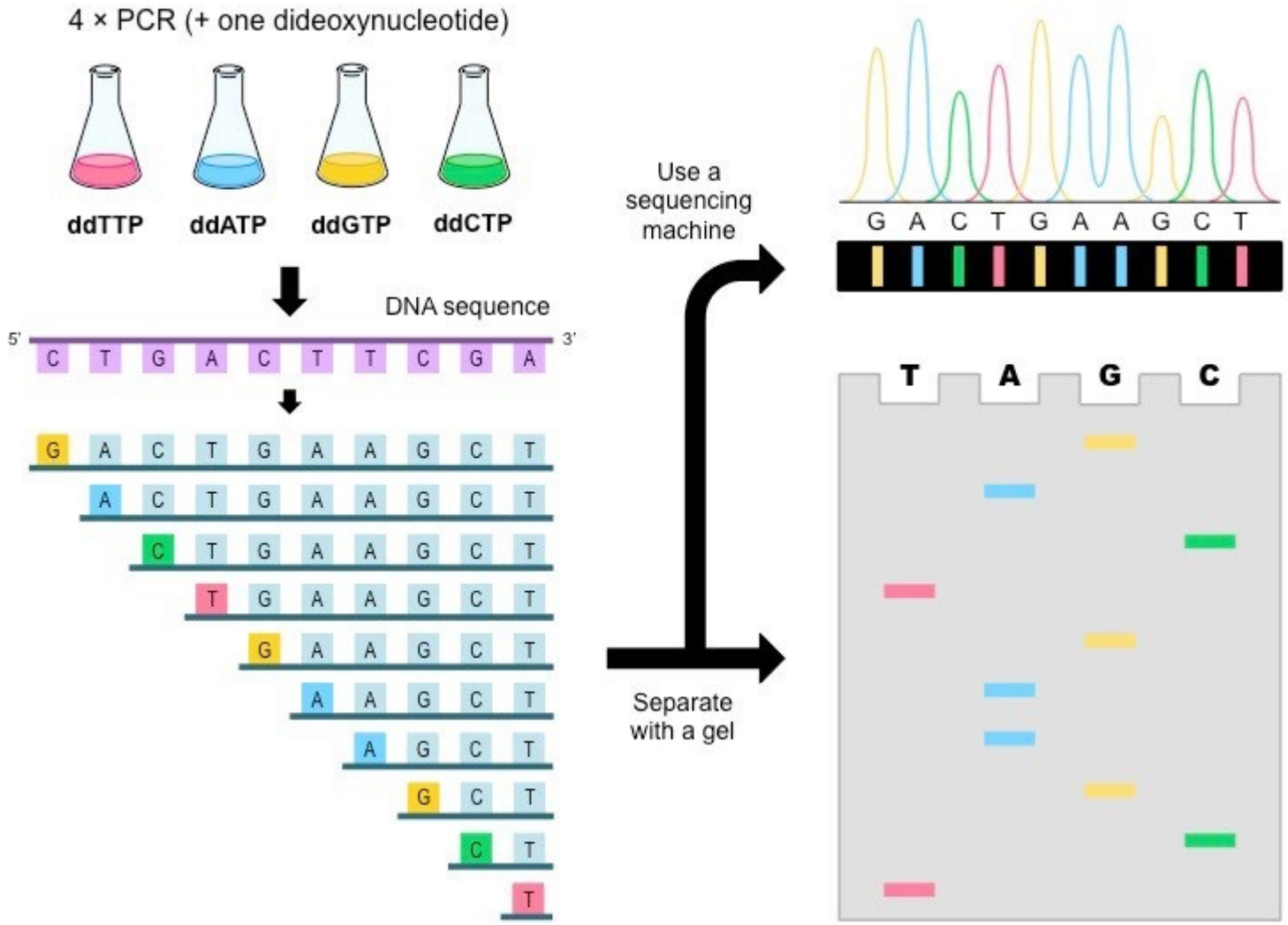
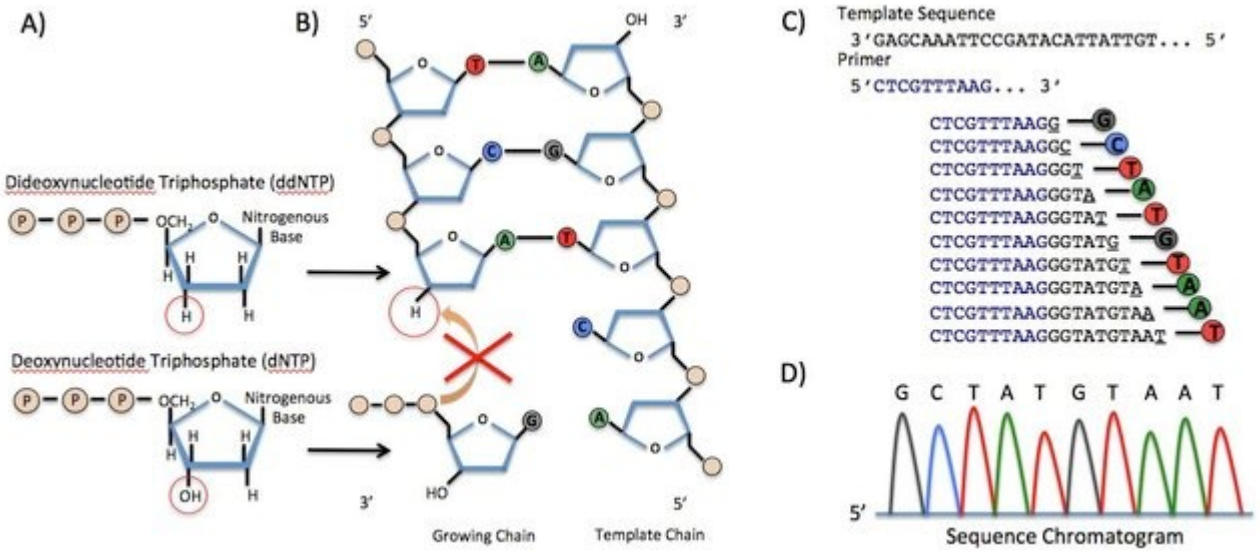
DNA Sequencing



Sanger Sequencing



Dye Terminations

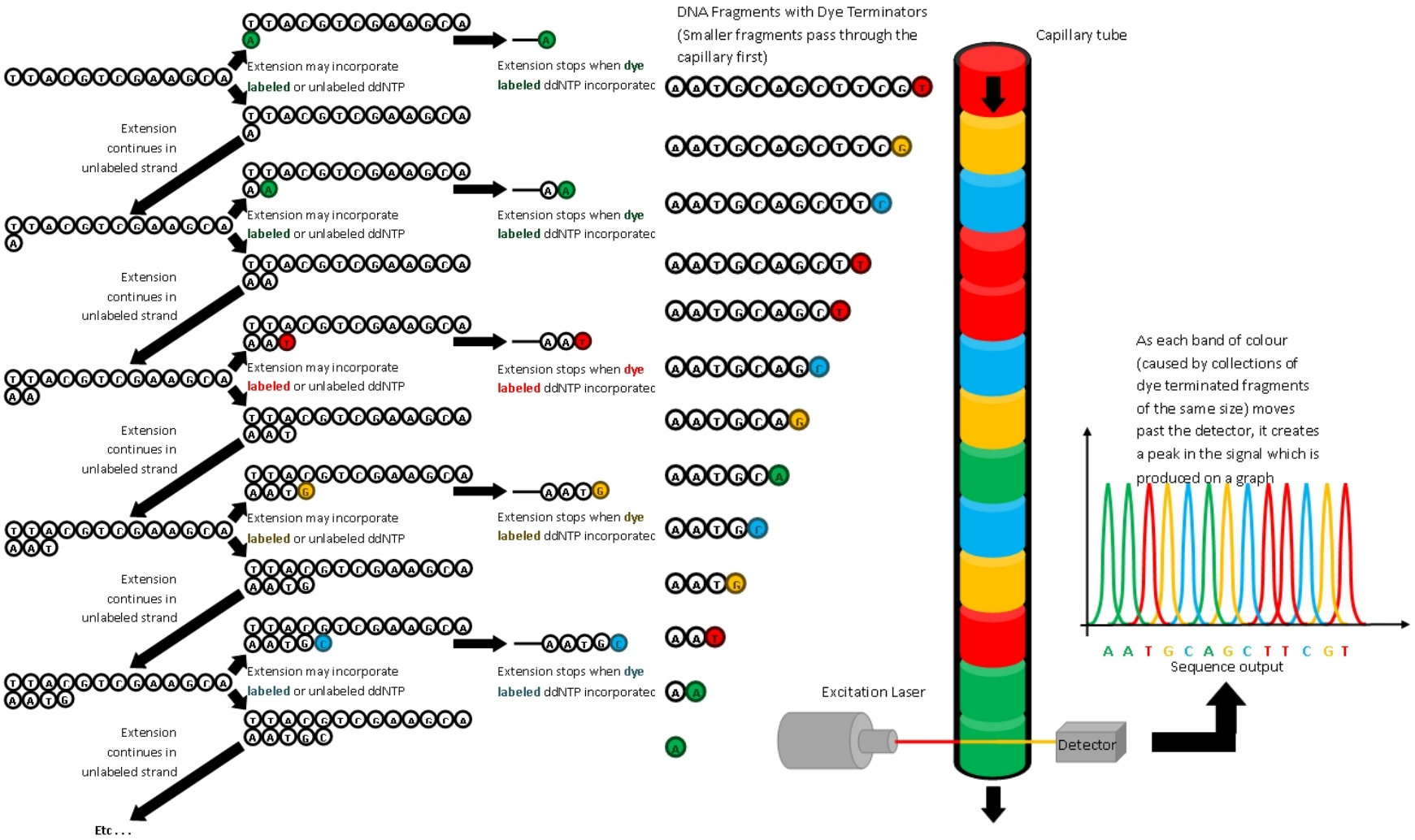
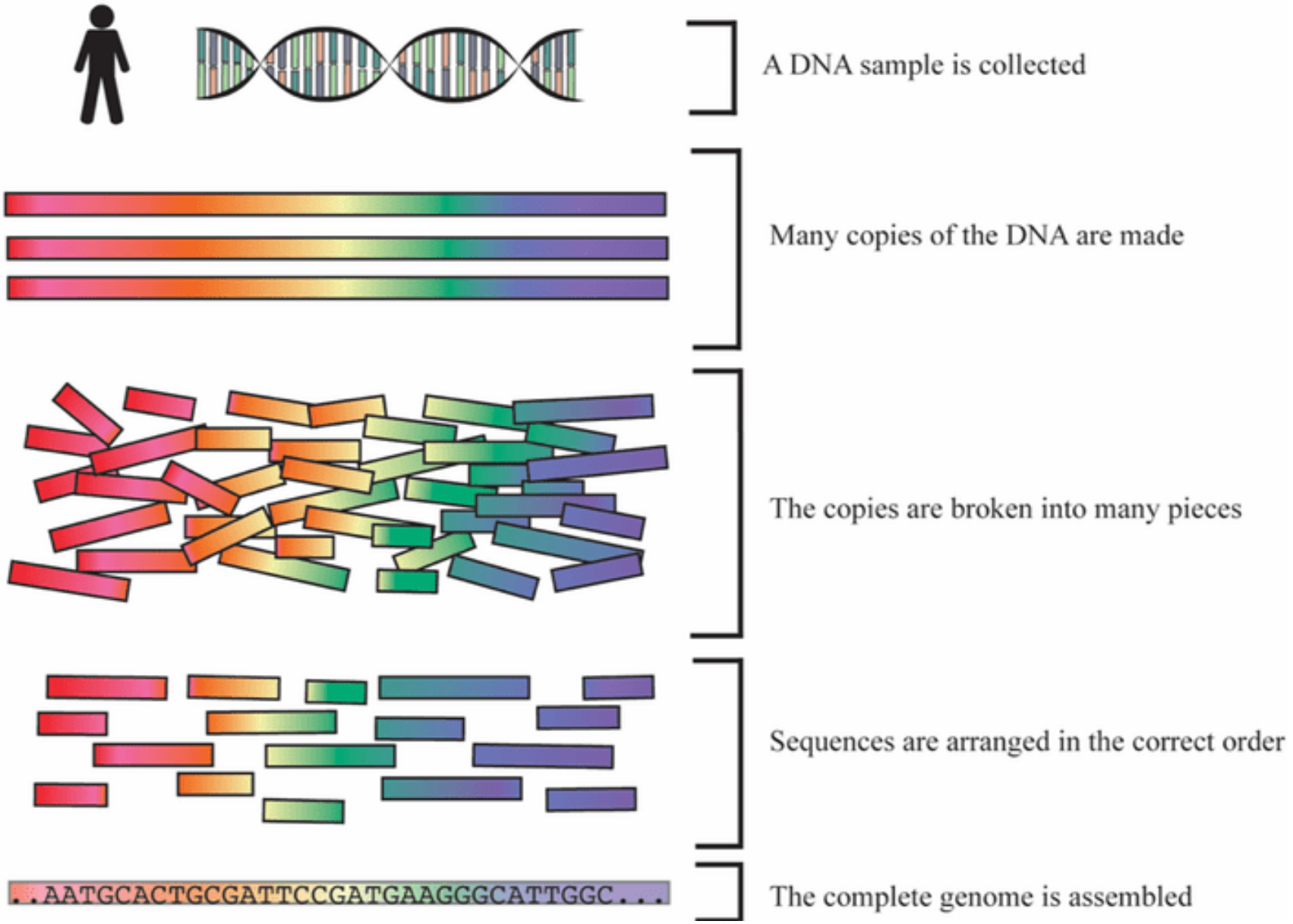
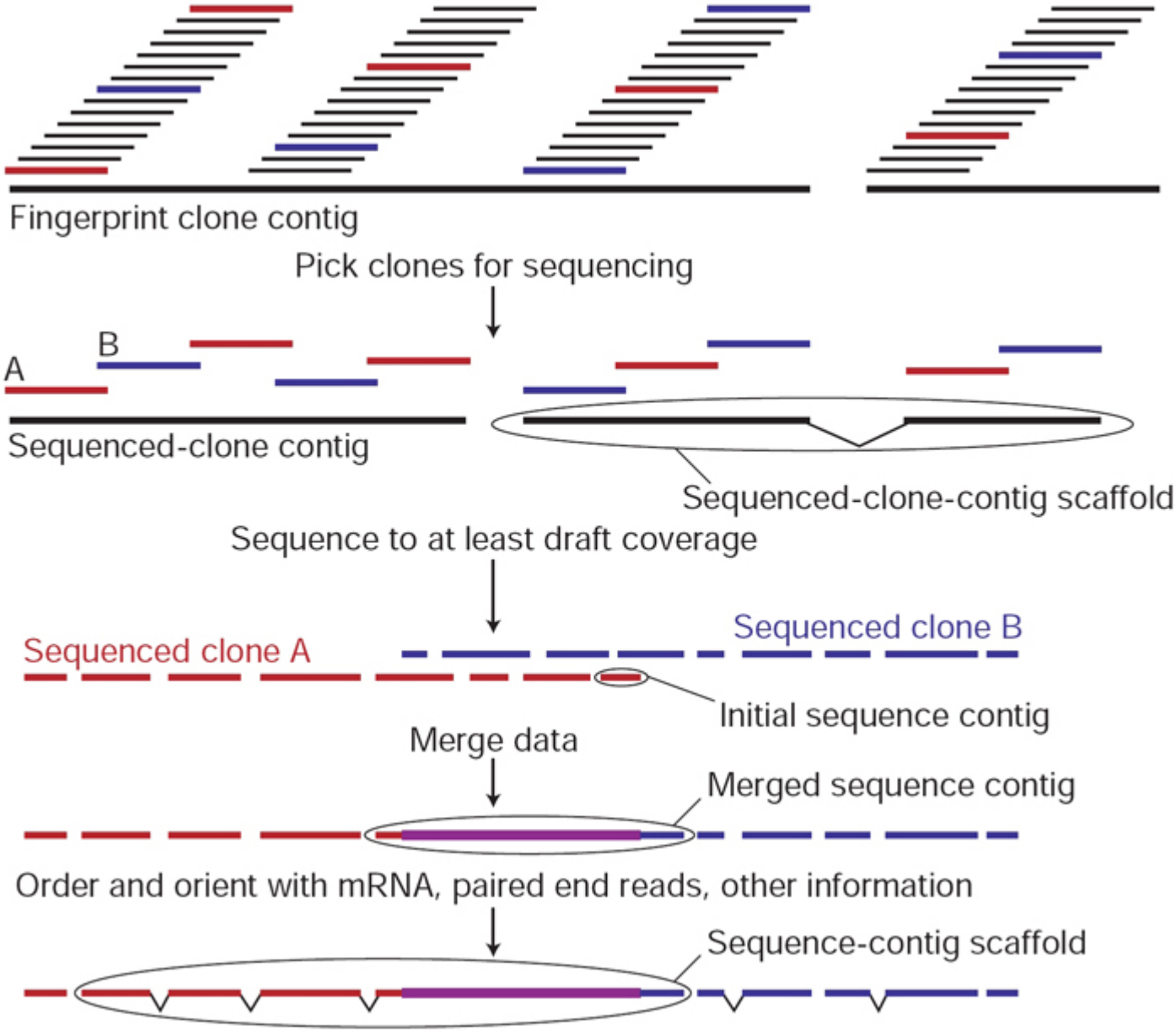
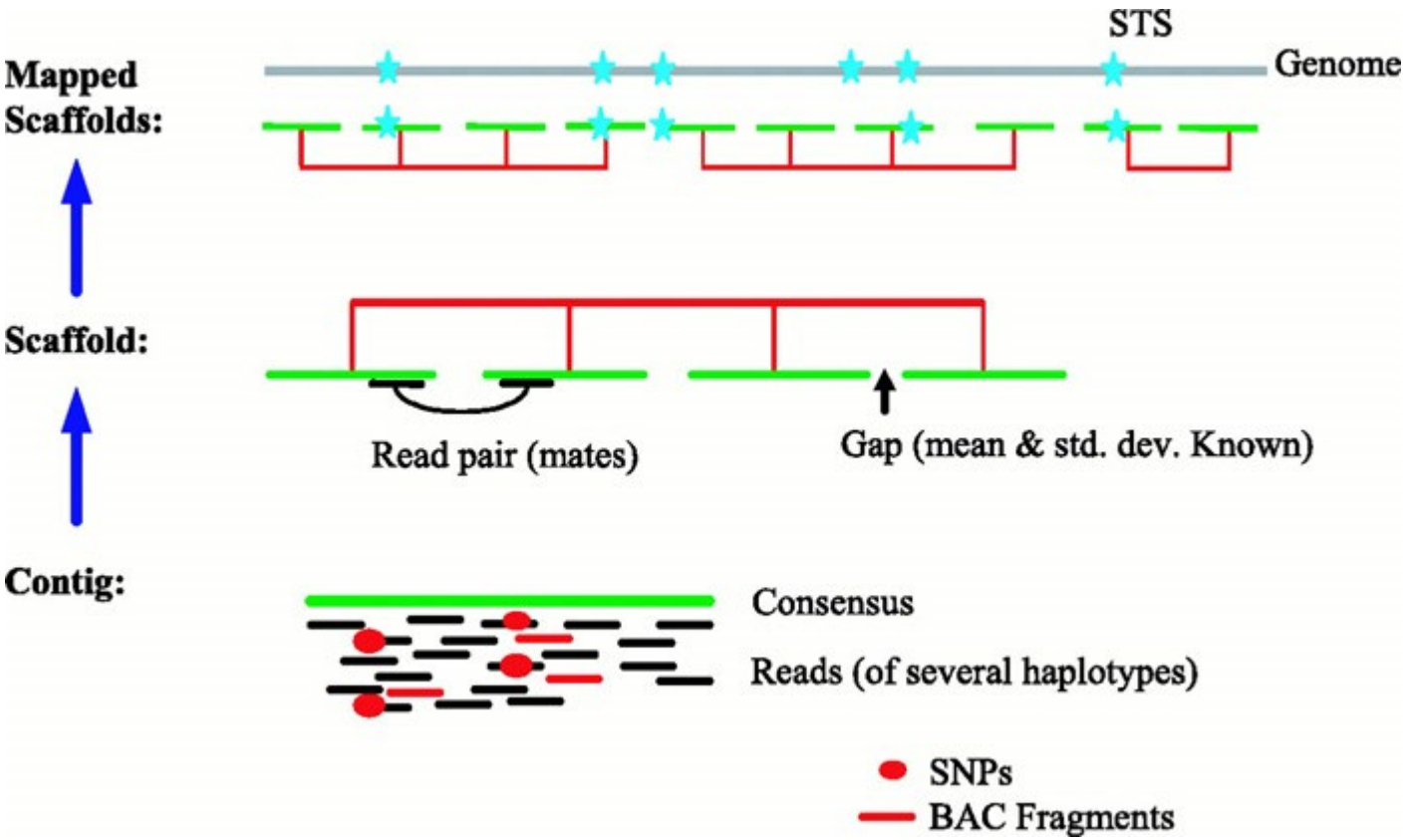


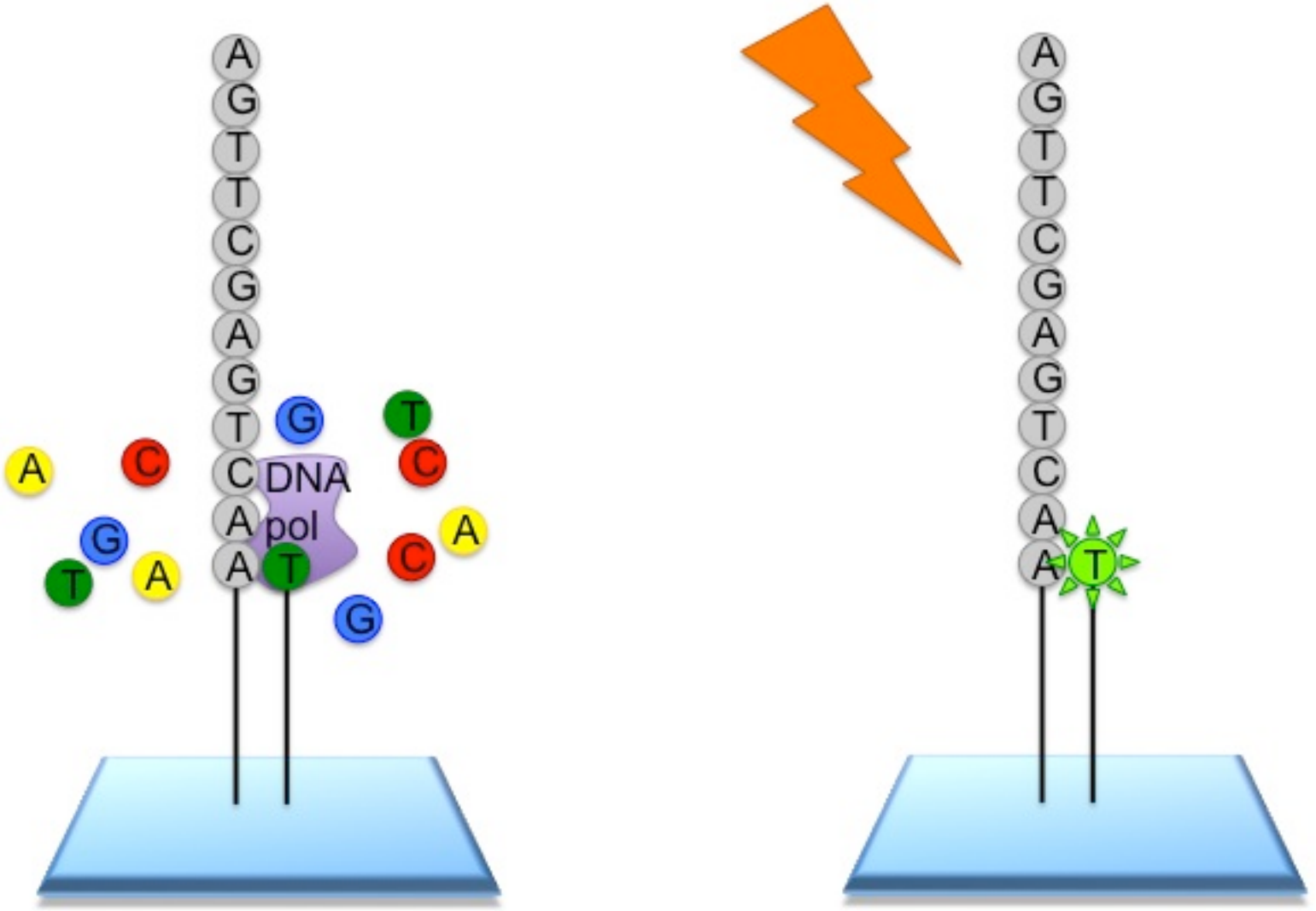
Figure 2: Shotgun Whole-Genome Sequencing

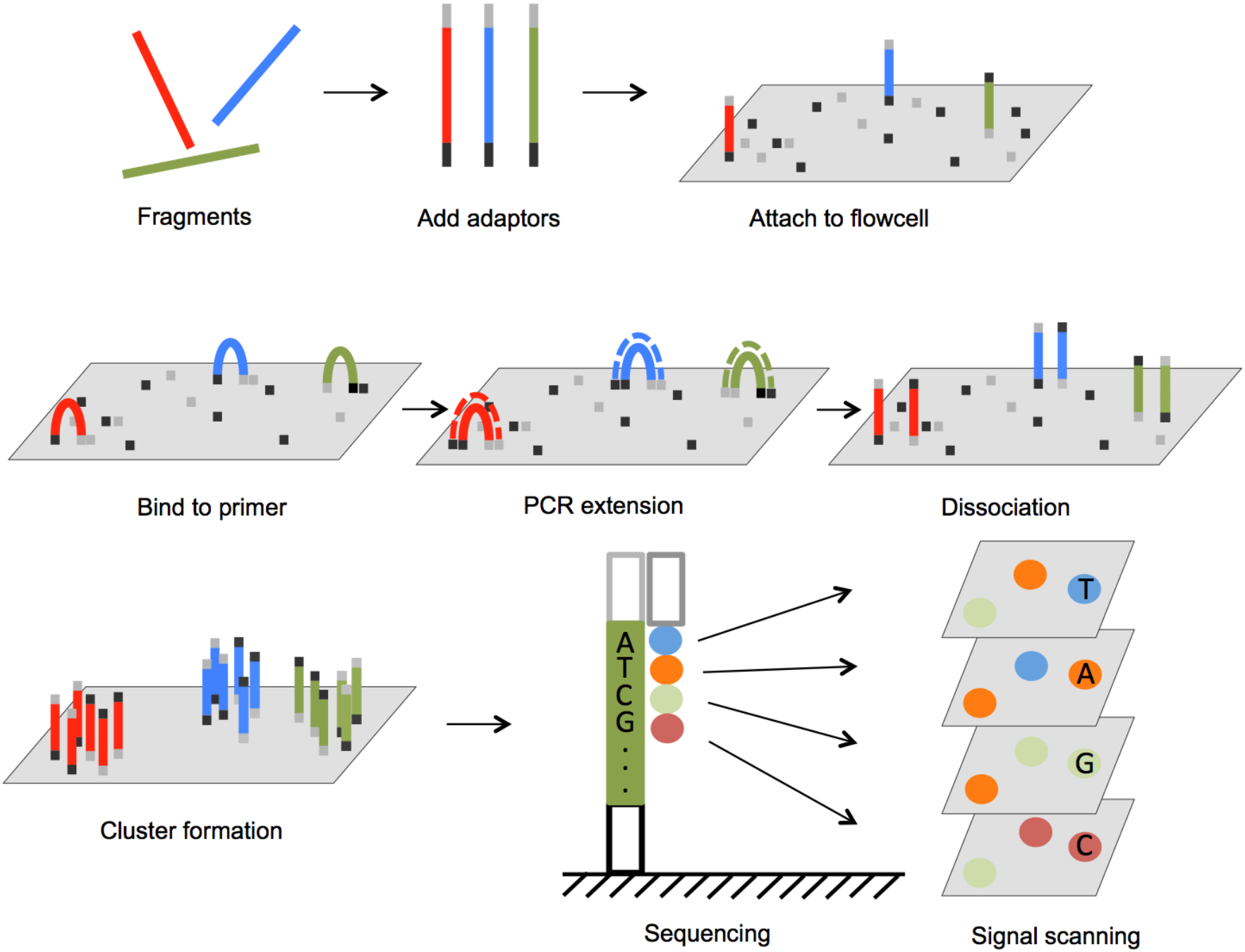



Fingerprint clone contig
Pick clones for sequencing
A
B
Sequenced-clone contig
Sequenced-clone-contig scaffold
Sequence to at least draft coverage
Sequenced clone A
Sequenced clone B
Initial sequence contig
Merge data
Merged sequence contig
Order and orient with mRNA, paired end reads, other information
Sequence-contig scaffold


STS
Genome
Mapped
Scaffolds:
Scaffold:
Read pair (mates)
Gap (mean & std. dev. Known)
Contig:
Consensus
Reads (of several haplotypes)
SNPs
BAC Fragments

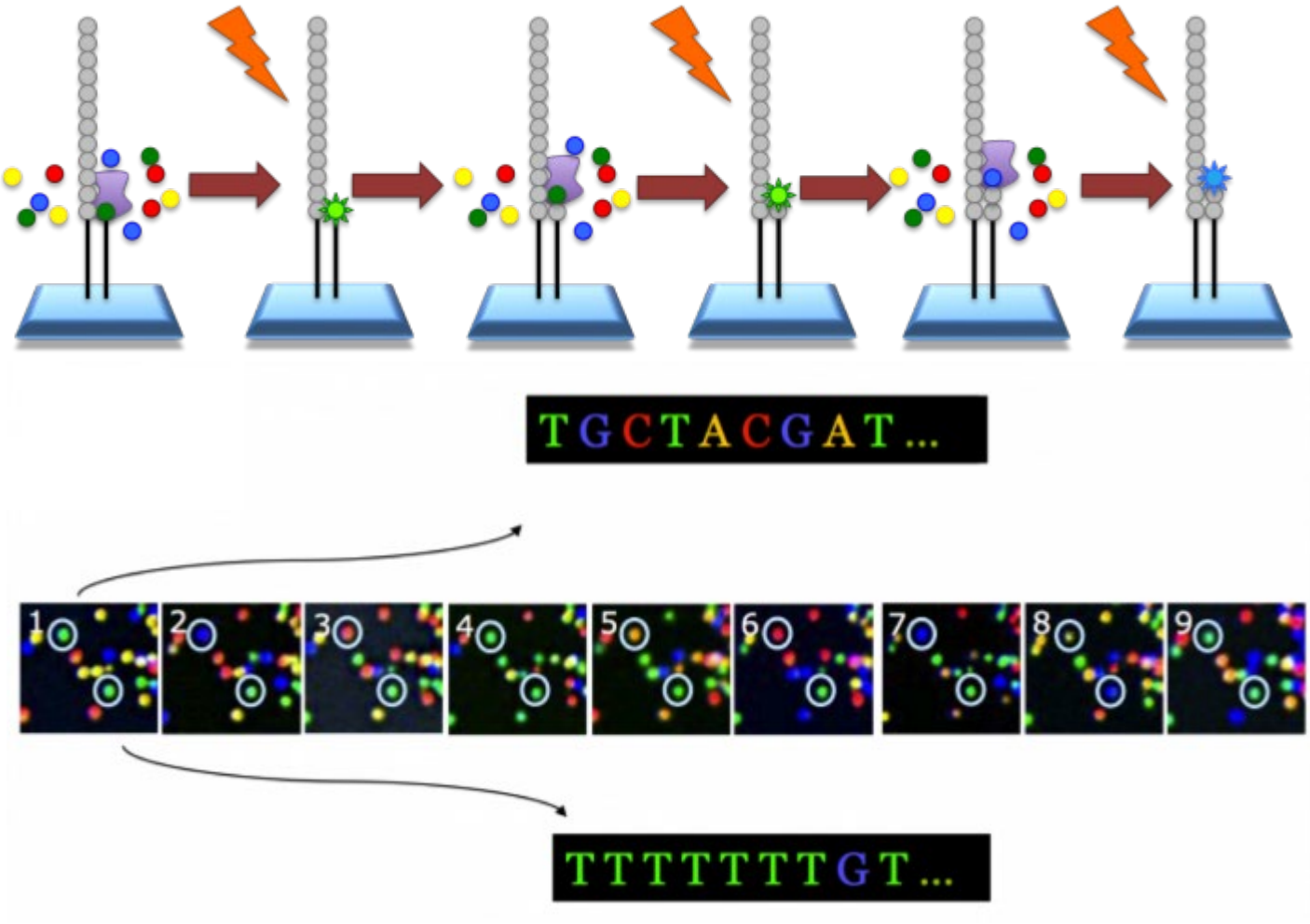
NGS Illumina




Fragments
Add adaptors
Attach to flowcell
Bind to primer
PCR extension
Dissociation
Cluster formation
ATCG...
Sequencing
T
A
G
C
Signal scanning

NGS Illumina

https://www.youtube.com/watch?v=fCd6B5HRaZ8



The four main advantages of NGS over classical Sanger sequencing are:

speed
cost
sample size
accuracy

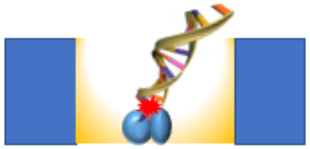
NGS is significantly cheaper, quicker, needs significantly less DNA and is more accurate and reliable than Sanger sequencing.

NGS is quicker than Sanger sequencing in two ways. Firstly, the chemical reaction may be combined with the signal detection in some versions of NGS, whereas in Sanger sequencing these are two separate processes. Secondly and more significantly, only one read (maximum ~1kb) can be taken at a time in Sanger sequencing, whereas NGS is massively parallel, allowing 300Gb of DNA to be read on a single run on a single chip.
The first human genome sequence cost in the region of £300M. Using modern Sanger sequencing methods, aided by data from the known sequence, a full human genome would still cost £6M. Sequencing a human genome with Illumina today would cost only £6,000.

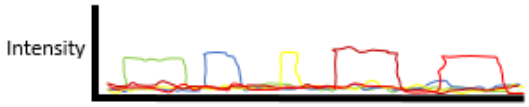
Third Generation Sequencing

PacBio SMRT seq

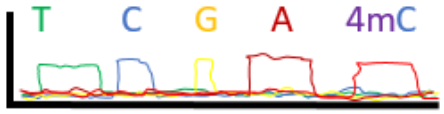
DNA passes thru polymerase in an illuminated volume



Raw output is fluorescent signal of the nucleotide incorporation, specific to each nucleotide

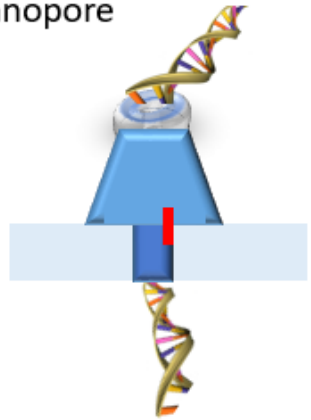


A,C,T,G have known pulse durations, which are used to infer methylated nucleotides

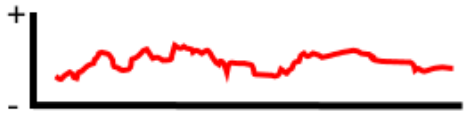


Oxford Nanopore

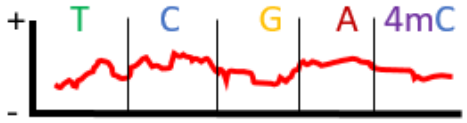
DNA passes thru nanopore



Raw output is electrical signal caused by nucleotide blocking ion flow in nanopore

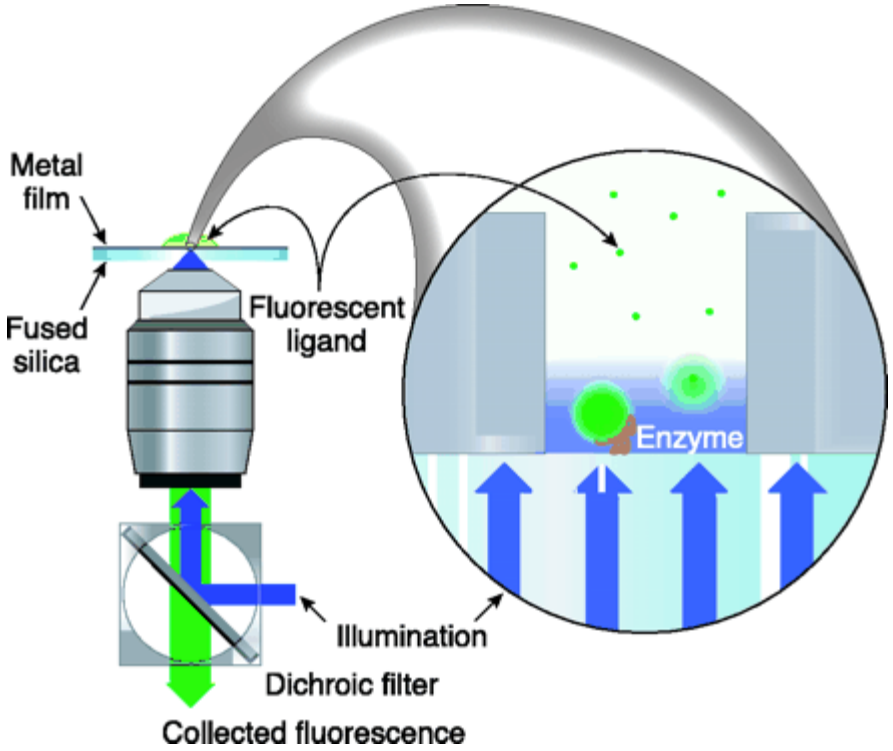


Each nucleotide has a specific electric “signature”



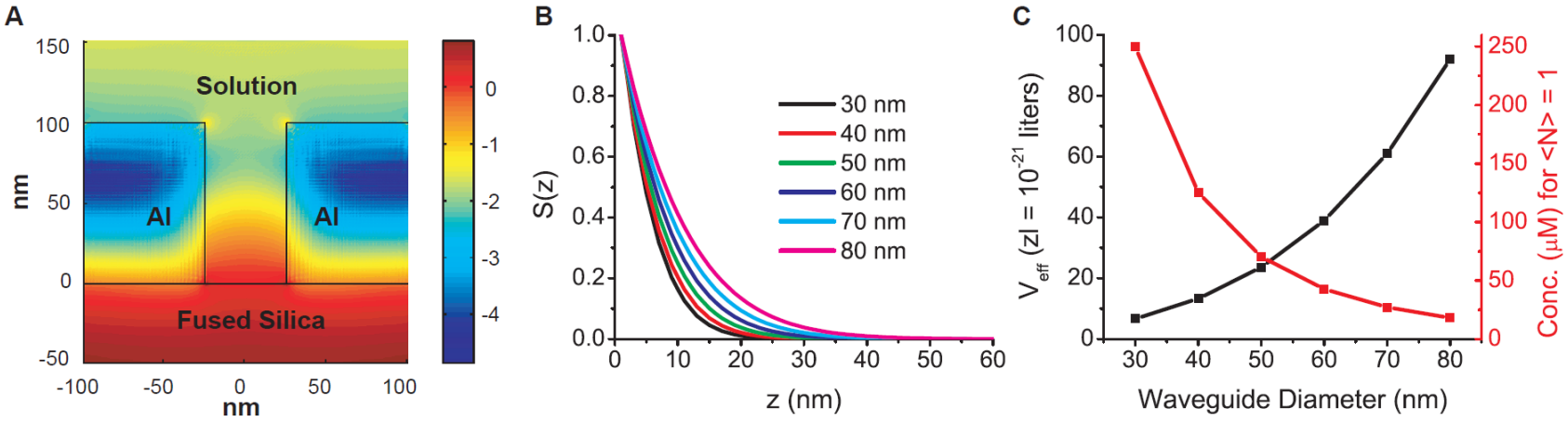
Zero Mode Waveguide

https://www.youtube.com/watch?v=NHCJ8PtYCFc



SCIENCE VOL 299 31 JANUARY 2003

Zero Mode Waveguide



50 x 50 x 10 nm^3 = 2.5 x 10^4 x 10^{-21} cc = 2.5 x 10^{-20} L= 25 zeptoliter

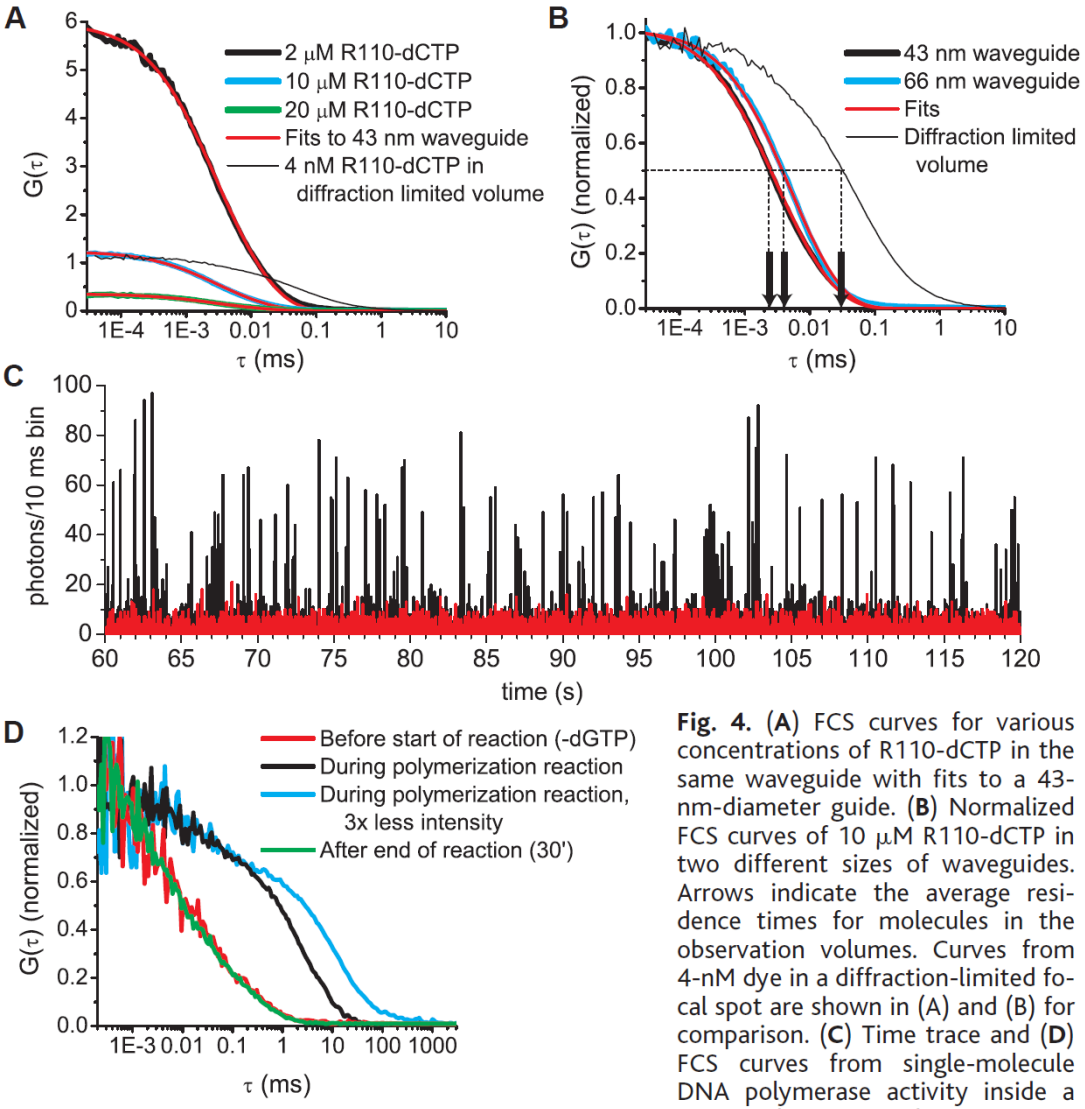
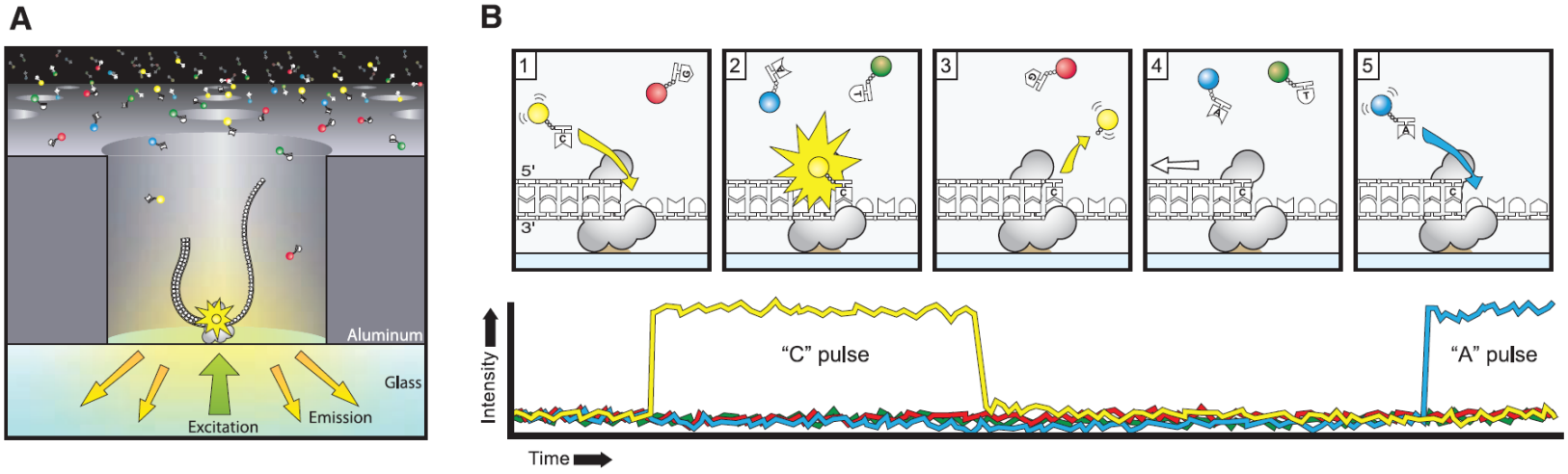


Fig. 4. (**A**) FCS curves for various concentrations of R110-dCTP in the same waveguide with fits to a 43-nm-diameter guide. (**B**) Normalized FCS curves of 10 μM R110-dCTP in two different sizes of waveguides. Arrows indicate the average residence times for molecules in the observation volumes. Curves from 4-nM dye in a diffraction-limited focal spot are shown in (A) and (B) for comparison. (**C**) Time trace and (**D**) FCS curves from single-molecule DNA polymerase activity inside a zero-mode waveguide. Incorporation events and subsequent photobleaching of coumarin-dCTP appear as distinct fluorescence bursts in the black time trace (10-ms time bins). This results in a long-time shoulder in the corresponding FCS curves during polymerization (black and blue curves) in (D). Decreasing the intensity results in slower photobleaching as seen by the longer residence time in the blue curve. The red curves in (C) and (D) are the corresponding negative controls (absence of one native nucleotide) in the same waveguide before initiation of the reaction. The green curve in (D) is the control after the reaction has stopped.

Real-Time DNA Sequencing from Single Polymerase Molecules

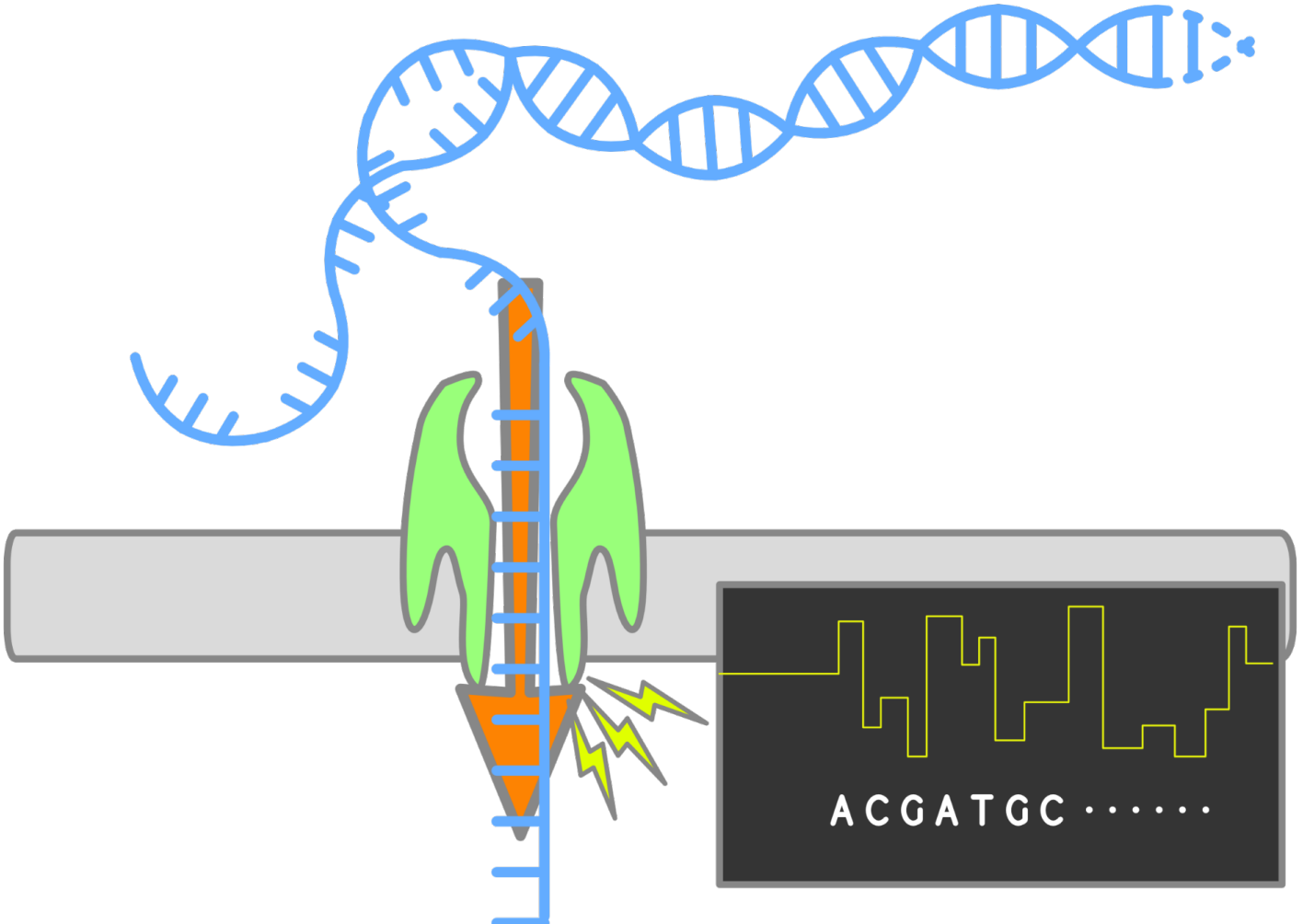


SCIENCE VOL 323 2 JANUARY 2009

Nanopore Sequencing

https://www.youtube.com/watch?v=RcP85JHLmnI

https://www.youtube.com/watch?v=qzusVw4Dp8w



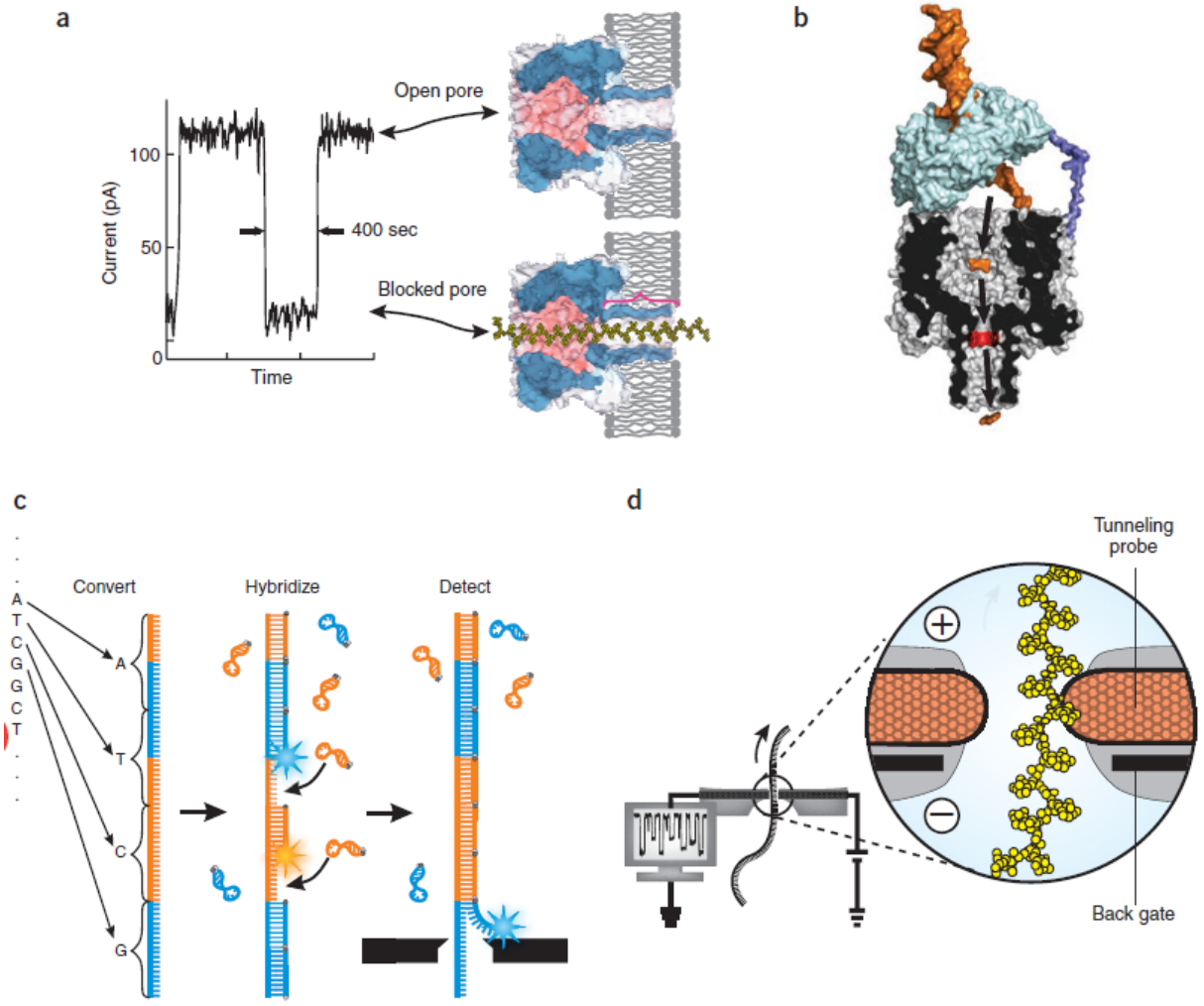

a
Open pore
100
50
0
Current (pA)
400 sec
Blocked pore
Time
b
c
Convert
Hybridize
Detect
A
T
C
G
G
C
T
A
T
C
G
d
Tunneling probe
Back gate

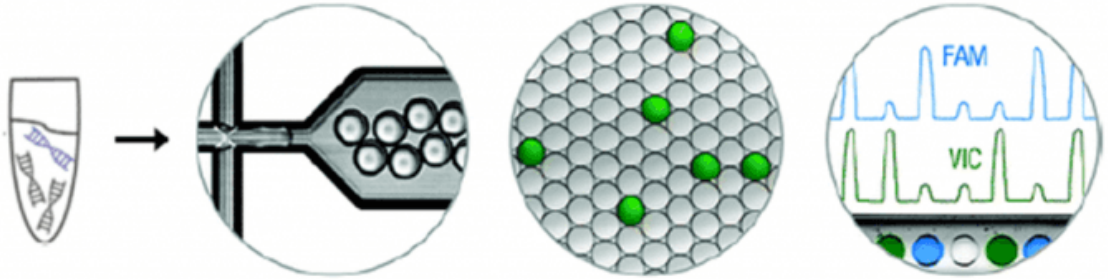
Table 2 Comparison of first-, second-, and third-generation genomic sequencing

	First generation	Second generation	Third generation
Fundamental technology	Size-separation of specifically end-labeled DNA fragments	Wash-and-scan SBS	Single molecule real time sequencing
Resolution	Averaged across many copies of the DNA molecule	Averaged across many copies of the DNA molecule	Single DNA molecule
Current raw read accuracy	High	High	Lower
Current read length	Moderate (800-1000 bp)	Short (generally much shorter than Sanger sequencing)	> 1000 bp
Current throughput	Low	High	High
Current cost	High cost per base, Low cost per run	Low cost per base, High cost per run	Low cost per base, High cost per run
RNA-sequencing method	cDNA sequencing	cDNA sequencing	Direct RNA sequencing
Time to result	Hours	Days	< 1 day
Sample preparation	Moderately complex, PCR amplification is not required	Complex, PCR amplification is required	Various
Data analysis	Routine	Complex (due to large data volumes & short reads)	Complex
Primary results	Base calls with quality values	Base calls with quality values	Base calls with quality values

Adapted from Schadt, et al. Hum Mol Genet 2010[13]

Digital PCR

Droplet digital PCR



Bio-Rad QX100

1. MAKE	2. CYCLE	3. READ	
Sample is partitioned into 20,000 droplets	Run PCR cycles in all droplets simultaneously	Measure fluorescence intensity in each droplet	Calculate concentration from number of positive droplets

Genome Size

AT = 615.3830
GC = 616.3711
the mean relative weight of one nucleotide pair is 615.8771

1au = 1.660539×10^{-27} kg
one nucleotide pair would be 1.023×10^{-9} pg

human female and male diploid genome sizes are 6.406×10^{9} bp

one cell contains 6.550 and 6.436 pg of DNA

10-30 pg RNA per cell
mRNA = 1-5% of total RNA (rRNA, rRNA)

Single Cell Analysis

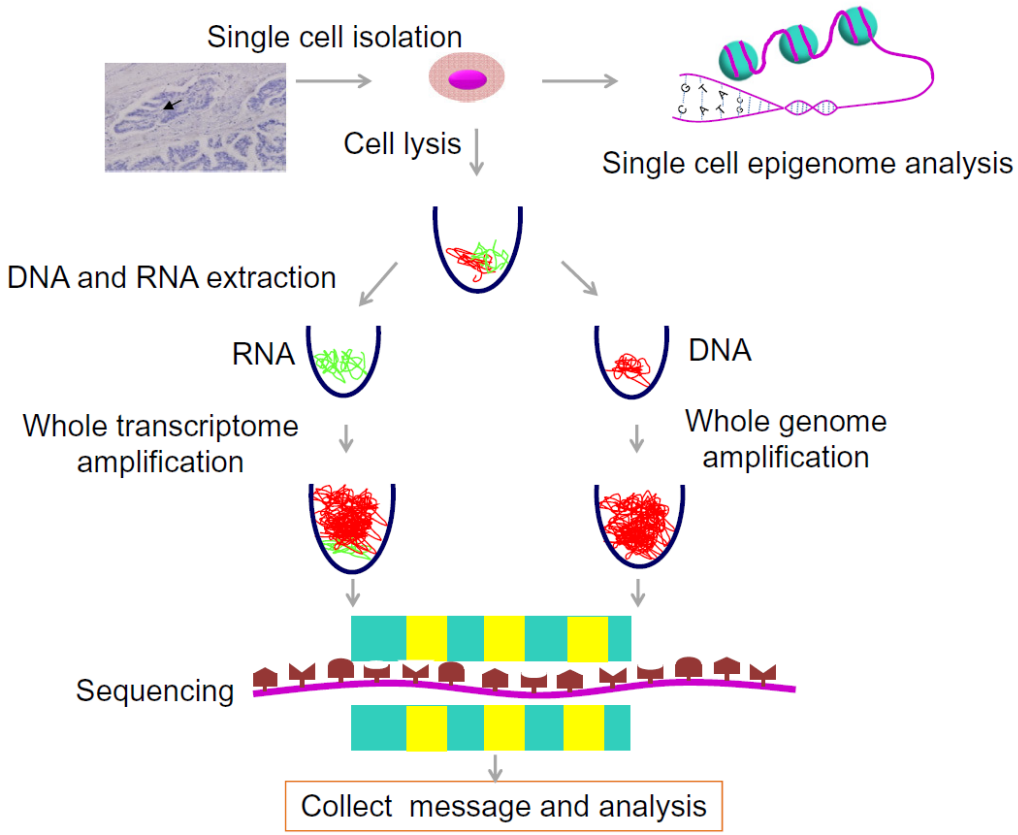
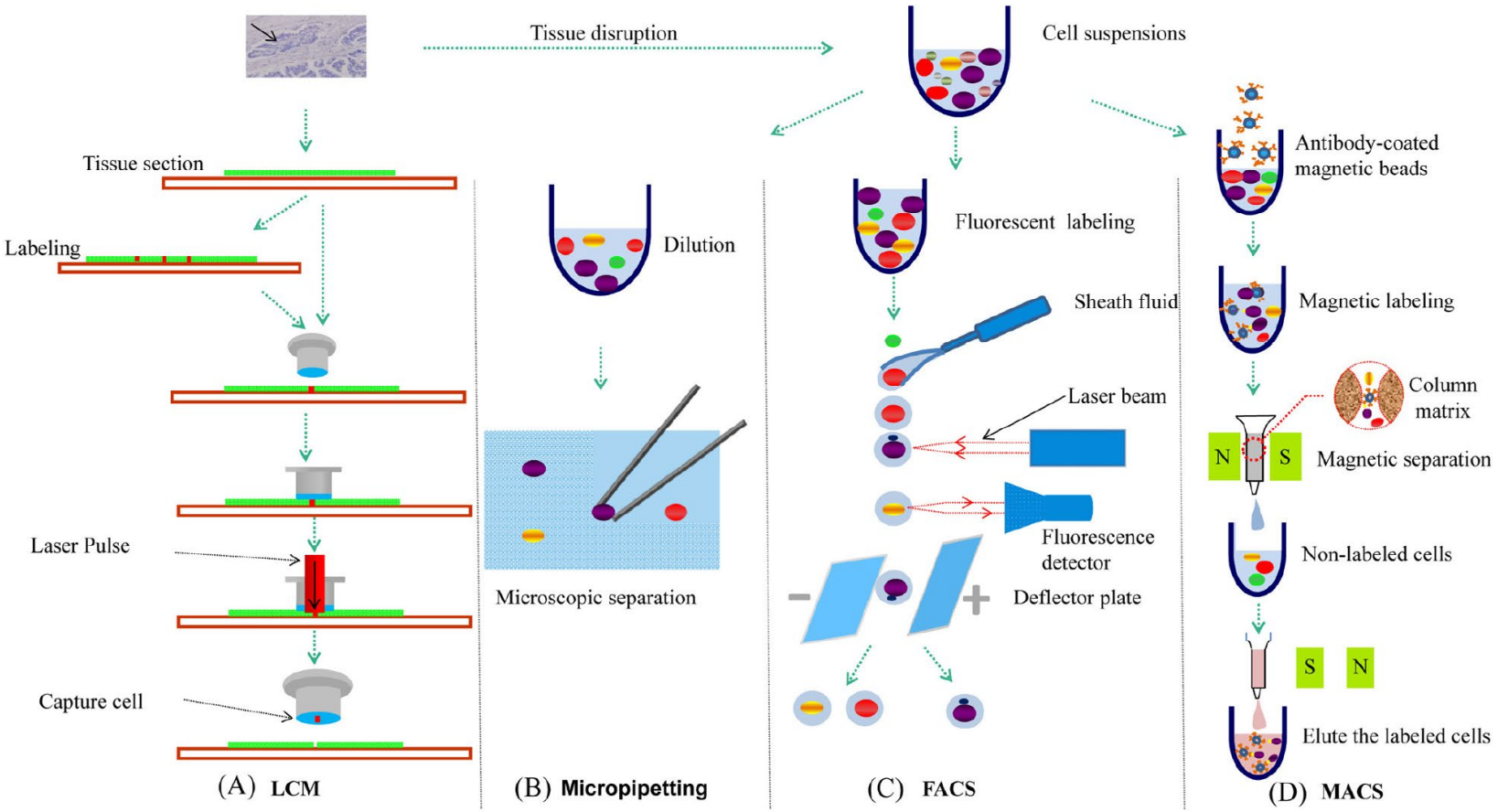
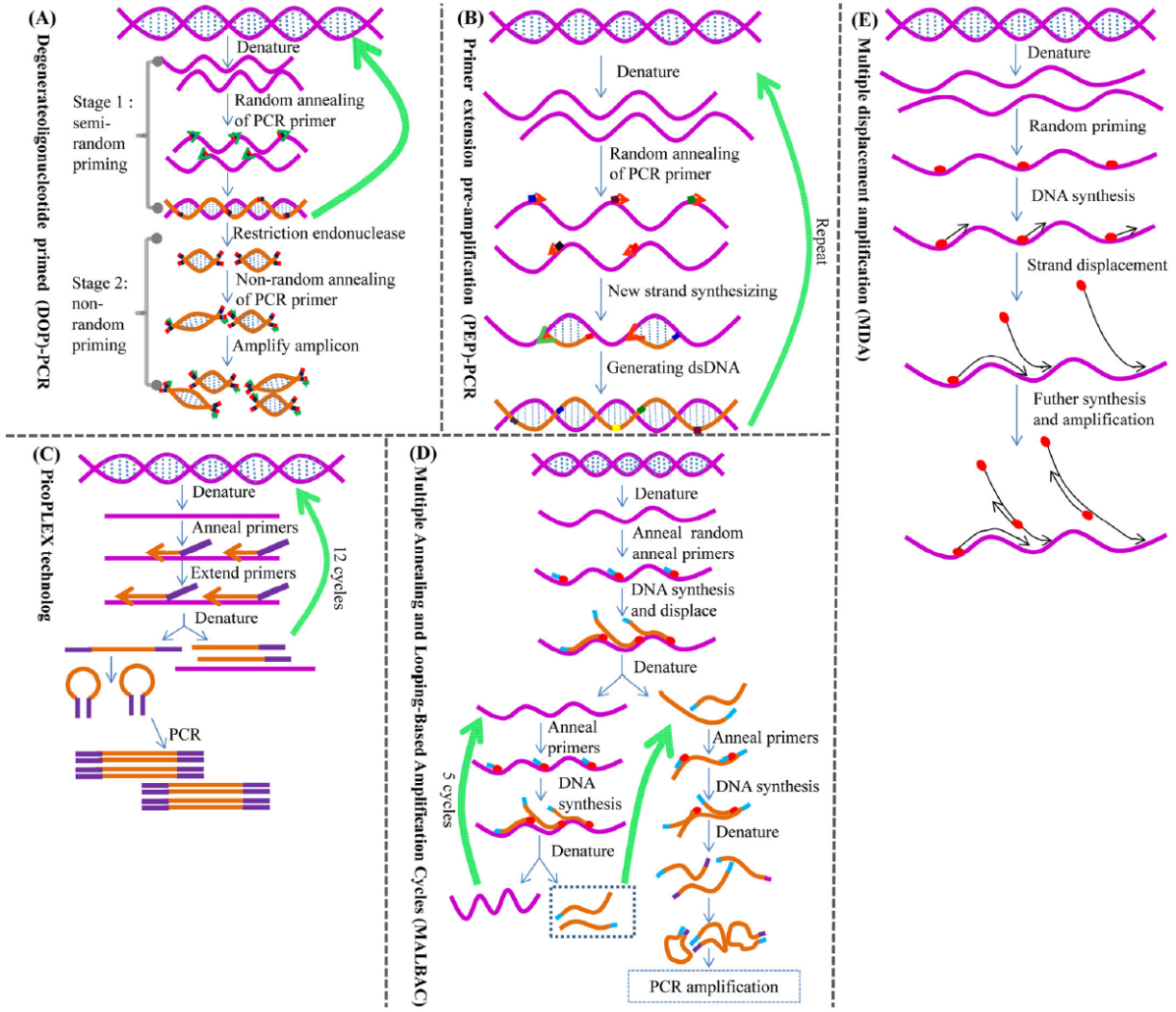


Fig. 1. Single-cell sequencing of a tumor cell. A tumor specimen is obtained by surgical excision and single cells are isolated by one of the several methods shown in Fig. 2. The individual cancer cell can be used for epigenome sequencing directly or lysed to extract the genetic material (DNA and RNA), which is in turn amplified by the methods shown in Fig. 3. Then, the amplified DNA and RNA are sequenced by single-cell sequencing technology and the result data are analyzed to provide insights into the molecular mechanisms underlying intratumor heterogeneity.

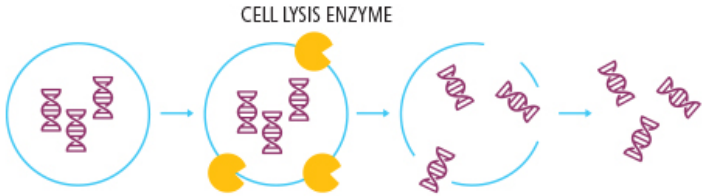
Isolation of Single Cells



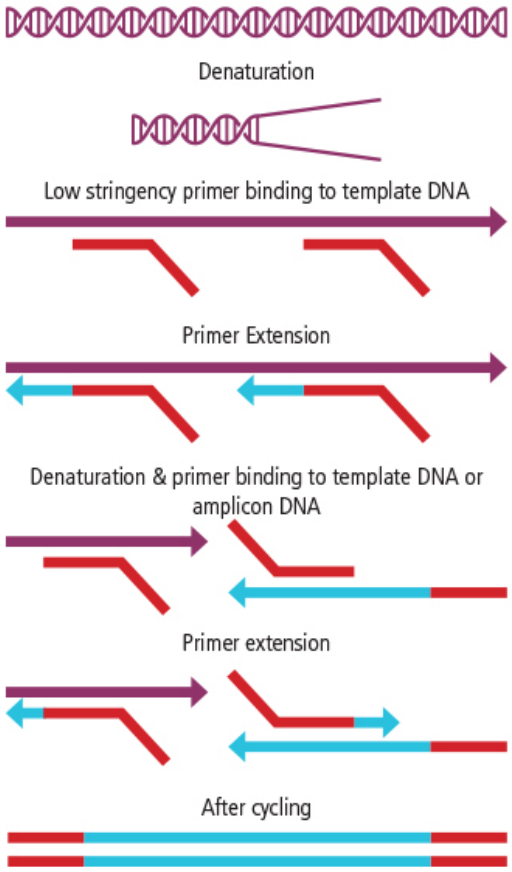
Whole Genome Amplifications



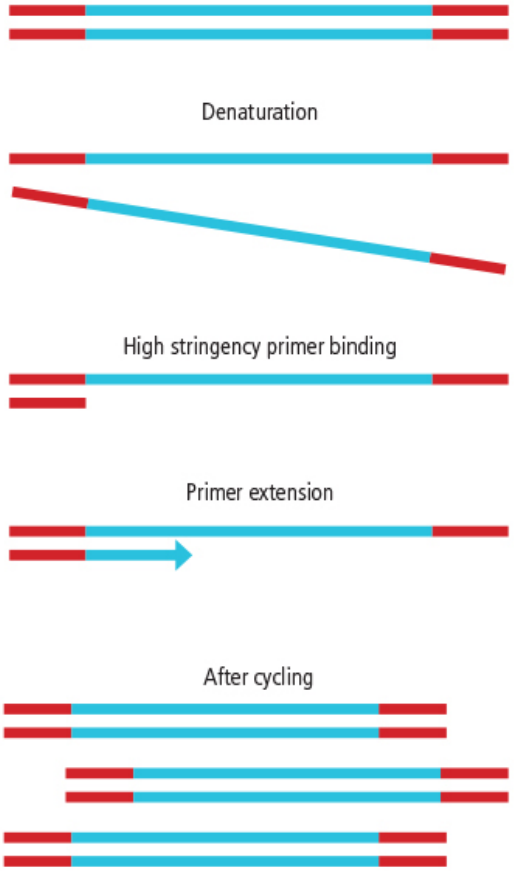
STEP 1: CELL LYSIS



STEP 2: LOW STRINGENCY PCR CYCLES



STEP 3: HIGH STRINGENCY PCR CYCLES



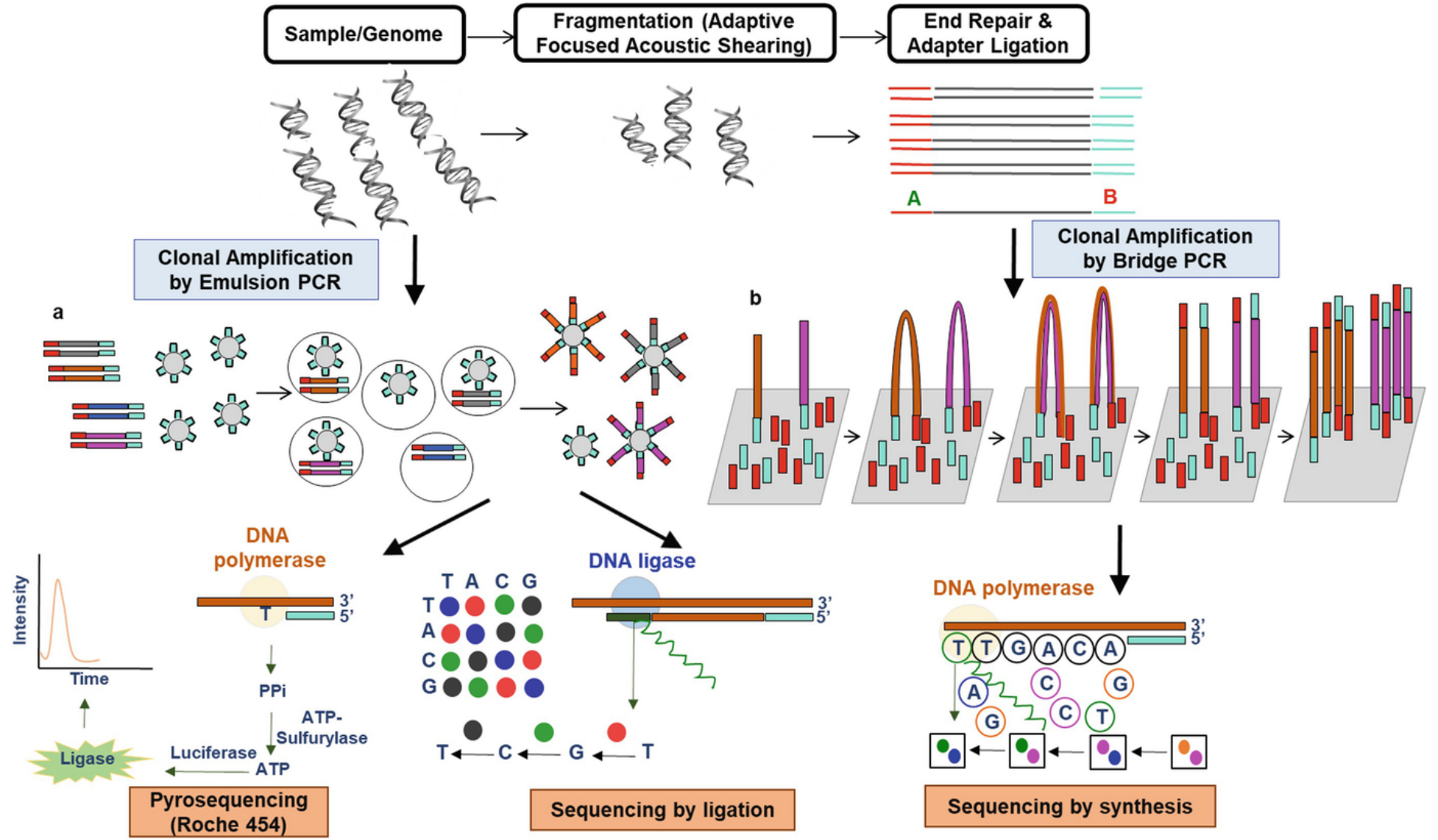

Sample/Genome
Fragmentation (Adaptive Focused Acoustic Shearing)
End Repair & Adapter Ligation
A
B
Clonal Amplification by Emulsion PCR
Clonal Amplification by Bridge PCR
a
b
DNA polymerase
DNA ligase
DNA polymerase
Intensity
Time
PPi
ATP-Sulfurylase
Luciferase
ATP
Ligase
Pyrosequencing (Roche 454)
Sequencing by ligation
Sequencing by synthesis

Table 2
Methods for amplifying RNA for whole transcriptome analysis.

Method	Year	Principle	Coverage	Transcript lengths	Limitation	Ref
Tang's method	2009	Traditional PCR	64%	0.5~3 kb	3′ end bias	[6,14,70]
STRT-seq	2011	Modified PCR	5–25%	0.75~2.0 kb	Strong 3′ end bias	[3,69,73,79]
CEL-seq	2012	*In vitro* transcription	49%	Average 1.0 kb	Strong 3′ end bias, usually targets the last exons highly	[14,69,75,76,80]
Smart-seq	2012	Modified PCR	Nearly full-length	Average 1.5 kb	Cannot capture partially reverse-transcribed mRNA	[69,80–82]
PMA	2013	Phi29 DNA polymerase	Full-length	All sizes	Slight 5′ end bias	[69,78]
SMA	2013	Traditional PCR	Full-length	All sizes	Slight 5′ end bias	[69,78]
Quartz–seq	2013	*In vitro* transcription	81%	Average 2.5 kb	5′ end bias, cannot amplify cDNA with higher GC content	[69,76,83]
Smart-seq2	2014	Modified PCR	Nearly full-length	~2 kb	Cannot capture poly A^- RNA and reflects strand-specific mRNA	[6,69,84]

Abbreviations: STRT, single-cell tagged reverse transcription; CEL-seq, cell expression by linear amplification and sequencing; Smart-seq, switching mechanism at the 5′ end of the RNA template sequencing; PMA, Phi29 DNA polymerase-based mRNA transcriptome amplification; SMA, semi-random primed PCR-based mRNA transcriptome amplification procedure.

Integrative single-cell analysis

Tim Stuart[1] *and Rahul Satija*[1,2]*

Abstract | The recent maturation of single-cell RNA sequencing (scRNA-seq) technologies has coincided with transformative new methods to profile genetic, epigenetic, spatial, proteomic and lineage information in individual cells. This provides unique opportunities, alongside computational challenges, for integrative methods that can jointly learn across multiple types of data. Integrated analysis can discover relationships across cellular modalities, learn a holistic representation of the cell state, and enable the pooling of data sets produced across individuals and technologies. In this Review, we discuss the recent advances in the collection and integration of different data types at single-cell resolution with a focus on the integration of gene expression data with other types of single-cell measurement.

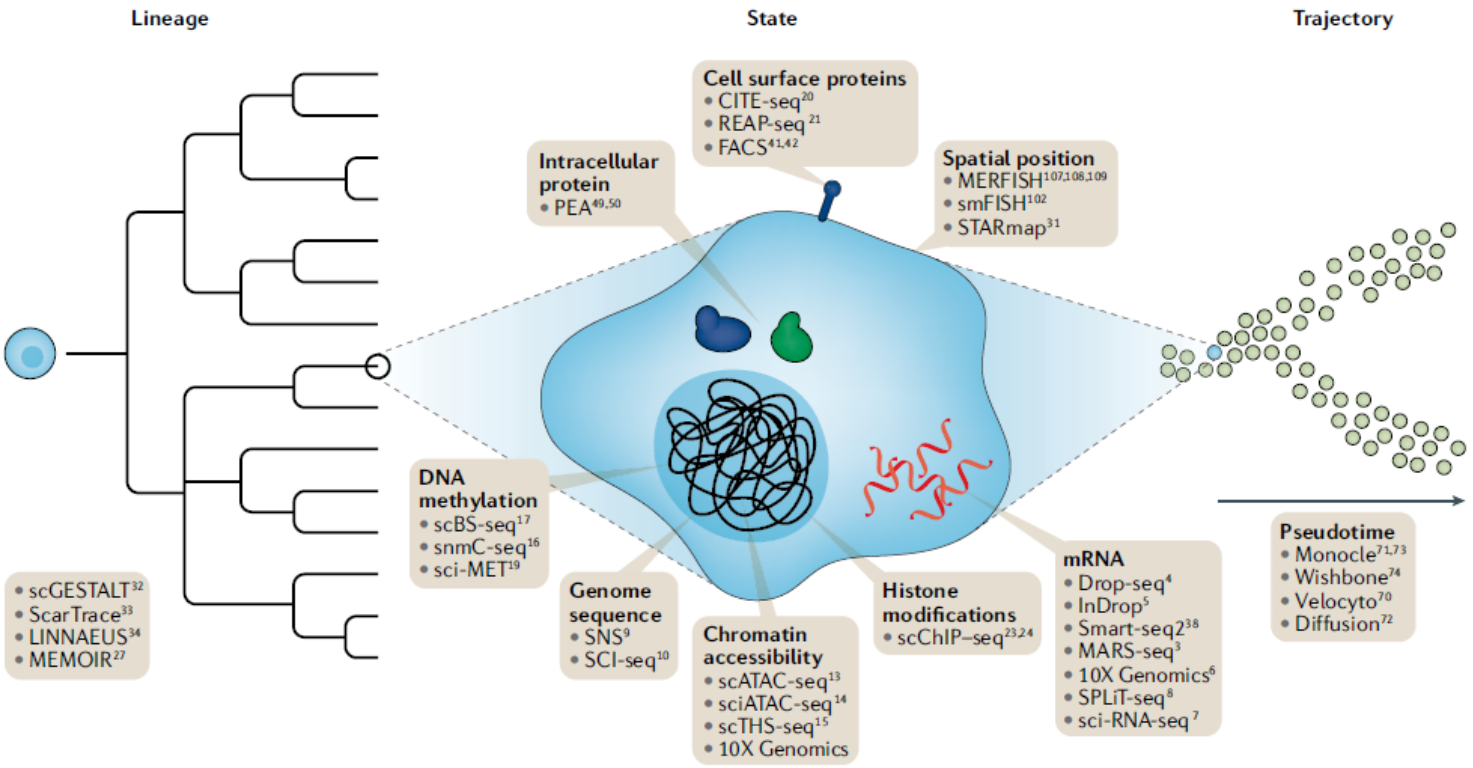
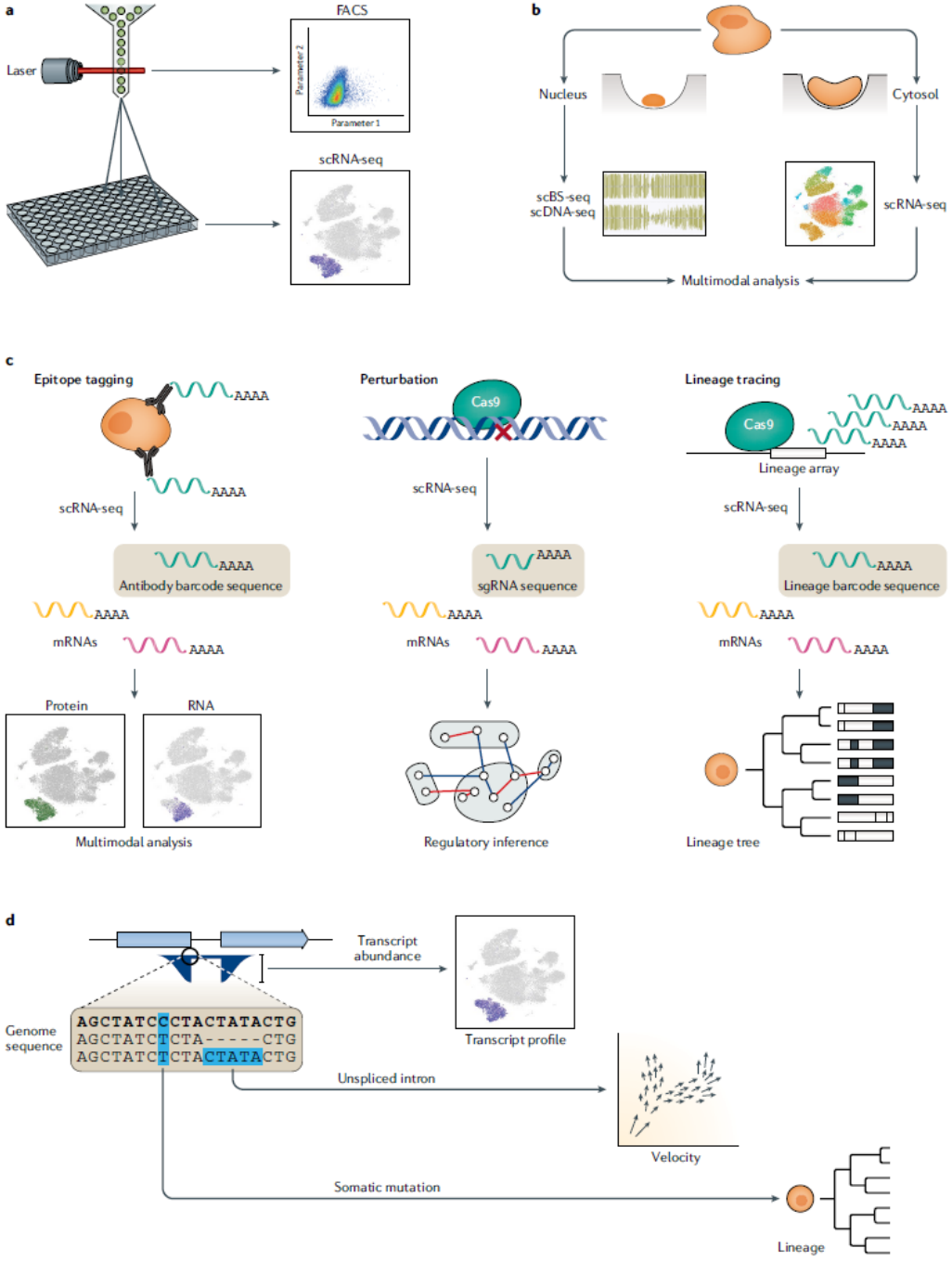
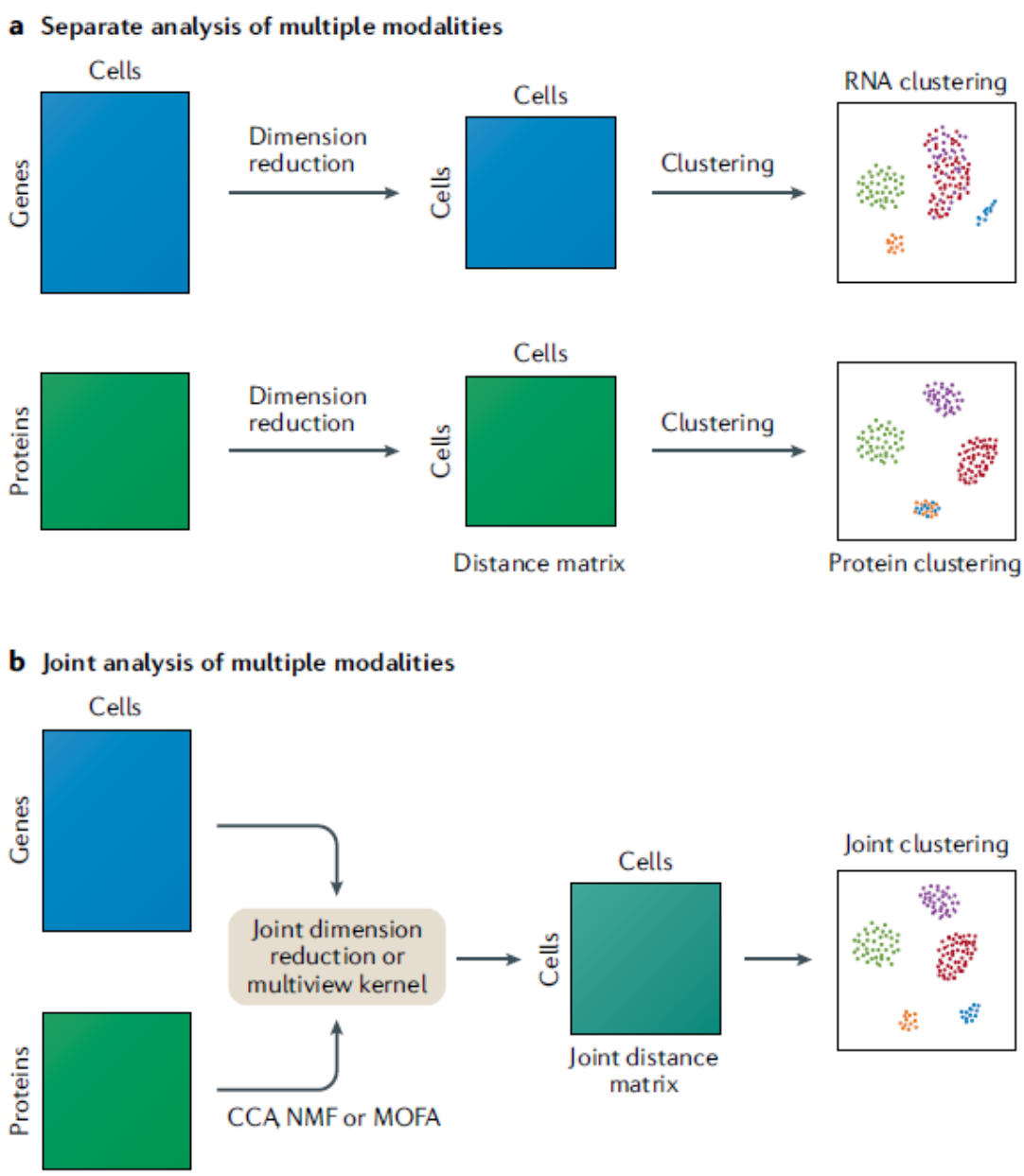


Table 1 | **Current experimental methods for unimodal and multimodal single-cell measurements**

Data types	Method name	Feature throughput	Cell throughput	Refs
Unimodal				
mRNA	Drop-seq	Whole transcriptome	1,000–10,000	4
	InDrop	Whole transcriptome	1,000–10,000	5
	10X Genomics	Whole transcriptome	1,000–10,000	6
	Smart-seq2	Whole transcriptome	100–300	38
	MARS-seq	Whole transcriptome	100–300	3
	CEL-seq	Whole transcriptome	100–300	1
	SPLiT-seq	Whole transcriptome	≥50,000	8
	sci-RNA-seq	Whole transcriptome	≥50,000	7
Genome sequence	SNS	Whole genome	10–100	9
	SCI-seq	Whole genome	10,000–20,000	10
Chromatin accessibility	scATAC-seq	Whole genome	1,000–2,000	13
	sciATAC-seq	Whole genome	10,000–20,000	14
	scTHS-seq	Whole genome	10,000–20,000	15
DNA methylation	scBS-seq	Whole genome	5–20	17
	snmC-seq	Whole genome	1,000–5,000	16
	sci-MET	Whole genome	1,000–5,000	19
	scRRBS	Reduced representation genome	1–10	18
Histone modifications	scChIP–seq	Whole genome + single modification	1,000–10,000	24
Chromosome conformation	scHi-C-seq	Whole genome	1–10	26
Multimodal				
Histone modifications + spatial	NA	Single locus + single modification	10–100	23
mRNA + lineage	scGESTALT	Whole transcriptome	1,000–10,000	32
	ScarTrace	Whole transcriptome	1,000–10,000	33
	LINNAEUS	Whole transcriptome	1,000–10,000	34
Lineage + spatial	MEMOIR	NA	10–100	27
mRNA + spatial	osmFISH	10–50 RNAs	1,000–5,000	35
	STARmap	20–1,000 RNAs	100–30,000	31
	MERFISH	100–1,000 RNAs	100–40,000	100
	seqFish	125–250 RNAs	100–20,000	29
mRNA + cell surface protein	CITE-seq	Whole transcriptome + proteins	1,000–10,000	20
	REAP-seq	Whole transcriptome + proteins	1,000–10,000	21
mRNA + chromatin accessibility	sci-CAR	Whole transcriptome + whole genome	1,000–20,000	48
mRNA + DNA methylation	scM&T-seq	Whole genome	50–100	46
mRNA + genomic DNA	G&T-seq	Whole genome + whole transcriptome	50–200	44
mRNA + intracellular protein	NA	96 mRNAs + 38 proteins	50–100	50
		82 mRNAs + 75 proteins	50–200	49
DNA methylation + chromatin accessibility	scNOMe-seq	Whole genome	10–20	11


a
Laser
FACS
Parameter 2
Parameter 1
scRNA-seq
b
Nucleus
Cytosol
scBS-seq
scDNA-seq
scRNA-seq
Multimodal analysis
c
Epitope tagging
AAAA
scRNA-seq
Antibody barcode sequence
mRNAs
Protein
RNA
Multimodal analysis
Perturbation
Cas9
scRNA-seq
sgRNA sequence
mRNAs
Regulatory inference
Lineage tracing
Cas9
Lineage array
scRNA-seq
Lineage barcode sequence
mRNAs
Lineage tree
d
Transcript abundance
Transcript profile
Genome sequence
AGCTATCCCTACTATACTG
AGCTATCTCTA-----CTG
AGCTATCTCTACTATACTG
Unspliced intron
Velocity
Somatic mutation
Lineage


a Separate analysis of multiple modalities
Cells
Genes
Dimension reduction
Cells
Cells
Clustering
RNA clustering
Proteins
Dimension reduction
Cells
Cells
Clustering
Distance matrix
Protein clustering
b Joint analysis of multiple modalities
Cells
Genes
Joint dimension reduction or multiview kernel
Cells
Cells
Joint clustering
Joint distance matrix
Proteins
CCA, NMF or MOFA

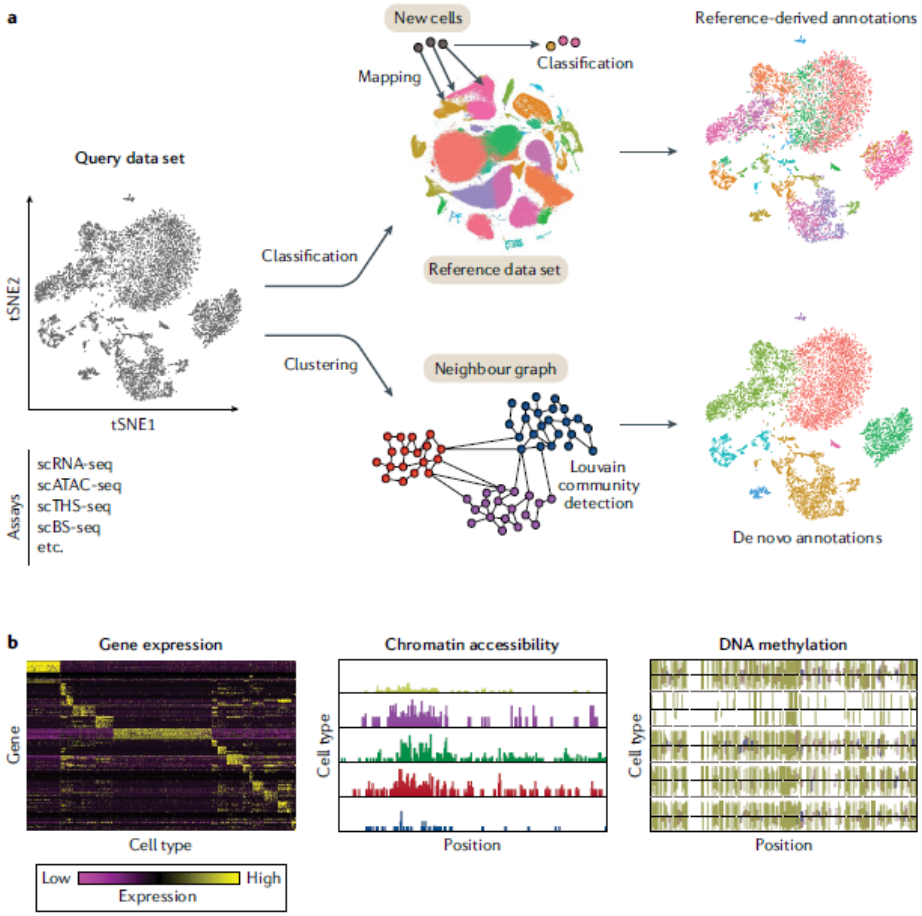


Fig. 5 | **Clustering and classification of cells. a** | The classification of cells in a new data set can be achieved in one of two ways. Cells can be classified on the basis of existing cell-type annotations in a reference data set derived from a similar population of cells, or unsupervised clustering can be performed to identify groups of similar cells. By classifying cells according to reference annotations, more subtle divisions between groups of cells may be identified by leveraging the structure present in a higher quality data set. **b** | By grouping cells on the basis of common properties, through either a reference-based classification or de novo clustering, the differences between groups of cells can be further analysed in different ways depending on the type of data measured in the cells. Gene expression data can be used to identify differentially expressed genes between cell types or clusters, and chromatin accessibility can be used to identify differentially accessible regions and enriched DNA motifs. Similarly, single-cell DNA methylation data can be used to identify differentially methylated regions between cell types or clusters. scATAC-seq, single-cell assay for transposase-accessible chromatin using sequencing; scBS-seq, single-cell bisulfite sequencing; scRNA-seq, single-cell RNA sequencing; scTHS-seq, single-cell transposome hypersensitivity site sequencing; tSNE, *t*-distributed stochastic neighbour embedding.

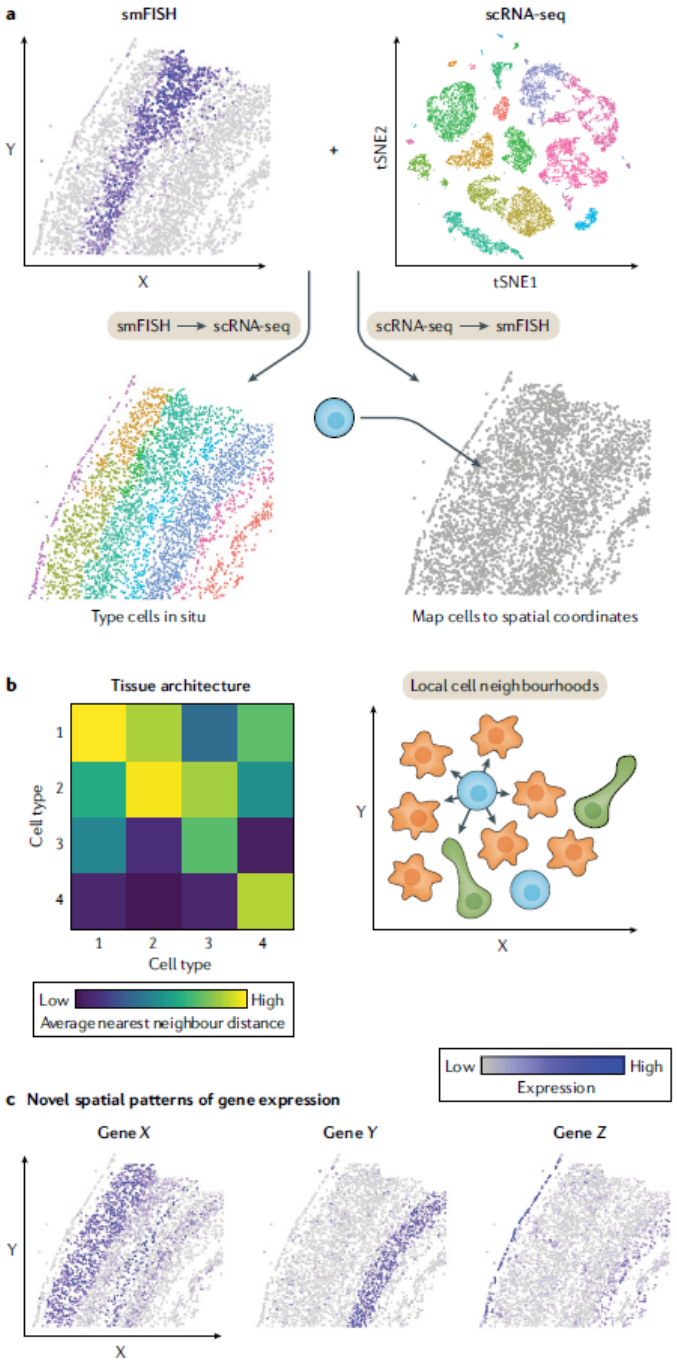
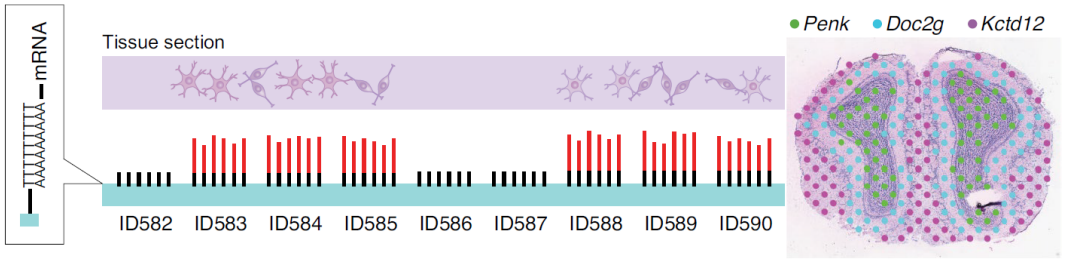


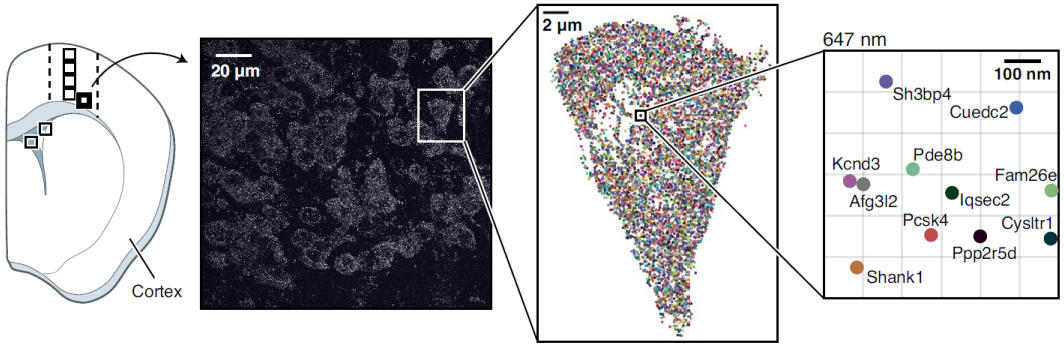
Fig. 6 | **Integration of spatial single-cell data.** **a** | The integration of single-molecule fluorescence in situ hybridization (smFISH) data with single-cell RNA sequencing (scRNA-seq) data can be done in two ways: mapping smFISH-profiled cells onto scRNA-seq clusters or mapping scRNA-seq-profiled cells onto spatially resolved smFISH data. Mapping smFISH cells onto scRNA-seq data allows the transfer of cell-type classifications derived from transcriptome-wide gene expression measurements to be transferred to the spatially resolved cells (left panel), whereas mapping scRNA-seq data onto smFISH-profiled spatial coordinates can allow scRNA-seq data from dissociated cells to be placed back into their spatial context (right panel). **b** | Following spatial integration, tissue architectures can be analysed to determine the cellular composition of tissues and the spatial relationships between cell types. One way of assessing how cell types are spatially organized in the tissue is to look at the local neighbourhood surrounding cells of each type. By measuring the average spatial distance between cell types, it is possible to learn characteristics about the tissue architecture, including which cell types are dispersed throughout the tissue and which cell types often form local neighbourhoods with another cell type, indicating a possible interaction. **c** | By mapping scRNA-seq-profiled cells onto spatially resolved coordinates through the integration with smFISH data, spatial patterns of gene expression can be predicted for any gene measured in the scRNA-seq data set. Through these predictions, novel spatial patterns of gene expression may be identified through the analysis of genes that were not profiled by smFISH. tSNE, *t*-distributed stochastic neighbour embedding.

Method of the Year 2020: spatially resolved transcriptomics

Spatially resolved transcriptomics methods are changing the way we understand complex tissues.



Researchers in Sweden developed an approach in which fixed, stained tissue is imaged, permeabilized and the mRNAs attach to an array of barcoded oligos. The RNAs are reverse-transcribed; the cDNAs are sequenced and yield spatially resolved transcriptomic information. Credit: Adapted with permission from ref. [4], AAAS



Spatial techniques help with atlas-building by localizing expressed genes. Here, seqFISH+ was used to measure 10,000 genes in mouse cortex. Credit: Cai lab, Caltech, I. Strazhnik; adapted with permission from ref. [6], Springer Nature.

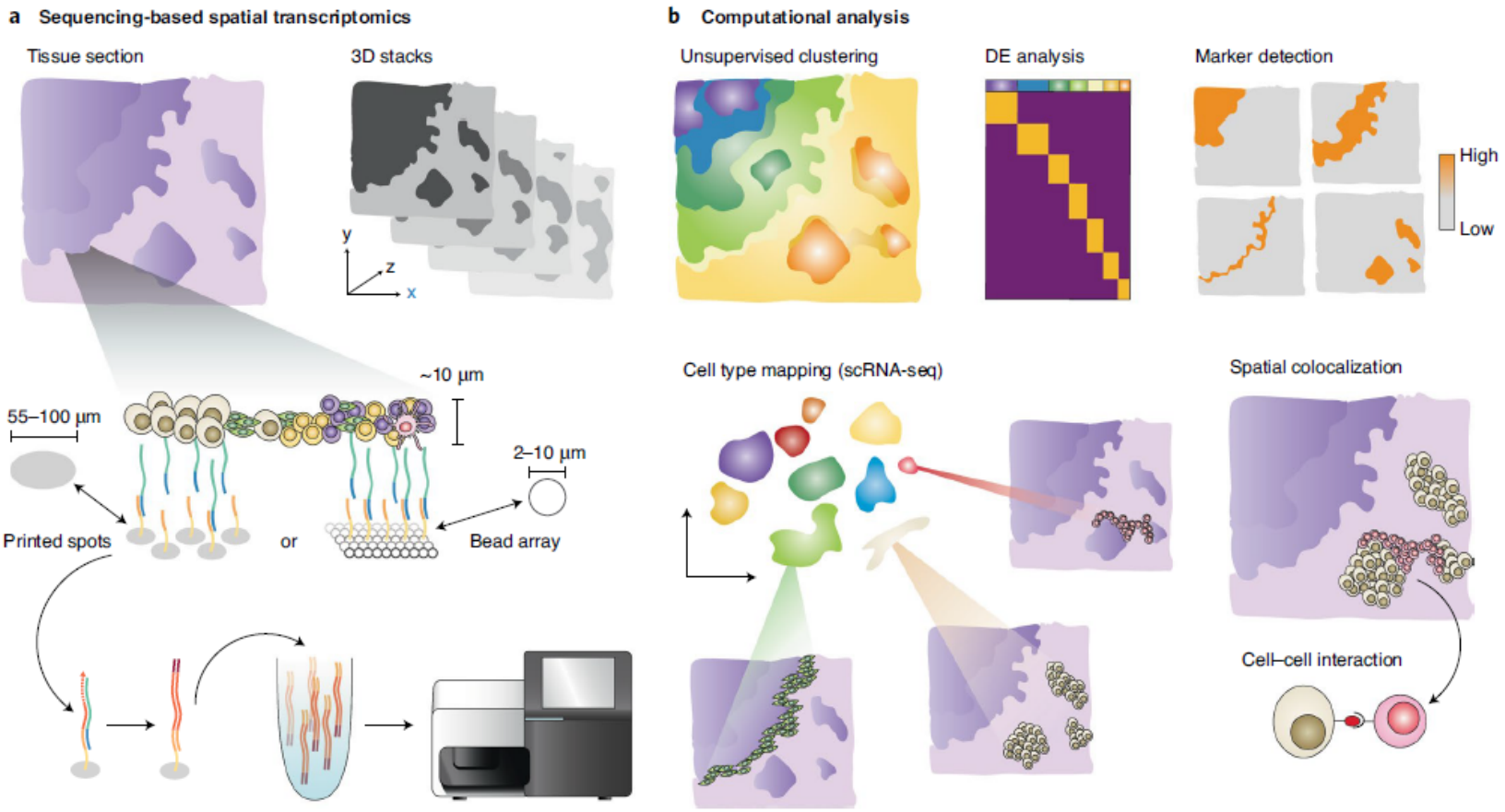


Fig. 1 | Experimental and computational methods for sequencing-based spatial transcriptomics. a, Sequencing-based, spatially resolved transcriptomics methods overview. Thin tissue sections are placed atop an array of spatially barcoded features, either densely packed beads (Slide-seq or HDST) or printed spots (ST or Visium). Multiple sections can be processed to create 3D models of tissue specimens. Permeabilization of the tissue releases the mRNA from the cells, allowing it to diffuse toward the surface and hybridize to the spatially barcoded features. Hybridized material is subsequently processed into cDNA and sequenced using next-generation sequencing. **b**, Computational analysis. A common approach is to define regions with distinct transcriptional profiles using unsupervised clustering and further characterize these regions with differential expression (DE) analysis[25–27]. Alternatively, region-specific markers can be found by querying genes expressed with a high spatial autocorrelation. Cell types and cell states can be mapped on the tissue through integration with scRNA-seq data[27–29], which in turn facilitates interaction analyses of colocalized cells.

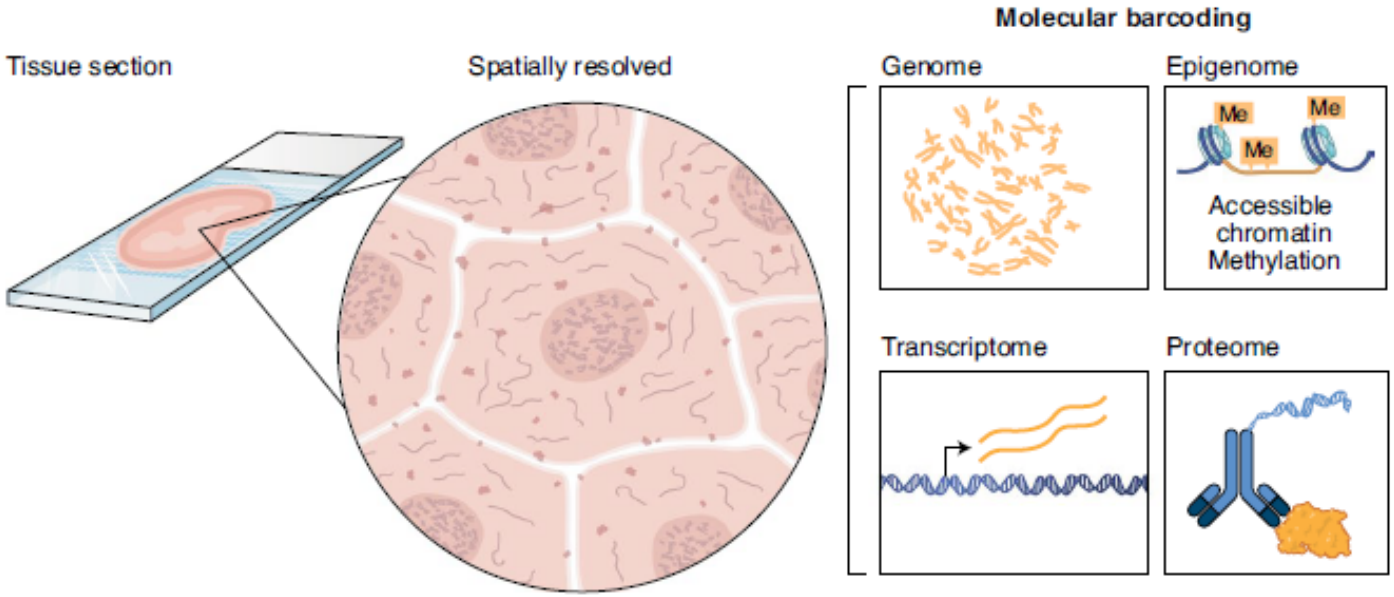


Fig. 2 | Next-generation spatial barcoding of biomolecules from tissue sections for spatial multi-omics.

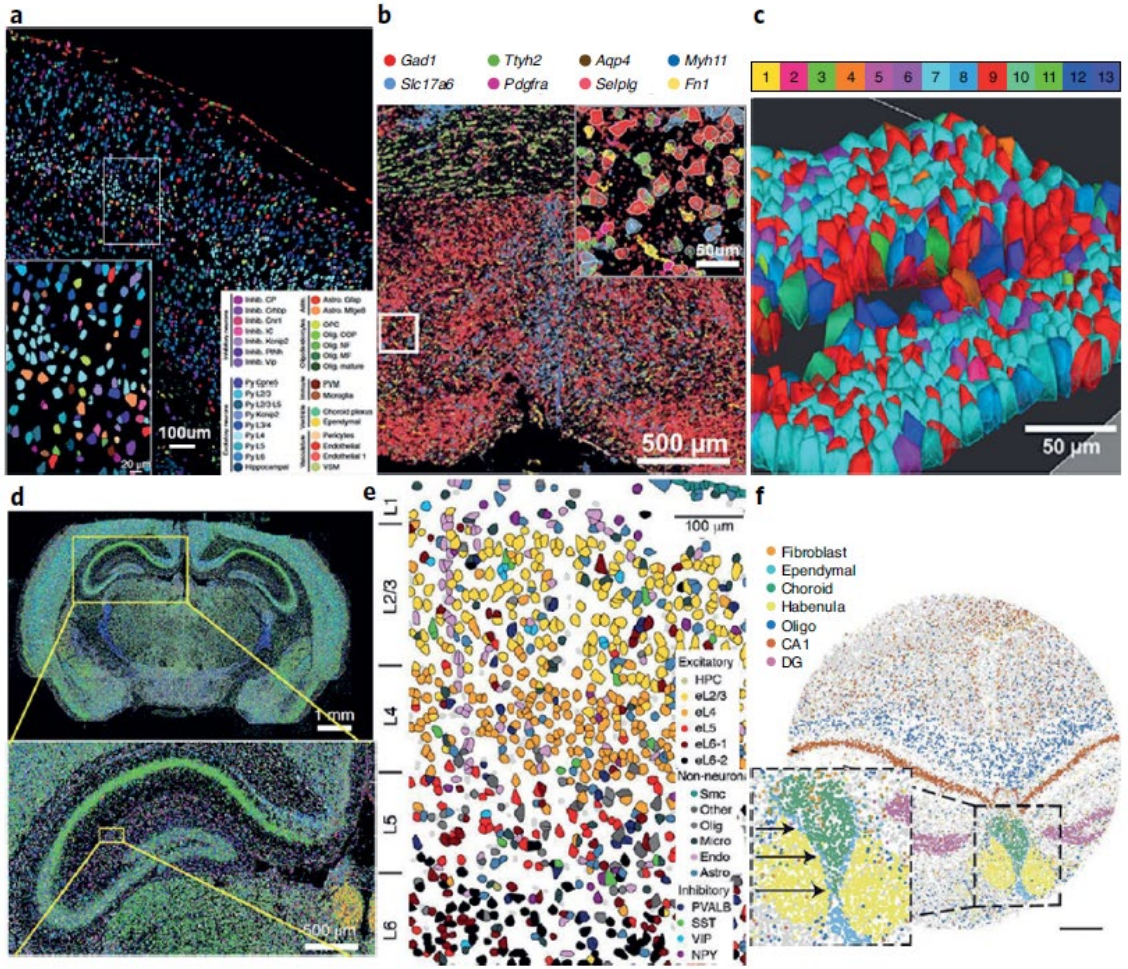


Fig. 1 | Cell type mapping with spatially resolved transcriptomics. a, An osmFISH map of 31 cell types in mouse somatosensory cortex was created by serial hybridization of smFISH probes recognizing 39 genes (adapted with permission from ref. [8], Springer Nature). **b**, A MERFISH-derived map of mouse hypothalamic preoptic area was created with a panel of 155 genes. Inset shows high-magnification image of individual cells from area indicated in white box (adapted with permission from ref. [10], AAAS). **c**, seqFISH-derived de novo cell typing of 13 cell types of mouse dentate gyrus (adapted with permission from ref. [13], Elsevier). **d**, pciSeq, a method of padlock probe–based in situ sequencing, was used to determine cell type composition of mouse hippocampus (adapted with permission from ref. [14], Springer Nature). **e**, STARmap, a padlock probe–based method, was applied to mouse visual cortex to map 16 cell types (adapted with permission from ref. [15], AAAS). **f**, Slide-seq was used to generate a map of hippocampus (adapted with permission from ref. [17], AAAS).

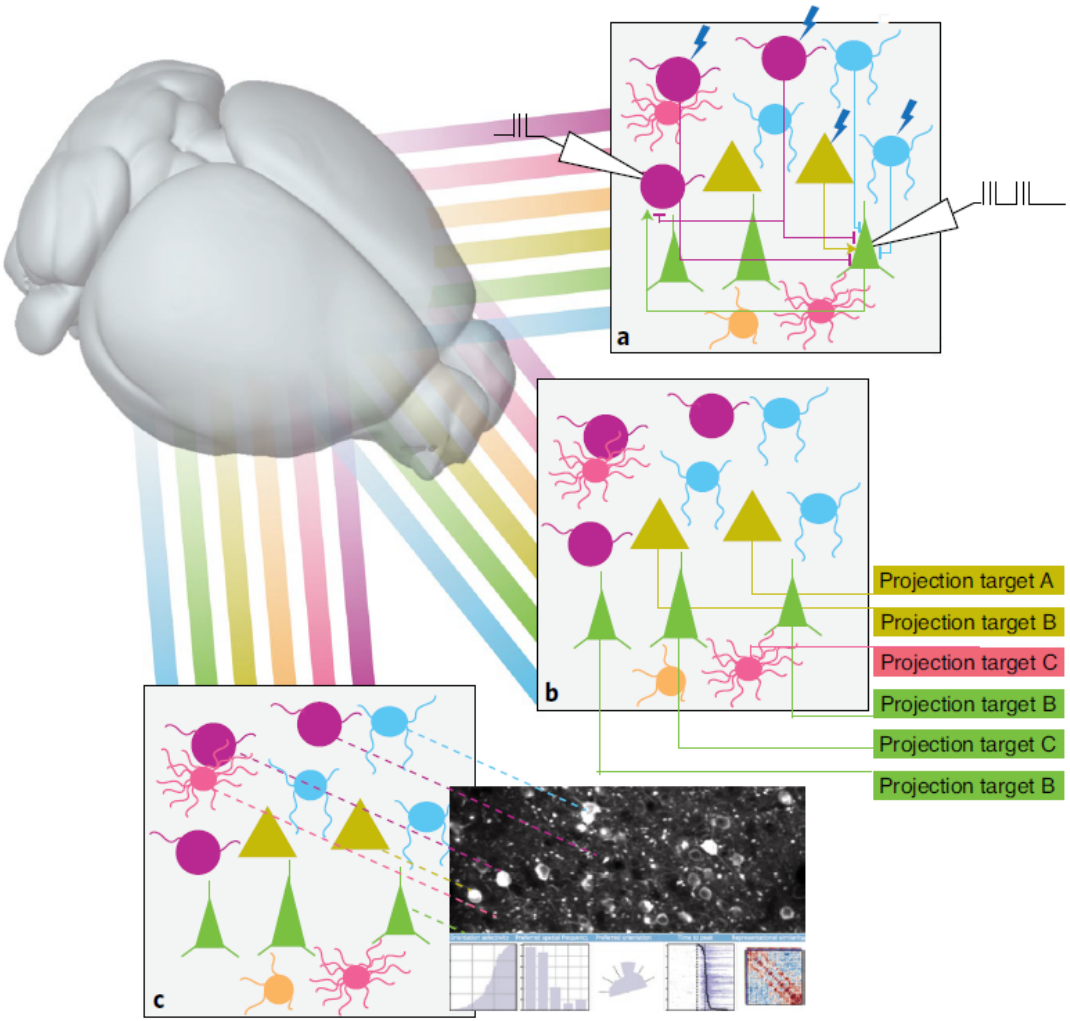


Fig. 2 | Spatially resolved transcriptomics linking transcriptomic cell types with connectivity and circuit function. a, Determining cell type identities after multi-patch and optogenetic stimulation experiments allows high-throughput measurements of cell-type-dependent local connectivity. **b**, Barcoded anterograde or retrograde viral labeling can add long-range projection information to cell type maps. **c**, Spatially resolved transcriptomics following in vivo activity measurements such as calcium imaging would add cell type identities to circuit analysis.

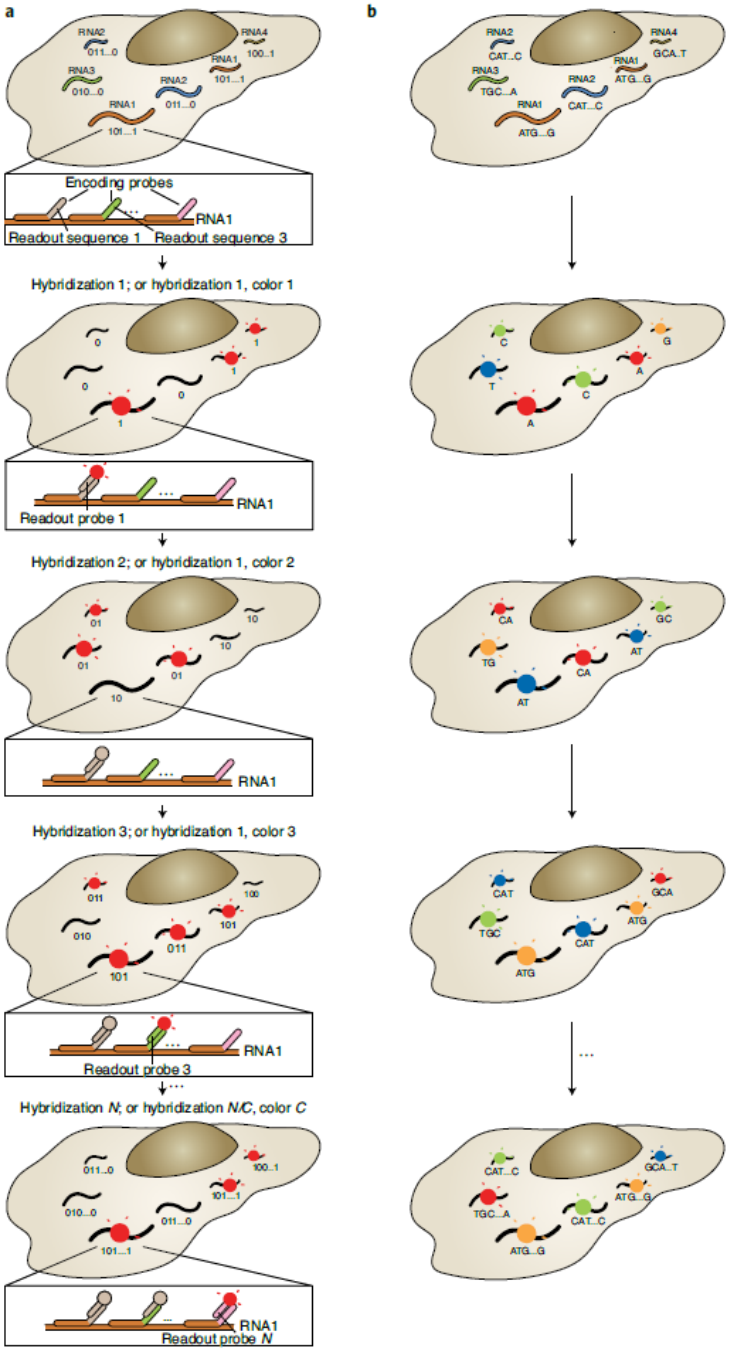


Fig. 1 | Imaging-based single cell genomics and transcriptomics approaches. a, A multiplex FISH scheme. For simplicity, a MERFISH scheme is used as an illustrative example. A binary barcoding scheme is shown together with how the barcodes are imprinted on the RNAs with a library of encoding probes and detected bit-by-bit with distinct readout probes. **b**, A simplified scheme for in situ sequencing. The nucleic acid sequence could either be the endogenous sequences of the genes or barcode sequences assigned to the genes, which are detected nucleotide-by-nucleotide via sequencing. Depending on the different sequencing chemistries used (not depicted here), the colors of the circles may indicate the identity of individual nucleotide or adjacent dinucleotide pairs.

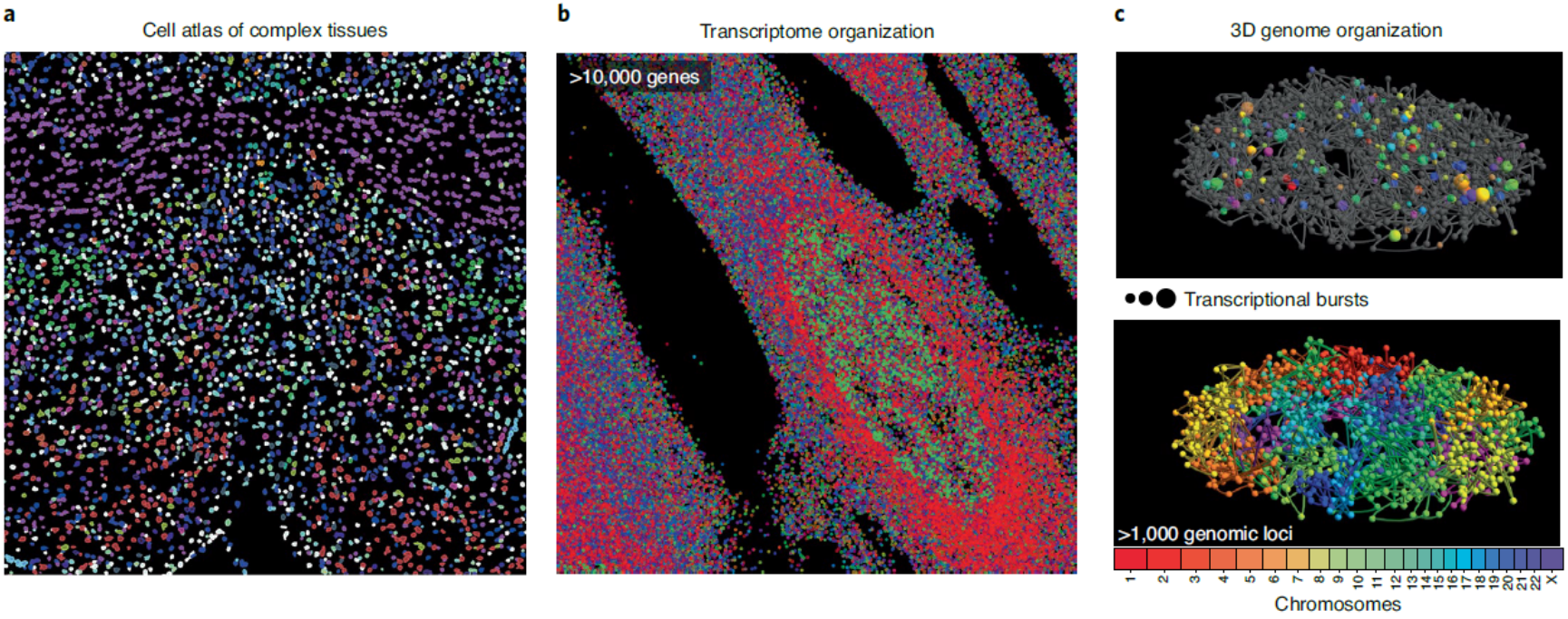
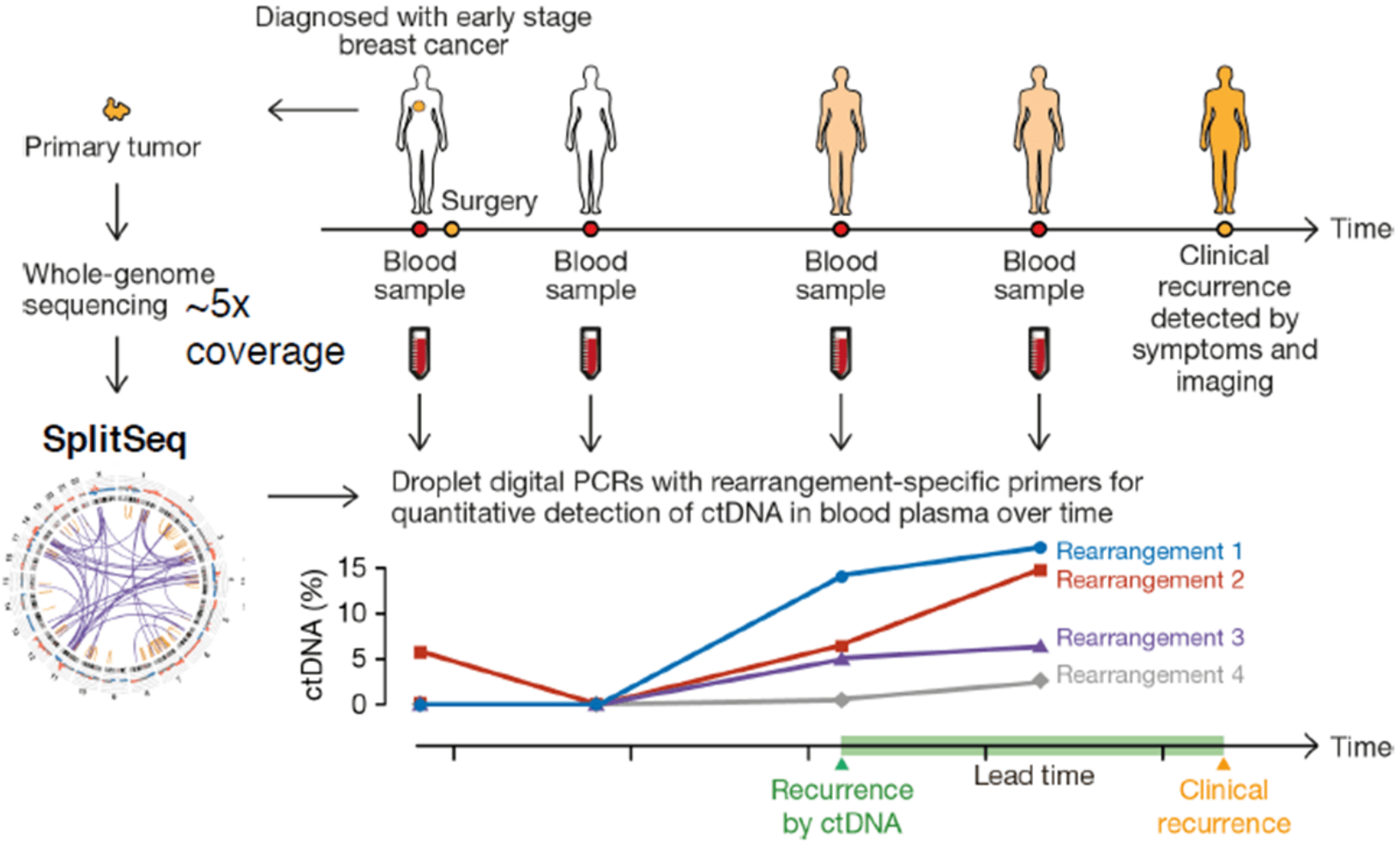


Fig. 2 | Three example application areas of imaging-based single-cell genomics and transcriptomics. a, Cell atlas of complex tissues. A cell-type map of a region in the mouse hypothalamus[10] is shown as an example, with different types of cells shown in different colors. **b**, Intracellular organization of the transcriptome. MERFISH image of >10,000 genes in individual cells[6] is shown as an example, with RNA molecules from different genes shown in different colors. **c**, 3D organization of the genome. MERFISH image of >1,000 genomic loci (bottom) and transcription bursts of >1,000 genes (top) in a single nucleus[11] is shown as an example.


Diagnosed with early stage breast cancer
Primary tumor
Surgery
Time
Whole-genome sequencing
~5x coverage
Blood sample
Blood sample
Blood sample
Blood sample
Clinical recurrence detected by symptoms and imaging
SplitSeq
Droplet digital PCRs with rearrangement-specific primers for quantitative detection of ctDNA in blood plasma over time
ctDNA (%)
15
10
5
0
Rearrangement 1
Rearrangement 2
Rearrangement 3
Rearrangement 4
Time
Lead time
Recurrence by ctDNA
Clinical recurrence

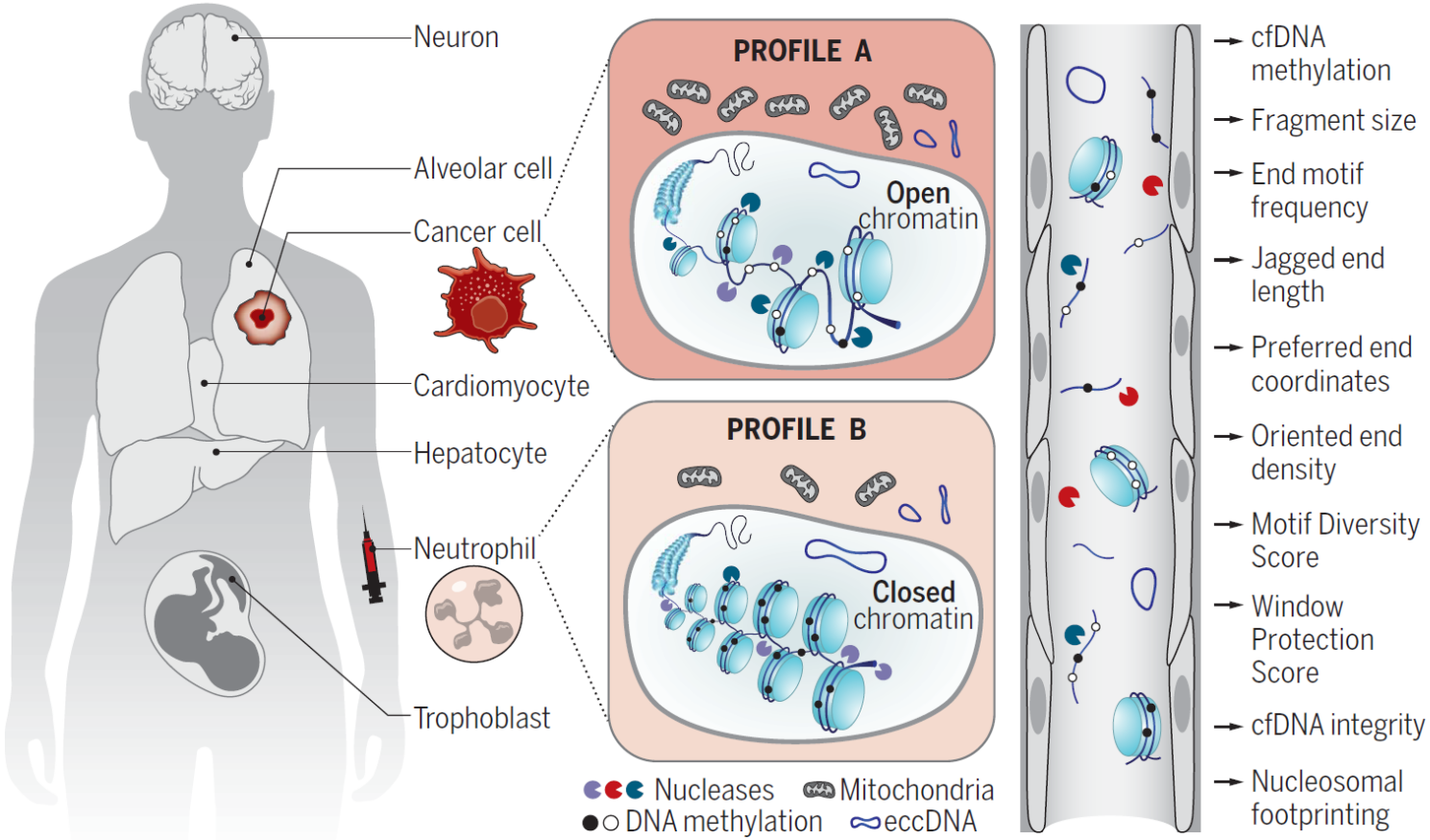
REVIEW SUMMARY

Lo *et al.*, *Science* **372**, 144 (2021) 9 April 2021

MEDICINE

Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies

Y. M. Dennis Lo*, Diana S. C. Han, Peiyong Jiang, Rossa W. K. Chiu



Cell-free DNA contains molecular features characteristic of their tissues of origin. Different tissues, including cancer cells and trophoblasts in pregnancy, contribute cell-free DNA to the circulation. The cell-free DNA molecules may bear methylation states reflective of the cell of origin. The DNA molecule size, fragment end locations, and end motifs are influenced by the nucleosome organization, chromatin structure, nuclease content, and gene expression of the tissue of origin. Parameters that can be measured to quantify these characteristics are shown on the right.

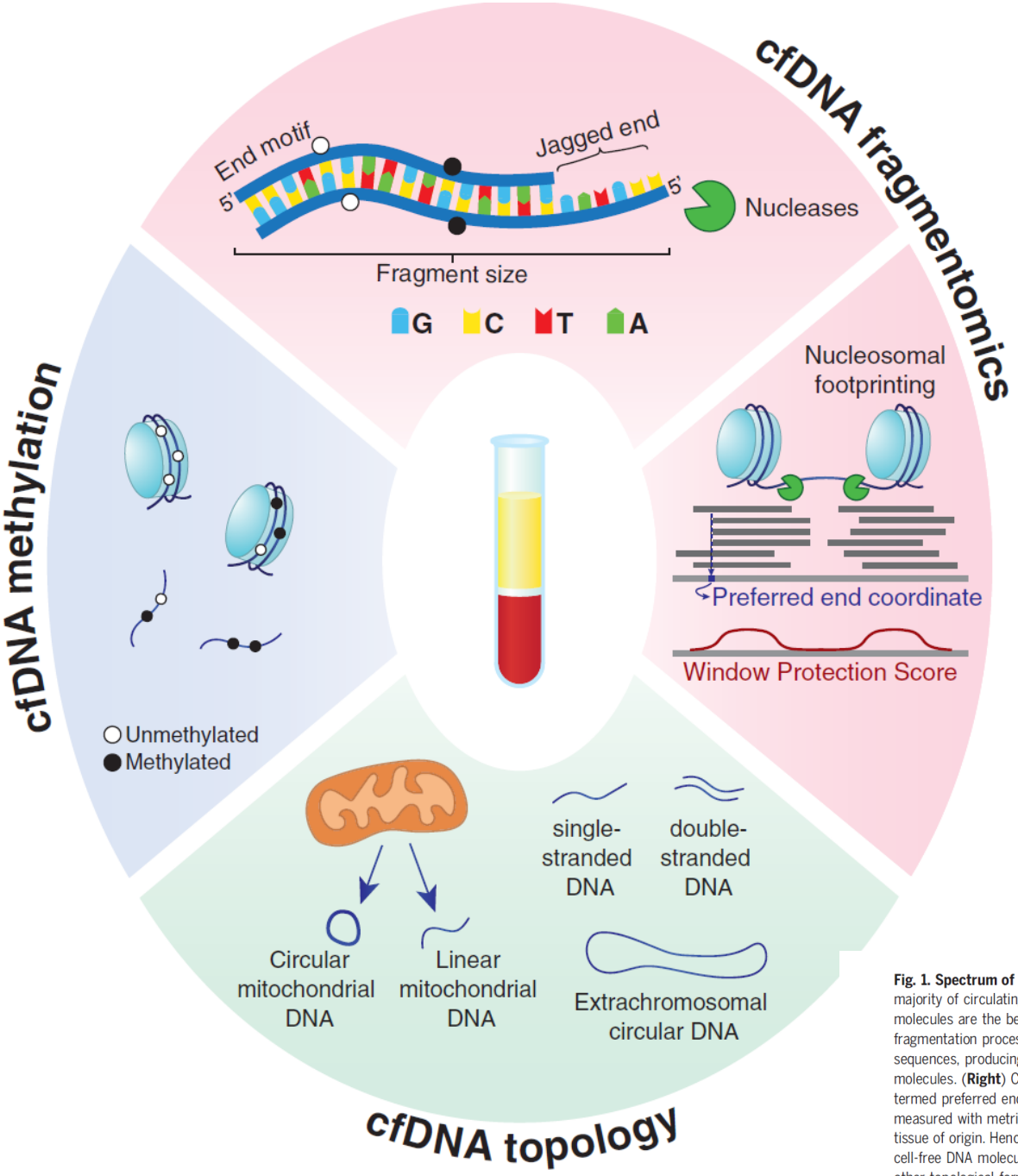


Fig. 1. Spectrum of nongenetic signatures identified among circulating DNA molecules. (**Top**) The majority of circulating DNA molecules are short linear fragments. Double-stranded circulating DNA molecules are the best studied. The ends of the cell-free DNA molecules may be blunt or jagged. The fragmentation process involves nuclease digestion. Certain nucleases have a predilection for certain sequences, producing overrepresentation of characteristic motif sequences at the ends of circulating DNA molecules. (**Right**) Cell-free DNA ending sites tend to cluster at particular locations in the genome, termed preferred end sites or coordinates. Cell-free DNA coverage across the genome is uneven, can be measured with metrics such as the window protection score, and reflects nucleosome positioning of the tissue of origin. Hence, nucleosome footprints could be inferred from cell-free DNA data. (**Left**) CpG sites on cell-free DNA molecules may be methylated or unmethylated. (**Bottom**) Besides linear DNA fragments, other topological forms of circulating DNA exist, such as circular mitochondrial DNA and eccDNA.

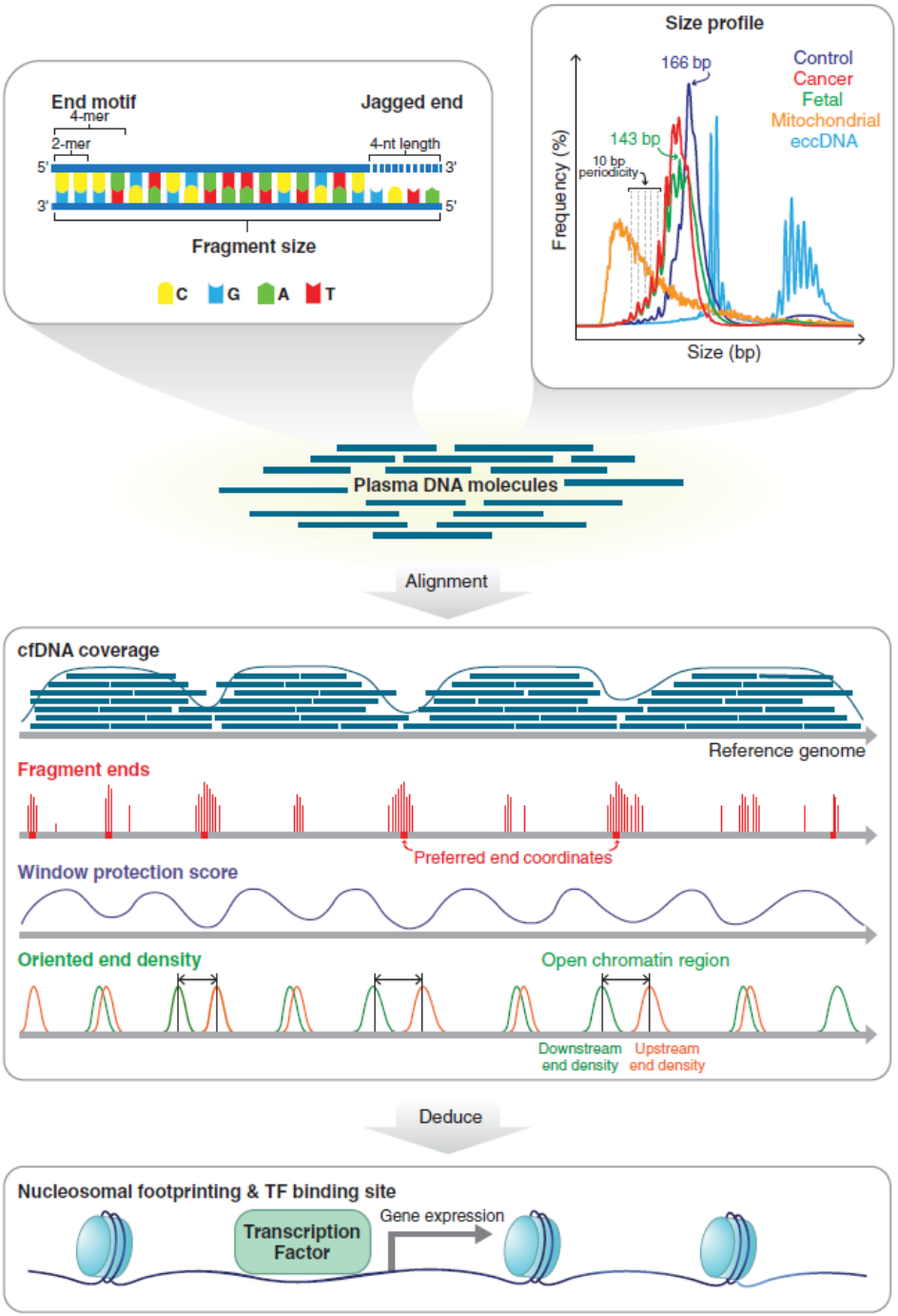


Fig. 2. Fragmentomic analysis of cell-free DNA. (**Top left**) Double-stranded cell-free DNA may have blunt ends or jagged ends. The ending sequences reveal overrepresentation of certain motifs that could be characterized by a range of nucleotides, such as 2-nucleotide oligomer (2-mer) or 4-mer motifs. (**Top right**) Circulating DNA have known size profiles. Nuclear-derived cell-free DNA in healthy individuals shows a modal size of 166 bp. Cell-free fetal DNA in maternal plasma and tumor-derived DNA in plasma of cancer patients have shorter modal sizes of approximately 143 bp. Below approximately 143 bp, size peaks exhibiting a 10-bp periodicity can be seen. Linear DNA fragments derived from mitochondria are much shorter in size and do not exhibit the 10-bp periodicity. eccDNA in plasma exhibits a larger size distribution, with 10-bp periodicity. (**Middle**) When pools of circulating DNA are aligned to the human genome, their ends tend to cluster at genomic locations, called preferred end sites, which could be variable between DNA molecules that originate from different tissues. The window protection score, which is calculated as the number of complete fragments minus the number of fragment endpoints within a given window size, conveys information about DNA protection from digestion, which could be used to infer nucleosome positioning. (**Bottom**) The genomic coverage and directional information of the cell-free DNA ending locations—namely upstream end or downstream end—are reflective of the chromatin structure of the tissue of origin. TF, transcription factor.

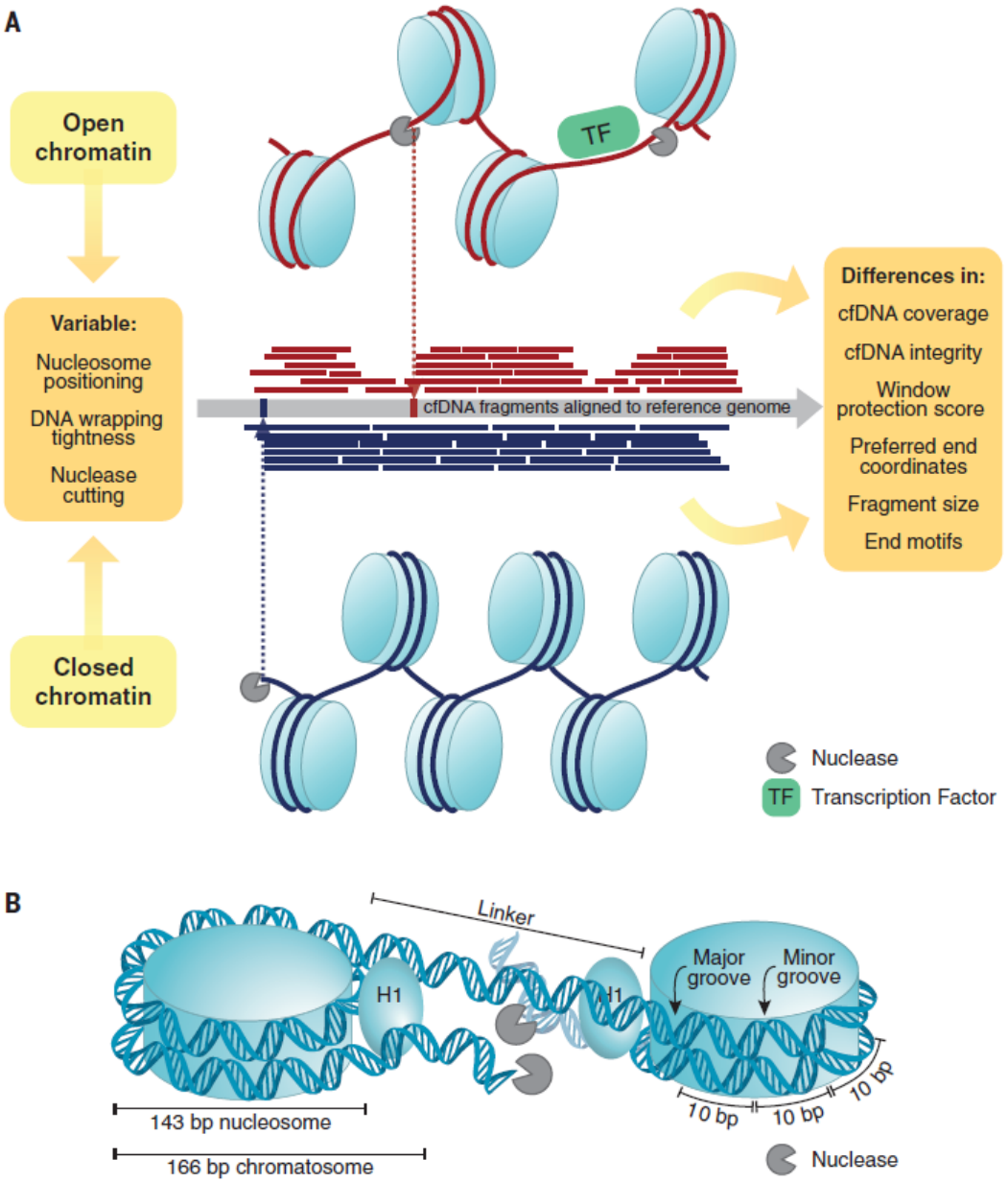


Fig. 3. Chromatin structure influences circulating DNA fragment characteristics. (**A**) Open or closed chromatin regions differ in their nucleosome repeat lengths and gene expression levels (for example, being affected by the presence of TF binding sites), which in turn may increase or decrease nuclease accessibility. These variables are manifested in the fragmentomic features of cell-free DNA molecules. The majority of circulating DNA fragments are mononucleosomal in length, and nucleosomal packaging affects the cell-free DNA size. The various characteristics can be measured by use of metrics that are listed on the right. cfDNA, cell-free DNA. (**B**) The characteristic 10-bp periodic peaks of short cell-free DNA fragment size frequencies are thought to be related to the coiling of the helical DNA around the nucleosome core. There is evidence suggesting that the linker region may be more susceptible to nuclease digestion.

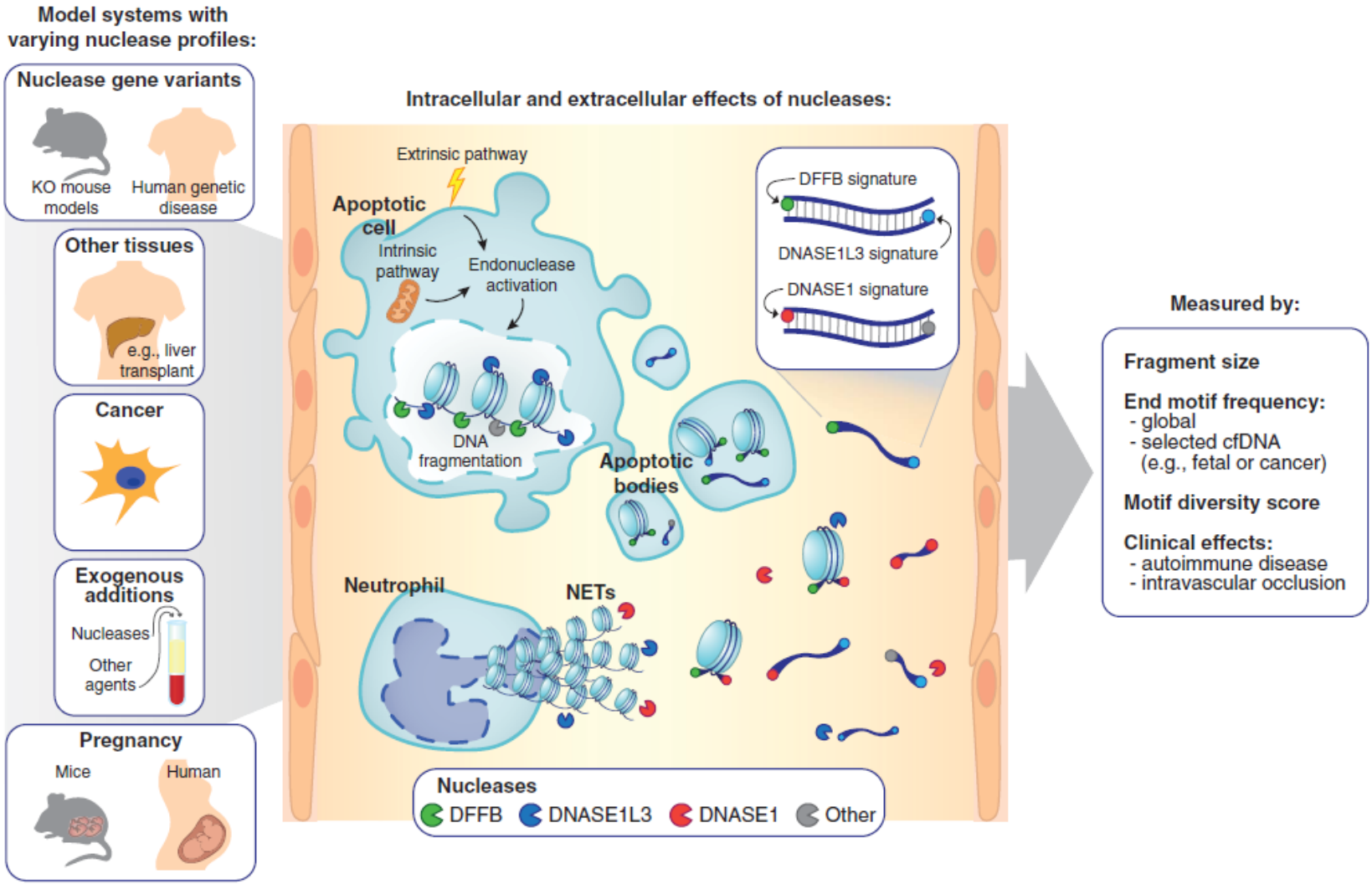
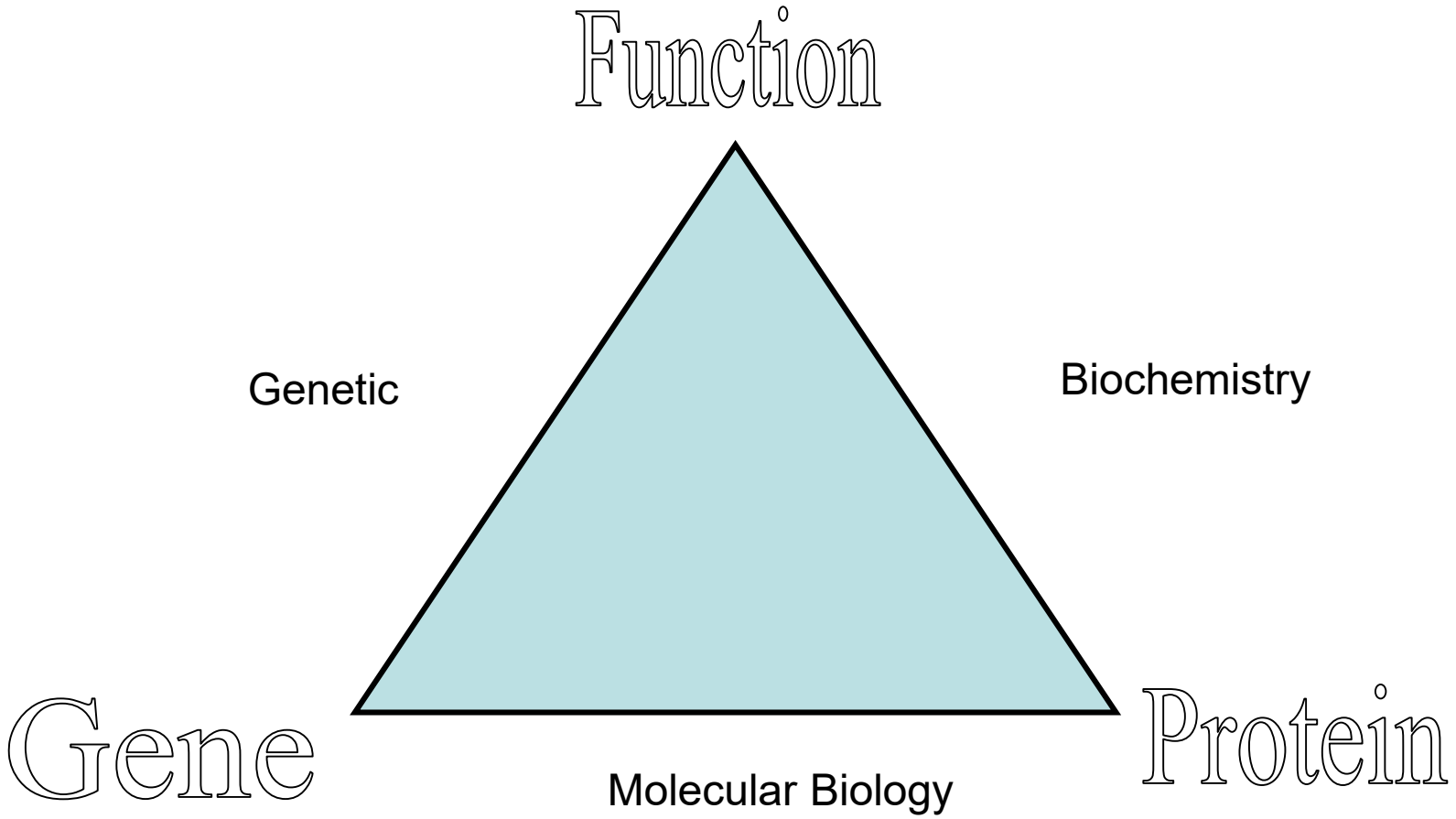
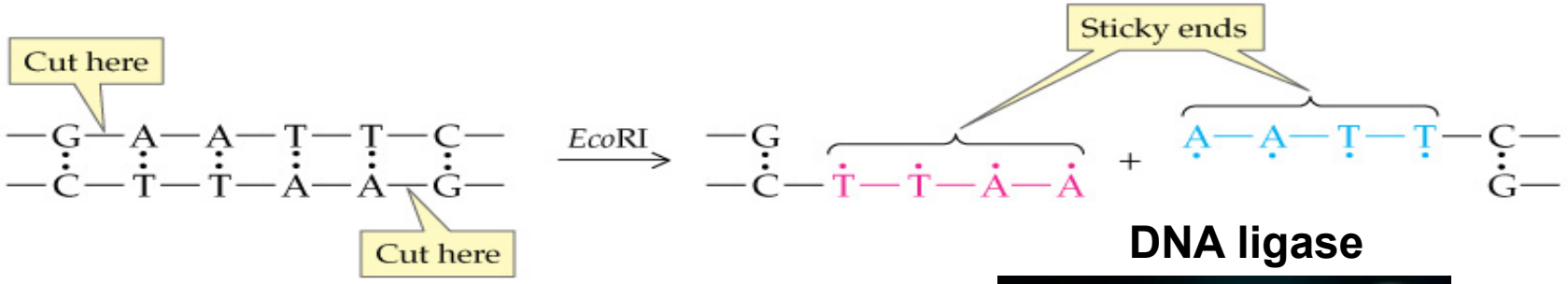


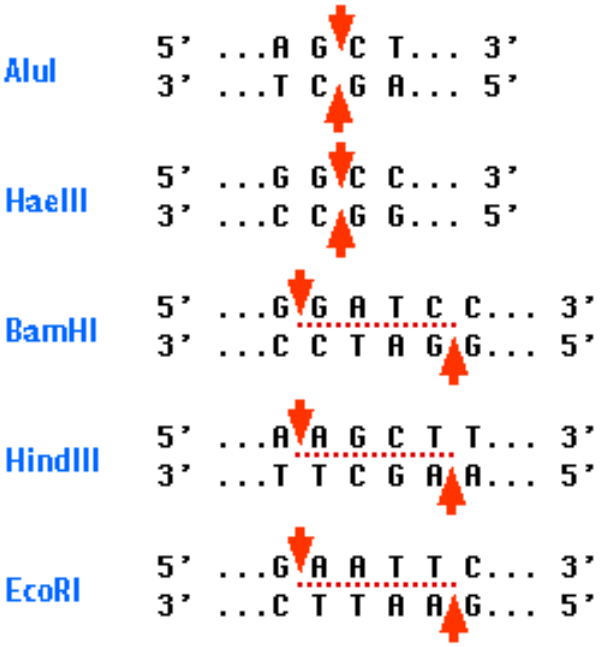
Fig. 4. Link between nuclease biology and cell-free DNA fragment characteristics. (**Bottom middle**) Several nucleases are involved in the fragmentation process of circulating DNA. (**Left**) Experimental approaches that have been used include mice with one or more nuclease genes deleted; tissue-specific nuclease expression levels inferred from human pregnancies, mouse pregnancy models, human cancers, and organ transplants; and exogenous addition of nucleases or other agents to influence cell-free DNA generation or nuclease activities. (**Middle**) DFFB and DNASE1L3 are activated in apoptotic cells, and DNASE1L3 and DNASE1 are also secreted in plasma, playing a role in the further fragmentation of plasma cell-free DNA and NETs. The preferences of different nucleases in cutting different sequences leave characteristic end-motif signatures in cell-free DNA. (**Right**) Metrics that can be measured to study the effects of nucleases include fragment size and end motif frequency in the total (global) or selected (fetal or cancer) circulating DNA population. The pathogenic effects of nuclease biology and circulating DNA have been linked with the development of autoimmune diseases and intravascular occlusions.

Recombinant DNA





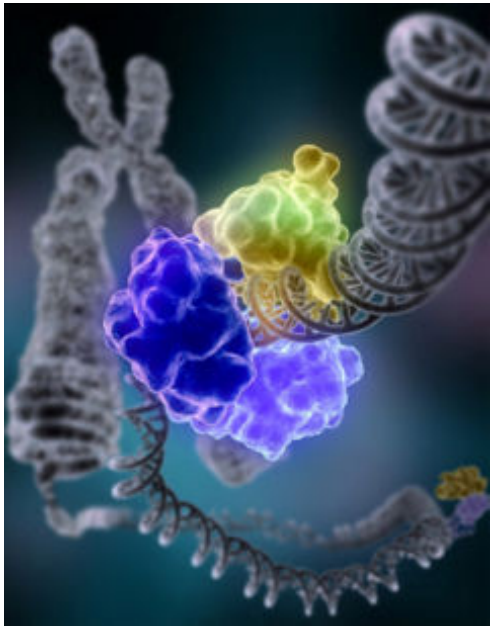
Restriction Enzyme

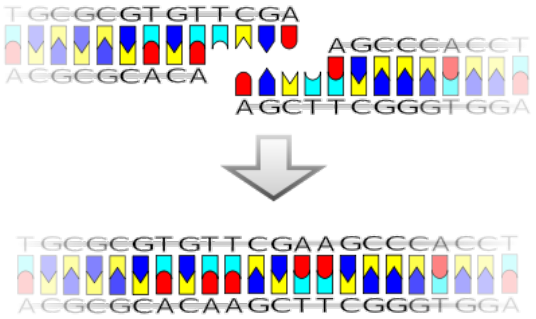


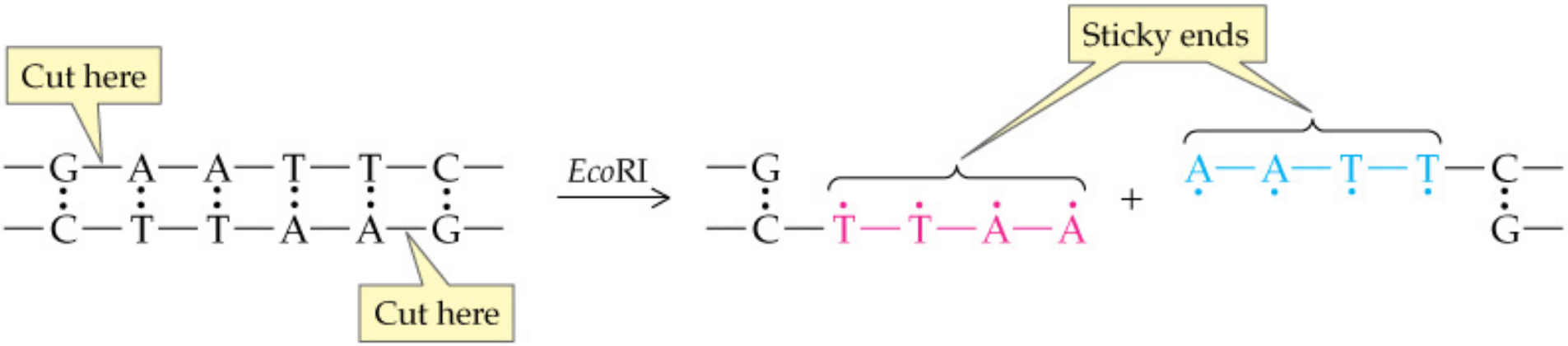
AluI and **HaeIII** produce blunt ends

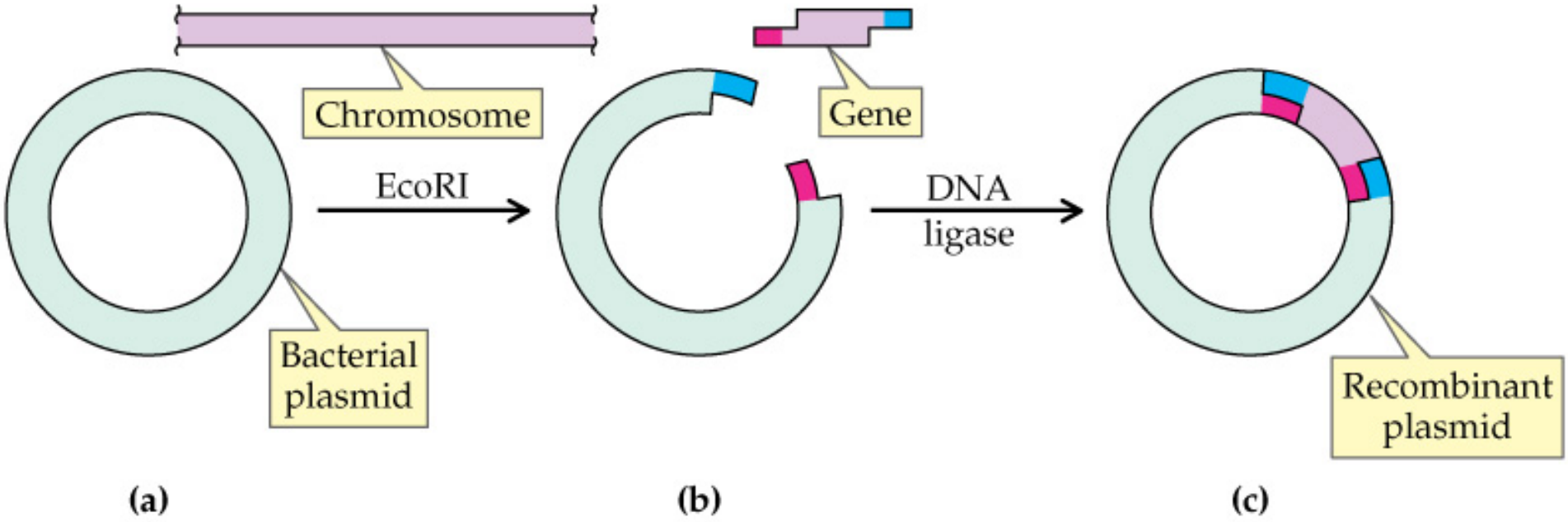
BamHI **HindIII** and **EcoRI** produce "sticky" ends

DNA ligase



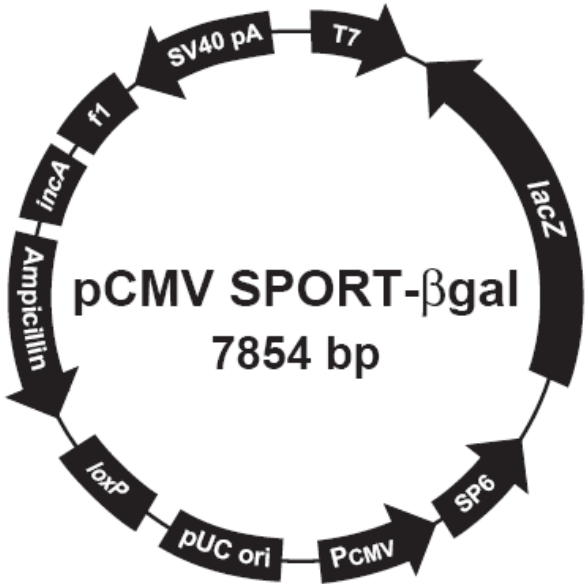


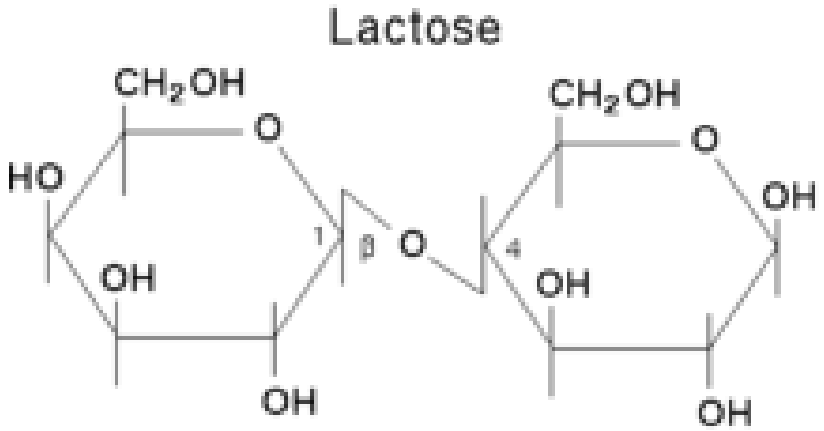

Cut here
G A A T T C
C T T A A G
Cut here
EcoRI
Sticky ends
G
C T T A A
+
A A T T C
G


Chromosome
Gene
EcoRI
DNA
ligase
Bacterial
plasmid
Recombinant
plasmid
(a)
(b)
(c)

β-Galactosidase

The enzyme that splits lactose into glucose and galactose. Coded by a gene (lacZ) in the lac operon of Escherichia coli.

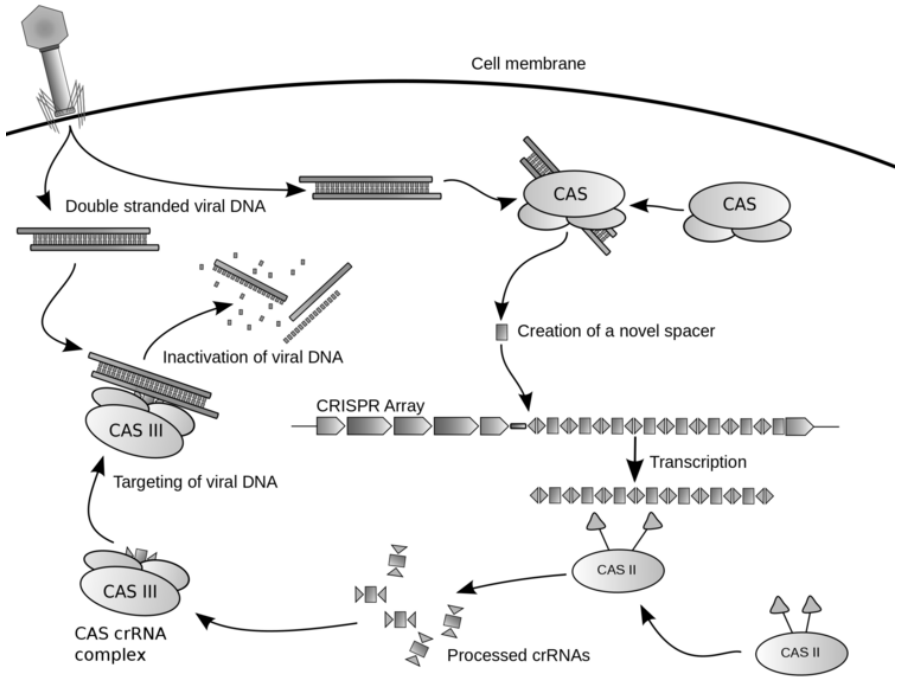




PUC is a family of plasmids that have an ampicillin resistance gene and more importantly a *lacZ* gene. A functional lacZ gene will produce the protein β - galactosidase. Bacterial colonies in which β - galactosidase is produced, will form blue colonies in the presence of the substrate 5 - bromo - 4 - chloro - 3 - indolyl - b - D - galactoside or as it is more commonly referred to, X-gal.

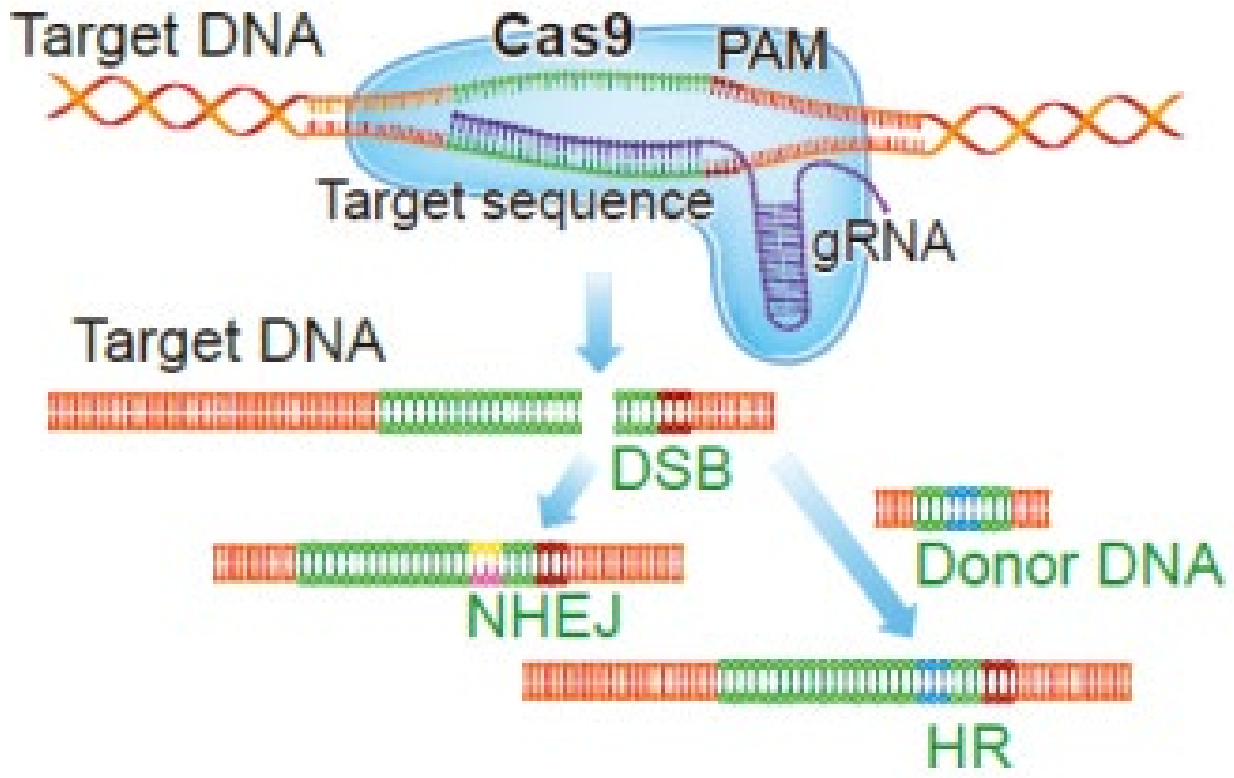
CRISPR CAS9

CRISPRs (**clustered regularly interspaced short palindromic repeats**) are segments of prokaryotic DNA containin g short repetitions of base sequences. Each repetition is followed by short segments of "spacer DNA" from previous exposures to a bacterial virus or



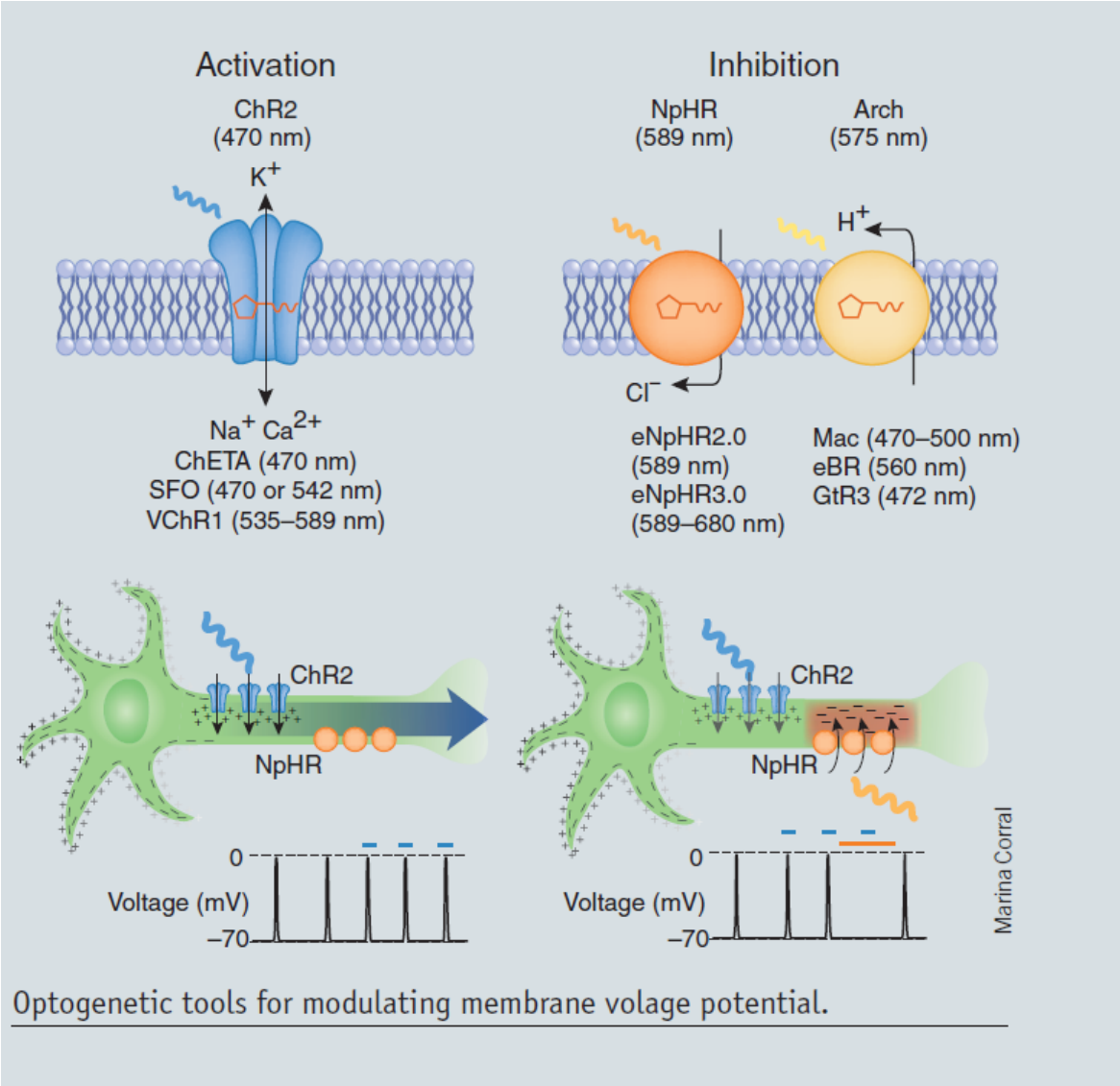
Movie

- https://youtu.be/2pp17E4E-O8


Target DNA
Cas9
PAM
Target sequence
gRNA
Target DNA
DSB
NHEJ
Donor DNA
HR

Optogenetics

https://www.youtube.com/watch?v=I64X7vHSHOE



Optogenetic tools for modulating membrane volage potential.

•Eukaryotic cells are about 1000 times larger than bacteria cells and also have a membrane enclosed nucleus containing their DNA, and several other internal structures known as organelles.

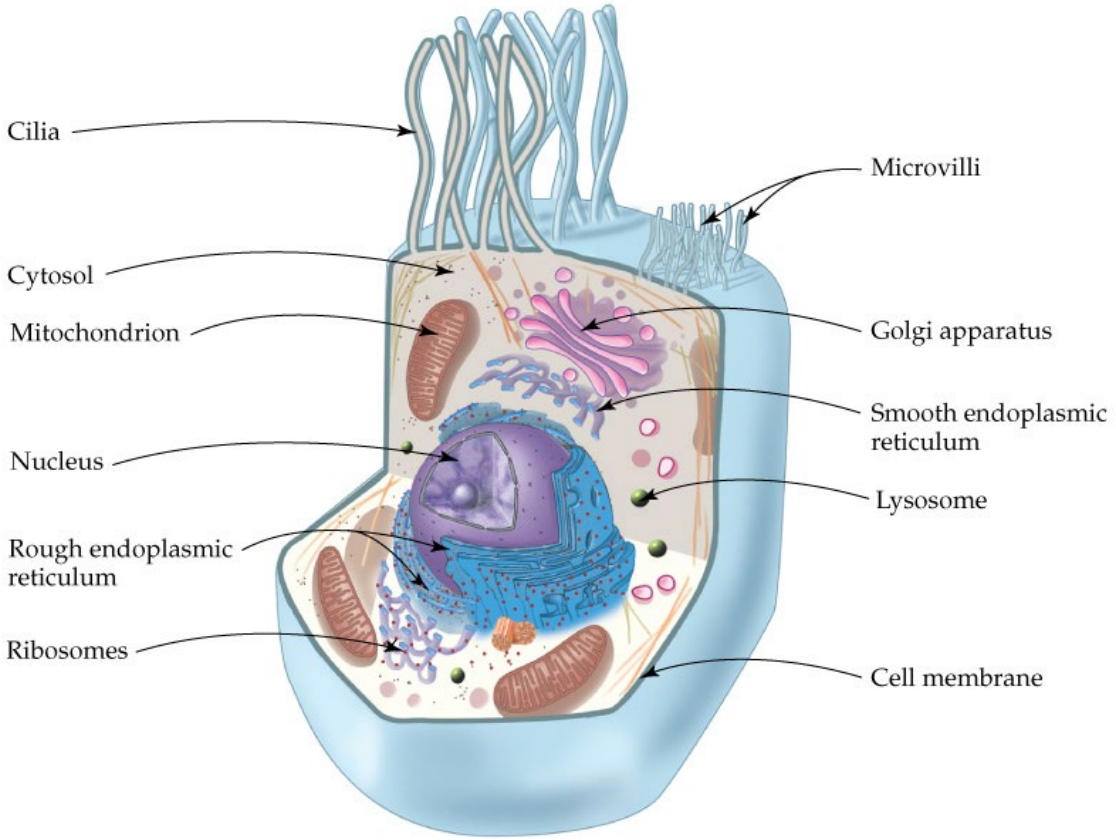
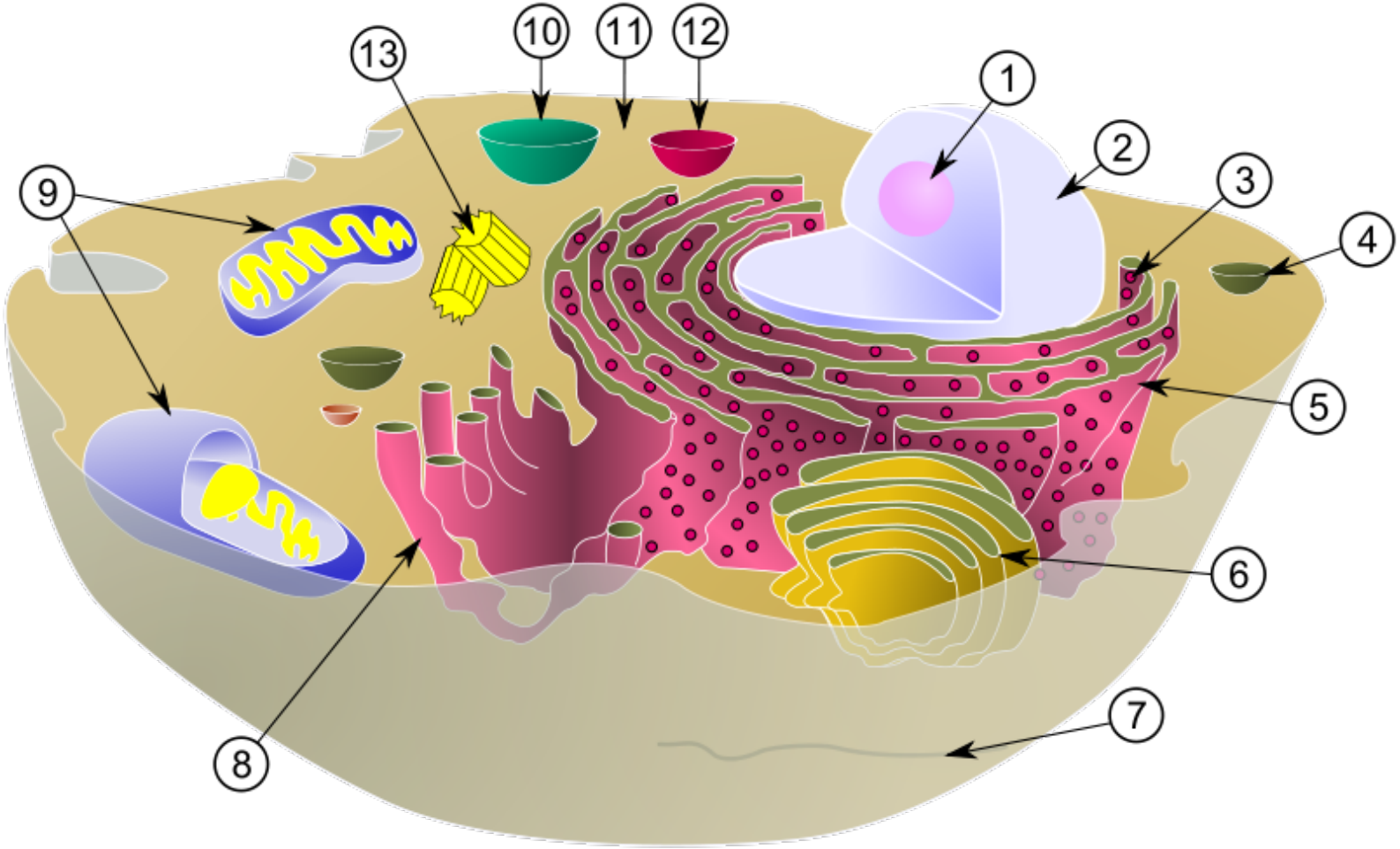
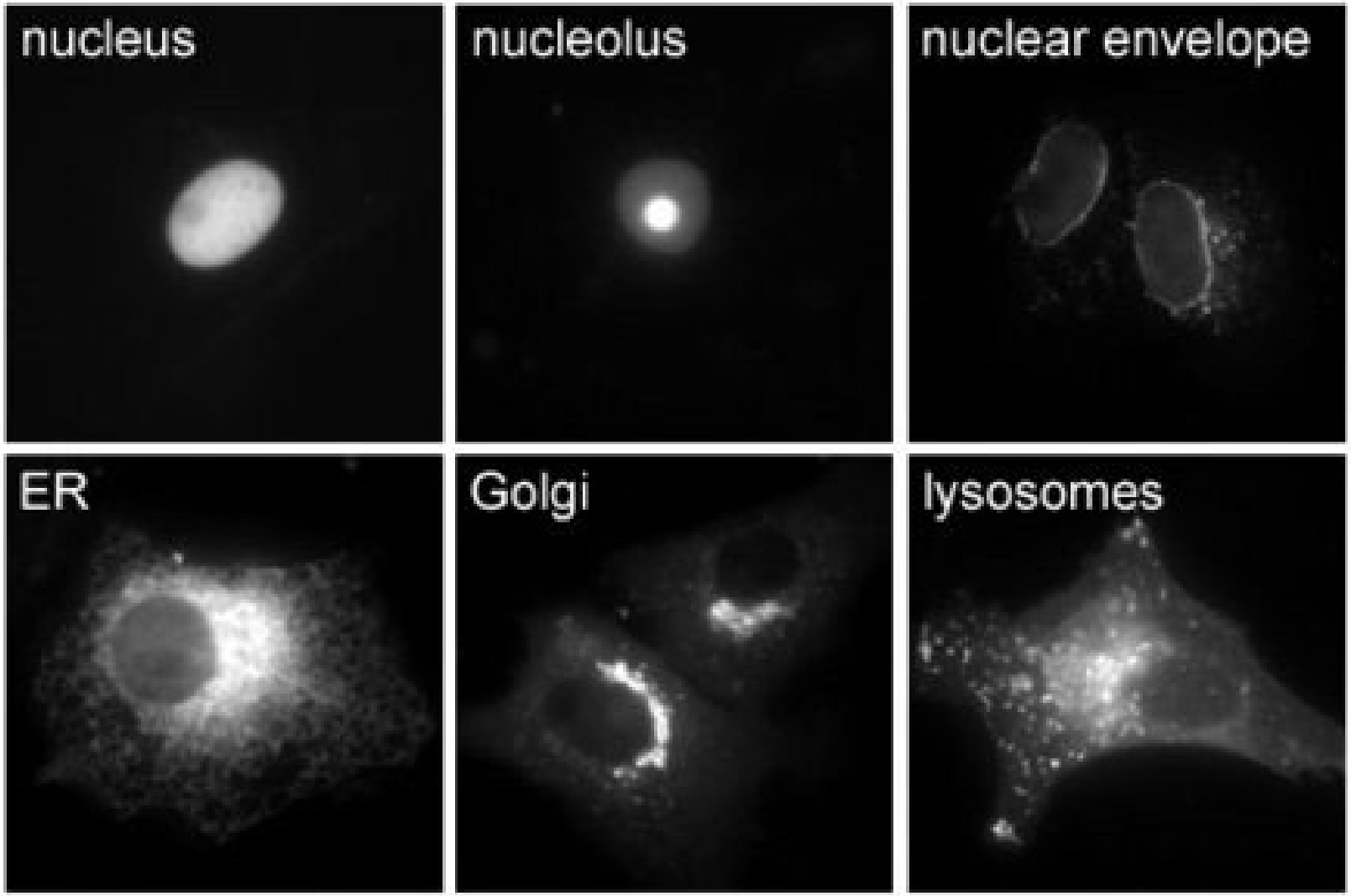
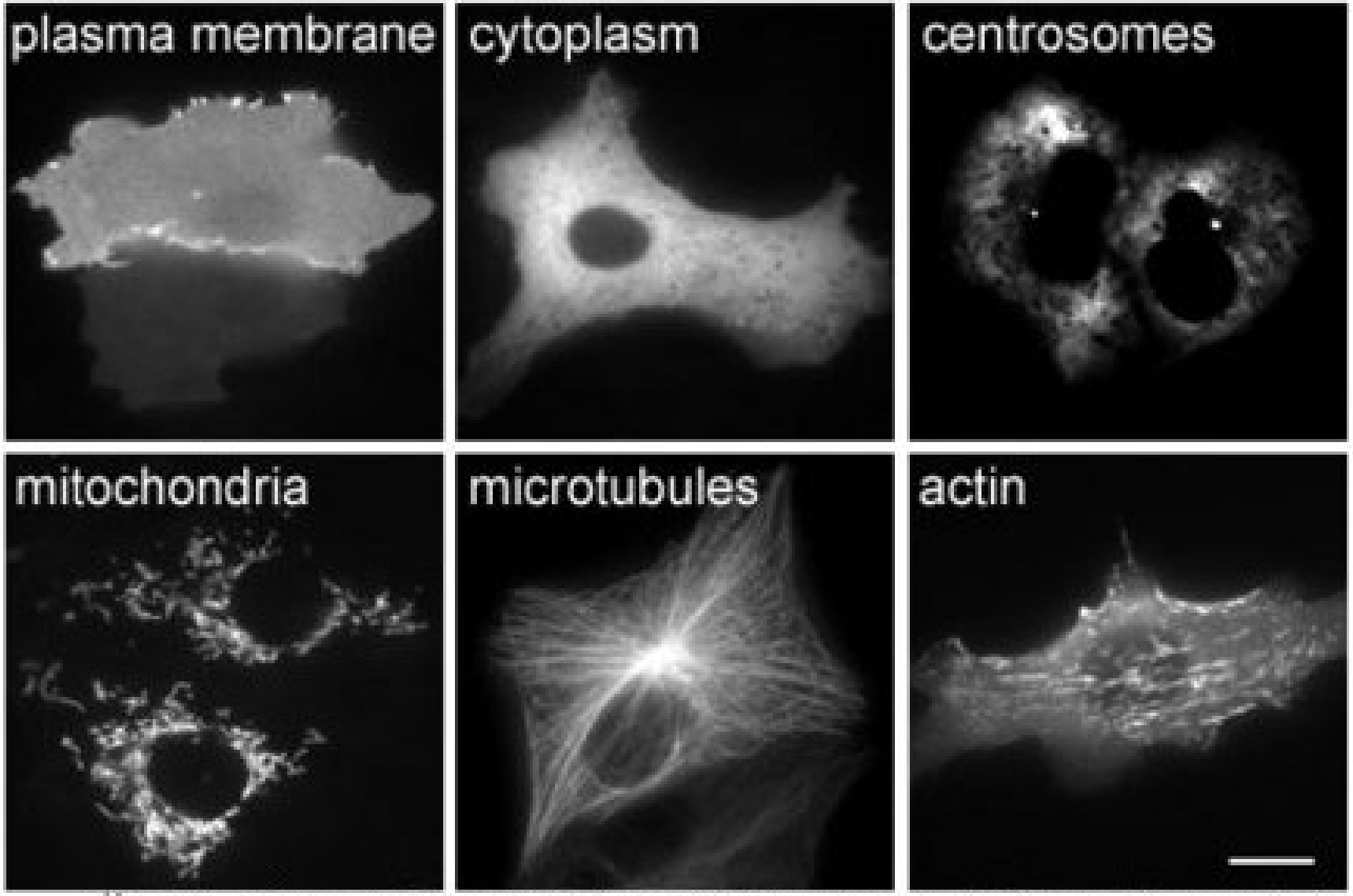


Fig 21.3 A generalized eukaryotic cell.



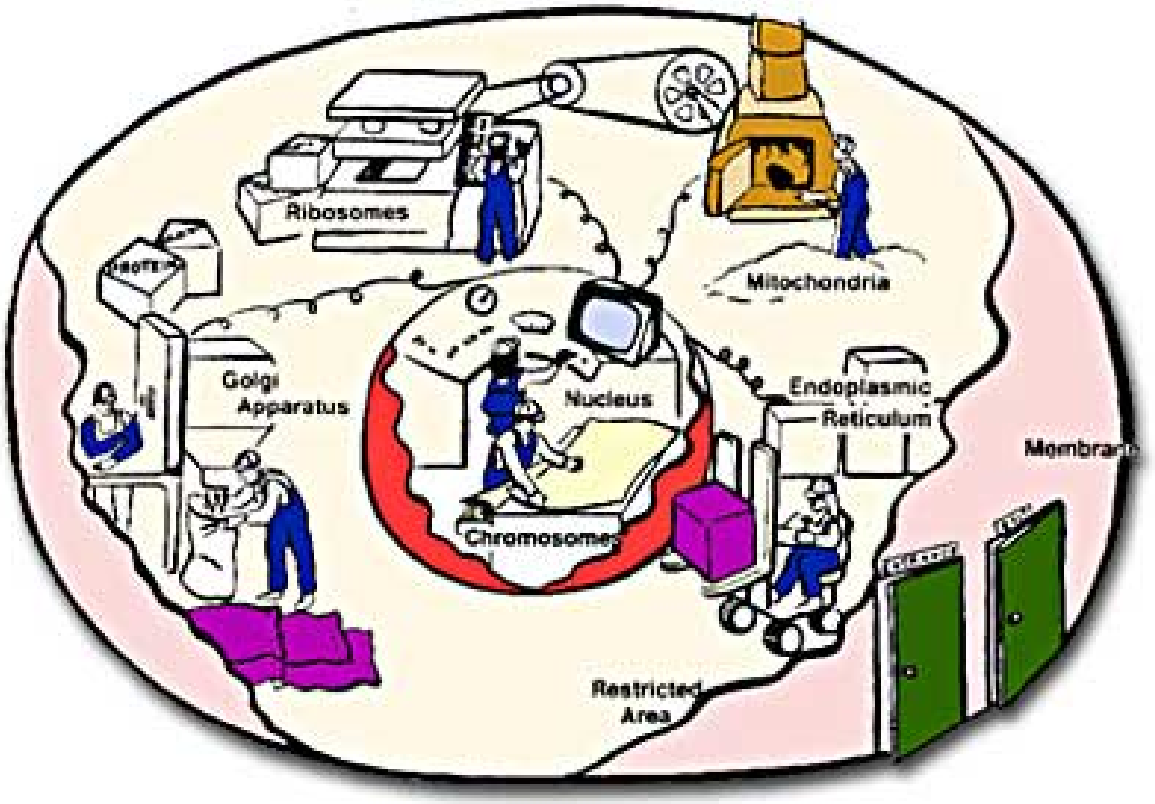
•Schematic showing the cytoplasm, with its components (or *organelles*), of a typical animal cell. Organelles: (1) nucleolus (2) nucleus (3) ribosome (4) vesicle (5) rough endoplasmic reticulum (6) Golgi apparatus (7) cytoskeleton (8) smooth endoplasmic reticulum (9) mitochondria (10) vacuole (11) cytosol (12) lysosome (13) centriole.


nucleus
nucleolus
nuclear envelope
ER
Golgi
lysosomes



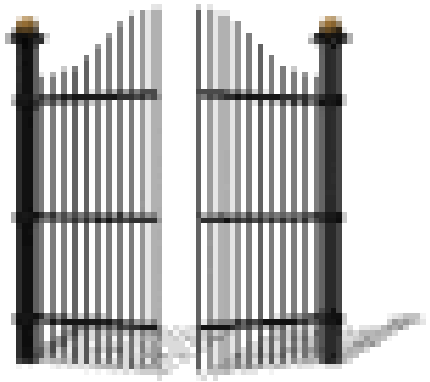
with friendly permission of Jeremy Simpson and Rainer Pepperkok

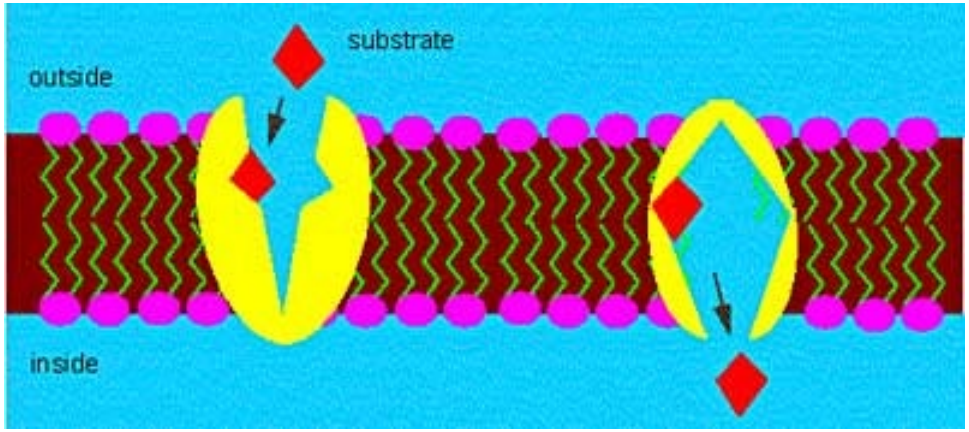
A Busy Factory



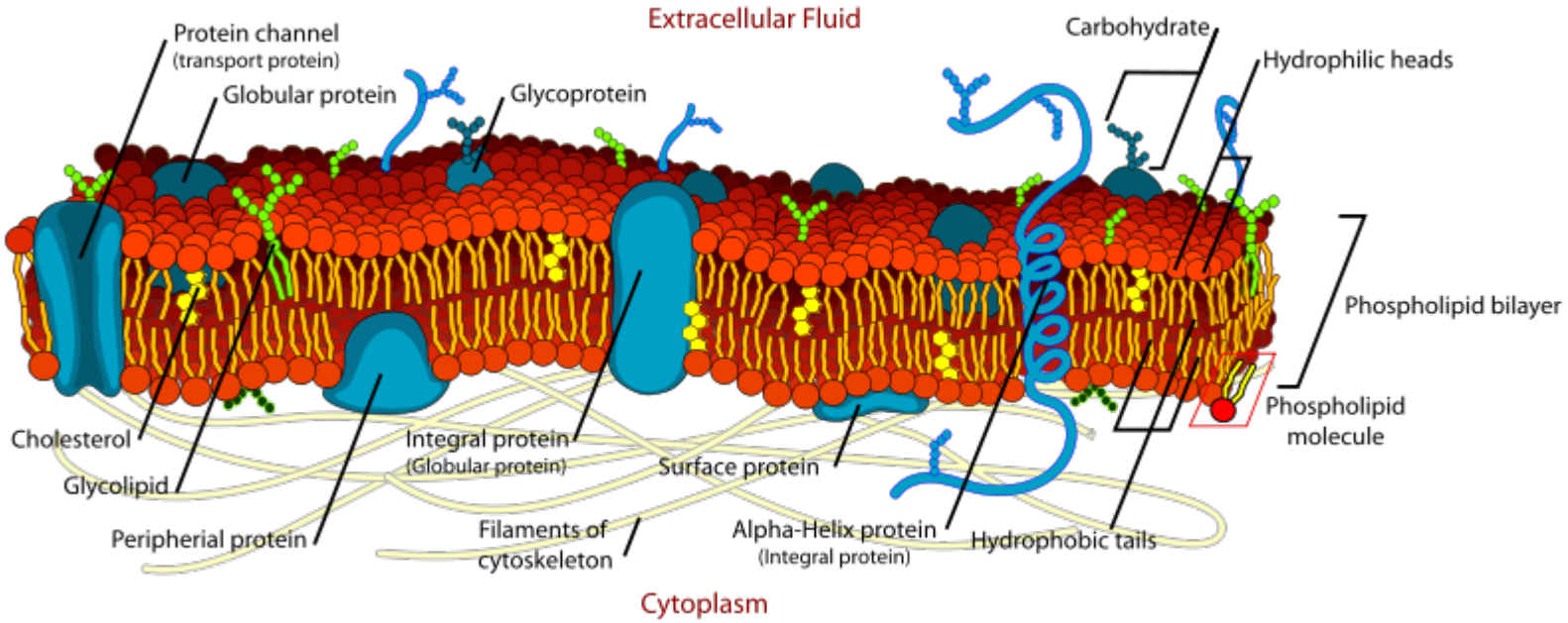
A cell can be thought of as a "factory," with different departments each performing specialized tasks.

The Plasma Membrane





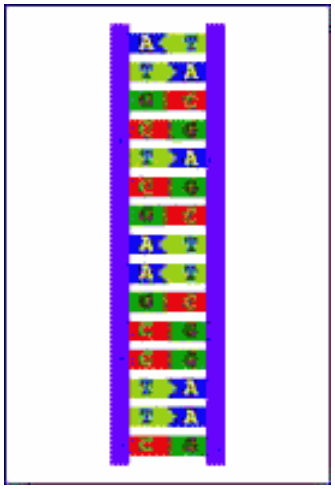
Cell Membrane



Characteristic diffusivities

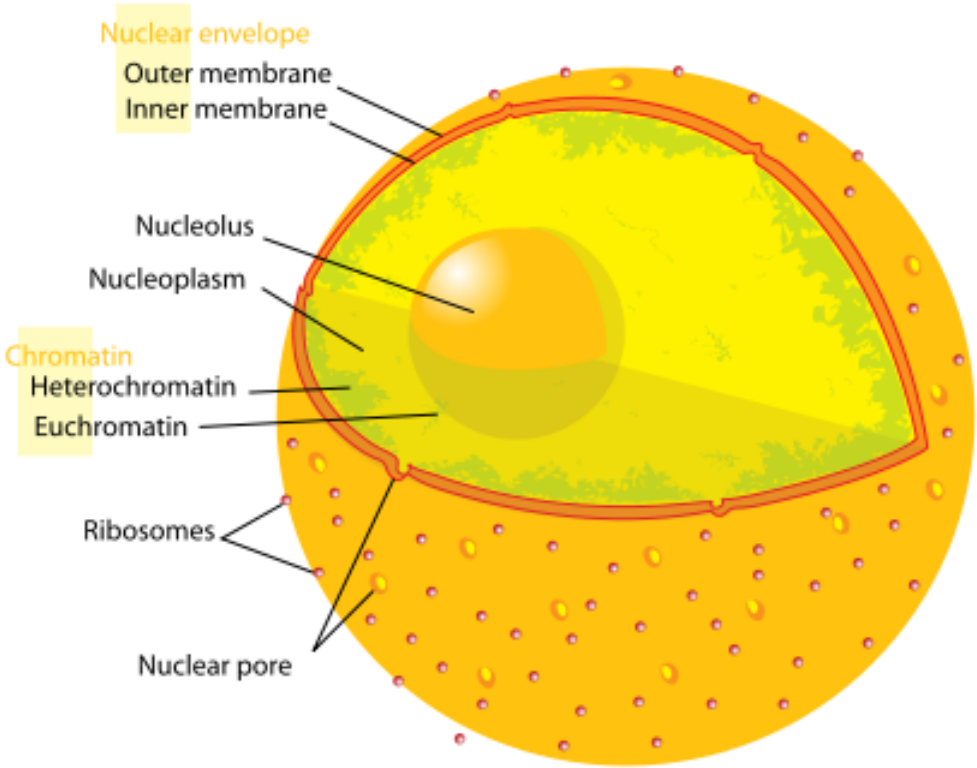
Particle	Typical size	Diffusion constant
Solute ion	10^{-1} nm	2×10^3 $\mu m^2/s$
Small protein	5 nm	40 $\mu m^2/s$
Virus	100 nm	2 $\mu m^2/s$
Bacterium	1 μm	0.2 $\mu m^2/s$
Mammalian/human cell	10 μm	0.02 $\mu m^2/s$

The Nucleus



The cell factory contains a large inventory of blueprints dating all the way to its founding. Some of these blueprints are out of date, and some are for parts and products that are no longer made. Part of your job would entail sorting through everything, finding the correct blueprints, copying them, and sending the copies out to the assembly line at the correct time.

Nucleus

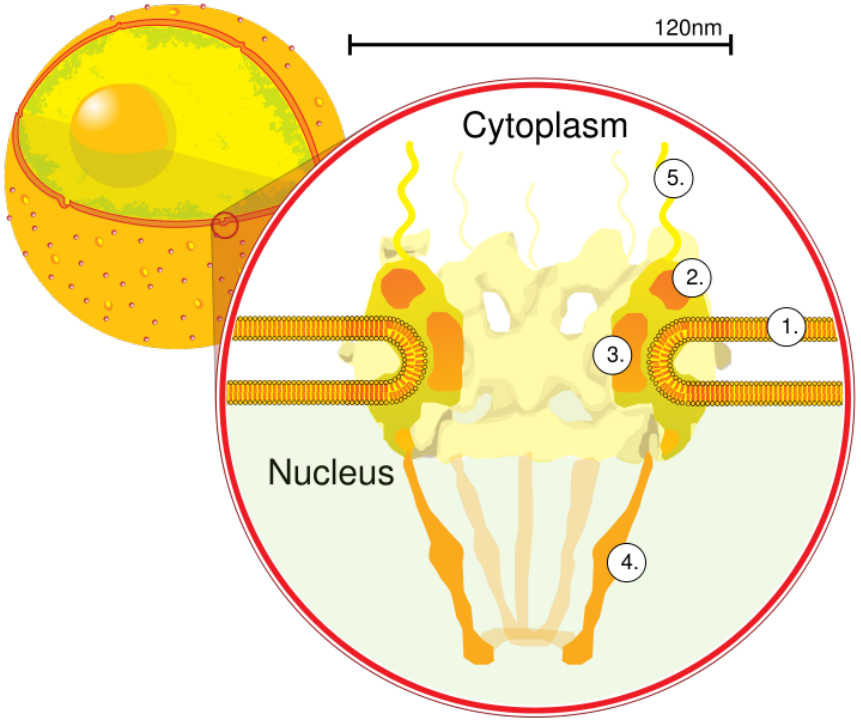


•In cell biology, the **nucleus** is a membrane-enclosed organelle found in most eukaryotic cells. It contains most of the cell's genetic material, organized as multiple long linear DNA molecules in complex with a large variety of proteins such as histones to form chromosomes. The genes within these chromosomes make up the cell's nuclear genome. The function of the nucleus is to maintain the integrity of these genes and to control the activities of the cell by regulating gene expression.

In cell biology, the **nucleolus** (plural *nucleoli*) is a "sub-organelle" of the cell nucleus, which itself is an organelle. A main function of the nucleolus is the production and assembly of ribosome components

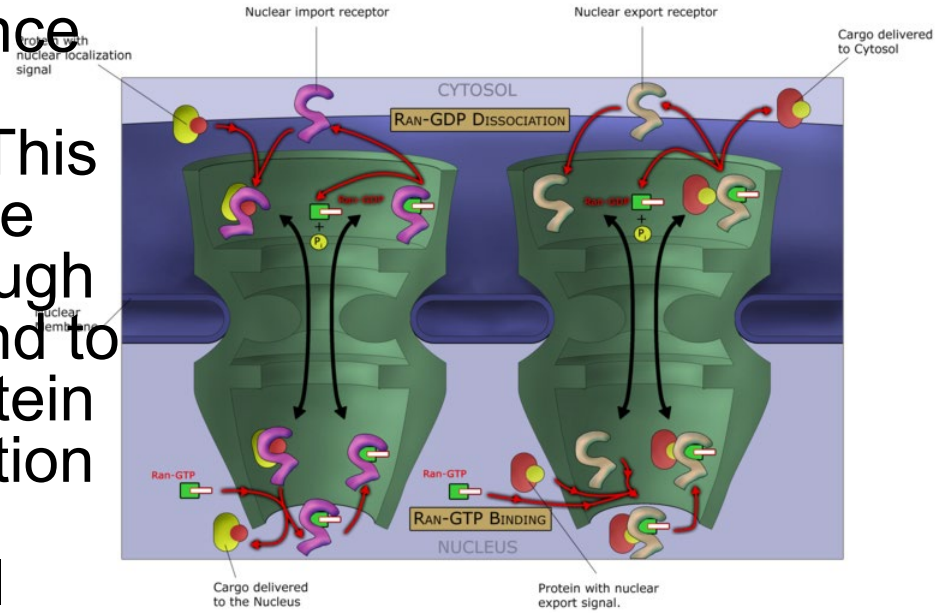
Nuclear pores

Nuclear pores, which provide aqueous channels through the envelope, are composed of multiple proteins, collectively referred to as nucleoporins. The pores are 100 nm in total diameter; however, the gap through which molecules freely diffuse is only about 9 nm wide, due to the presence of regulatory systems within the center of the pore. This size allows the free passage of small water-soluble molecules while preventing larger molecules, such as nucleic acids and proteins, from inappropriately entering or exiting the nucleus. These large molecules must be actively transported into the nucleus instead. The nucleus of a typical mammalian cell will have about 3000 to 4000 pores throughout its envelope

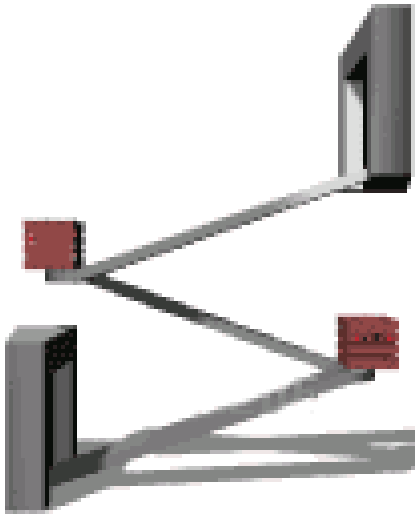


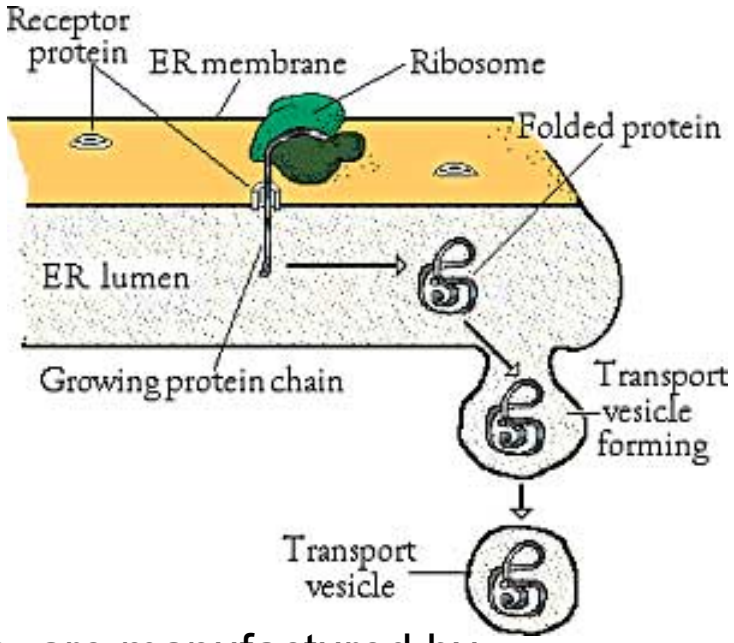
Nuclear localizing sequence (NLS)

- A **nuclear localizing sequence** (**NLS**) is an amino acid sequence which acts like a 'tag' on the exposed surface of a protein. This sequence is used to confine the protein to the cell nucleus through the **Nuclear Pore Complex** and to direct a newly synthesized protein into the nucleus via its recognition by cytosolic nuclear transport receptors. Typically, this signal consists of a few short sequences of positively charged lysines or arginines. Typically the NLS will have a sequence (NH2)-Pro-Pro-Lys-Lys-Lys-Arg-Lys-Val-(COOH).



The Ribosomes and the ER



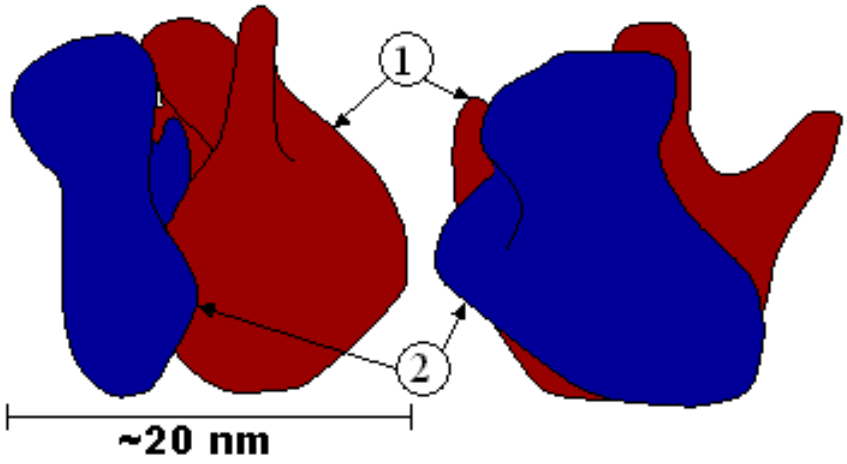


Ribosomes, the workers that build proteins, are manufactured by the nucleolus. They consist of two separate subunits: a large, lower subunit and a small, upper subunit. Ribosomes attach to the rough ER . Now let's take a look at how final processing occurs

The cell has its own assembly line and workers. Within the cytoplasm is a series of large, flattened membranes that fold back and forth on each other and have a very large surface area. This collection of membranes is called the **ENDOPLASMIC RETICULUM**, or**ER**.

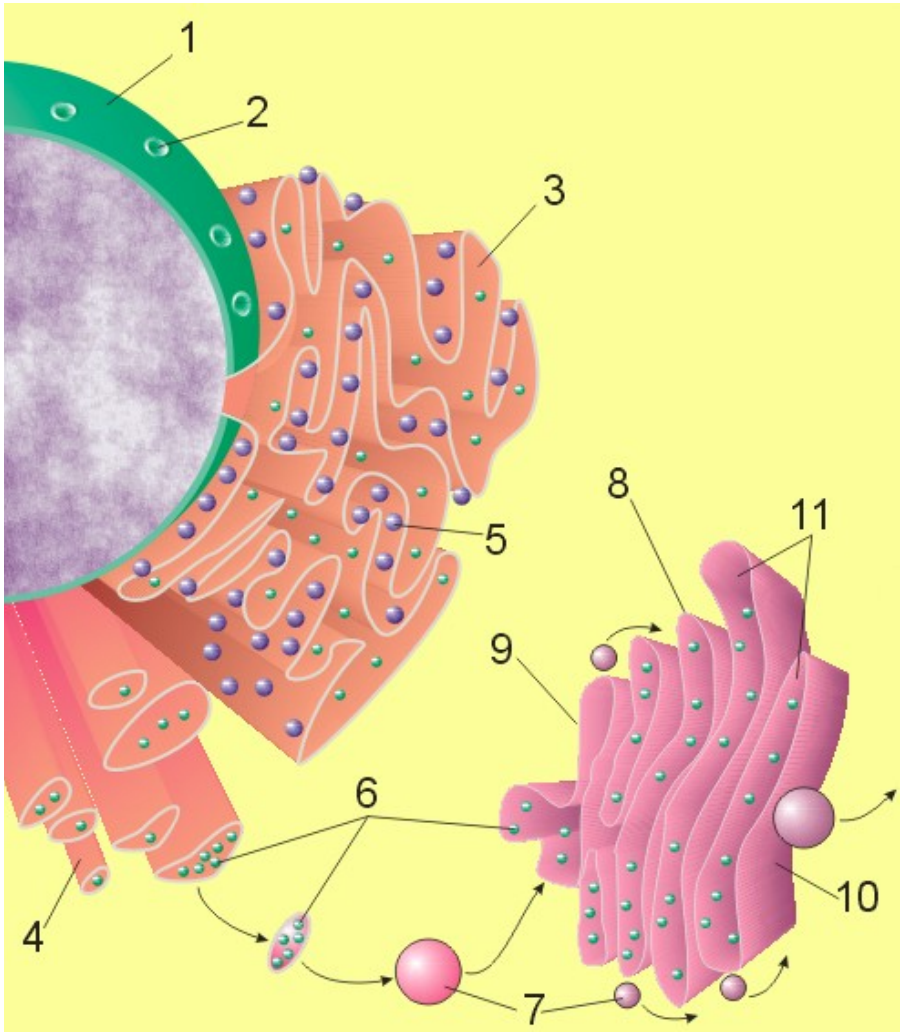
Ribosome

A **ribosome** is a small, dense organelle in cells that assembles proteins. Ribosomes are about 20nm in diameter and are composed of 65% ribosomal RNA and 35% ribosomal proteins (known as a Ribonucleoprotein or RNP). It translates messenger RNA (mRNA) to build a polypeptide chain (e.g., a protein) using amino acids delivered by Transfer RNA (tRNA). It can be thought of as a giant enzyme that builds a protein from a set of genetic instructions. Ribosomes can float freely in the cytoplasm (the internal fluid of the cell) or bound to the endoplasmic reticulum, or to the nuclear envelope.



Endoplasmic Reticulum

The **endoplasmic reticulum** or **ER** is an organelle found in all eukaryotic cells that is an interconnected network of tubules, vesicles and cisternae that is responsible for several specialized functions: Protein translation, folding, and transport of proteins to be used in the cell membrane (e.g., transmembrane receptors and other integral membrane proteins), or to be secreted (exocytosed) from the cell (e.g., digestive enzymes); sequestration of calcium; and production and storage of glycogen, steroids, and other macromolecules.[1] The endoplasmic reticulum is part of the endomembrane system. The basic structure and composition of the ER membrane is similar to the plasma membrane.



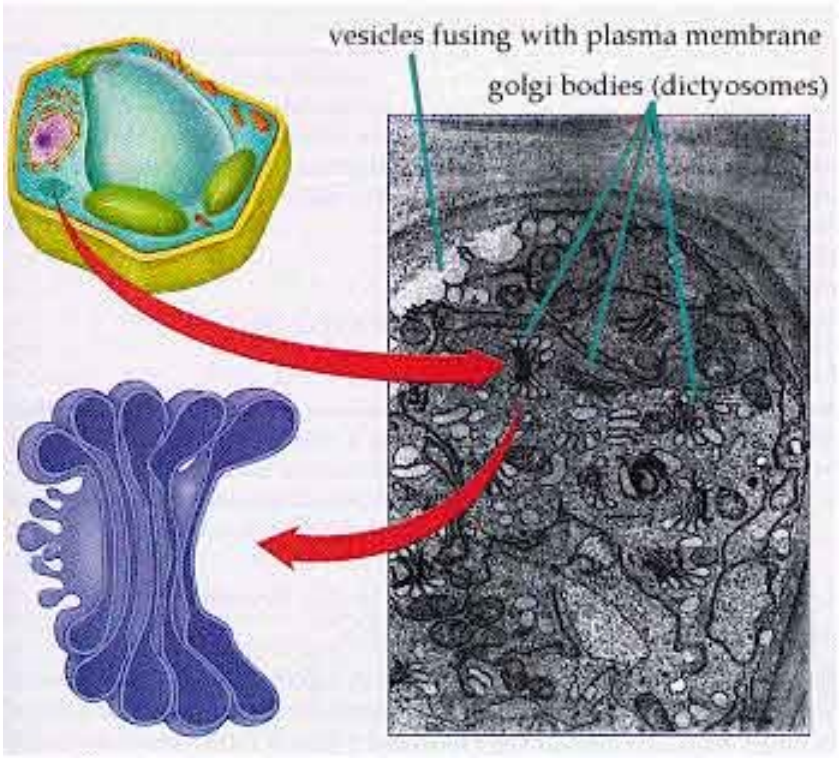
Rough endoplasmic reticulum

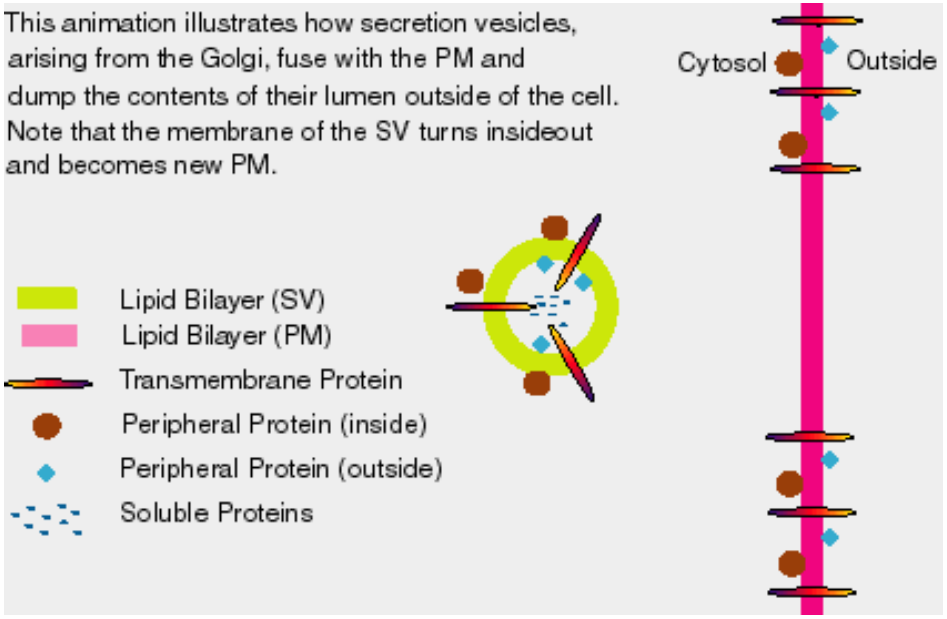
- The surface of the rough endoplasmic reticulum is studded with protein-manufacturing ribosomes giving it a "rough" appearance. But it should be noted that these ribosomes are not resident of the endoplasmic reticulum incessantly. The ribosomes only bind to the ER once it begins to synthesize a protein destined for sorting. The membrane of the rough endoplasmic reticulum is continuous with the outer layer of the nuclear envelope. Although there is no continuous membrane between the rough ER and the Golgi apparatus, membrane bound vesicles shuttle proteins between these two compartments. The rough endoplasmic reticulum works in concert with the Golgi complex to target new proteins to their proper destinations

Smooth endoplasmic reticulum

- The smooth endoplasmic reticulum has functions in several metabolic processes, including synthesis of lipids, metabolism of carbohydrates and calcium concentration, and attachment of receptors on cell membrane proteins. It is connected to the nuclear envelope. Smooth endoplasmic reticulum is found in a variety of cell types (both animal and plant) and it serves different functions in each. It consists of tubules and vesicles that branch forming a network. In some cells there are dilated areas like the sacs of rough endoplasmic reticulum. The network of smooth endoplasmic reticulum allows increased surface area for the action or storage of key enzymes and the products of these enzymes. The smooth endoplasmic reticulum is known for its storage of calcium ions in muscle cells.

The Golgi Apparatus

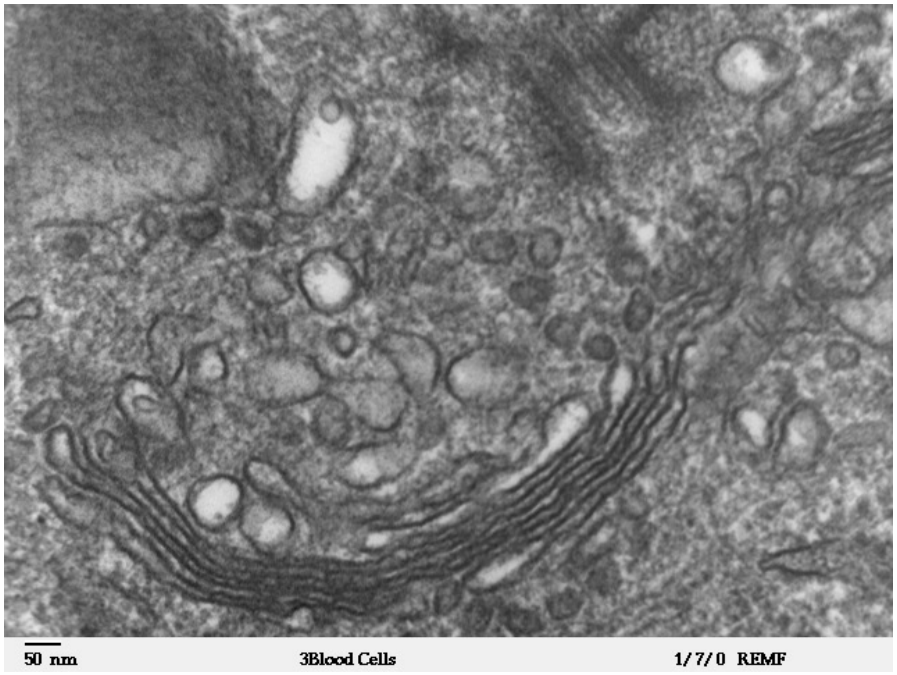




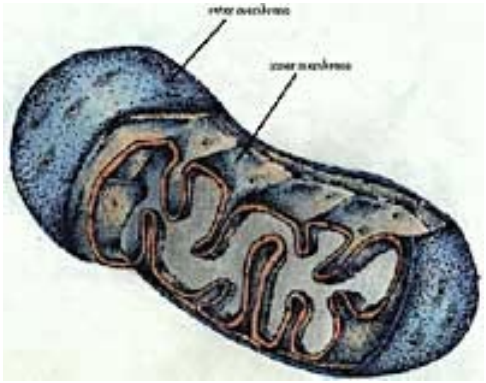
The Golgi apparatus is analogous to the finishing and packing room in a factory. Once the ribosome finishes manufacturing a protein in the rough ER, the protein needs to be prepared for use or export. Special enzymes will trim off any extra amino acids, and then the unfinished protein moves through channels in the smooth ER.

Golgi apparatus

The **Golgi apparatus** (also called the **Golgi body**, **Golgi complex**, or **dictyosome**) is an organelle found in typical eukaryotic cells. It was identified in 1898 by the Italian physician Camillo Golgi and was named after him. The primary function of the Golgi apparatus is to process and package macromolecules synthesised by the cell, primarily proteins and lipids. The Golgi apparatus forms a part of the endomembrane system present in eukaryotic cells.



Mitochondria

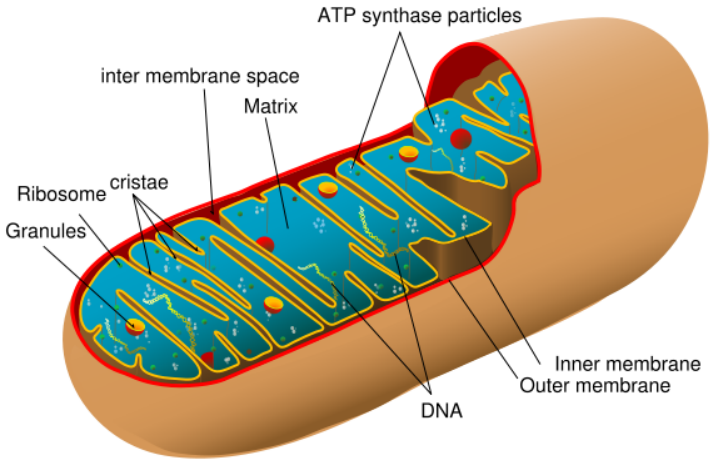


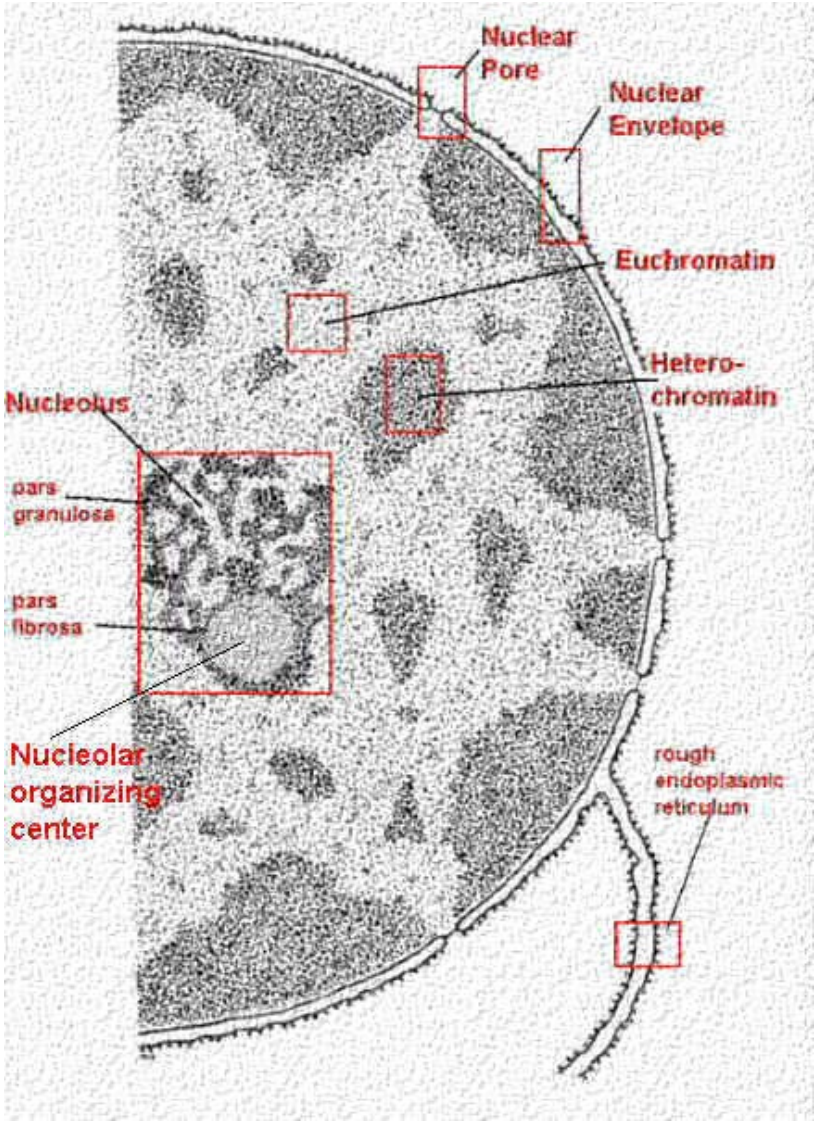
- ATP -
a Source of
ENERGY

Like our factory's power plant, mitochondria and chloroplasts transform one form of energy to another. Remember that nearly all the energy used by living things on Earth comes from the Sun. This section discusses how energy is made available for cell processes.

Mitochondrion

- In cell biology, a **mitochondrion** is a membrane-enclosed organelle, found in most eukaryotic cells.Mitochondria are sometimes described as "cellular power plants," because they convert NADH and NADPH into energy in the form of ATP via the process of oxidative phosphorylation. A typical eukaryotic cell contains about 2,000 mitochondria, which occupy roughly one fifth of its total volume. Mitochondria contain DNA that is independent of the DNA located in the cell nucleus. According to the endosymbiotic theory, mitochondria are descended from free-living prokaryotes.





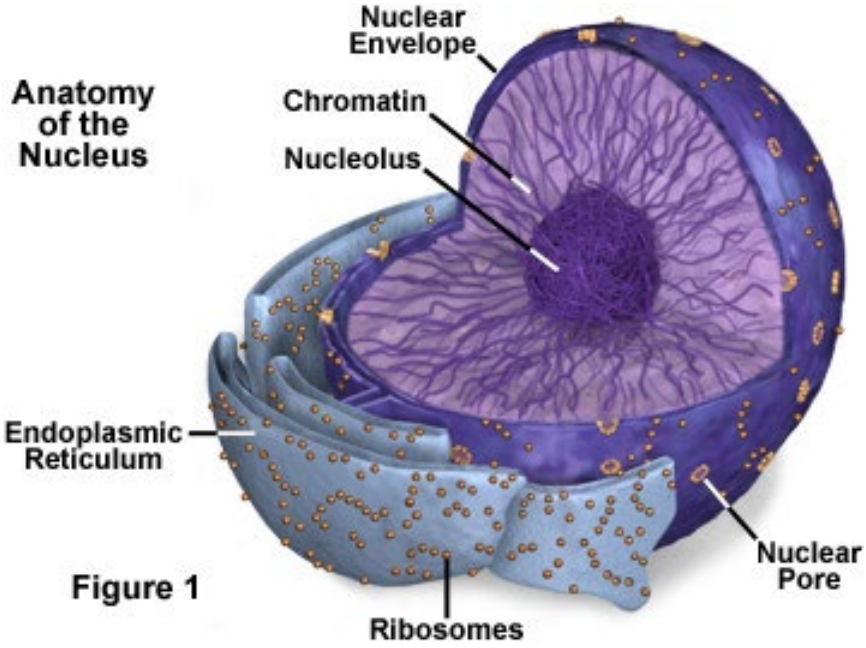


Figure 1

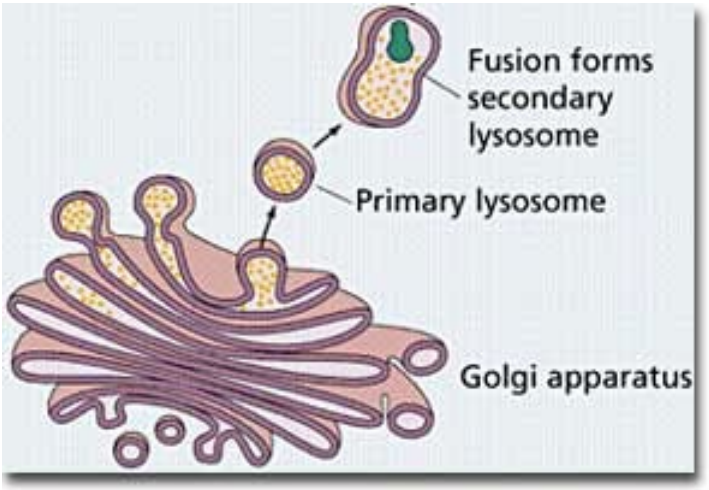
The main roles of the nucleolus are to synthesize rRNA and assemble ribosomes

The main function of the cell nucleus is to control gene expression and mediate the replication of DNA during the cell cycle

Lysosomes

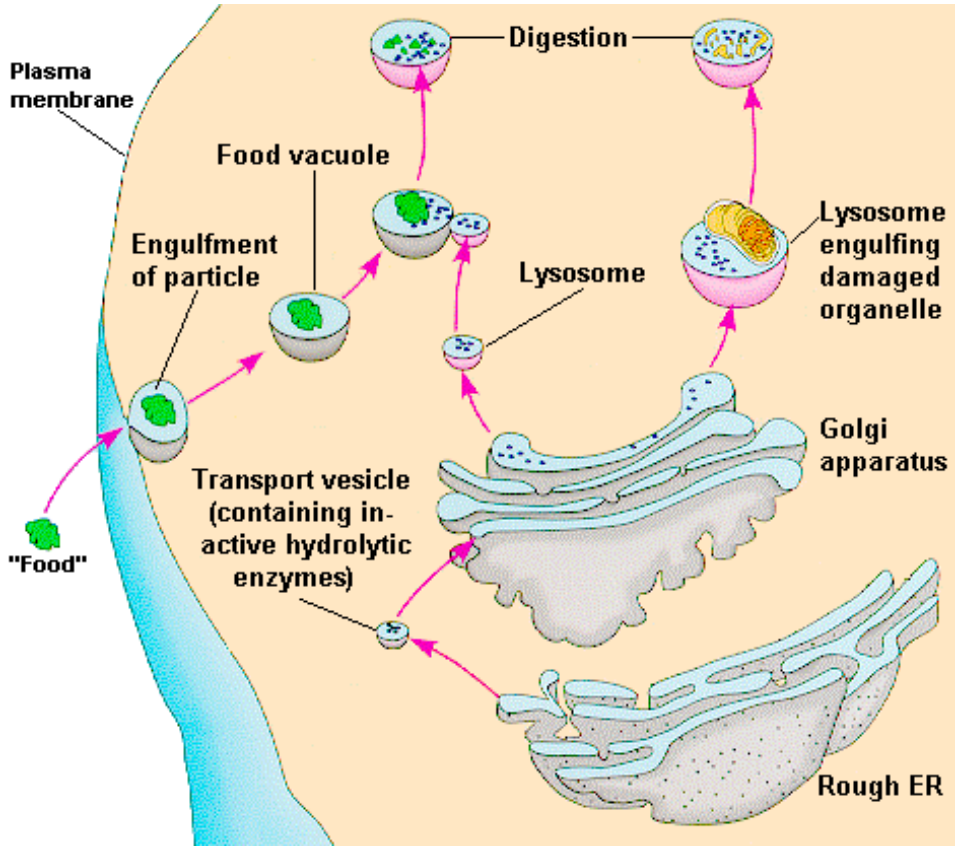
- **Lysosomes** are organelles that contain digestive enzymes (acid hydrolases). They digest excess or worn out organelles, food particles, and engulfed viruses or bacteria. The membrane surrounding a lysosome prevents the digestive enzymes inside from destroying the cell. Lysosomes fuse with vacuoles and dispense their enzymes into the vacuoles, digesting their contents. They are built in the Golgi apparatus. The name *lysosome* derives from the Greek words *lysis*, which means dissolution or destruction, and *soma*, which means body. They are frequently nicknamed "suicide-bags" or "suicide-sacs" by cell biologists due to their role in autolysis.

Lysosomes



Lysosomes are responsible for the breakdown and absorption of materials taken in by the cell. Often, a cell engulfs a foreign substance through **ENDOCYTOSIS**, another form of active transport. During endocytosis, the cell membrane puckers up, forms a pouch around materials outside the cell, and pinches off to become a vesicle. If the contents need to be destroyed, lysosomes combine with the vesicle and release their enzymes.

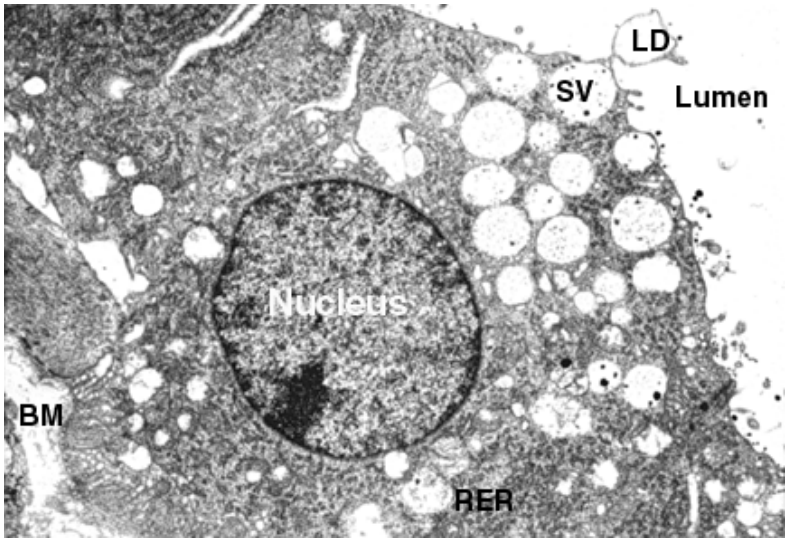
Lysosome

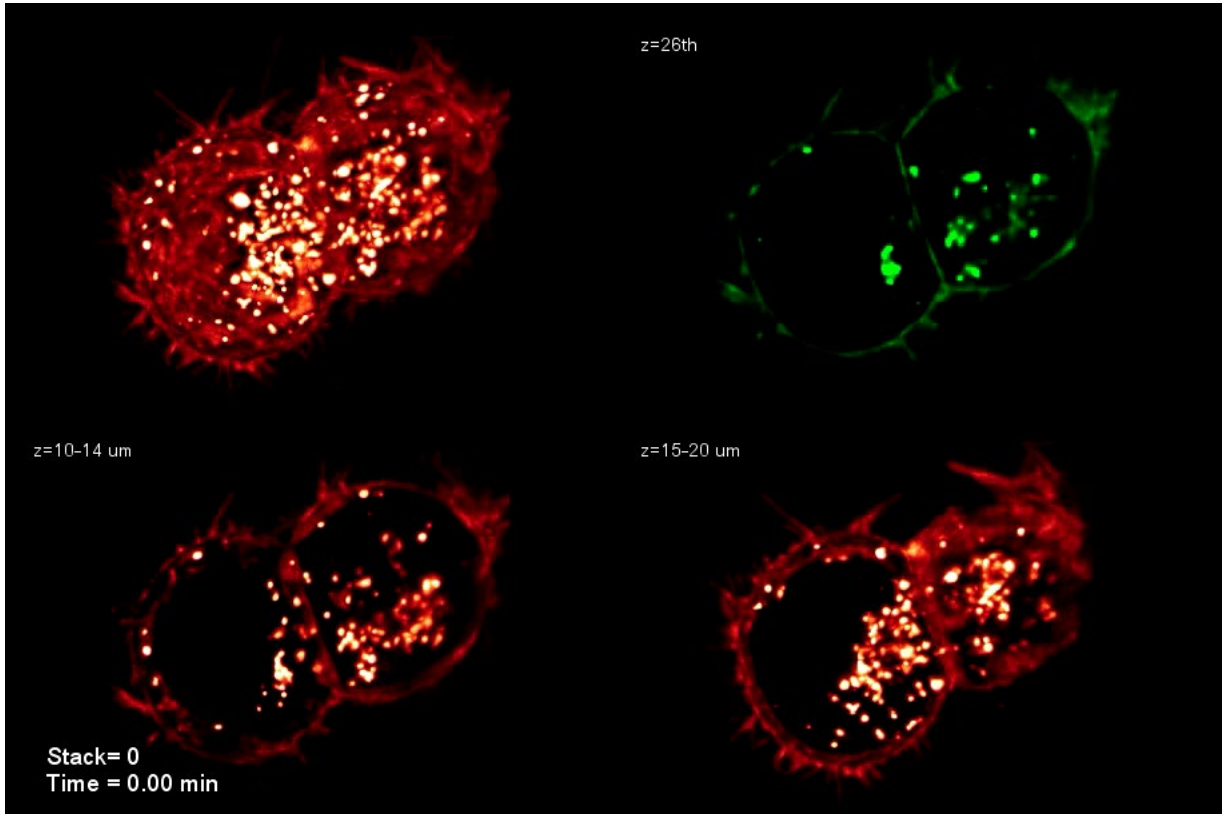

Digestion
Plasma membrane
Food vacuole
Engulfment of particle
Lysosome
Lysosome engulfing damaged organelle
Golgi apparatus
Transport vesicle (containing in-active hydrolytic enzymes)
"Food"
Rough ER

Vesicle

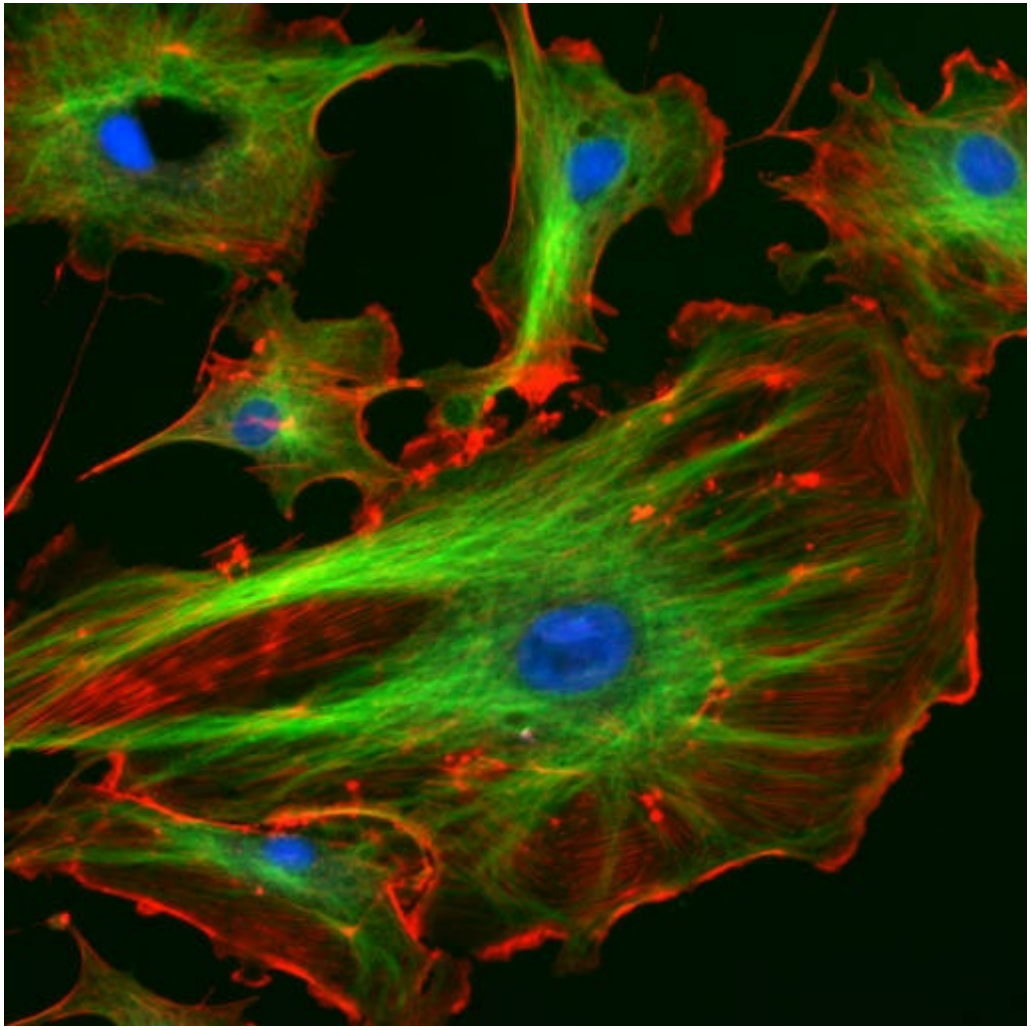
In cell biology, a **vesicle** is a relatively small and enclosed compartment, separated from the cytosol by at least one lipid bilayer. If there is only one lipid bilayer, they are called *unilamellar* vesicles; otherwise they are called *multilamellar*. Vesicles store, transport, or digest cellular products and waste.

This biomembrane enclosing the vesicle is similar to that of the plasma membrane. Because it is separated from the cytosol, the intravesicular environment can be made to be different from the cytosolic environment. Vesicles are a basic tool of the cell for organizing metabolism, transport, enzyme storage, as well as being chemical reaction chambers. Many vesicles are made in the Golgi apparatus, but also in the endoplasmic reticulum, or are made from parts of the plasma membrane.



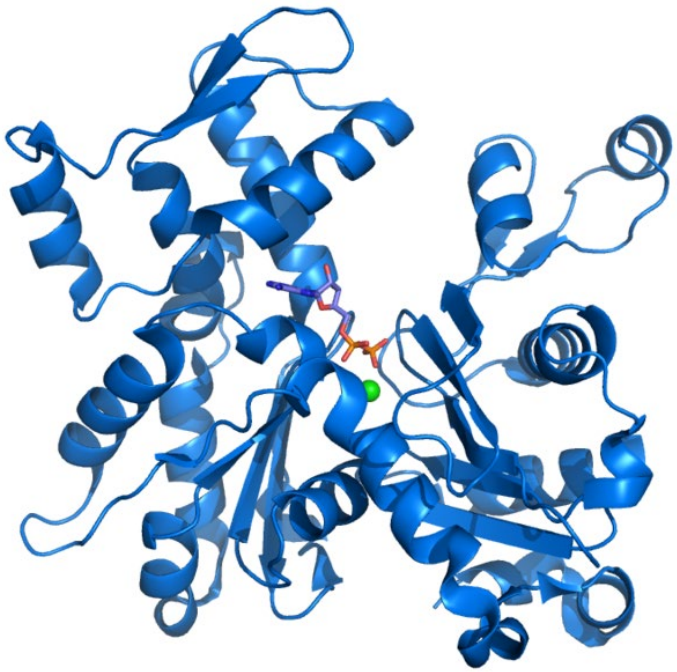

z=26th
z=10-14 um
z=15-20 um
Stack= 0
Time = 0.00 min

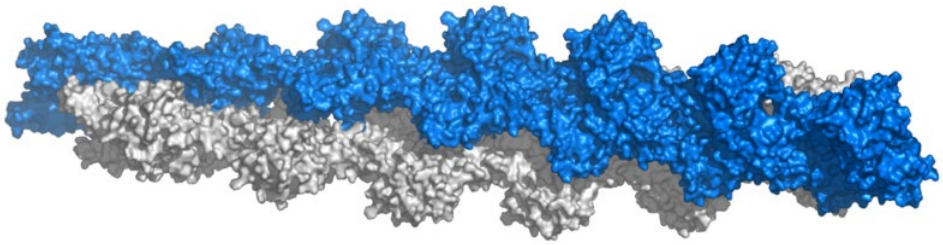
Cytoskeleton



The eukaryotic cytoskeleton. Actin filaments are shown in red, microtubules in green, and the nuclei are in blue.

Actin

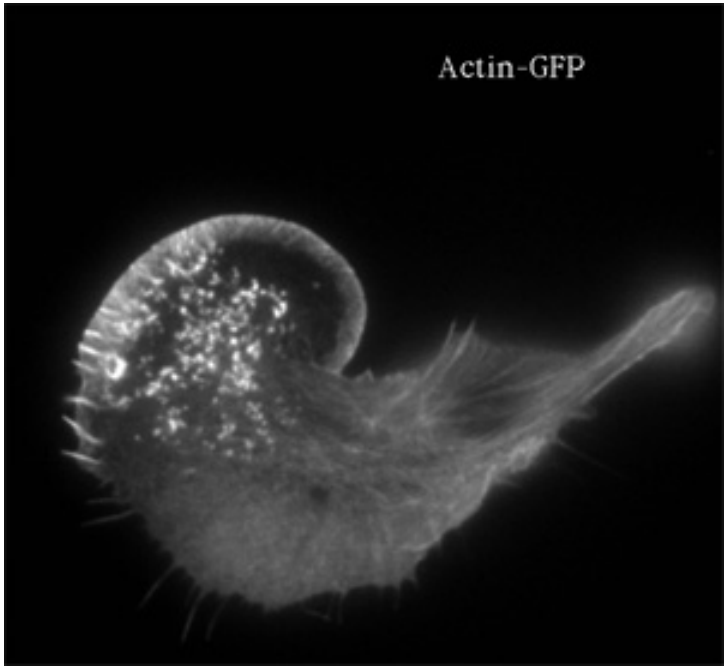


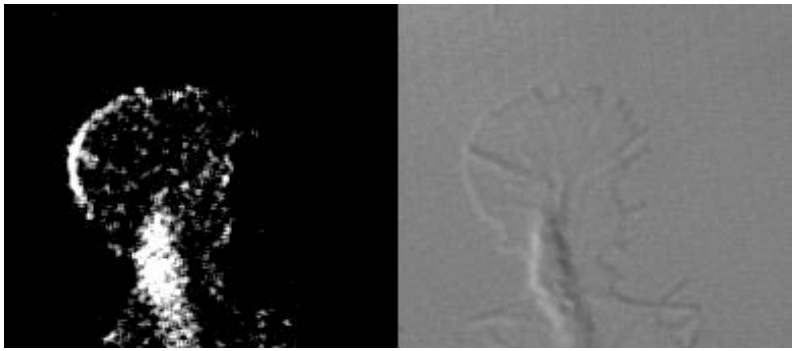


- **Actin** is a globular structural, 42 kDa, protein that polymerizes in a helical fashion to form **actin filaments** (or **microfilaments**). These form the cytoskeleton, a three-dimensional network inside the eukaryotic cell. Actin filaments provide mechanical support for the cell, determine its shape, and enable movement of the cell through lamellipodia, filopodia, or pseudopodia. Actin filaments, along with myosin, have an essential role in muscular contraction. In the cytosol, actin is predominantly bound to ATP, but can also bind to ADP. An ATP-actin complex polymerizes faster and dissociates slower than an ADP-actin complex.

Lamellipodia

- The **lamellipodium** is a cytoskeletal actin projection on the mobile edge of the cell. It contains a two-dimensional actin mesh; the whole structure pulls the cell across a substrate. Within the lamellipodia are ribs of actin called microspikes, which, when they spread beyond the lamellipodium frontier, are called filopodia (Small, et all, 2002). The lamellipodium is born of actin nucleation in the plasma membrane of the cell (Alberts, et al, 2002) and is the primary area of actin incorporation or microfilament formation of the cell. Lamellipodia range from 1µm to 5µm in breadth and are approximately 0.2µm thick.Lamellipodia are found primarily in very mobile cells, crawling at a speeds of 10-20µm/minute over epithelial surfaces..
- The tip of the lamellipodium is the site where exocytosis occurs in migrating mammalian cells as part of their clathrin-mediated endocytic cycle.

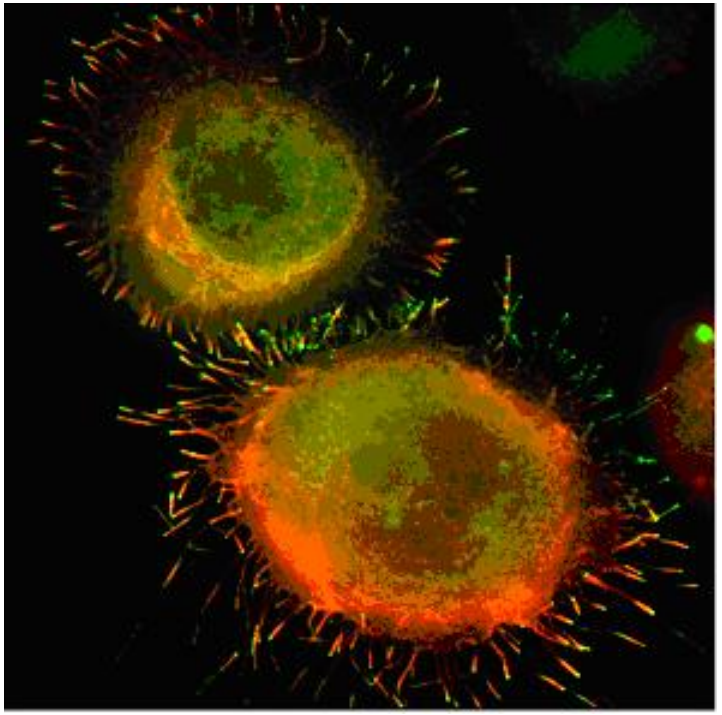




http://www.microscopyu.com/moviegallery/livecellimaging/3t3/t1/3t3-dslwmp1.html

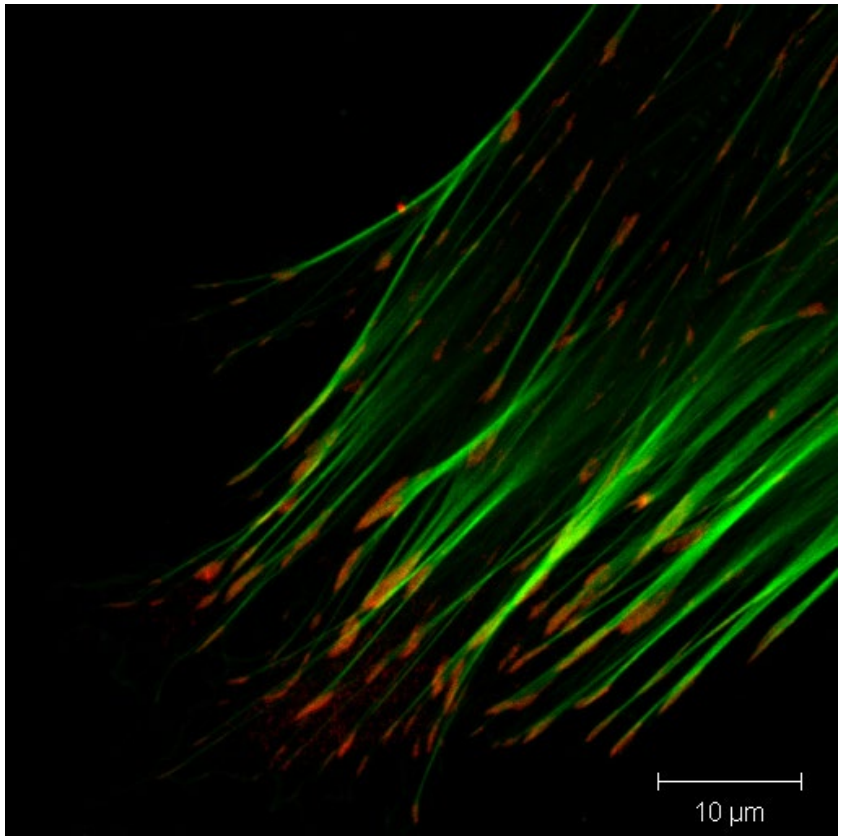
Filopodia

The **filopodia** are slender cytoplasmic projections, similar to lamellipodia, which extend from the leading edge of migrating cells. They contain actin filaments cross-linked into bundles by actin-binding proteins, e.g. fimbrin. Filopodia form focal adhesions with the substratum, linking it to the cell surface. A cell migrates along a surface by extending filopodia at the leading edge. The filopodia attach to the substratum further down the migratory pathway, then contraction of stress fibres retracts the rear of the cell to move the cell forwards.



Focal adhesion

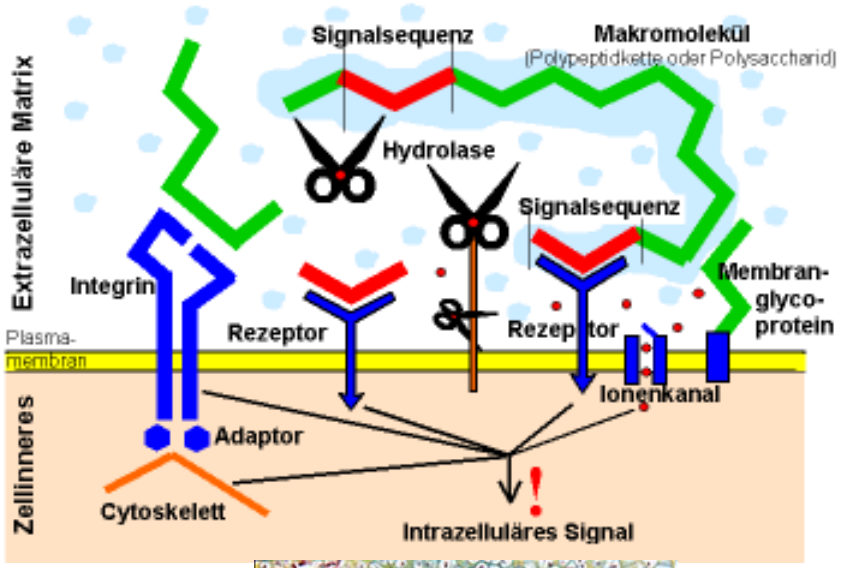
- In cell biology, '**Focal Adhesions'** are specific types of large macromolecular assemblies through which both mechanical force and regulatory signals are transmitted. More precisely, **FAs** can be considered as sub-cellular macromolecules that mediate the regulatory effects (e.g. cell anchorage) of extracellular matrix (ECM) adhesion on cell behavior.

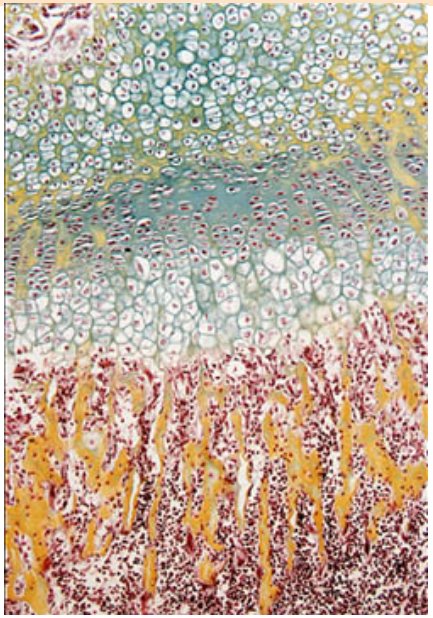


Extra Cellular Matrix

The ECM's main components are various glycoproteins, proteoglycans and hyaluronic acid. In most animals, the most abundant glycoproteins in the ECM are collagens.

ECM also contains many other components: proteins such as fibrin, elastin, fibronectins, laminins, and nidogens, and minerals such as hydroxylapatite, or fluids such as blood plasma or serum with secreted free flowing antigens.



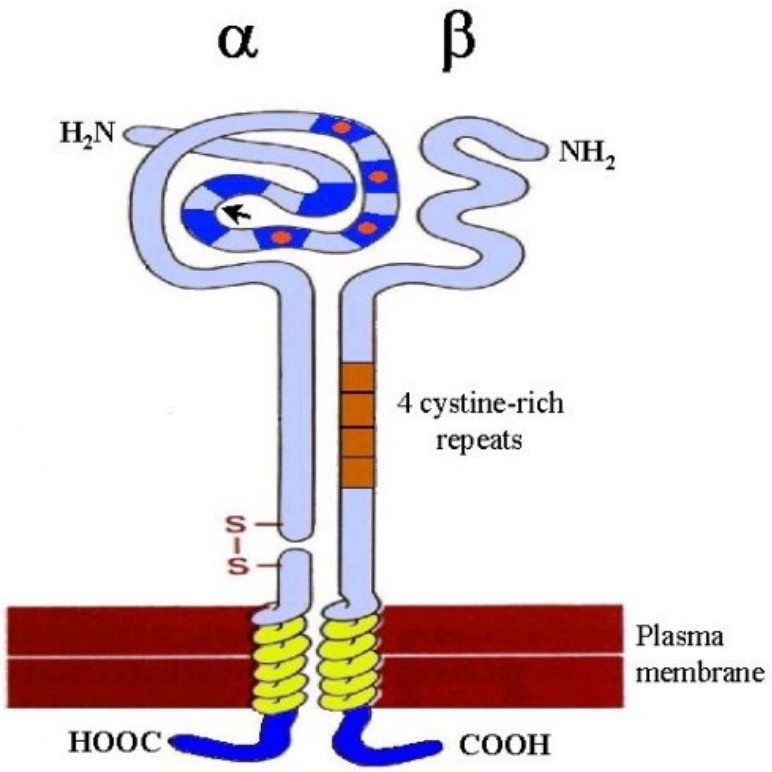


Integrin

An **integrin**, or **integrin receptor**, is an integral membrane protein in the plasma membrane of cells. It plays a role in the attachment of a cell to the extracellular matrix (ECM) and to other cells, and in signal transduction from the ECM to the cell. There are many types of integrin, and many cells have multiple types on their surface. Integrins are of vital importance to all metazoans, from humans to sponges.

Schematic drawing of a typical integrin dimer

Arrow shows the region where an I domain is inserted in some α subunits. Not all α subunits are posttranslationally cleaved. Internal disulphide bonds within subunits are not shown. Dark blue regions in the head segment of the α subunit represent homologous repeats. Those with the EF-hand consensus sequence are marked with red circles to denote binding sites for divalent metal ion.



Endocytosis