Regulating Genes

We have been discussing the structure of DNA and its role in protein synthesis. We have seen that DNA stores the information about how to assemble a protein, and that RNA molecules are used to transcribe and translate that information to direct the synthesis of specific proteins.

We have also discussed briefly how mutations alter DNA sequences and can affect gene expression.

We also know that each cell of an organism has exactly the same DNA, yet we have many different types of cells and tissues within an organism. Not all of our DNA information is used in each cell, and not all information is used all of the time. The process of cell differentiation, in which cells become specialized for their specific function involves selectively activating some genes and repressing others. Many genes in multicellular organisms are activated only at one stage of development, do their job, and function no more. The effects of these genes are not reversible.

How does a cell " "know" what DNA is needed and when? What controls gene activity? Trying to find answers to these questions is part of the subject of **gene regulation**.

Some of the answers to how genes are regulated are coming from work on recombinant DNA research; some from looking at genetics and, in particular, mutant strains of species. Much is coming from our increasing knowledge of cancers and the failure of the body to control cell division in cancer formation. Another current research interest is **stem cells** – the cell lines that lead to the development of precise tissue types, such as skin, immune system or blood cells. At some point in development, stem cells are "programmed", do their job, and, as a part of their programming, may even lead to programmed cell death. Gene regulation is a very active area of research in developmental biology, the biology of aging, genetic diseases research, and cancer research.

Gene control is exerted chemically by molecules that interact with DNA, RNA and/or the polypeptide chains. Both hormone signal molecules and regulatory proteins have effects on gene expression. Genes can also be chemically modified to either enhance or inhibit their readability. Gene controls can be positive – inducing gene activity, or negative – repressing gene activity. Genes can be regulated at any step of gene expression.

Regulating Gene Expression in Prokaryotic Organisms

The early work on gene regulation was done with prokaryotes. It is easier to study activity in prokaryotes because they are less genetically complex, and absent a nucleus, the DNA is accessible to all components of the cell. Much of the research in gene regulation has been accomplished with *Escherichia coli*, the common intestinal bacterium.

Before we go too far, however, we need our vocabulary. Recall that the typical gene codes for a polypeptide that is used to help the cell function in some way, or codes for some structural protein. A gene that codes for such proteins is a **structural gene**.

Other genes control how much of a polypeptide gets formed and when it gets formed. Such genes are **regulatory genes.**

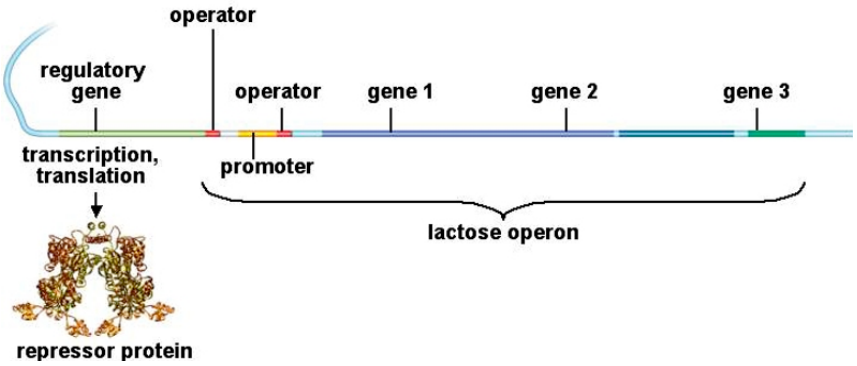
- Some regulatory genes code for small polypeptides that control how other genes get expressed. These polypeptides are called **transcription factors.** There are a number of different types of transcription factors.
- Another type of regulatory gene is a piece of DNA that a transcription factor binds to. These **regulatory sites** of DNA do not actually code for any protein.

The Operon of the Prokaryotic Cell

An active gene (or group of genes) includes the DNA that will be transcribed, the structural gene, along with a promoter and operator. This complex is known as the **operon** and was described in 1961 by Francois Jacob and Jacques Monod. An operon has three parts: **promoter**, **operator** and **structural gene** plus a **regulatory gene** that activates or represses the operon.

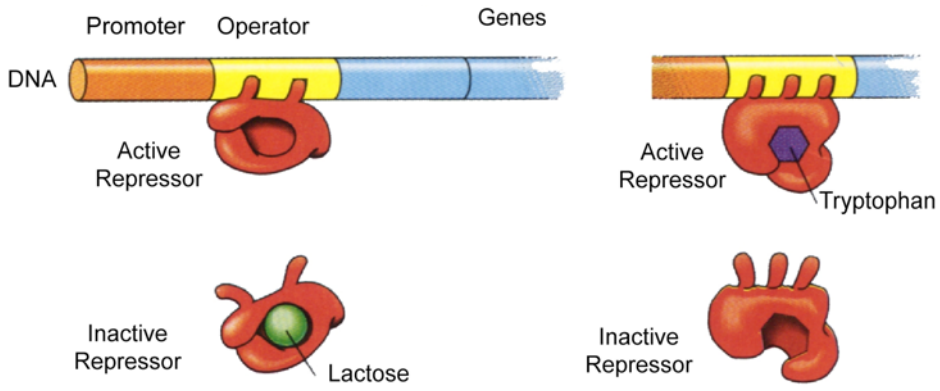
Operon

- **Promoter**
  Recognized by RNA polymerase as the place to start transcription
- **Operator**
  Controls RNA polymerase's access to the promoter, and is usually located within the promoter or between the promoter and the transcribable gene (or genes)
- **Structural (Transcribable) Gene**
  Codes for the needed protein



Regulatory Gene

- A regulatory gene codes for a repressor protein. The regulatory gene is located apart from the operon.
- Repressors typically work with **controller** molecules. A repressor can be active when attached to its controller molecule or deactivated when attached to a controller molecule. A controller molecule is typically a substance in the cell.
  - In an **inducible operon**, the repressor actively blocks the gene from transcription. The controller molecule attaches to the repressor removing it from the operator and transcription proceeds.
  - In a **repressible operon**, the repressor is not blocking gene transcription unless the controller molecule binds to the repressor. When that happens, the repressor, with its controller attached, actively blocks transcription.

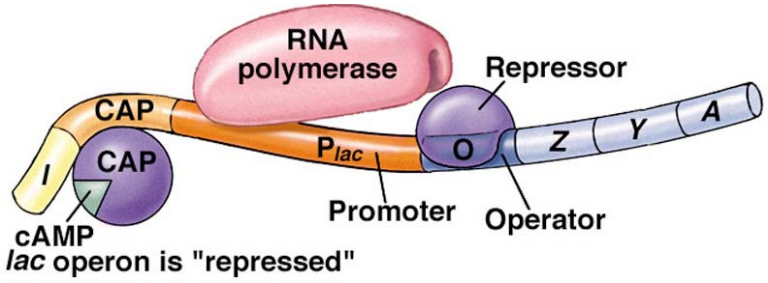


Inducible Operon Repressible Operon

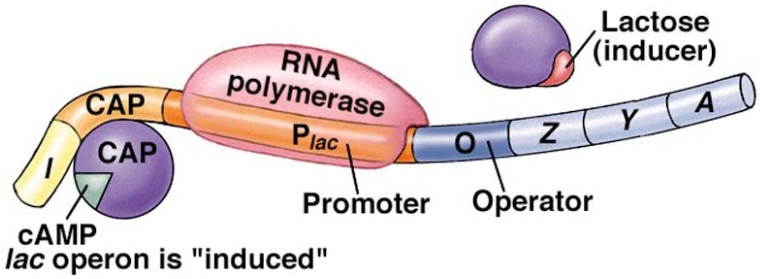
The Lactose Operon – An Inducible Operon

To get an idea of how genes get regulated, we will look at the Lactose operon, described by Jacob and Monod in *E. coli*. The lactose operon contains the three genes that code for the enzymes that degrade lactose.

In the absence of lactose, the controller substance, the repressor inhibits transcription by blocking RNA polymerase from attaching to the promoter.

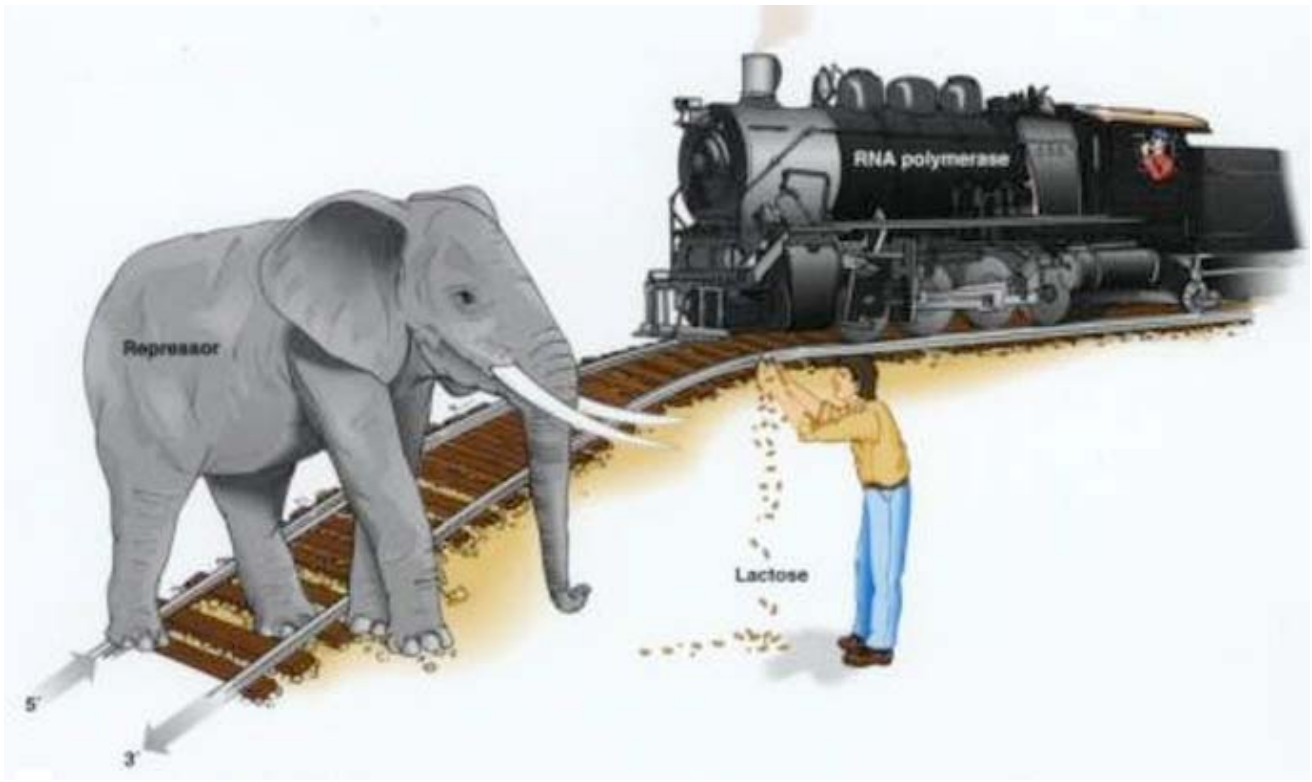


When the substrate, allolactose (an isomer of lactose), attaches to the repressor protein that sits on the operator region of the gene and removes the repressor, the promoter is available and the genes that code for the three enzymes to digest lactose are transcribed.



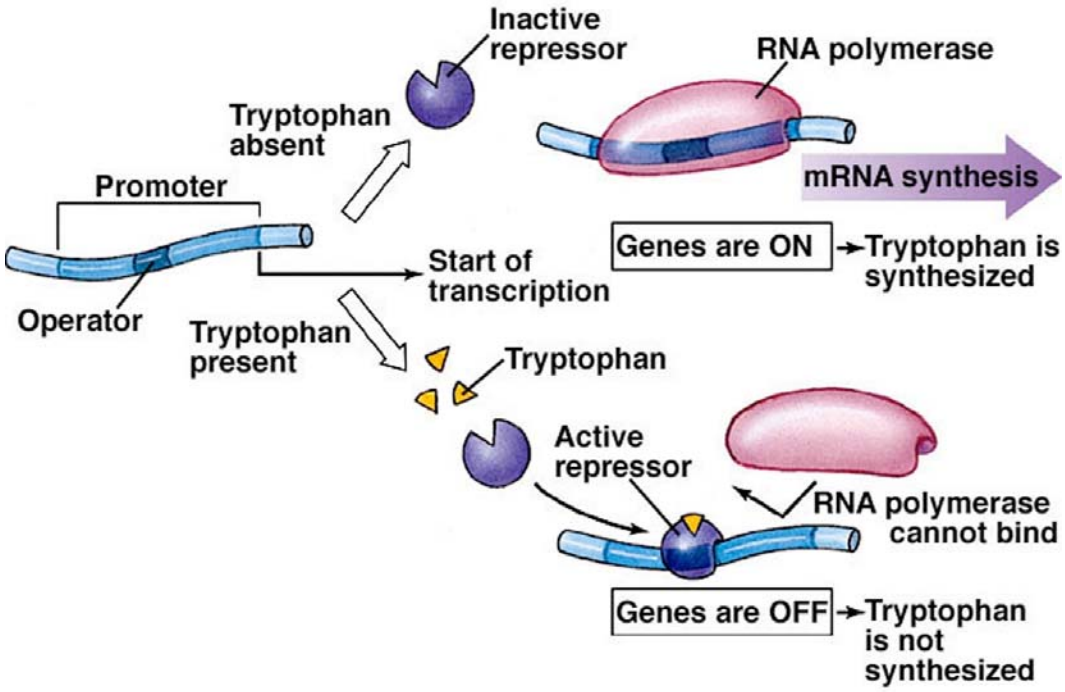
When the enzymes are synthesized, the lactose is degraded, including the allolactose molecules that are attached to the repressor. When lactose is no longer available to bind to the repressor protein, the repressor shuts down the promoter (by sitting back on the operator), which stops transcription. This is a negative control mechanism, because the promoter is blocked from activating the operator by the repressor.

One view of the Lactose Operon



The Tryptophan Operon – A Repressible Operon

Lactose in a bacterium's environment binds to the lactose gene repressor removing its block of transcription. Products in the environment can work in reverse too, by **inhibiting** rather than promoting transcription. If a bacterium has sufficient tryptophan in its environment, some of the tryptophan will bind to the tryptophan synthesis gene repressor molecule, but in contrast to lactose that removes the repressor, the tryptophan synthesis repressor needs tryptophan to fit into position and block transcription.

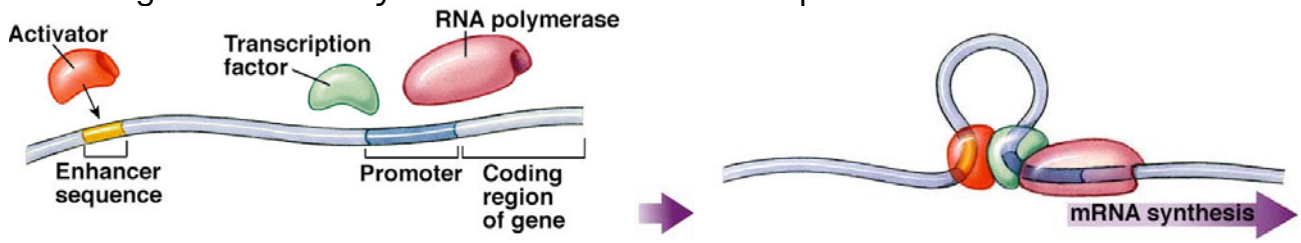


Eukaryotic Gene Transcription Complex

Like the prokaryotic gene, the eukaryotic gene has a number of regions, each important to transcription. The components of the eukaryotic transcription complex include:

1. **Control elements** that consist of:
   - A specific **promoter** region within the control elements, which indicates the starting point for transcription.

   - A region called the **enhancer** that stimulates the binding of RNA polymerase to the promoter region. The enhancer region is comprised of non-coding DNA that binds to transcription factors called **activators**. Activators fold the DNA so that the enhancers are brought to the promoter region of the gene where they bind to additional transcription factors



   - **Silencers** are control elements that can inhibit transcription. A transcription factor that binds to a silencer control element and blocks transcription is called a **repressor**.

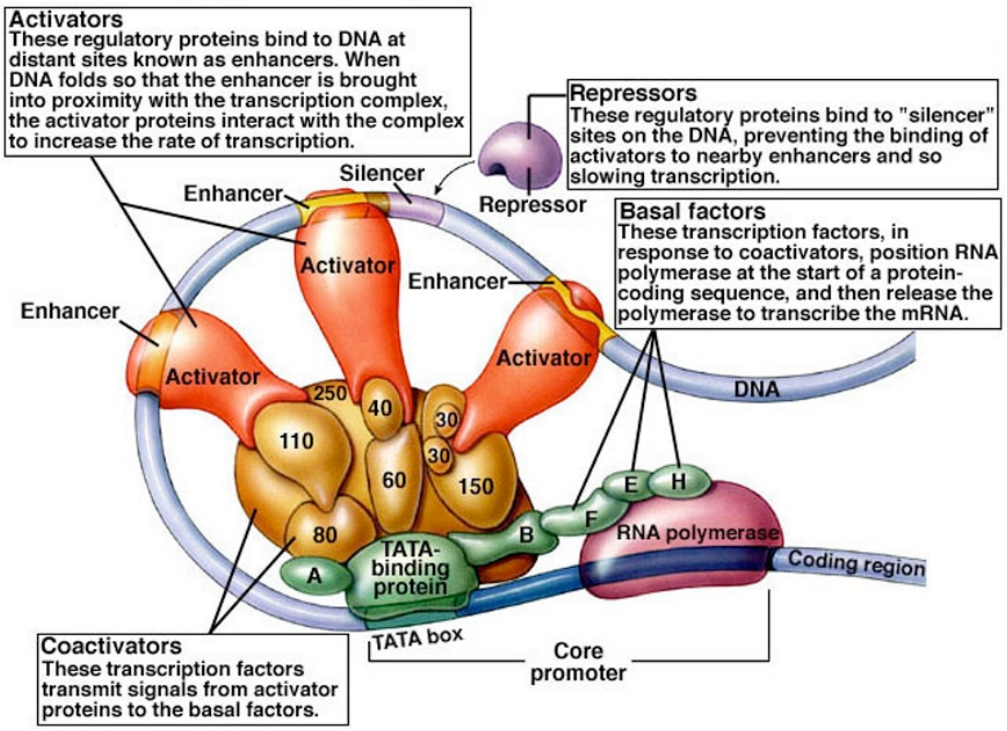
2. The **codable gene** including introns and exons
3. **Termination Signals** which end transcription

Transcription Initiation Complex

Transcription factors bind to the enhancer (or silencer region of the gene if we are going to repress transcription instead) where activator proteins have attached. Hundreds of transcription factors have been identified in eukaryotes, and most likely are the direct control of transcription.

The binding of activators to the enhancer results in **bending the DNA** molecule so that the activator proteins are brought more closely to the promoter region. This serves to attract more transcription factors to form **a transcription initiation complex** into which **RNA polymerase** can fit at the promoter region of the gene. When everything comes together, we get transcription.

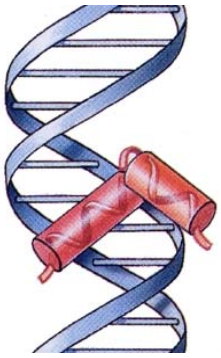
If a repressor has attached to the silencer region near the enhancers, activators are prevented from binding to the enhancers, and transcription is repressed.



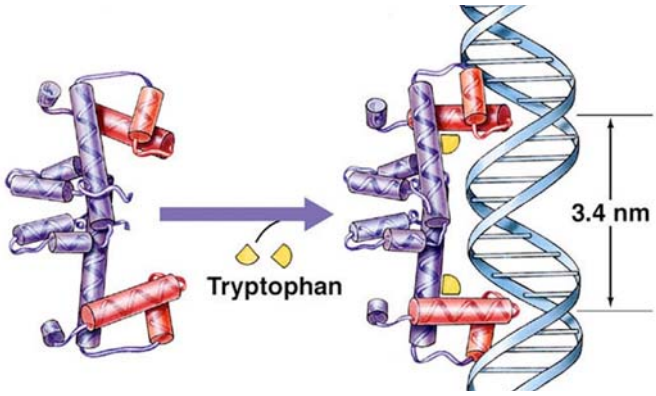
Transcription Factor Binding Sites

Transcription factors are characterized by specific binding sites that fit into the DNA molecule at the appropriate location. As expected, the binding sites have specific structural elements, called motifs or domains, each of which is cleverly named.

One of the more common binding motifs is the helix-turn-helix motif, a protein helix with a bend in it. Regulatory proteins generally have pairs of helix-turn-helix motifs for more strength. Often, repressor molecules attach to the binding motif, changing its shape so it can attach to the DNA molecule repressing its access to RNA polymerase.



Helix-Turn-Helix on DNA

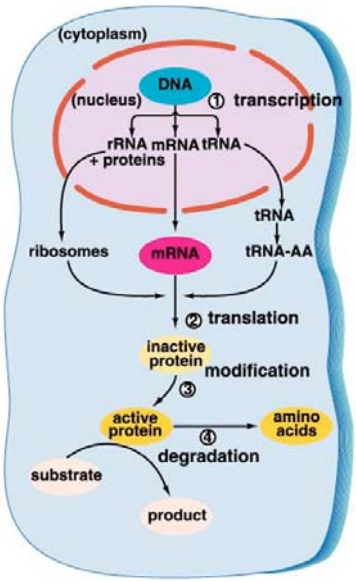


Tryptophan binding motif of paired helix-turn-helix

Regulating Gene Expression in Eukaryotic Organisms

Gene expression starts with transcription and "ends" with an enzyme catalyzing a particular chemical reaction, or with a structural/metabolic protein. The expression of a gene can be controlled at any level of gene activity.

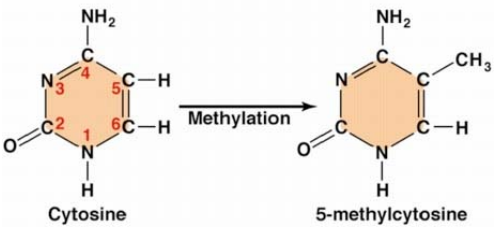
- Making the DNA readable
- Transcription
  - Activating transcription
  - Rate of transcription
- Processing the mRNA
  - Selective intron removal
- Translation
  - Stability of mRNA can be blocked in the cytoplasm
  - mRNA access to translation
- Post-translation Protein Modification
- Enzyme Activity and Feedback Inhibition

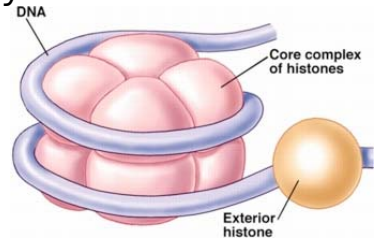


DNA Controls – Pre-transcription

Chemical Modification of DNA

Inactive DNA contains nucleotides (especially cytosine) that have **methyl groups** (-CH_3) attached. (The Barr body is an example of a chromosome that is highly methylated.) Most methylated DNA will remain inactive during differentiation and cell divisions. Methylation keeps some genes permanently turned off.





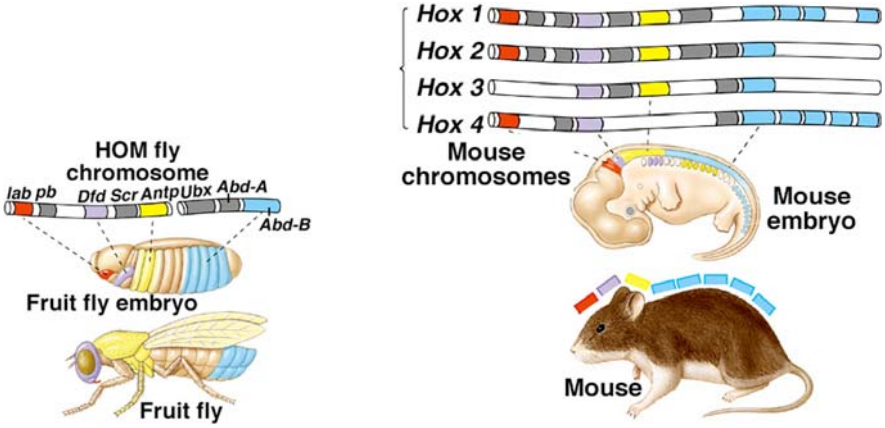
Adding an **acetyl group** (-$COCH_3$) to the histone proteins associated with DNA helps transcription. Histones with acetyl groups bind more loosely to DNA .

Transcription Controls

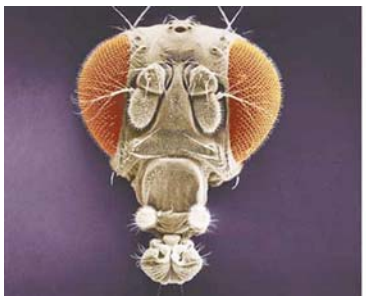
Pattern Formation and Homeotic Genes

During development different cells become organized in predictable patterns, some of which are determined by organization of the egg cell cytoplasm, some by cell position within the embryo. Plant cell differentiation is almost exclusively positional. Their position in the meristems fixes their ultimate position and function. Animals have pattern formation, too, and often once pattern is induced, it is fixed. Cells transplanted from one position to another in the embryo may develop into the structure to which they were patterned prior to the transplant. Chemical signals are used extensively during pattern formation. The signals trigger cell responses that affect gene transcription of what are called "master genes", genes that control developmental sequences.

Homeotic genes are a group of master genes that determine body-part identity and pattern development. There is a remarkable similarity of homeotic genes among different animals. The HOM genes of the fruit fly and the set of HOX genes in the mouse determine similar pattern development.



Mutations in the homeotic genes produce organisms with misplaced body parts.

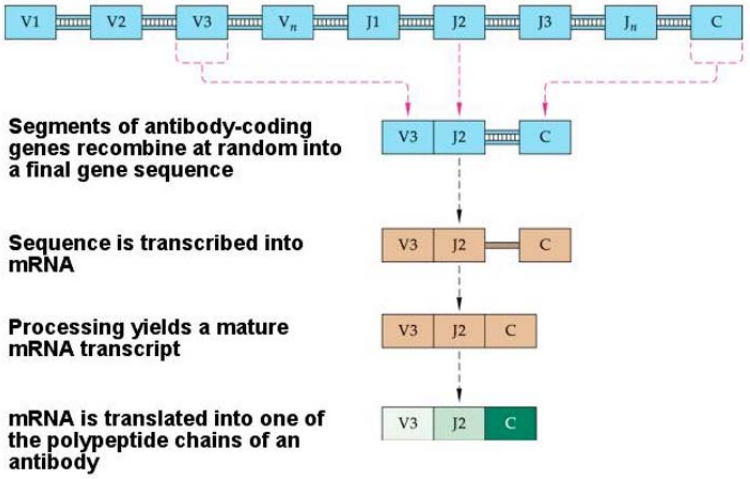

Normal Fly


Mutant Antennapedia Gene

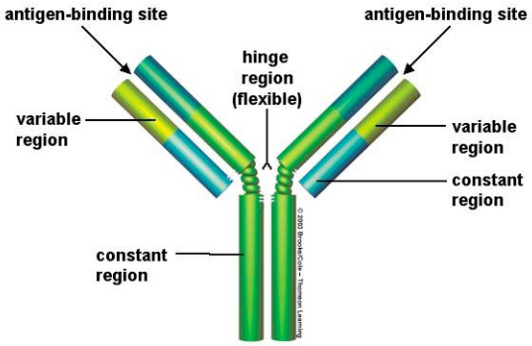

Eye Induction Mutant
Targeted Gene Expression

Gene Rearrangement

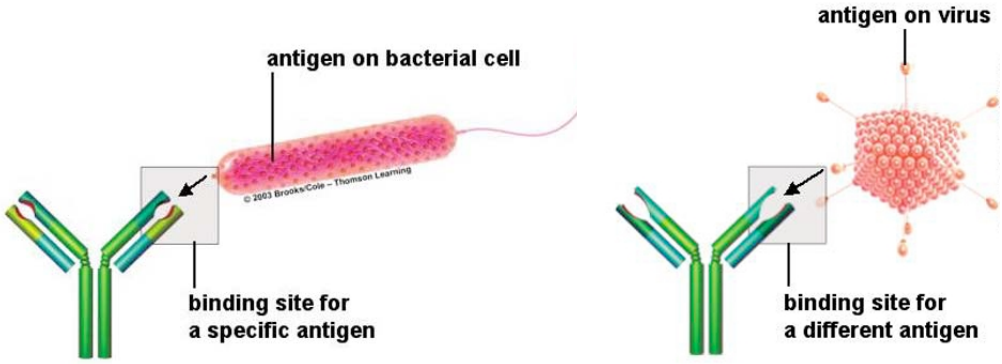
Some genes have a number of different possible arrangements for their nucleotide sequence. The genes that code for antibody formation have multiple possibilities. For each antibody, a core sequence is coded, but additional code, making each antibody unique is pretty much randomly determined.



Coding for Antibodies



Common Antibody Structure



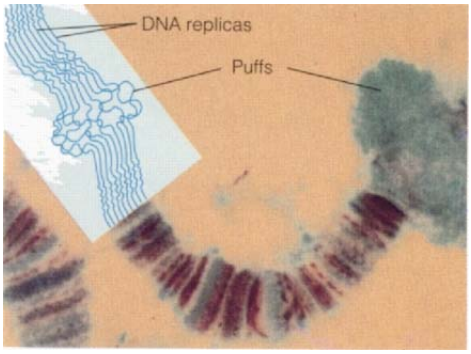
Two Specific Antibodies

Chromosome Inactivation

The formation of the Barr body, which inactivates of the two X chromosomes in females, prevents transcription of the deactivated X chromosome. As discussed with genetics and gene expression, the selection of which X gets condensed for a given cell line appears to be random, with a resultant mosaic pattern X chromosome gene expression in the individual. Barr body formation appears to be needed to regulate and maintain equivalent gene expression in males and females. A single gene is responsible for the deactivation of the X-chromosome, and is, itself the target of methylation to block further transcription of its RNA material.

Gene Amplification

It is possible to get multiple copies of genes via **gene amplification**, a process in which a gene, or portion of chromosome, or even entire chromosomes get multiplied many, many times. Salivary glands of many flies have gene amplification of chromosomes forming **polytene** chromosomes, which associated puffs.



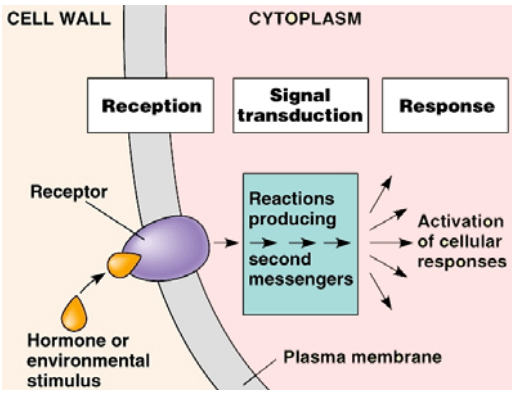
Fruit Fly Polytene Chromosome

Gene amplification occurs in ovum cells of vertebrates when the genes for ribosomal RNA (rRNA) get replicated millions of times to ensure that the cytoplasm will have the many, many ribosomes needed for protein synthesis activities in early development.

Cancer cells can also have gene amplification, where genes with resistance to the chemotherapy drugs are replicated thousands of times. Increasing the drug concentration results in increasing the resistance of the cancer cell population to the drug by selecting for those cells that have amplified genes for drug resistance.

Chemical Activators

Hormones (or other chemicals) can function as signal molecules that trigger **signal transduction pathways** in cells. Signal transduction pathways often result in the synthesis of **transcription factors**. Signal transduction pathways are equally important for chemical messaging in plants and animals. Which transcription factors get synthesized to activate genes depends on the signal molecule, and the transduction pathways triggered.

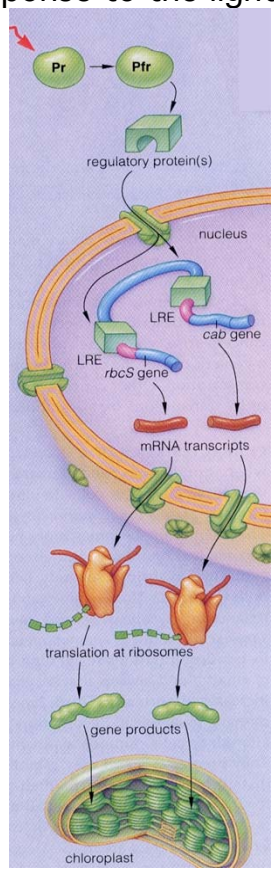


- The hormone, **ecdysone**, found in flies, stimulates the transcription of the genes for the production of **saliva** on the **polytene chromosomes.** Large amounts of saliva are needed to moisten the copious amounts of food the larvae consume.
- **Albumin synthesis** in bird eggs is promoted by an estrogen-protein complex that binds near the enhancer region of the albumin gene. Estrogen is only produced during the breeding season, so no albumin is synthesized when it is not needed.
- The activation of the **molting gene** in insect larvae is controlled by a hormone that activates a regulatory protein on the gene.
- The **androgen receptor** in human males is essential for testosterone to function.

Environmental Activators

Plants in particular have responses to environmental signal activators. Many plant growth activities are activated by light, or the absence of light. **Phytochrome** is a light-sensitive pigment that undergoes a conformational change to trigger signal transduction pathways that affect transcription factors. Phytochrome is involved in flowering of plants, chlorophyll synthesis and etiolation (rapid elongation of cells in the absence of light. Phytochrome is also responsible for signaling anthocyanin production in leaves prior to abscission, which gives the red pigments to leaves in autumn.

Chlorophyll synthesis in response to the light activation of phytochrome



mRNA Processing Gene Regulation

When mRNA is processed, we can get different functional mRNA transcripts depending on which part of the primary transcript is determined to be introns and which exons. These differences can be accomplished via **alternative splicing** by the spliceosomes, thereby controlling gene activity.

Post Transcription Controls of Gene Activity

Duration of the mRNA Transcript

The length of time a mRNA transcript can be read prior to degradation will affect the amount of final product, which will affect cell activity. Some mRNA lasts for weeks; some for hours.

Translation Control

Blocking mRNA attachment to Ribosomes

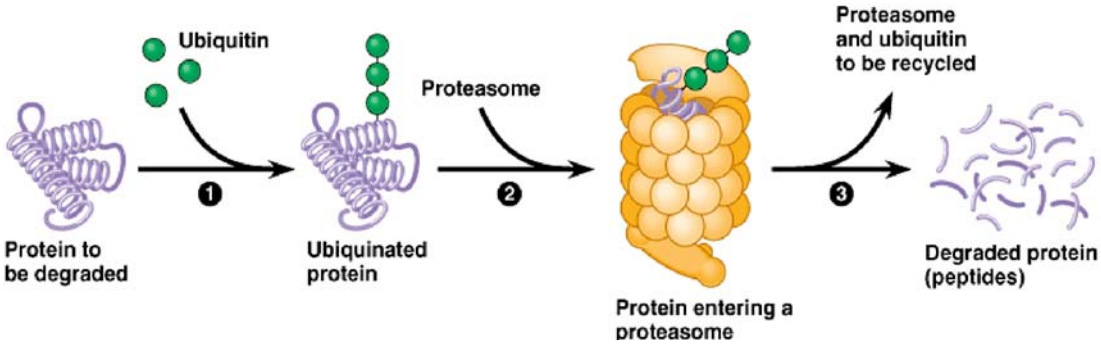
Translation can be controlled by regulatory proteins that block mRNA attachment to the small subunit of the ribosome. In humans, translation of the protein ferritin, an iron carrier, is blocked unless iron is present in the cell.

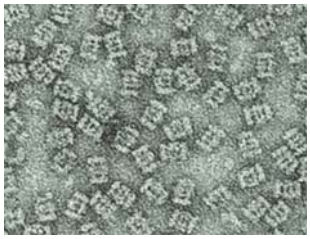
Post Translation Processing

After a polypeptide is synthesized it can be altered by processing. Many metabolic proteins are non-functional until activated by other molecules. Hydrolytic enzymes, in particular, are synthesized in inactive forms; membrane recognition proteins have additional molecules attached to them before they function.

Protein Stability

Proteins targeted for degradation are tagged by a tiny protein that is recognized by huge protease enzymes called **proteasomes.** The tagged protein is readily degraded within the proteasome. In cystic fibrosis, the chloride ion channel protein gets tagged and degraded before reaching the plasma membrane.

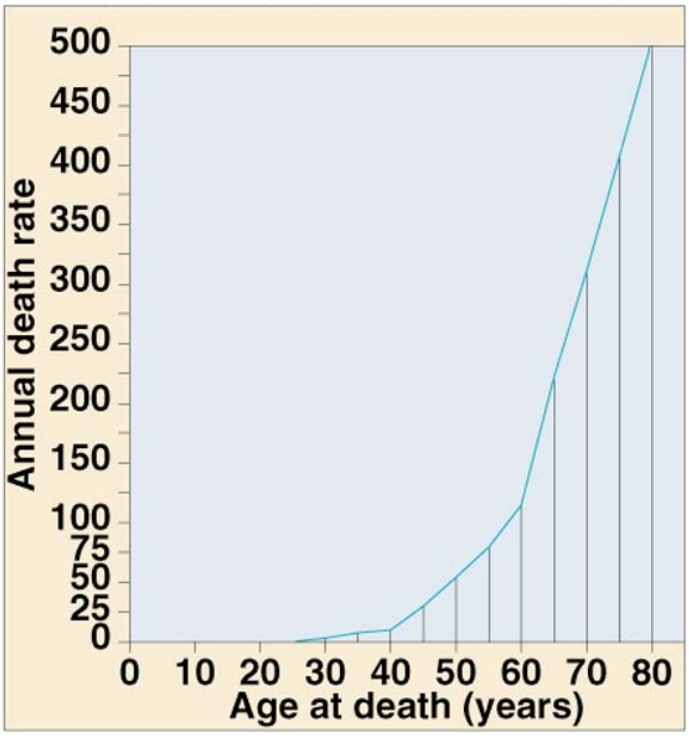


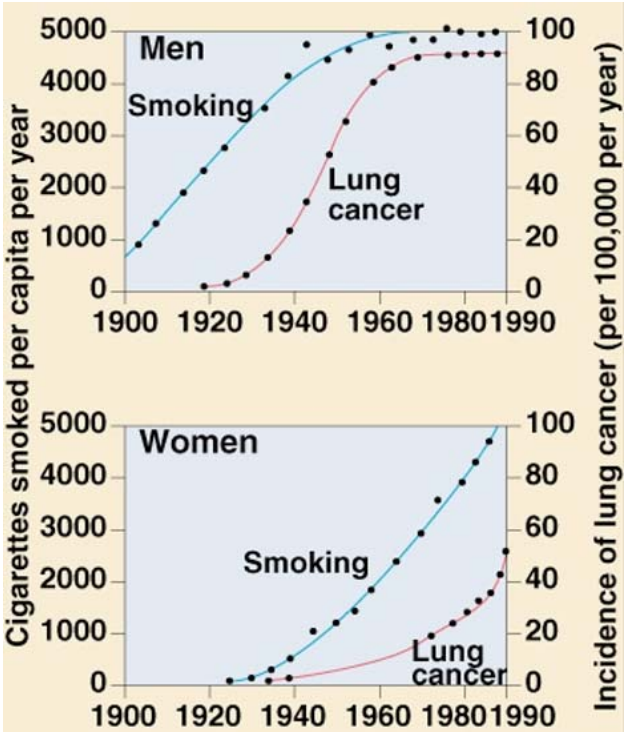


Proteasomes

Cancer and Gene Regulation

Cancer is a disease of uncontrolled and invasive cell reproduction. The current estimates are that 1/3 of the children born now will get some form of cancer in their lifetime. Lung cancer is still a major killer, and the cause of most lung cancers is straightforward: smoking. The three most common cancers are breast cancer (an assortment of cancers), prostate cancer and colon cancer. One in eight will get breast cancer. Almost any male who lives long enough will have prostate cancer.



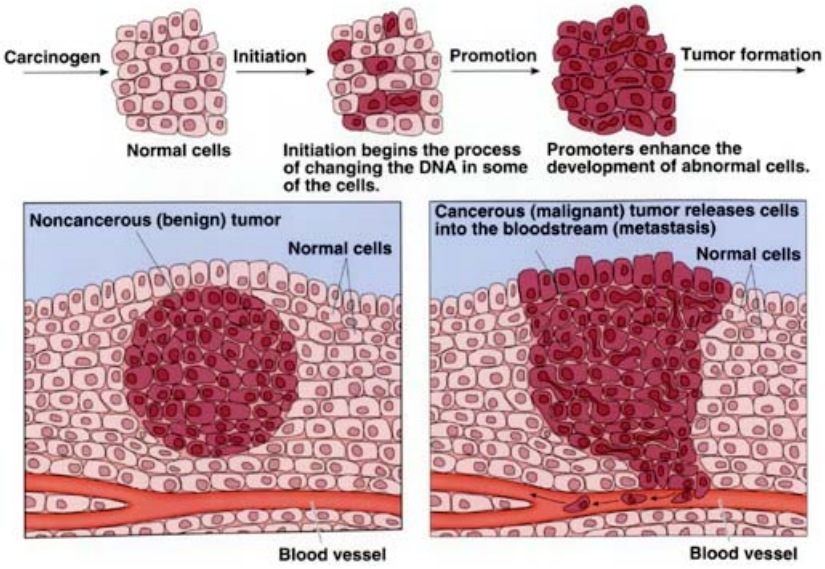


Some cancers develop when the gene regulators are defective, and in all cancers, gene expression is defective. For example, normal cell division has a number of checkpoint controls that ensure that division proceeds correctly. Like everything else, the checkpoint molecules are coded for by genes, and genes can mutate and genes can be suppressed. When checkpoint controls go awry, abnormal division results.

Cancer cells have abnormal plasma membranes and abnormal cytoplasm. Cancer cells divide rapidly and ignore overcrowding inhibition signals. They can make masses of cells called tumors. Once cell controls are not in effect, rapidly dividing cancer cells lose normal positioning and adhesion properties, too. Cancer cells can metastasize – migrate to new areas of the body and start growths in different tissues. Spread of cancer makes it more difficult to treat.

From what we know today, the steps in cancer development include:

- Exposure to a carcinogen from the environment by ingestion, inhalation, etc., naturally or via contamination
- Entry of the carcinogen into a cell
- Initiation of cancer via sufficient DNA changes in cell division control genes
- Promotion and enhancement of cancer via cell transformation
- Tumor formation and uncontrolled cell growth



No one knows why any one person gets cancer. Some cancers are familial, and probably genetic. Some cancers are related to the environment, especially the smoker's environment. A substance that causes a change in DNA that can lead to cancer is a **carcinogen** and exposure to a carcinogen is the first step in cancer.

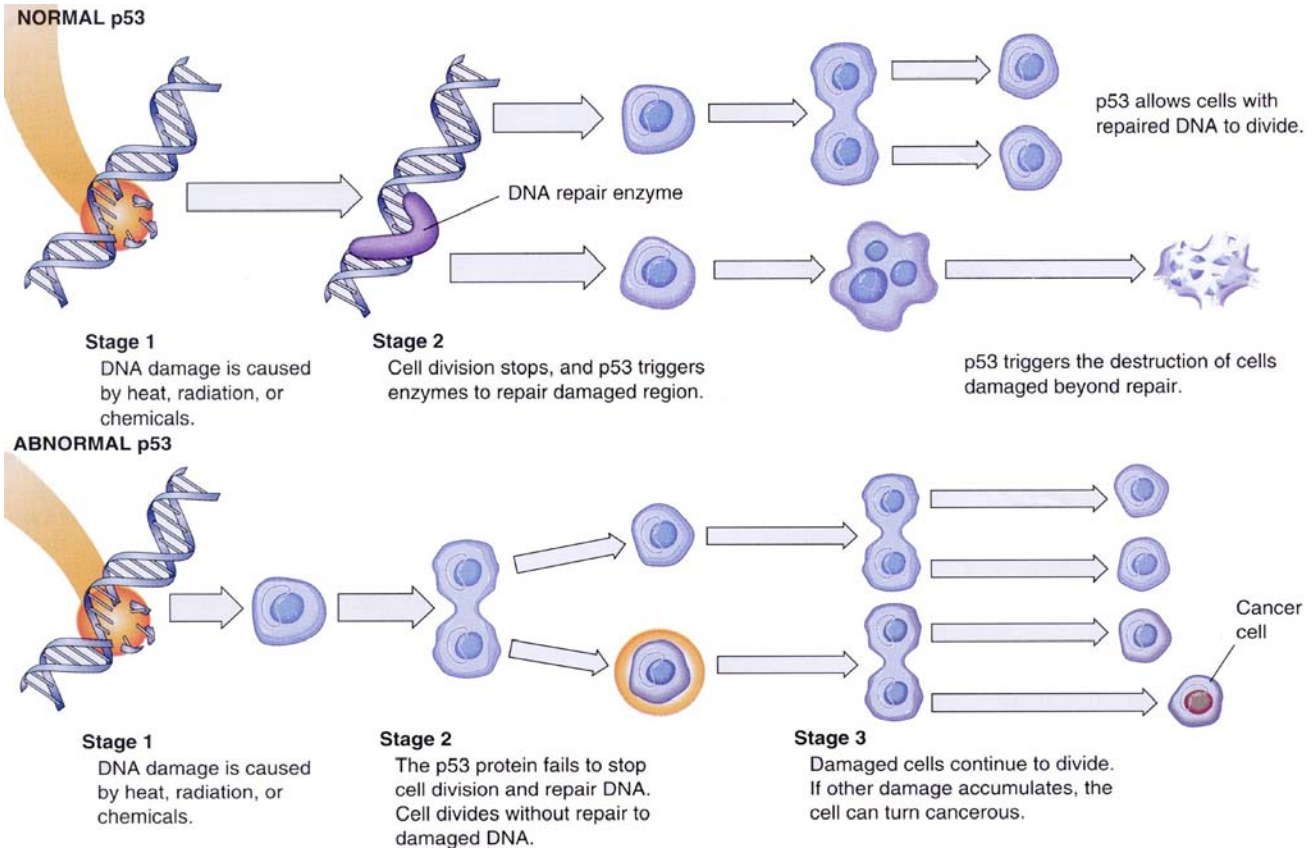
How does one get cancer?

Most believe that the onset of cancer is an accumulation of mutations rather than one single alteration. This correlates with the increase in many cancers with aging. A gene that has the potential to induce cancer is called an **oncogene**.

Any number of things in our surroundings can activate oncogenes. Chemicals that do so are called carcinogens. Radiation and the combustion products from tobacco are two of the most common carcinogens. Asbestos and some heavy metals in particulate form are also carcinogens. Many steroids in higher than normal concentrations are carcinogenic, and a high fat, low fiber diet is also suspected as being cancer promoting. Some viruses promote cancer formation.

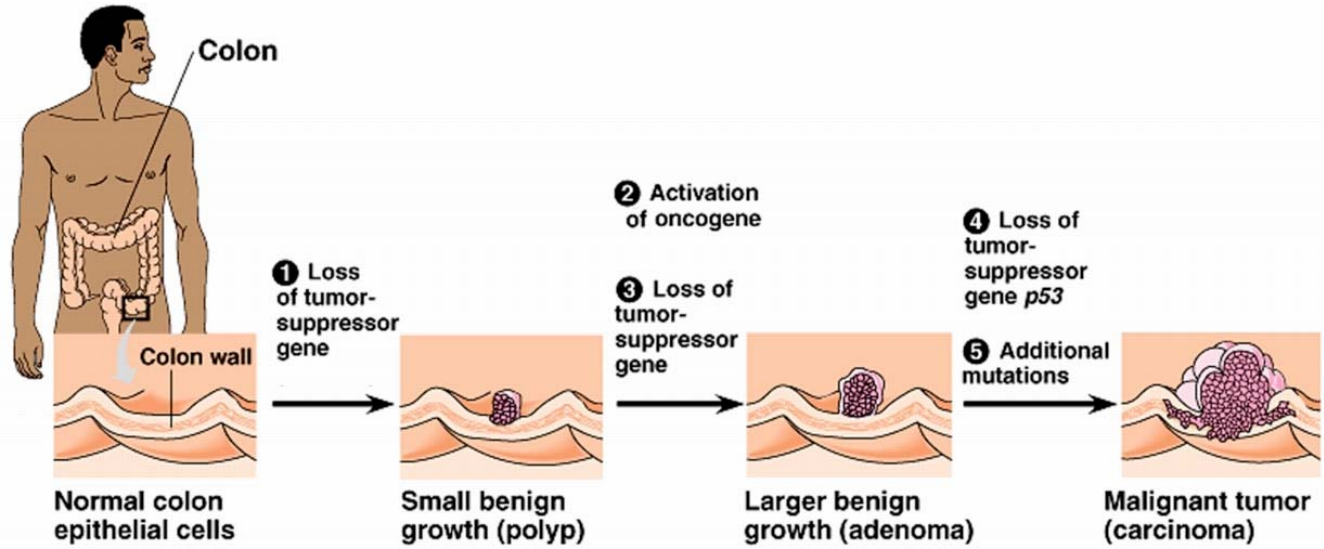
One specific protein product of an oncogene that is important in cell division is the **P53 tumor suppressor gene**. P53 is a transcription factor for genes that keep a cell's DNA repaired and genes that delay the cell's rate of cell division so that there is time for DNA repair. If the cell is in bad shape, P53 activates cell suicide genes to prevent the harmful mutations from being passed on. Such cell death is called **apoptosis**, and is genetically programmed. When p53 is defective or missing, cancers are more likely.

P53 and Cancer



Cancer and the Accumulated Mutations

We still cannot answer why one person will and another will not get cancer after exposure to the same potential carcinogens, but in some familial cancers several mutations have been identified in polyp cells that can become colon cancer tumors, including mutations in APC (a gene involved in cell migration and adhesion), Ras (tumor-suppressing gene) and p53.



Major Factors That Increase the Risk of Cancer

Factor	Examples of Implicated Cancers	Comments
Heredity	Retinoblastoma (childhood eye cancer) Osteosarcoma (childhood bone cancer)	Most cancers are not caused by heredity alone. Persons having family histories of certain cancers should follow physicians' recommendations.
Tumor viruses	Liver cancer Adult T cell leukemia/lymphoma Cervical cancer	Five viruses are initiators of certain cancers. (See Table 19.1)
Tobacco use	Lung cancer Cancers of the oral cavity, esophagus, and larynx Cancers of the kidney and bladder	Cigarette smoking is responsible for approximately one-third of all cancers. Nonsmokers have an increased risk of smoking-related cancers if they regularly breathe in sidestream smoke.
Alcohol consumption	Cancers of the oral cavity, esophagus, and larynx Breast cancer	The combined use of alcohol and tobacco leads to a greatly increased risk of these cancers. The mechanism of action in breast cancer is not yet known.
Industrial hazards	Lung cancer	Certain fibers, such as asbestos, chemicals such as benzene and arsenic, and wood and coal dust are prominent industrial hazards.
Ultraviolet radiation from the sun	Skin cancers	Those at greatest risk are fair-skinned persons who burn easily. However, everyone is at risk and should wear sunscreens and protective clothing when in the sun for extended periods of time. All types of UV radiation in tanning beds (UVA, UVB, & UVC) are harmful and may lead to skin cancer.
Ionizing radiation	Related to location and type of exposure	Eliminate unnecessary medical X rays to lower cancer risk. Infants and children are particularly susceptible to the damaging effects of ionizing radiation. Check your home to detect high levels of radon gas.
Hormones (estrogen and possibly testosterone)	Breast, cervical, ovarian, and prostate cancers	Estrogen-only and estrogen-progesterone hormone replacement therapies both increase the risk of breast cancer. Oral contraceptives increase the risk of breast cancer and cervical cancer, while reducing the risk of ovarian cancer. The role of testosterone in prostate cancer is unclear.
Diet	Breast and prostate cancers (weak association with high-fat diets), stomach and esophageal cancers (nitrites).	Nitrites found in salt-cured, salt-pickled, and smoked foods increase the risk of cancer.