Special-topic Lecture Bioinformatics: Modeling Cell Fate

Leistungspunkte/Credit points: 5 (V2/Ü1)

This course is taught in English language.

The material (from books and original literature) are provided online at the course website:

http://gepard.bioinformatik.uni-saarland.de/teaching/ss-2013/stl-bioinformatics-modcellfate-ss13

Biological topics to be covered:

This course will enter into details of three selected topics in current cell biology:

(1) Cell cycle

(2) Stem cell differentiation

(3) Cancerogenesis

Bioinformatics content

- microarray expression analysis
- DNA methylation analysis
- GO and pathway annotation
- interaction networks
- application of clustering techniques
- construction of gene-regulatory networks
- stochastic simulations

Aim of this lecture, „Lernziele“

(1) The aim of this course is not to fully cover these three topics but to **enter deeply** into various **details** of these fields.

(2) This course should train you to **analyze** original biological **data** using modern bioinformatics tools.

(3) You shoud also become familiar with the biological processes (**pathways**) controlling cellular adaptation / cell fate.

Tutorial

We will handout 6 **biweekly assignments**.
Groups of up to two students can hand in a solved assignment.

Send your **solutions** by e-mail to the responsible tutors :

Mohamed Hamed, Ruslan Akulenko, Christian Spaniol

until the time+date indicated on the assignment sheet.

The **tutorial** on Thursday 12 am - 1 pm will provide help to understand the papers, prepare the student presentations and the assignment solutions.

Schein condition 1

Only those students can get a „Schein“ who have obtained more than 50% of the points for all assignments.

Schein = pass 3 written tests

Schein condition 2

The successful participation in the lecture course („Schein“) will be certified upon fulfilling Schein condition 1 and upon successful completion of 3 written 45 minute tests.

Each test roughly covers the content of one of the three lecture topics.

Dates: probably at the beginning of lectures V5, V9, V13.

All students registered for the course may participate in the tests.

2 out of 3 tests have to be passed.

The final grade on the Schein is the average of your 2 best tests.

Rounding scheme: (1.0 + 1.3 -> 1.0 ; 1.3 + 2.0 -> 1.7)

written tests

The tests will cover the lecture material (slides on the lecture website) and the theory behind the assignments for this topic.

In case of illness please send E-mail to:
kerstin.gronow-p@bioinformatik.uni-saarland.de and provide a medical certificate.

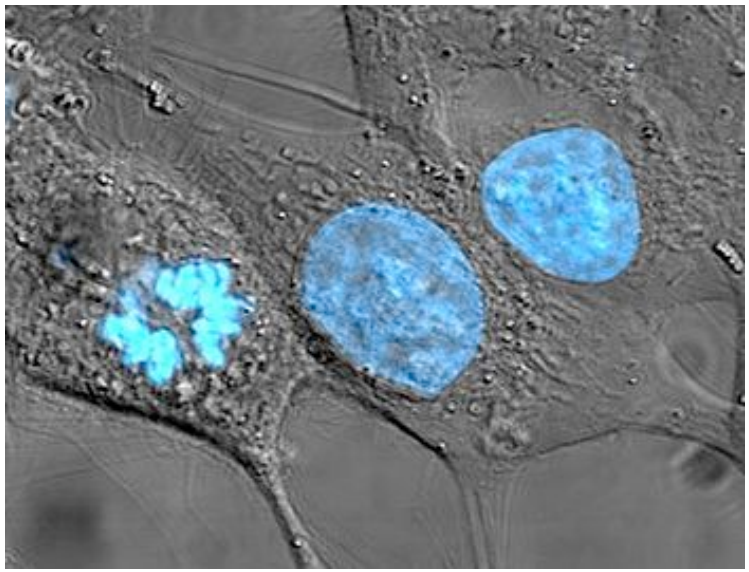
Those who miss or fail one test, will be given a second-chance oral exam.
If you fail or miss more than two tests, you cannot get a Schein.

Gene Transcription etc.

Basic terms that you should remember from an introductory genetics lecture ...
or that you should read up:

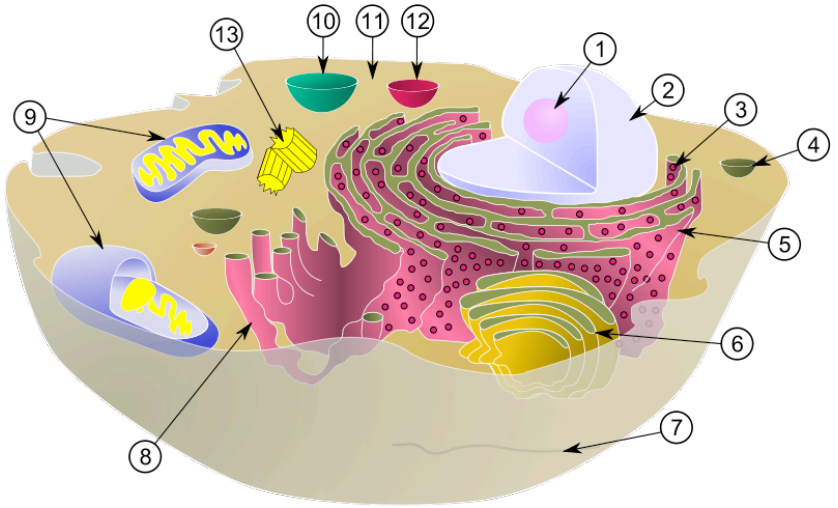
Genome
Genes
Introns, Exons
Nucleus
DNA-Polymerase
Transcription
mRNA
Splicing
Ribosome
tRNA
Translation

A biological cell



HeLa cells stained for DNA with the Blue Hoechst dye. The central and rightmost cell are in interphase, thus their entire nuclei are labeled. On the left a cell is going through mitosis and its DNA has condensed ready for division.

wikipedia.org

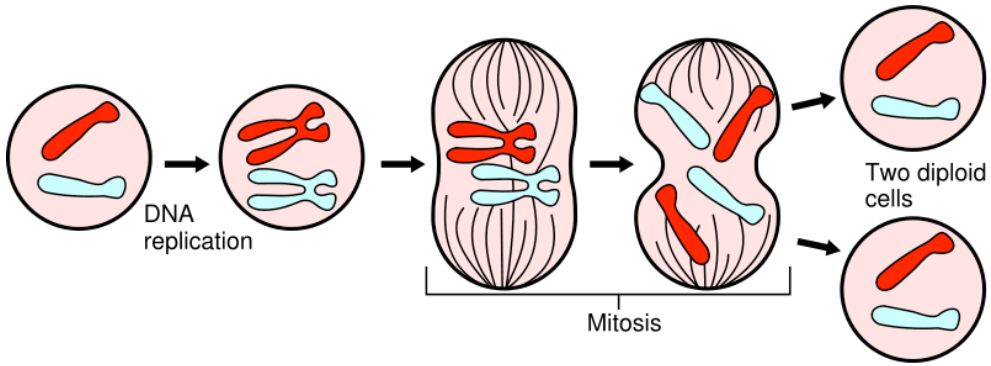


Schematic of typical animal cell, showing subcellular components.
Organelles: (1) nucleolus (2) nucleus (3) ribosome (4) vesicle (5) rough endoplasmic reticulum (ER) (6) Golgi apparatus (7) Cytoskeleton (8) smooth ER (9) mitochondria (10) vacuole (11) cytoplasm (12) lysosome (13) centrioles

cell cycle

The cell cycle, or cell-division cycle, is the series of events that takes place in a cell leading to its division and duplication (replication).

In cells without a nucleus (prokaryotes), the cell cycle occurs via a process termed **binary fission**.



Each turn of the cell cycle divides the chromosomes in a cell nucleus.

www.wikipedia.org

In cells with a nucleus (eukaryotes), the cell cycle can be divided in 2 brief periods:

interphase—during which the cell grows, accumulating nutrients needed for mitosis and duplicating its DNA—and

the **mitosis** (M) phase, during which the cell splits itself into two distinct cells, often called "daughter cells".

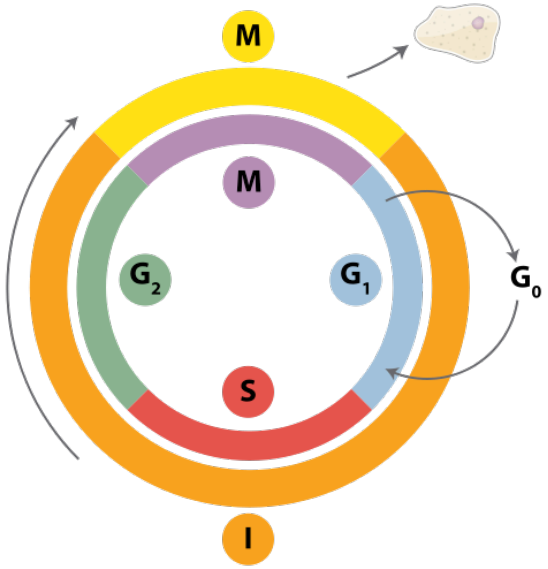
Phases

The cell cycle consists of **4 distinct phases**:
- G_1 phase,
- S phase (synthesis),
- G_2 phase
- and M phase (mitosis).

Interphase: combines G_1, S, and G_2

Activation of each phase is dependent on the proper progression and completion of the previous one.

Cells that have temporarily or reversibly stopped dividing are said to have entered a state of quiescence called **G_0 phase**.

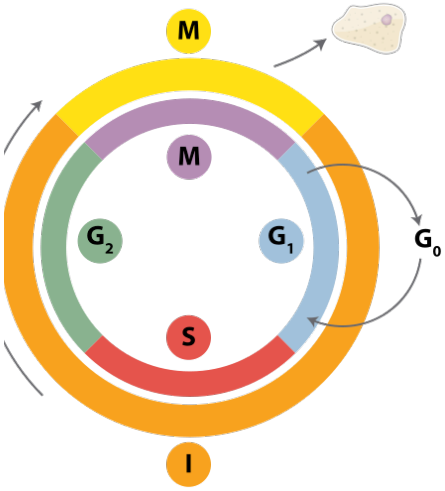


Schematic of the cell cycle.
Outer ring:
I = Interphase, M = Mitosis;
Inner ring:
M = Mitosis, G1 = Gap 1, G2 = Gap 2, S = Synthesis.

www.wikipedia.org

Activity during 4 phases

State	Phase	Abbreviation	Description
quiescent/ senescent	Gap 0	G_0	A resting phase where the cell has left the cycle and has stopped dividing.
Interphase	Gap 1	G_1	Cells increase in size in Gap 1. The *G_1 checkpoint* control mechanism ensures that everything is ready for DNA synthesis.
	Synthesis	S	DNA replication occurs during this phase.
	Gap 2	G_2	During the gap between DNA synthesis and mitosis, the cell will continue to grow. The *G_2 checkpoint* control mechanism ensures that everything is ready to enter the M (mitosis) phase and divide.
Cell division	Mitosis	M	Cell growth stops at this stage and cellular energy is focused on the orderly division into two daughter cells. A checkpoint in the middle of mitosis (*Metaphase Checkpoint*) ensures that the cell is ready to complete cell division.



M phase itself is composed of 2 tightly coupled processes:

- mitosis, in which the cell's chromosomes are divided between the two daughter cells, and

- cytokinesis, in which the cell's cytoplasm divides in half forming distinct cells.

www.wikipedia.org

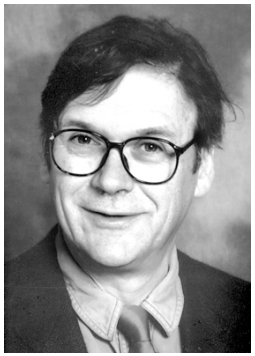
Regulation of the eukaryotic cell cycle

Regulation of the cell cycle involves processes crucial to the survival of a cell, including the detection and repair of genetic damage as well as the prevention of uncontrolled cell division.

The molecular events that control the cell cycle are ordered and directional.

Each process occurs in a sequential fashion.

It is impossible to "reverse" the cycle.



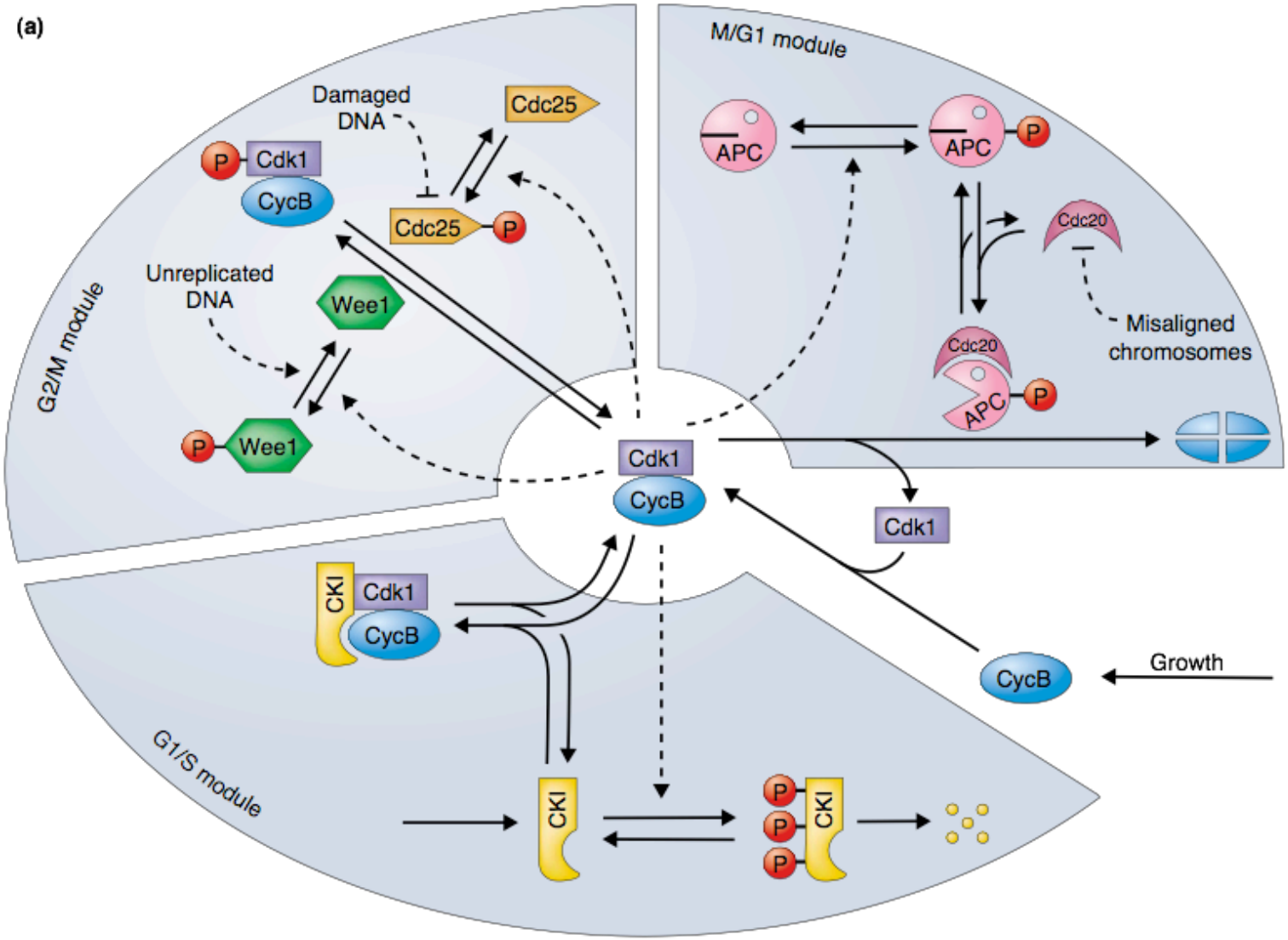
Leland Hartwell Tim Hunt Paul Nurse

Noble Price in Physiology/Medicine 2001 „for their discoveries of key regulators of the cell cycle“

Two key classes of regulatory molecules, **cyclins** and **cyclin-dependent kinases (CDKs),** determine a cell's progress through the cell cycle.

www.wikipedia.org

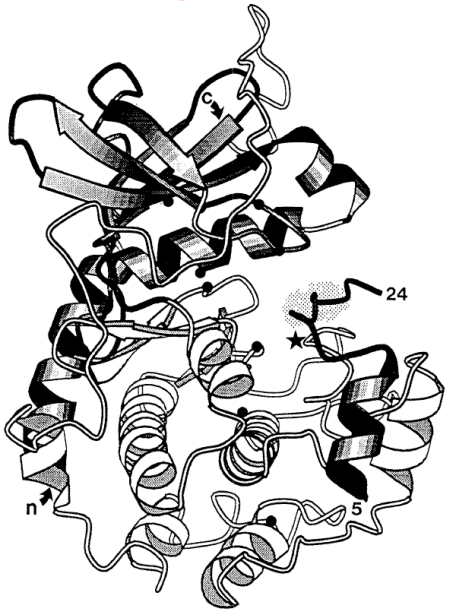
Cell cycle control model

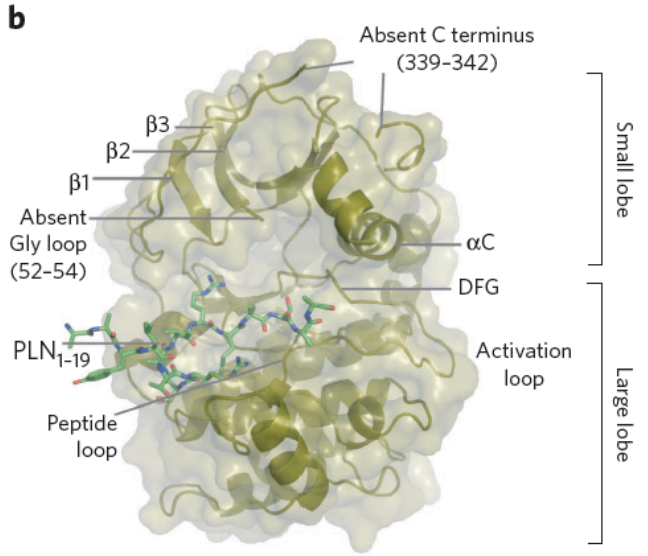


Tyson et al, *Curr. Op. Cell Biol.* **15** (2003) 221

protein kinase A: a model system for phosphate transfer

Susan S. Taylor
UC San Diego





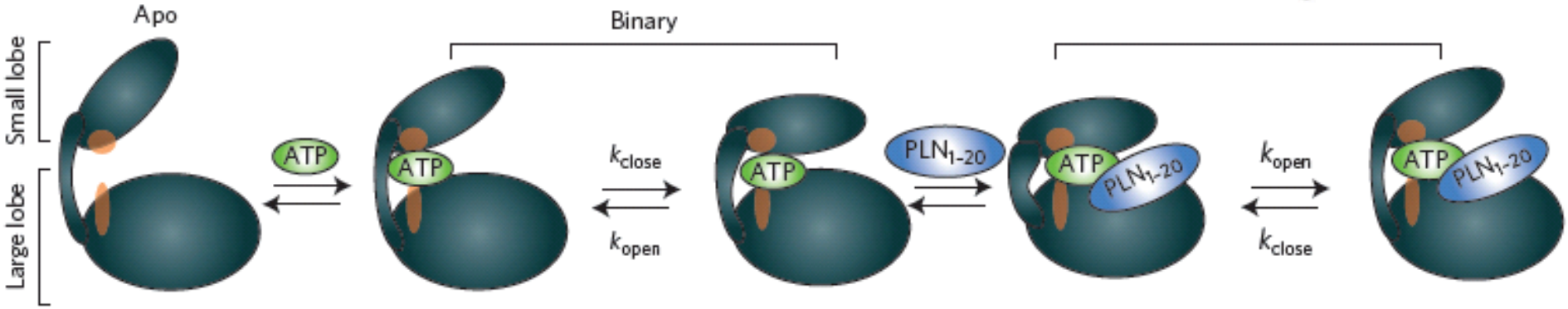
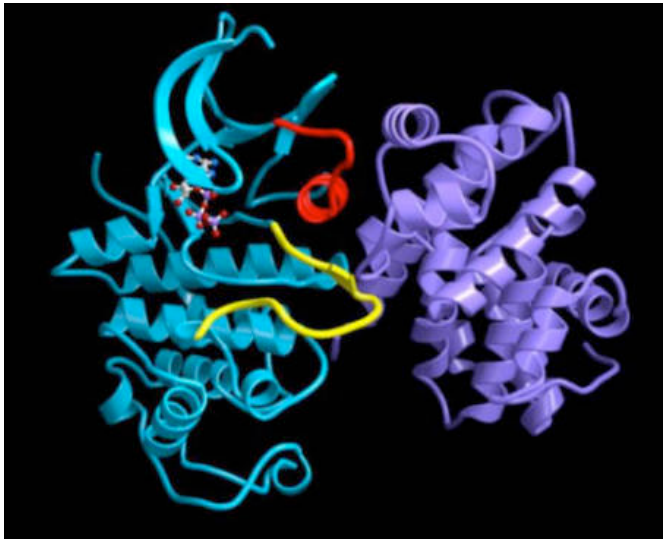


Figure 5 | Model for the mechanism of the formation of a catalytically competent ternary complex. The apo form contains the C-spine residues (red), which are disengaged from the two lobes. Nucleotide binding completes the C-spine architecture and induces the conformational changes throughout the enzyme. The conformational fluctuations (opening and closing) present in the ternary complex limit the rate of catalysis.

Masterson et al. *Nat Chem Biol.* 6, 825 (2010)

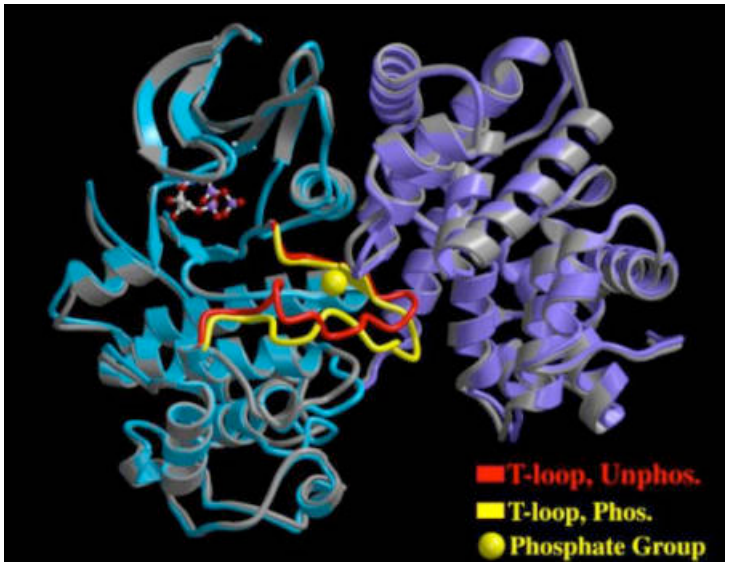
Taylor et al. Phil Trans R.Soc. B (1993)

Cyclin – cdk2 complex crystal structure



Cyclin A – cdk 2 complex
red: PSTAIRE motif
yellow: activation loop

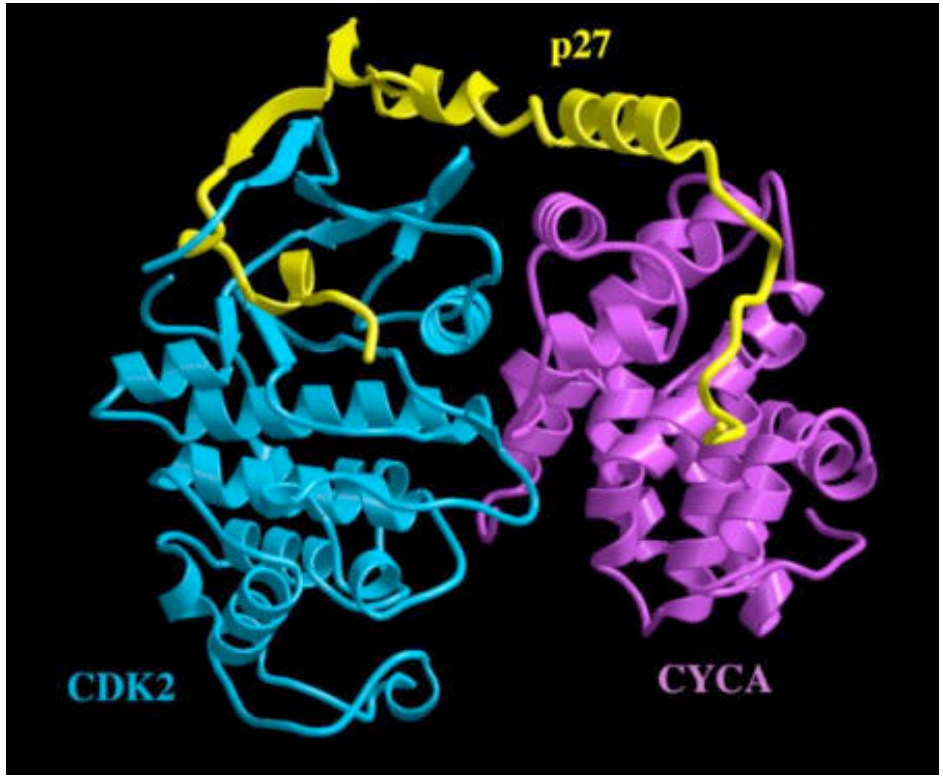
Nikola Pavletich
Memorial Sloan-Kettering Cancer Center



www.wikipedia.org

Cyclin A – cdk2 phosphorylated at Thr160

Crystal structure



p27(Kip1)-CyclinA-Cdk2 Complex

p27 (Kip1) is shown bound to the CyclinA-Cdk2 complex, provoking profound changes in the kinase active site and rendering it inactive.

p27 also interacts with the secondary substrate recognition site on the cyclin.

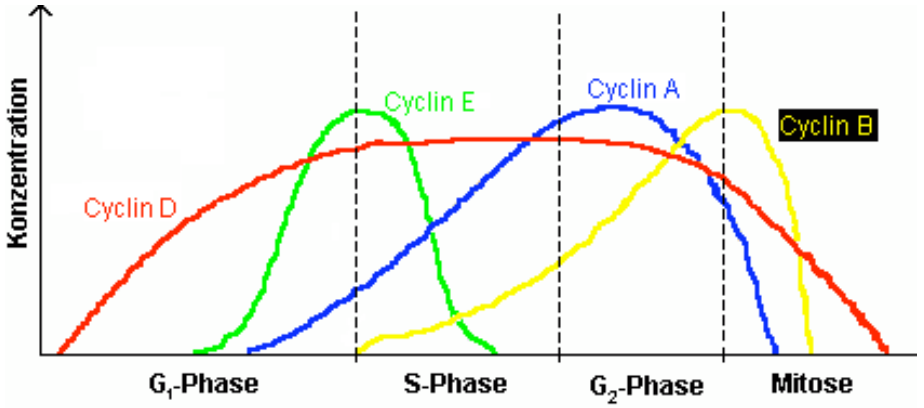
www.wikipedia.org

Cdk1-phosphorylation sites

Cdk1 substrates frequently contain multiple phosphorylation sites that are clustered in regions of intrinsic disorder.

Their exact position in the protein is often poorly conserved in evolution, indicating that precise positioning of phosphorylation is not required for regulation of the substrate.

Cdk1 interacts with nine different cyclins throughout the cell cycle.



Expression of human cyclins through the cell cycle.

Enserink and Kolodner
Cell Division 2010 **5**:11

www.wikipedia.org

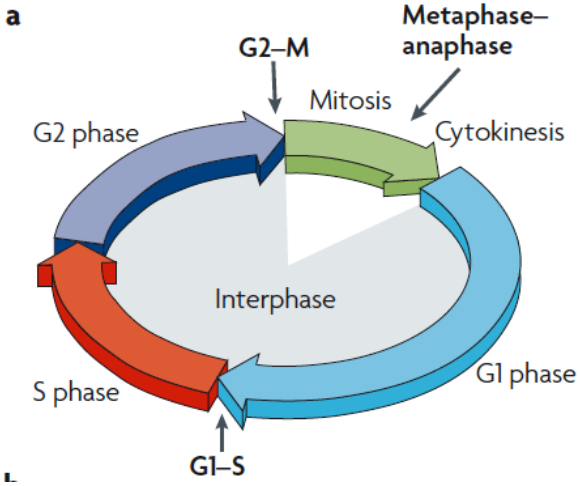
The classical model of cell-cycle control

OPINION

Cyclin-dependent kinases and cell-cycle transitions: does one fit all?

Helfrid Hochegger, Shunichi Takeda and Tim Hunt

Nature Reviews Molecular Cell Biology 9, 910-916 (2008)



Cyclin-dependent kinases (cDKs) trigger the transition from G_1 to S phase and from G_2 to M phase by phosphorylating distinct sets of substrates.

The metaphase-to-anaphase transition requires the ubiquitylation and proteasome-mediated degradation of mitotic B-type cyclins and various other proteins, and is triggered by the anaphase-promoting complex/cyclosome (APc/c) e3 ubiquitin ligase

Cell cycle checkpoints

Cell cycle **checkpoints** are control mechanisms that ensure the fidelity of cell division in eukaryotic cells.

These checkpoints verify whether the processes at each phase of the cell cycle have been accurately completed before progression into the next phase.

An important function of many checkpoints is to **assess DNA damage**, which is detected by sensor mechanisms.

When damage is found, the checkpoint uses a signal mechanism either to stall the cell cycle until **repairs** are made or, if repairs cannot be made, to target the cell for destruction via **apoptosis** (effector mechanism).

All the checkpoints that assess DNA damage appear to utilize the same sensor-signal-effector mechanism.

www.wikipedia.org

The Hallmarks of Cancer

Cell, Vol. 100, 57–70, January 7, 2000, Copyright ©2000 by Cell Press

The Hallmarks of Cancer

Douglas Hanahan* and Robert A. Weinberg†
*Department of Biochemistry and Biophysics and
Hormone Research Institute
University of California at San Francisco
San Francisco, California 94143
†Whitehead Institute for Biomedical Research and
Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02142

Robert A. Weinberg

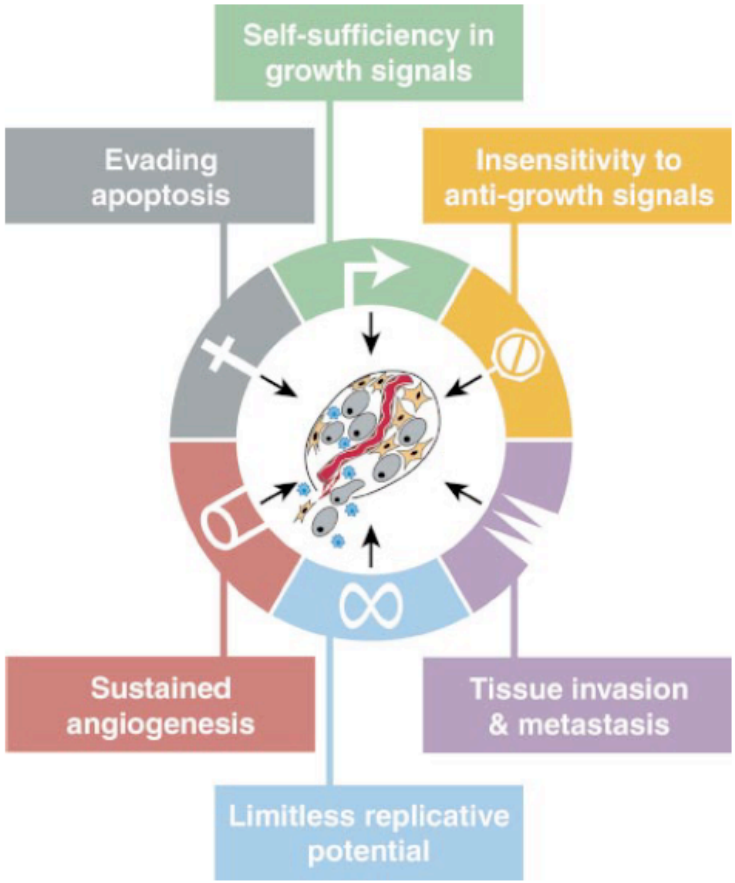


Figure 1. Acquired Capabilities of Cancer

The Hallmarks of Cancer

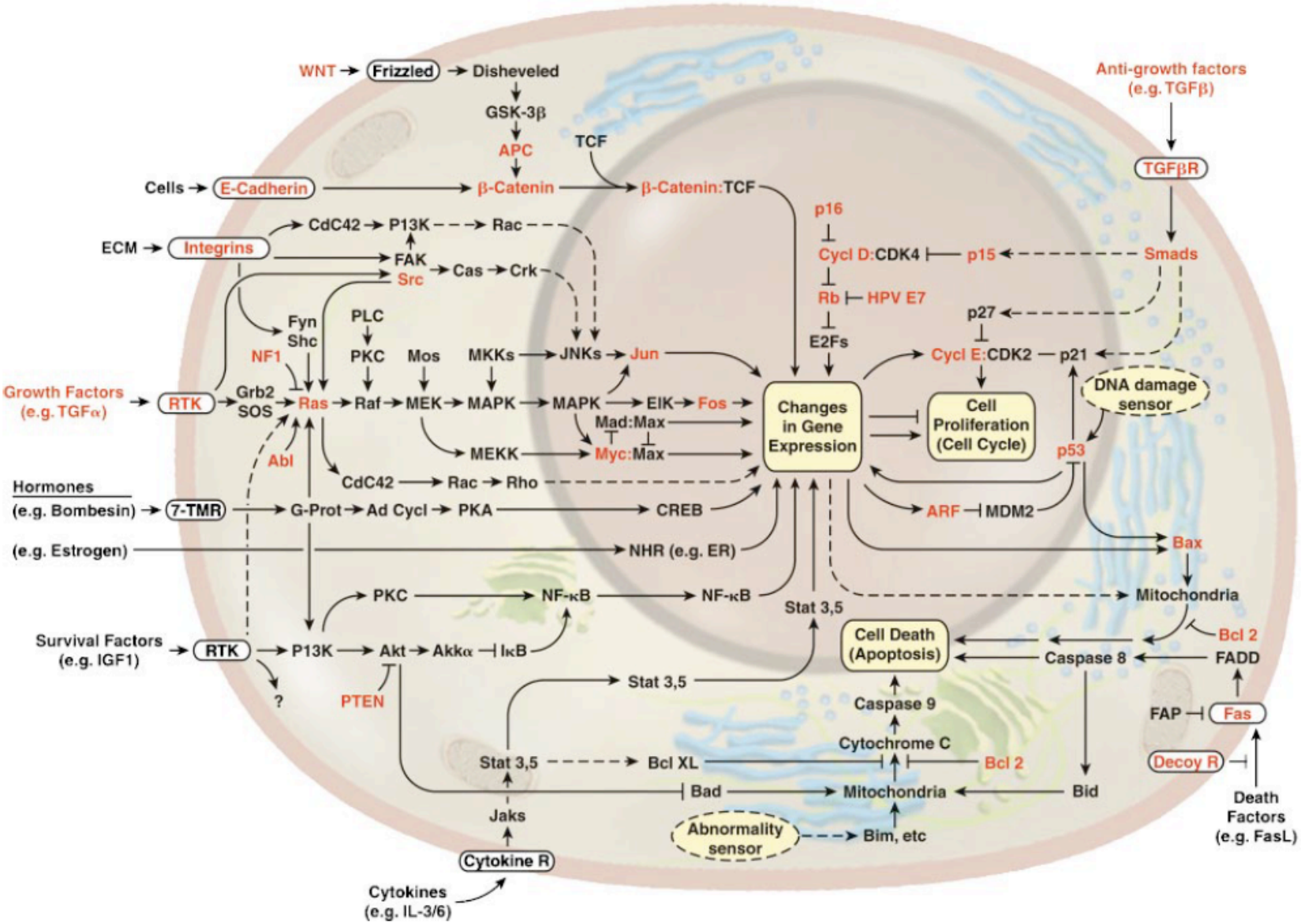


Figure 2. The Emergent Integrated Circuit of the Cell

The Hallmarks of Cancer

A

Component	Acquired Capability	Example of Mechanism
	Self-sufficiency in growth signals	Activate H-Ras oncogene
	Insensitivity to anti-growth signals	Lose retinoblastoma suppressor
	Evading apoptosis	Produce IGF survival factors
	Limitless replicative potential	Turn on telomerase
	Sustained angiogenesis	Produce VEGF inducer
	Tissue invasion & metastasis	Inactivate E-cadherin

B

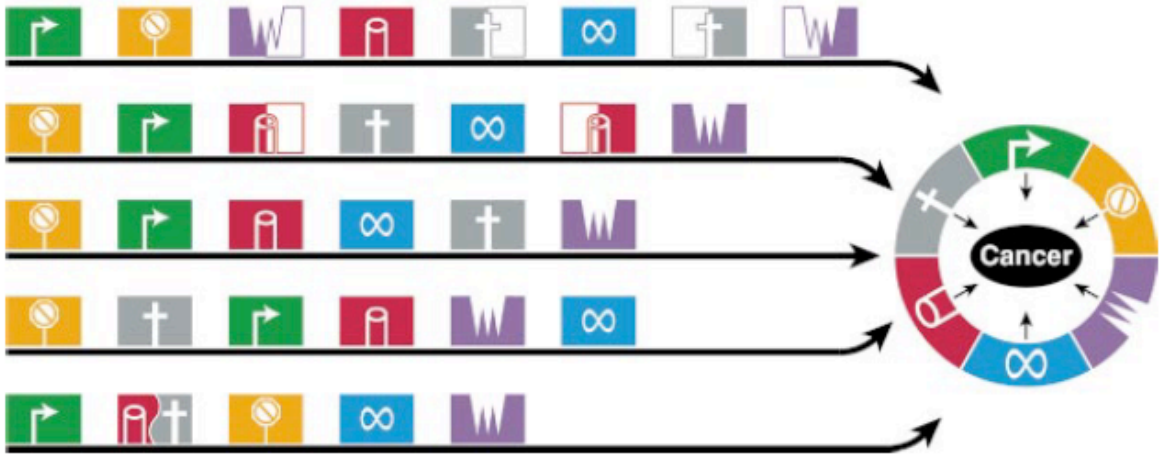
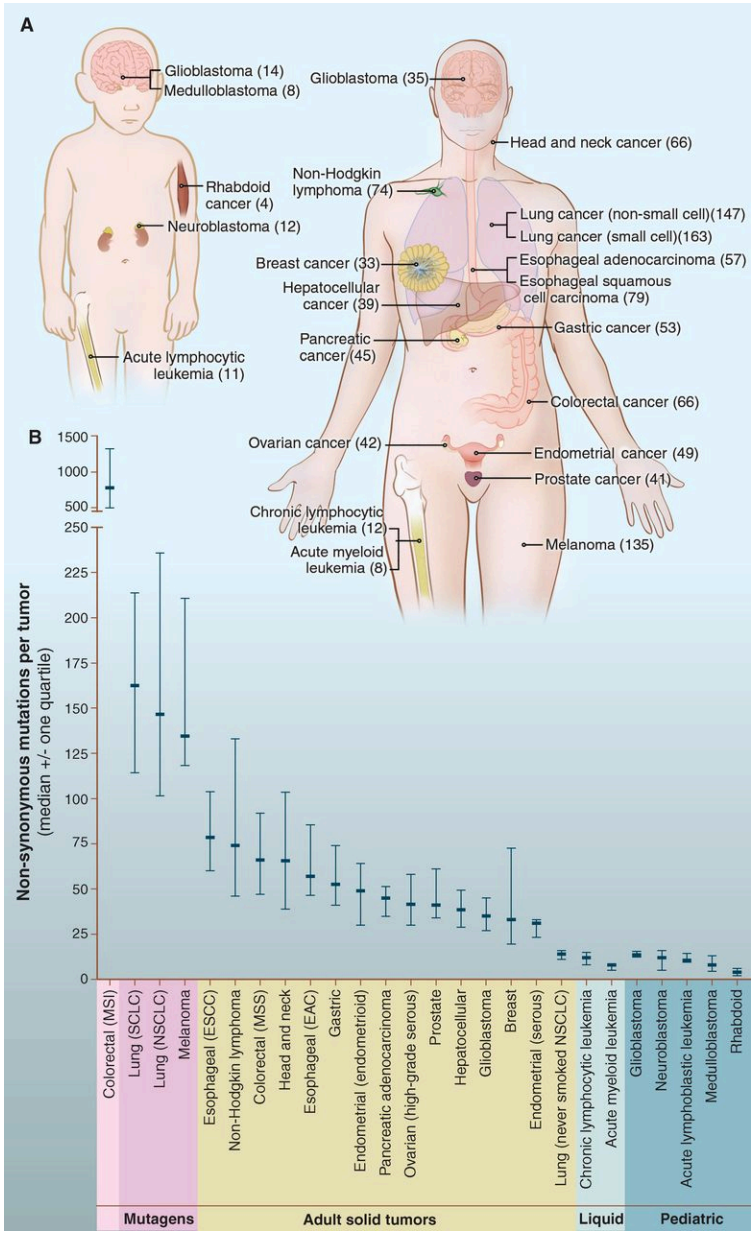


Figure 4. Parallel Pathways of Tumorigenesis

While we believe that virtually all cancers must acquire the same six hallmark capabilities (A), their means of doing so will vary significantly, both mechanistically (see text) and chronologically (B). Thus, the order in which these capabilities are acquired seems likely be quite variable across the spectrum of cancer types and subtypes. Moreover, in some tumors, a particular genetic lesion may confer several capabilities simultaneously, decreasing the number of distinct mutational steps required to complete tumorigenesis. Thus, loss of function of the p53 tumor suppressor can facilitate both angiogenesis and resistance to apoptosis (e.g., in the five-step pathway shown), as well as enabling the characteristic of genomic instability. In other tumors, a capability may only be acquired through the collaboration of two or more distinct genetic changes, thereby increasing the total number necessary for completion of tumor progression. Thus, in the eight-step pathway shown, invasion/metastasis and resistance to apoptosis are each acquired in two steps.

Number of somatic mutations in human cancers



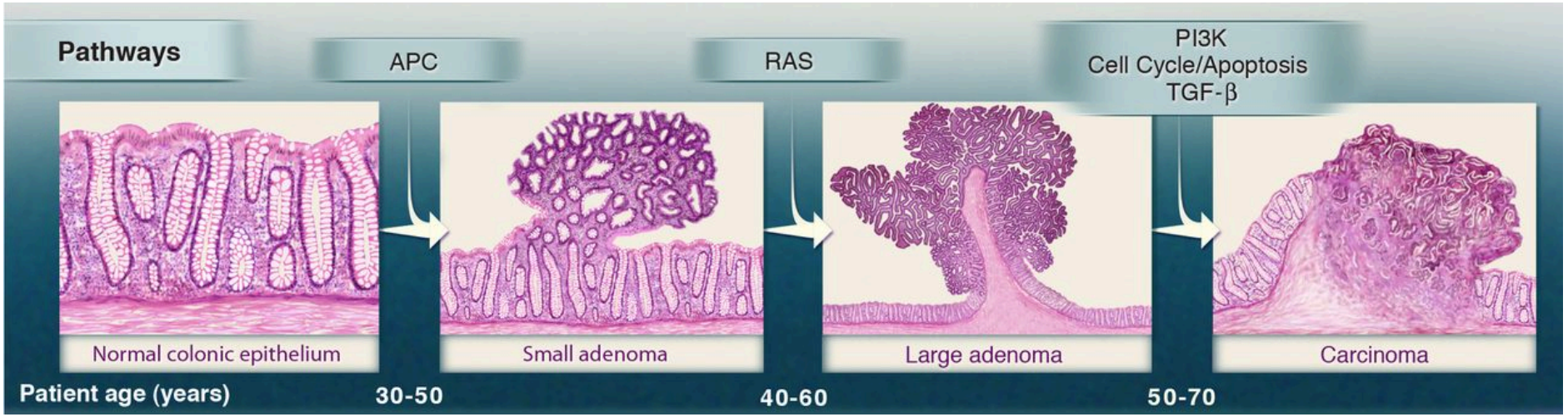
Top: children vs. adults

Numbers in parentheses : median number of nonsynonymous mutations per tumor.

MSI, microsatellite instability;
SCLC, small cell lung cancers;
NSCLC, non–small cell lung cancers;
ESCC, esophageal squamous cell carcinomas;
MSS, microsatellite stable;
EAC, esophageal adenocarcinomas.

B Vogelstein et al. Science 2013; 339:1546-1558

Progression of colorectal cancer



The major signaling pathways that drive tumorigenesis are shown at the transitions between each tumor stage.

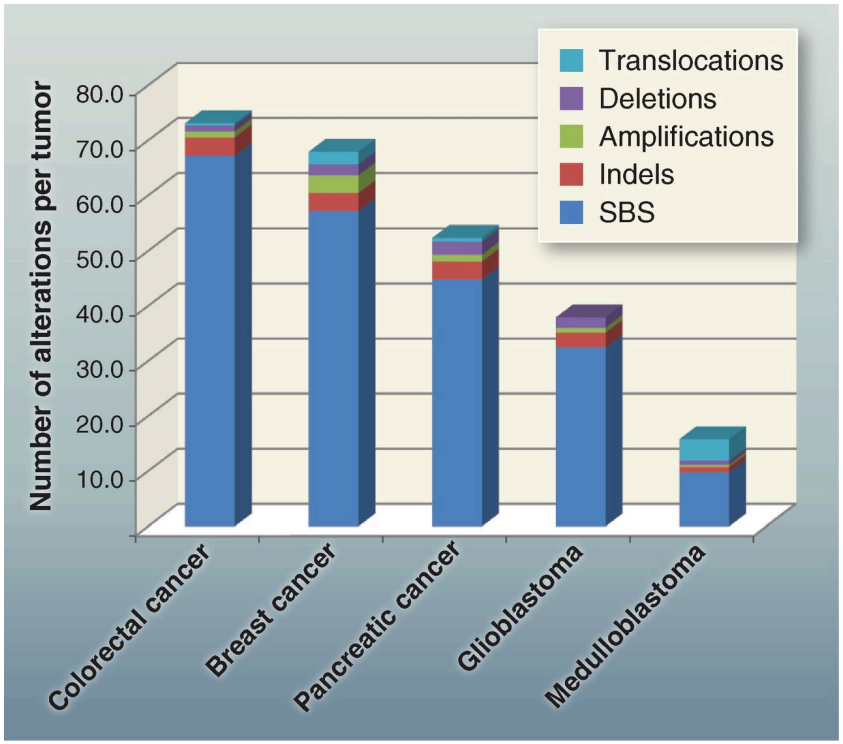
One of several driver genes that encode components of these pathways can be altered in any individual tumor.

Patient age indicates the time intervals during which the driver genes are usually mutated.

B Vogelstein et al. Science 2013; 339:1546-1558

TGF-β, transforming growth factor–β.

Alterations affecting protein-coding genes

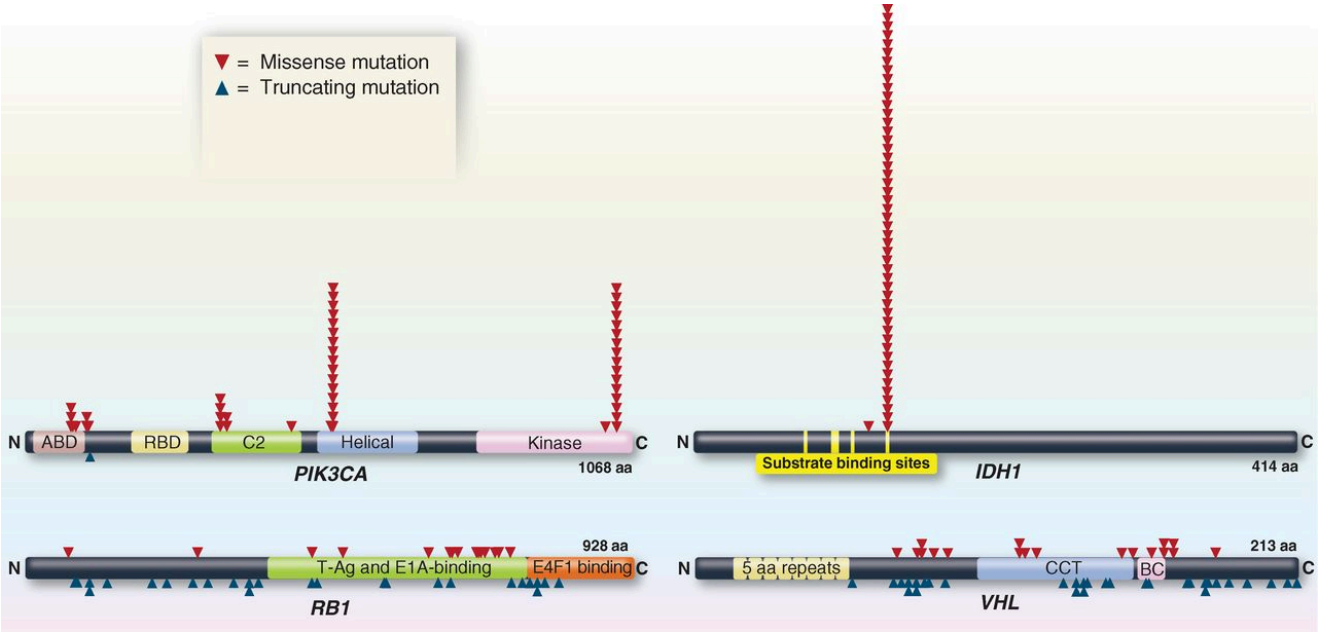


SBS: single-base substitutions (SBS),

Indels: small insertions and deletions,

B Vogelstein et al. Science 2013; 339:1546-1558

Mutations in oncogenes and tumor suppressor genes

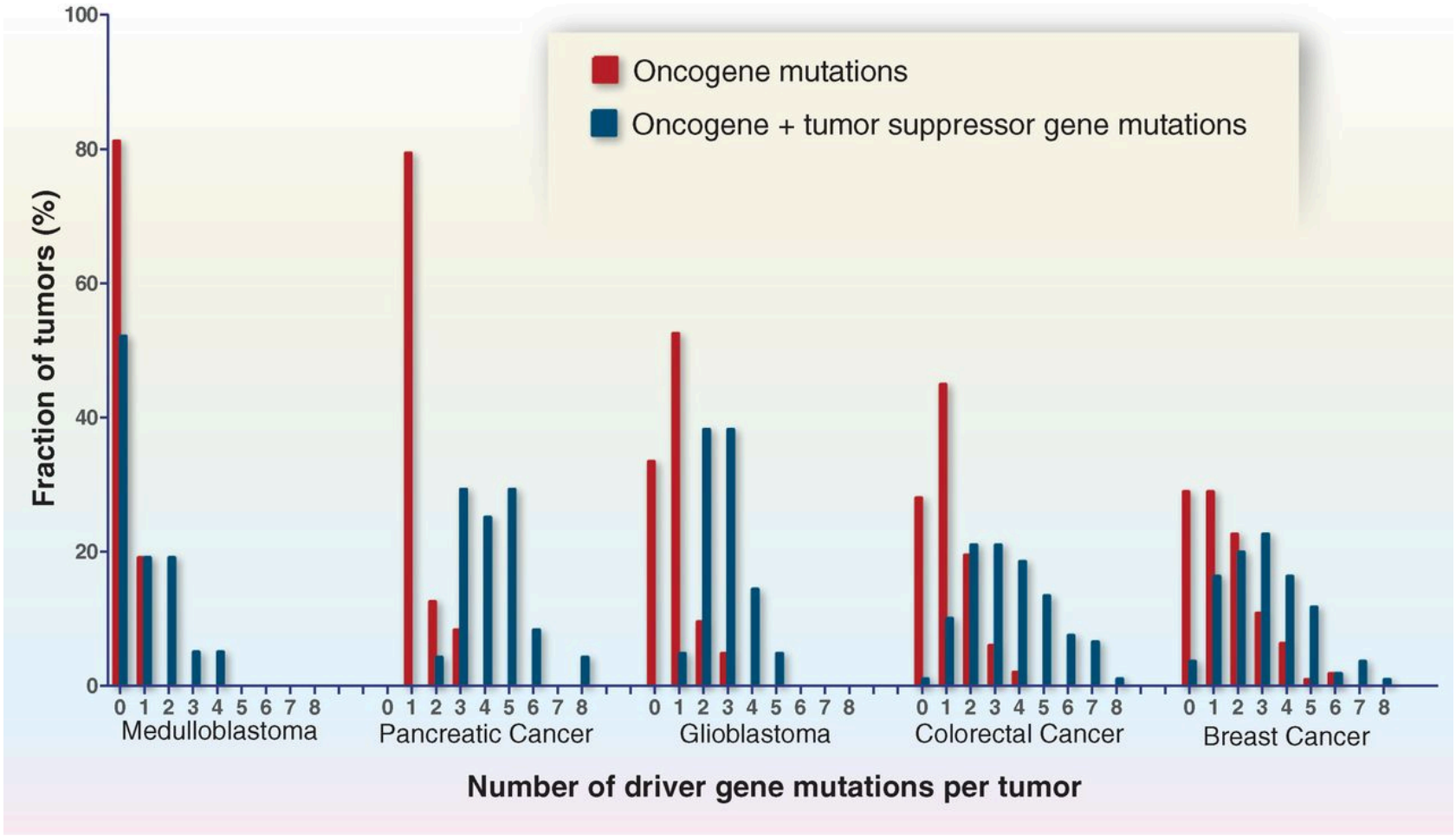


Oncogenes PIK3CA and IDH1: missense mutations accumulate at identical positions, (almost) no truncation mutations

tumor suppressor genes RB1 and VHL: truncating mutations and missense mutations spread over the entire genes

B Vogelstein et al. Science 2013; 339:1546-1558

Number of driver gene mutations per tumor



B Vogelstein et al. Science 2013; 339:1546-1558

Genetic heterogeneity in tumors

Example: primary tumor in the pancreas and its metastatic lesions in the liver.

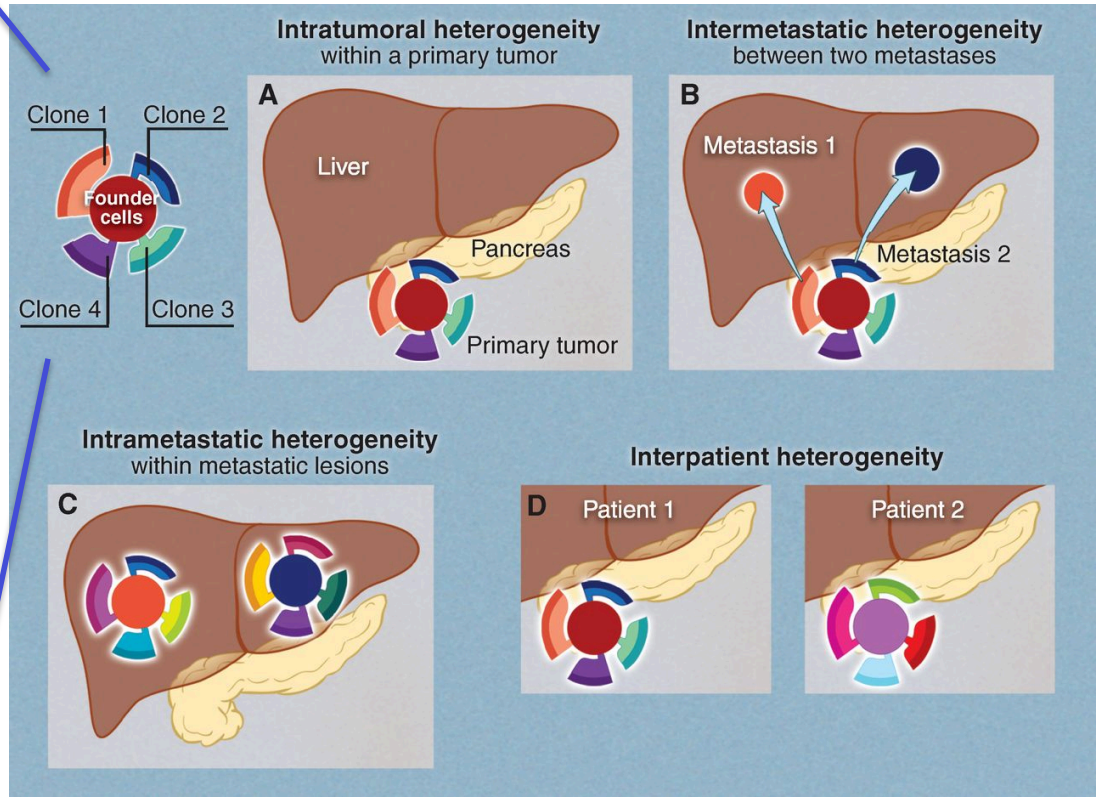
Mutations introduced during primary tumor cell growth result in clonal heterogeneity.

A typical tumor is represented by cells with a large fraction of the total mutations (founder cells) from which subclones are derived.

The differently colored regions in the subclones represent stages of evolution within a subclone.

heterogeneity among the cells of the primary tumor.

heterogeneity among different metastatic lesions in the same patient

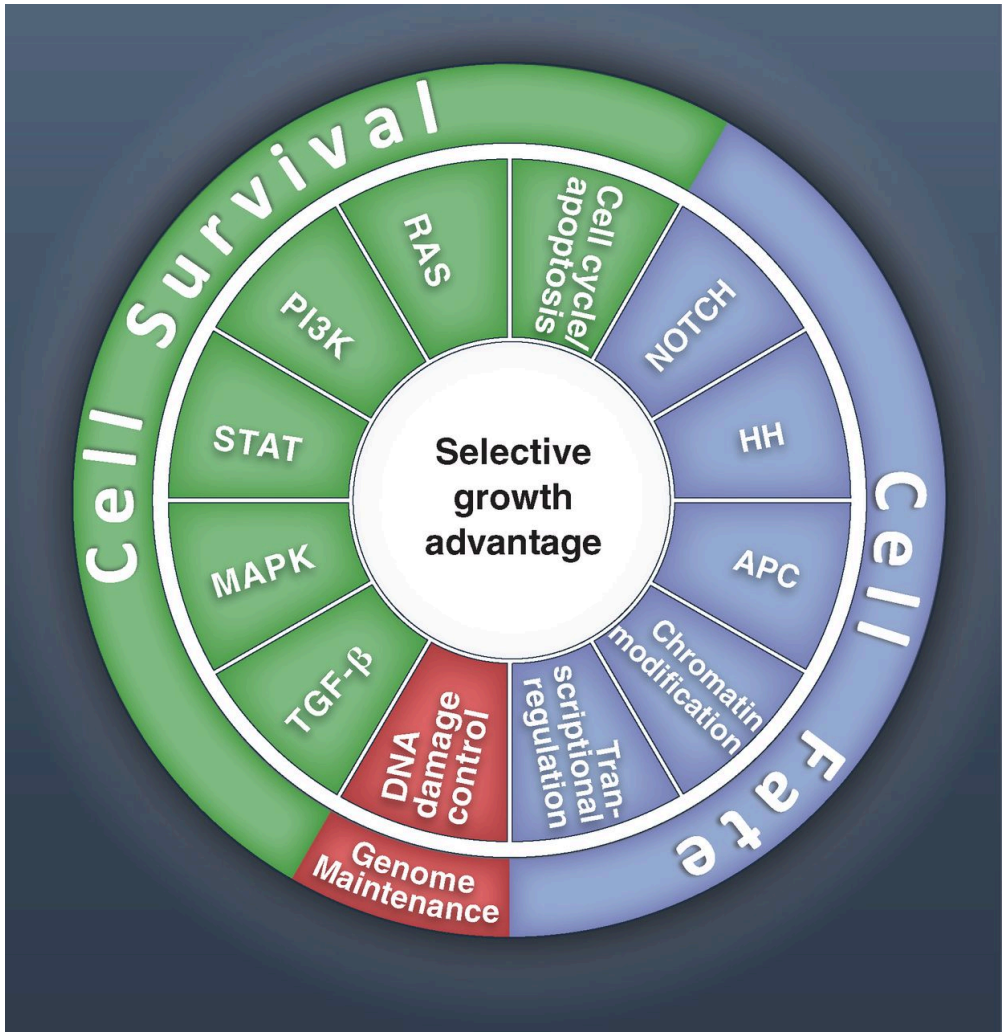


heterogeneity among the cells of each metastasis develops as the metastases grow

heterogeneity among the tumors of different patients. The mutations are almost completely distinct.

B Vogelstein et al. Science 2013; 339:1546-1558

Cancer driver genes belong to 12 pathways



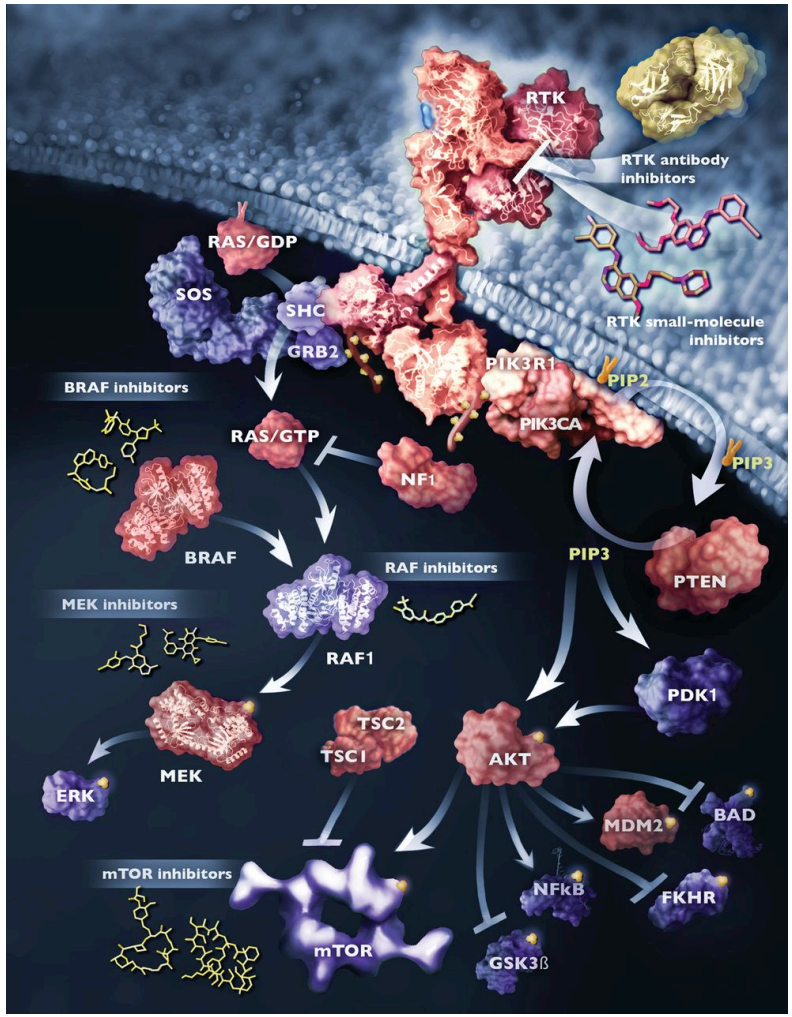
Cancer cell signaling pathways and the cellular processes they regulate.

All known driver genes can be classified into one or more of 12 pathways (middle ring) that confer a selective growth advantage (inner circle; see main text).

These pathways can themselves be further organized into three core cellular processes (outer ring).

B Vogelstein et al. Science 2013; 339:1546-1558

Signal transduction pathways affected by mutations in human cancer



Two representative pathways (RAS and PI3K) are illustrated.
The signal transducers are color coded: red indicates protein components encoded by the driver genes;
yellow balls : sites of phosphorylation.
Stick models: therapeutic agents that target some of the signal transducers.
RTK, receptor tyrosine kinase;
GDP, guanosine diphosphate;
MEK, MAPK kinase;
ERK, extracellular signal–regulated kinase;
NFkB, nuclear factor κB;
mTOR, mammalian target of rapamycin.

B Vogelstein et al. Science 2013; 339:1546-1558

Cellular differentiation

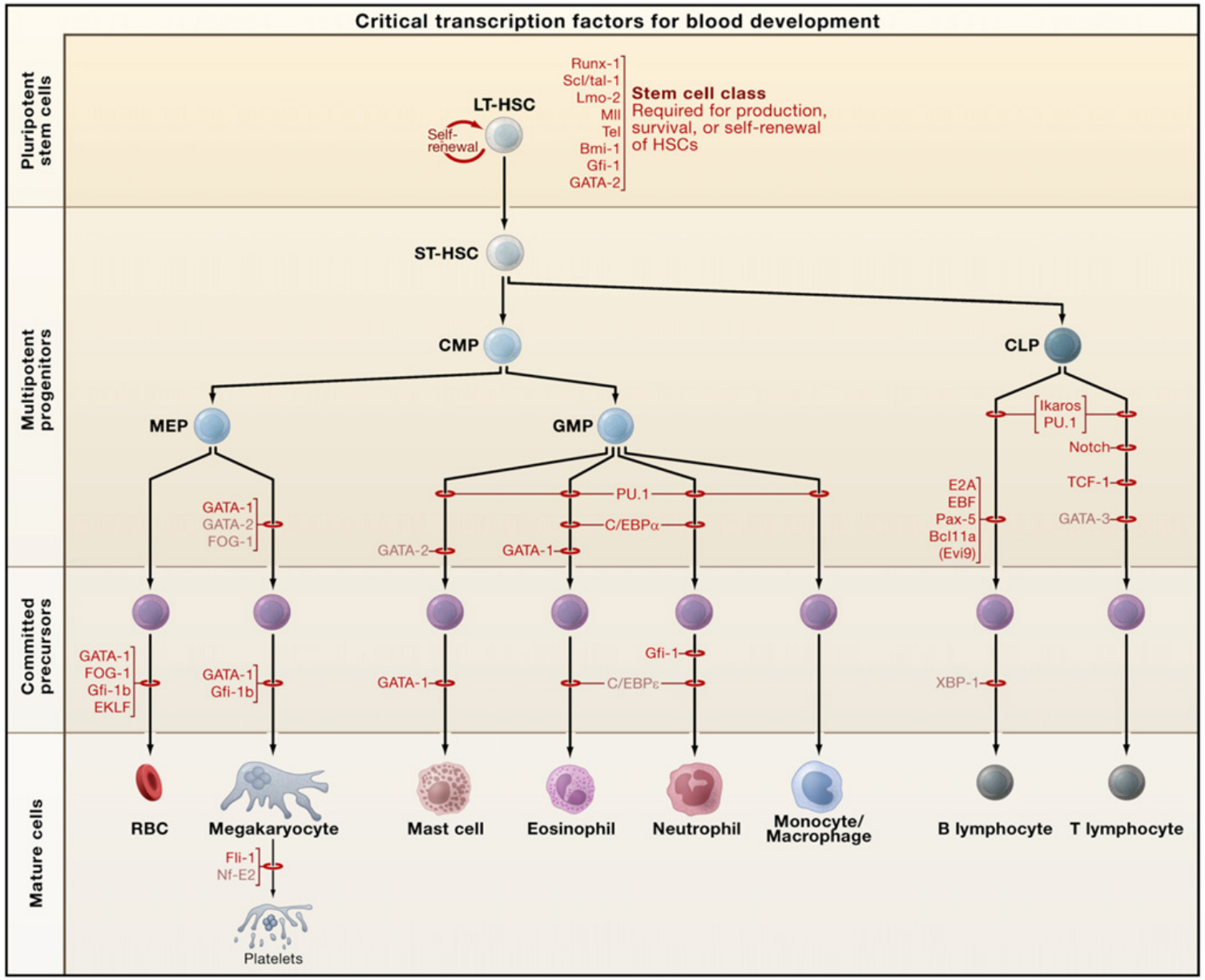
Differentiation is a key example of cell fate.

Differentiation does not depend on mutations.

So how does a cell know in which state it is?

-> This is controlled by epigenetic modifications of the genome

Hematopoiesis: development of blood cells



Orkin & Zon, Cell (2008) 132: 631–644.

What is epigenetics?

Epigenetics refers to **alternate phenotypic states** that are **not based** in **differences in genotype**, and are potentially reversible, but are generally stably maintained during cell division.

Examples: imprinting, twins, cancer vs. normal cells, differentiation, ...

Narrow interpretation of this concept : stable differential states of gene expression.

A much more expanded view of epigenetics has recently emerged in which multiple mechanisms interact to collectively establish
- alternate states of chromatin structure (open – packed/condensed),
- **histone modifications**,
- associated protein (e.g. histone) composition,
- transcriptional activity, and
- in mammals, **cytosine-5 DNA methylation** at CpG dinucleotides.

Laird, Hum Mol Gen 14, R65 (2005)

Basic principles of epigenetics: DNA methylation and histone modfications

The human genome contains 23 000 genes that must be expressed in specific cells at precise times.

Cells manage gene expression by wrapping DNA around clusters (octamers) of globular **histone** proteins to form **nucleosomes**.

These nucleosomes of DNA and histones are organized into **chromatin**, the building block of a chromosome.

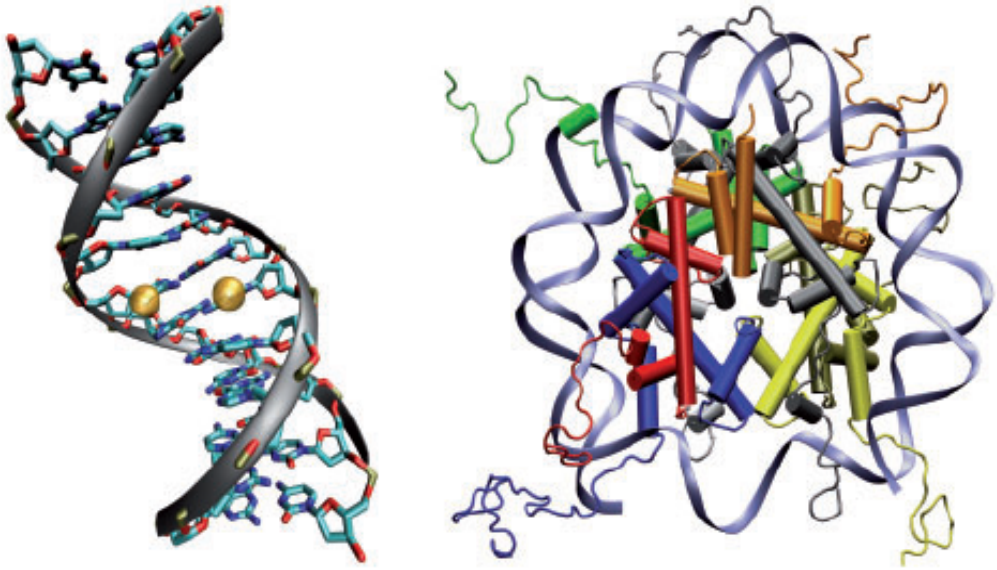


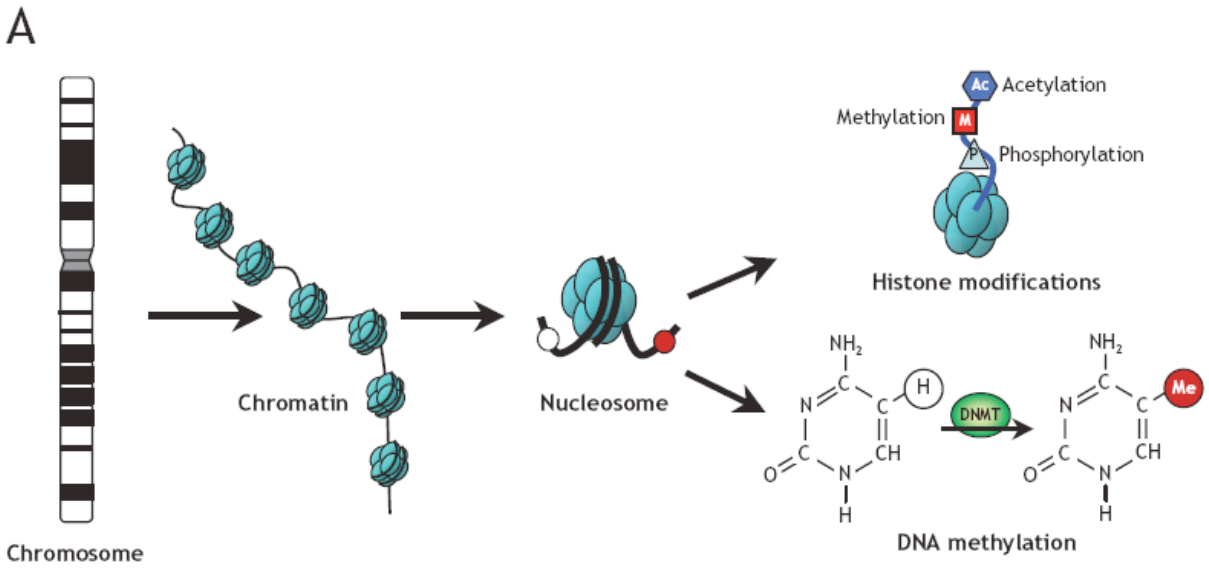
Fig. 1. Carriers of epigenetic information: DNA and nucleosome. The left panel shows a DNA double helix that is methylated symmetrically on both strands (orange spheres) at its center CpG (PDB structure: 329d). DNA methylation is the only epigenetic mechanism that directly targets the DNA. The right panel shows a nucleosome spindle consisting of eight histone proteins (center), around which two loops of DNA are wound (PDB structure: 1KX5). The nucleosome is subject to covalent modifications of its histones and to the binding of non-histone proteins.

Rodenhiser, Mann, CMAJ 174, 341 (2006)

Bock, Lengauer, Bioinformatics 24, 1 (2008)

Epigenetic modifications

Rodenhiser, Mann,
CMAJ 174, 341 (2006)



Reversible and site-specific **histone modifications** occur at multiple sites at the unstructured histone tails through **acetylation**, **methylation** and **phosphorylation**.

DNA methylation occurs at 5-position of cytosine residues within CpG pairs in a reaction catalyzed by DNA methyltransferases (DNMTs).

Together, these modifications provide a unique epigenetic signature that regulates chromatin organization and gene expression.

Cytosine methylation

Observation: 3-6 % of all cytosines are methylated in human DNA.

Mammalian genomes contain much fewer (only 20-25 %) of the CpG dinucleotide than is expected by the G+C content. This is typically explained in the following way:

As most CpGs serve as targets of DNA methyltransferases, they are usually methylated.

5-Methylcytosine, whose occurrence is almost completely restricted to CpG dinucleotides, can easily deaminate to thymine.

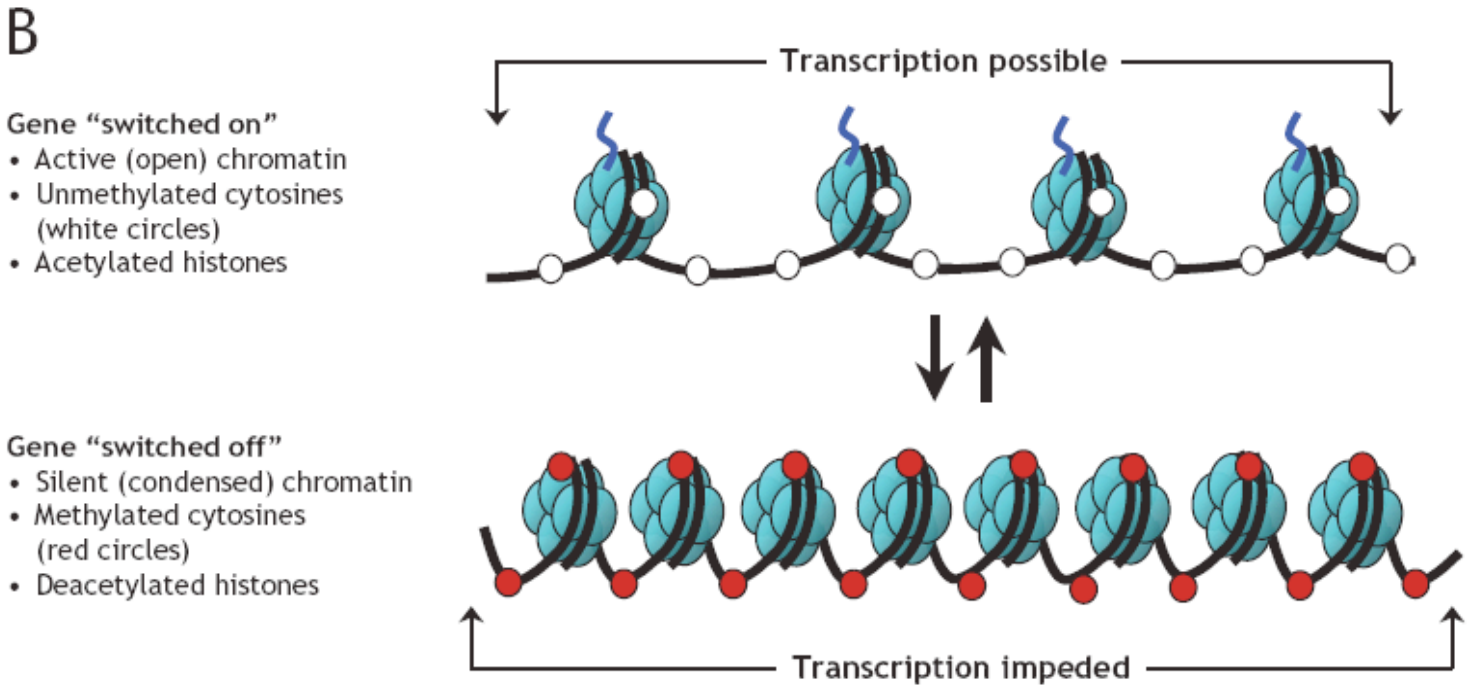
If this mutation is not repaired, the affected CpG is permanently converted to TpG (or CpA if the transition occurs on the reverse DNA strand).

Hence, methylCpGs represent mutational hot spots in the genome.
If such mutations occur in the germ line, they become heritable.

A constant loss of CpGs over thousands of generations can explain the scarcity of this special dinucleotide in the genomes of human and mouse.

Esteller, Nat. Rev. Gen. 8, 286 (2007)

effects in chromatin organization affect gene expression



Schematic of the reversible changes in chromatin organization that influence gene expression:

genes are expressed (switched on) when the chromatin is **open** (active), and they are inactivated (switched off) when the chromatin is **condensed** (silent).

White circles = unmethylated cytosines;

red circles = methylated cytosines.

Rodenhiser, Mann, CMAJ 174, 341 (2006)

Basic principles of epigenetics: DNA methylation and histone modfications

Changes to the structure of chromatin influence gene expression:
genes are inactivated (switched off) when the chromatin is condensed (silent),
and they are expressed (switched on) when chromatin is open (active).

These dynamic chromatin states are controlled by reversible epigenetic patterns of **DNA methylation** and **histone modifications**.

Interestingly, **repetitive** genomic sequences are **heavily methylated**, which means transcriptionally silenced.

Enzymes involved in this process include
- DNA methyltransferases (DNMTs),
- histone deacetylases (HDACs),
- histone acetylases,
- histone methyltransferases and the
- methyl-binding domain protein MECP2.

Rodenhiser, Mann, CMAJ 174, 341 (2006)

DNA methylation

Typically, unmethylated clusters of CpG pairs are located in **tissue-specific genes** and in essential **housekeeping genes**, which are involved in routine maintenance roles and are expressed in most tissues.

These clusters, or **CpG islands**, are targets for proteins that bind to unmethylated CpGs and initiate gene transcription.

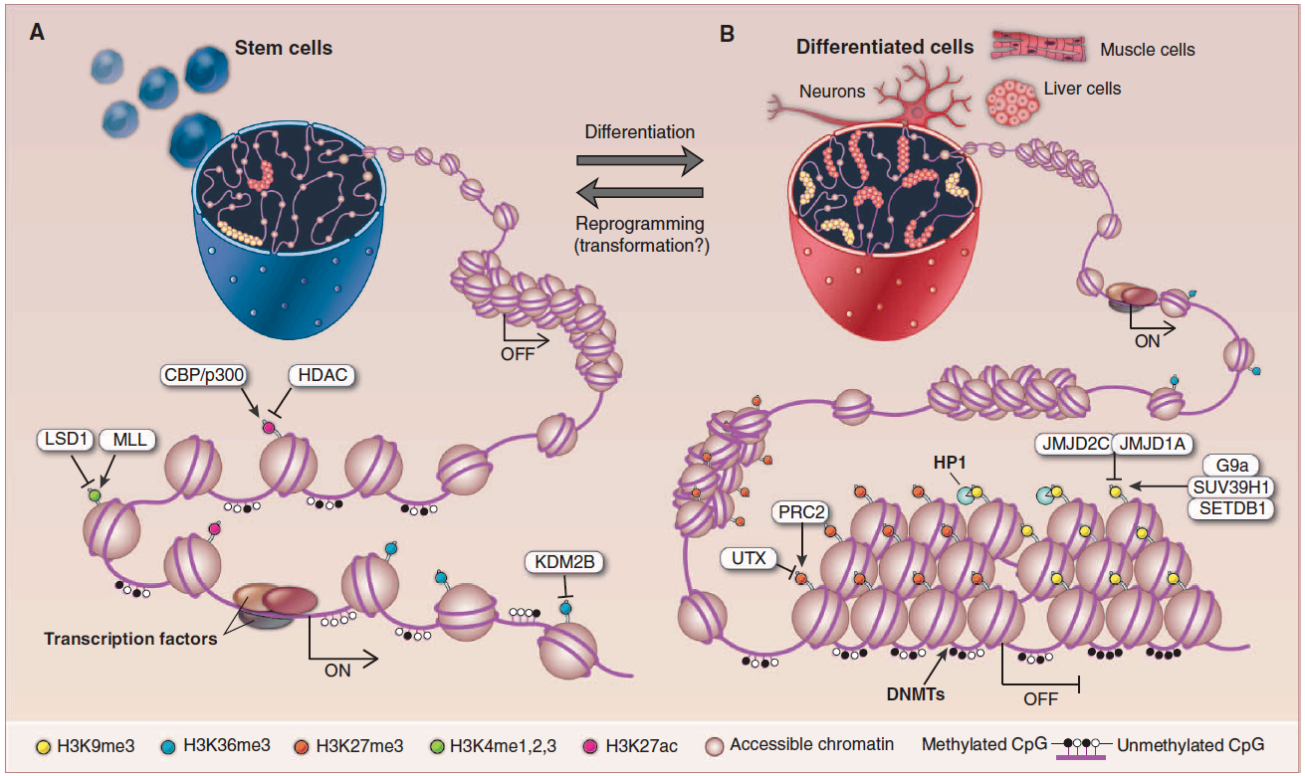
In contrast, methylated CpGs are generally associated with silent DNA, can block methylation-sensitive proteins and can be easily mutated.

The loss of normal DNA methylation patterns is the best understood epigenetic cause of disease.

In animal experiments, the removal of genes that encode DNMTs is lethal; in humans, overexpression of these enzymes has been linked to a variety of cancers.

Rodenhiser, Mann, CMAJ 174, 341 (2006)

Differentiation linked to alterations of chromatin structure



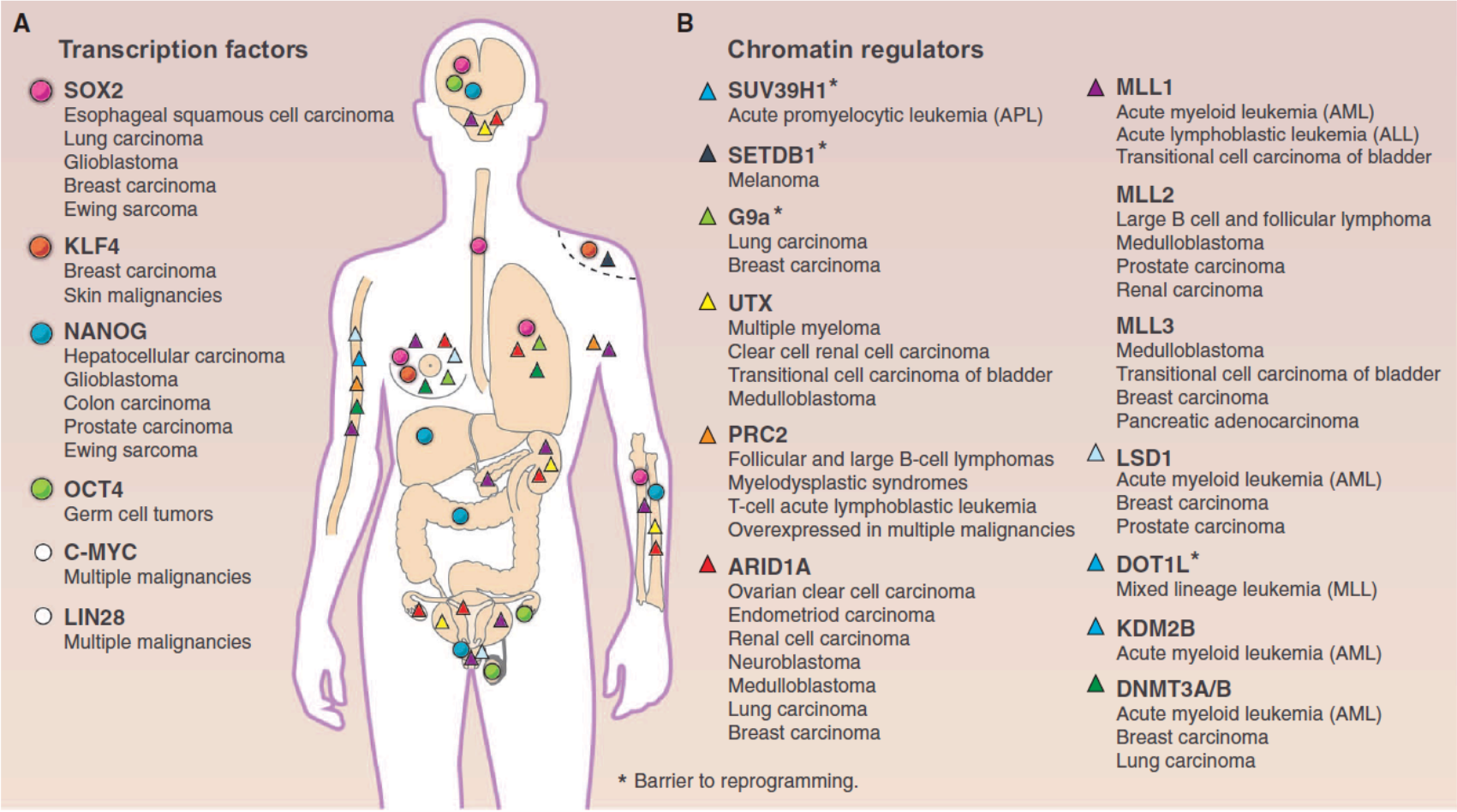
(A) In pluripotent cells, chromatin is hyperdynamic and globally accessible.

(B) Upon differentiation, inactive genomic regions may be sequestered by repressive chromatin enriched for characteristic histone modifications.

These global structures are regulated by DNA methylation, histone modifications, and numerous CRs whose expression levels are dynamically regulated through development.

ML Suva et al. Science 2013; 339:1567-1570

Genes involved in iPS nuclear programming and cancer



Genes include bona fide oncogenes and tumor suppressors that are directly affected by genetic alterations, as well as other genes with mechanistic roles in cancer.

ML Suva et al. Science 2013; 339:1567-1570

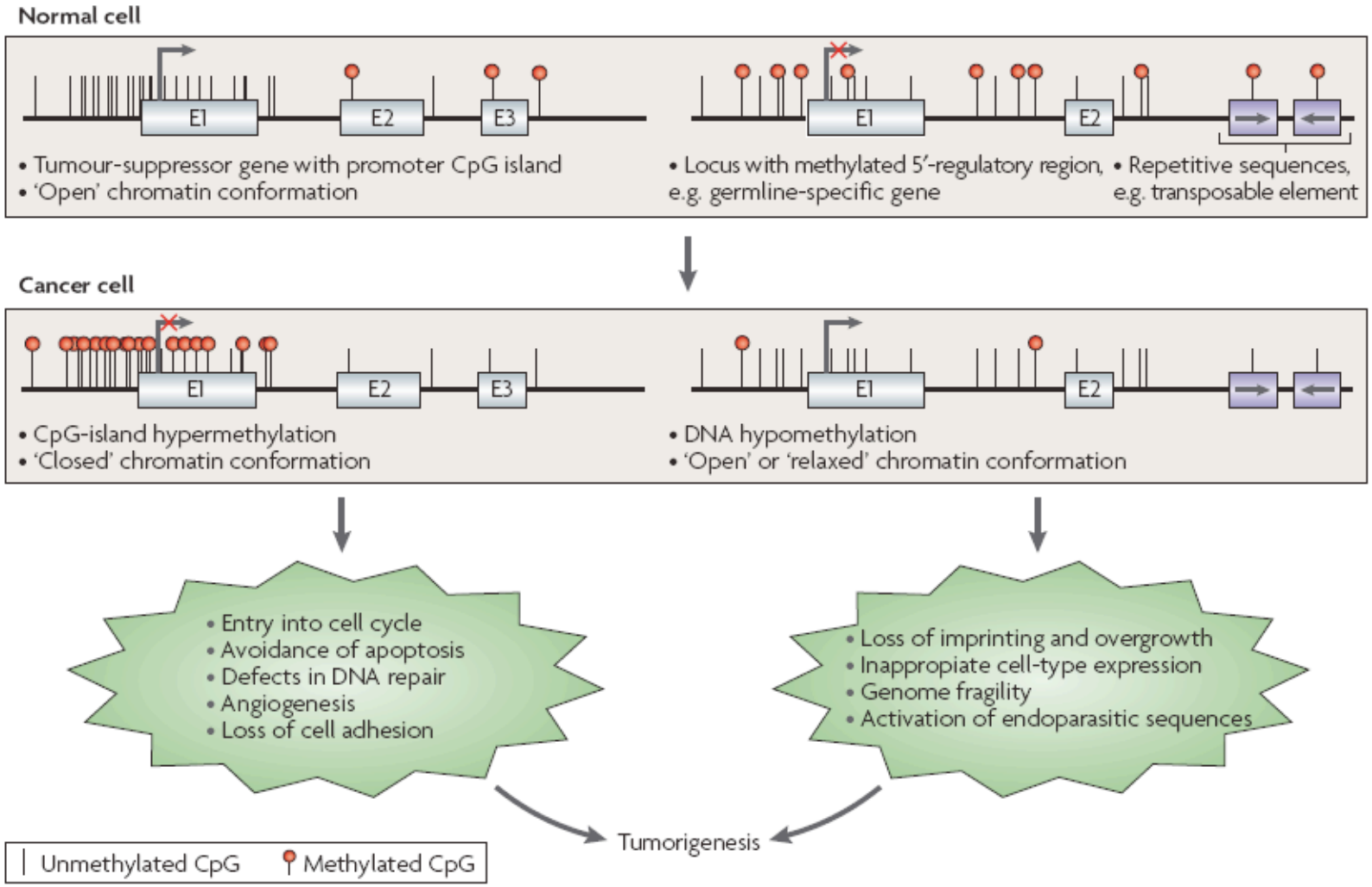


Figure 1 | **Altered DNA-methylation patterns in tumorigenesis.** The hypermethylation of CpG islands of tumour-suppressor genes is a common alteration in cancer cells, and leads to the transcriptional inactivation of these genes and the loss of their normal cellular functions. This contributes to many of the hallmarks of cancer cells. At the same time, the genome of the cancer cell undergoes global hypomethylation at repetitive sequences, and tissue-specific and imprinted genes can also show loss of DNA methylation. In some cases, this hypomethylation is known to contribute to cancer cell phenotypes, causing changes such as loss of imprinting, and might also contribute to the genomic instability that characterizes tumours. E, exon.

Esteller, Nat. Rev. Gen. 8, 286 (2007)

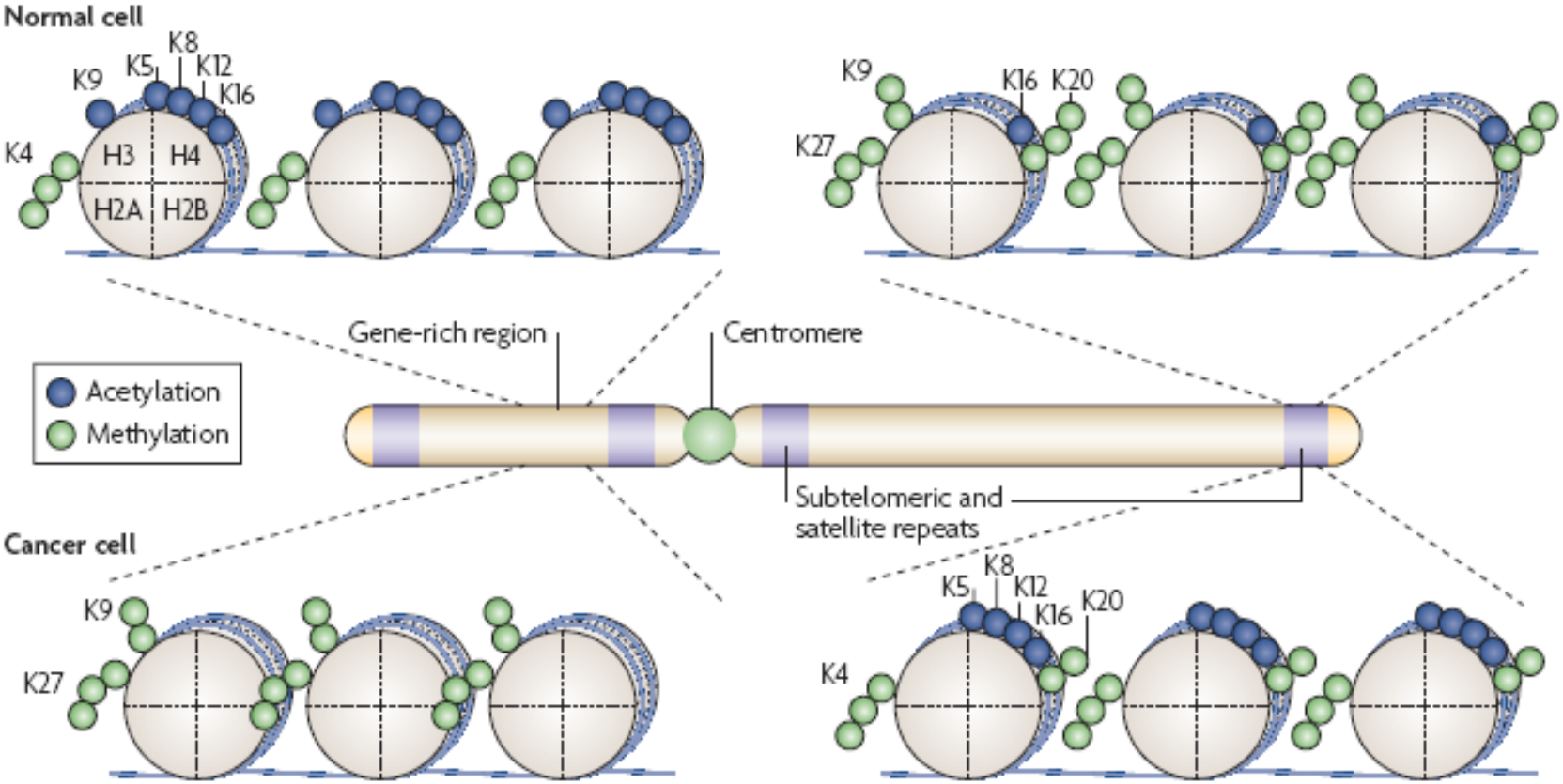


Figure 4 | **Histone-modification maps for a typical chromosome in normal and cancer cells.** Nucleosomal arrays are shown in the context of chromosomal location and transcriptional activity. Octamers consisting of histones H2A, H2B, H3 and H4 are represented as grey cylinders. Histone acetylation and methylation (di- and tri-) are shown. In 'normal' cells, genomic regions that include the promoters of tumour-suppressor genes are enriched in histone-modification marks associated with active transcription, such as acetylation of H3 and H4 lysine residues (for instance K5, K8, K9, K12 and K16) and trimethylation of K4 of H3. In the same cells, DNA repeats and other heterochromatic regions are characterized by trimethylation of K27 and dimethylation of K9 of H3, and trimethylation of K20 of H4, which function as repressive marks. In transformed cells, this scenario is disrupted by the loss of the 'active' histone-marks on tumour-suppressor gene promoters, and by the loss of repressive marks such as the trimethylation of K20 of H4 or trimethylation of K27 of histone H3 at subtelomeric DNA and other DNA repeats. This leads to a more 'relaxed' chromatin conformation in these regions.

Esteller, Nat. Rev. Gen. 8, 286 (2007)

Summary

Cells need to tightly control their exact position in the cell cycle and in development.

Control during cell cycle: checkpoints + Cdk / cyclin system

Control during development: different chromatin states / epigenetics

---

Cancerogenesis is determined by random apperance of driver mutations plus sofar poorly understood epigenetic changes.

Cellular differentiation and cancerogenesis involve similar players of the epigenetic machinery.

Next week: computational multi-scale model of an entire cell
JB Karr et al, Cell, 150, 389-401 (2012)