GENE EXPRESSION II: MECHANISMS FOR REGULATING TRANSCRIPTION; EPIGENETIC REGULATION OF CELLULAR INHERITANCE AND DIFFERENTIATION

Date: August 21, 2019 10:30 a.m. – 12:00 p.m.

KEY CONCEPTS AND LEARNING OBJECTIVES

1. Eukaryotes typically regulate gene transcription through the binding of multiple transcription factors (regulatory proteins).
   a. Explain the concept of combinatorial gene control.
   b. Describe how Myc, Max, and Mad binding to the E-box element can regulate the switch between cell division and cell differentiation.
   c. Explain the difference between gene silencing and gene repression.
   d. List some of the mechanisms of transcription factor activation
   d. Describe how different signal transduction pathways activate kinases that regulate the AP-1 transcription factors Jun and Fos.

2. Chromatin structure can regulate cell phenotype.
   a. Describe how the position of a gene in the heterochromatin or euchromatin of a chromatid influences its expression.
   b. Explain the role of the locus control region in regulating β-globin gene expression in red blood cells.
   c. Explain how the X-chromosome is inactivated in female cells.

3. Embryonic cells are able to differentiate into specialized cells, which imply a genetic memory passed on from parental cells to progeny cells that tell the progeny cell to take on a particular specialized phenotype (i.e. a parental liver cell produces progeny liver cells).
   a. Define the term epigenetic.
   b. Describe how a positive feedback loop can pass epigenetic information from a parent cell to its progeny.
   c. Explain how positive feedback transcription factor loops can generate cellular memory.
   d. Explain how DNA methylation affects gene expression, and how a pattern of DNA methylation can be passed from parental cell to progeny cells.
   e. Understand the function of CG islands in maintaining the expression of housekeeping genes.
   f. Explain how histone acetylated regions of the genome can be passed from parental cell to progeny cells.
   g. Understand how most genes in the DNA within the egg and sperm are stripped of methyl groups except for a small number of genes where methylation may be maintained.
   h. Explain how genomic imprinting occurs in a sex dependent manner
   j. Explain the concept of functionally haploid with respect to genetic imprinting
I. Combinatorial Gene Control is Normal for Eukaryotic Cells
A cell must be able to change its gene expression patterns through its life cycle. Whether the cell needs to divide, differentiate or die, it needs to turn on certain genes and turn off other genes. This is accomplished through the expression of multiple regulatory proteins that bind to the same regulatory element in a gene to either promote or repress transcription. This is known as combinatorial gene control.

Some important concepts surrounding combinatorial gene control

Figure 7-31 The integration of multiple inputs at a promoter.
This figure illustrates how multiple sets of transcription regulators (co-activators and co-repressors) can work together to influence transcription initiation at the promoter. It is not well understood how the cell achieves integration of multiple inputs, but it is likely that the final transcriptional activity of the gene results from a sum total of signals from the activators and repressors acting on the promoter.

A. Example of Combinatorial Gene Regulation: Mad, Max, and Myc
Myc, Mad and Max are a family of bHLH transcription factors that form homodimers and heterodimers. They bind to regulatory elements called E-boxes that are found in the promoter regions of genes involved in cell division. Max is constitutively expressed and forms a homodimer that binds to the E-box element. Max does not activate gene expression. When Max:Max is bound, the gene is "repressed". That is, it is not turned on nor is it completely off. Instead, there is a baseline level of transcriptional activity resulting in leaky, low level gene expression.

If the cell needs to divide, it will express Myc. Myc is a transcriptional activator. Myc binds Max and the Myc:Max heterodimer binds the E-box in the gene regulatory region of genes promoting cell division. Myc is only expressed when cells are transitioning from G1 to S phase of the cell cycle.
If a cell needs to differentiate, it will express Mad. Differentiation is the opposite of cell division. A cell that is dividing cannot differentiate as these processes are incompatible. If a cell is going to differentiate, it needs to completely shut down cell division. Mad:Max heterodimers bind to the E-box and silence expression of genes involved in cell division. This is a strong repressive signal that completely shuts off even leaky expression. Mad will recruit HDACs that deacetylate the histone tails and prevent TFIID from binding and initiating transcription.

Myc, Max and Mad are bHLH transcription factors. They function as dimers and the DNA binding domain contains basic amino acids that interact with the DNA via charge-charge-interactions. The dimerization domain is a leucine zipper motif with the leucines in the two $\alpha$-helices interacting via their hydrophobic edge.

This figure shows us a common way that the different regions transcription factors are illustrated in the literature. There is a leucine zipper (Zip), which is the dimerization domain; the helix-loop-helix (HLH) motif, which contains the DNA binding domain; and the basic amino acids ("b"), which are part of and/or close to the DNA binding domain to promote charge-charge interactions between the proteins and DNA. The activation domain is not labeled per se, but we can see that Max has a very short N-terminus compared to Mad and Myc. That is because Max has no activation domain. It cannot turn on expression or turn off expression, which is why it is said to repress expression. Mad has an activation domain, and it serves to silence gene expression while Myc’s activation domain promotes gene expression.

A summary of key characteristics of Mad, Max and Myc. Myc is an important oncogene and is often overexpressed in cancer cells. That makes logical sense – it promotes cell division and cancer is a disease of unregulated cell growth and proliferation.
II. Regulation of Gene Expression
A. Activation of transcription factors and induction of gene expression.

We have discussed how transcription factors binding to a regulatory site begin the cascade of events resulting in euchromatin formation and gene expression. Now we will consider how the transcriptional regulator was expressed in the first place.

Possible mechanisms for the activation of transcription factors.

Figure 7–32 Some ways in which the activity of transcription regulators is controlled inside eukaryotic cells. (A) The protein is synthesized only when needed and is rapidly degraded by proteolysis so that it does not accumulate. (B) Activation by ligand binding. (C) Activation by covalent modification. Phosphorylation is shown here, but many other modifications are possible. (D) Formation of a complex between a DNA-binding protein and a separate protein with a transcription-activating domain. (E) Unmasking of an activation domain by the phosphorylation of an inhibitor protein. (F) Stimulation of nuclear entry by removal of an inhibitory protein that otherwise keeps the regulatory protein from entering the nucleus. (G) Release of a transcription regulator from a membrane bilayer by regulated proteolysis.
1. Extracellular signals are a common mechanism for activating transcription factors. Generally speaking, extracellular signals often trigger the events that lead to transcription factor activation. These signals can be (1) autocrine signals, meaning that the cell produces a cytokine or other factor that is released from that cell, but then engages its own membrane receptors and induces signaling. (2) paracrine signals, meaning that a neighboring cell has produced a cytokine that will engage the cell surface receptors and start the signaling process. (3) cell-cell signals where an interaction occurs between two cells, with one cell expressing the ligand on its surface and the other cell expressing the receptor.

Regardless of how the receptor-ligand interaction is generated (autocrine, paracrine or cell-cell), it triggers a signal transduction cascade as shown in this cartoon. This is a kinase cascade where the receptor-ligand interaction causes activation of a kinase which then phosphorylates another target kinase, which phosphorylates another target kinase until the transcription factor is phosphorylated and translocates to the nucleus to activate transcription. This may seem convoluted, but it allows a small amount of a ligand to produce an amplified, powerful cell signal.

a. Example: Gene regulation through the c-Fos / c-Jun AP-1 element
The transcription factor Els is required in addition to the dimer transcription factor AP-1 (a heterodimer of c-Jun and c-Fos) to turn on transcription of many genes required for metastasis of cancer cells. The AP-1 element binds the transcription factors c-Jun (activated by the JNK kinase pathway) and c-Fos (activated by the ERK kinase pathway). Jun and Fos are bZIP proteins that heterodimerize through leucine zippers and bind to the AP-1 element in the gene regulatory region.
B. Changes in chromatin structure regulate gene expression

Gene expression can also be regulated by the conversion of euchromatin to heterochromatin. The best example is X-chromosome inactivation.

Figure 7-50 X-inactivation

Females have two copies of the X-chromosome while males have one copy. This creates a gene dosage issue in females, who could, theoretically, express too much product from X-encoded genes. To avoid this situation, one of the two X-chromosomes is inactivated in females. Early in development, at the 20-30 cell stage, the placenta sends out a signal for the cells to randomly inactivate one of the two X-chromosomes. A non-coding RNA will coat the X-chromosome and recruit HDACs and methylases to place inhibitory marks that convert the chromosome into heterochromatin permanently. This condensed X-chromosome can be visualized in cells and is called the Barr Body.

Selection of the X-chromosome (maternal or paternal) to be inactivated is random, but because it occurs at the 20-30 cell stage, the female embryo is a mosaic of cells, some with the paternal and some with the maternal X-chromosome inactivated. When a cell divides to make the next series of daughter cells, it maintains the knowledge of which X-chromosome was inactivated, and this information is passed on at each subsequent cell generation.

In actuality, only about 90% of the information on the X-chromosome is inactivated. There are a small number of genes that can still be expressed.
Another important example of regulation through heterochromatin formation occurs in hemoglobin expression. The globin genes are located on chromosome 11. β-globin is only transcribed in early red blood cells (RBCs), and in every other cell type, this region of chromosome 11 is maintained in the form of heterochromatin.

In early development, transcription factors bind to the locus control region (located upstream from the globin gene locus) and promote transcription of hemoglobin ε, which is found in the embryo. After 10 weeks gestation, this switches to transcription of γ-globin genes and finally to β-globin after birth. δ-globin is also produced, but only constitutes about 3% of all globin. At birth, the transcription factors required for ε and γ-globin synthesis are no longer produced. Thus, the transcription factors required to express the globin genes are only produced by RBCs, and within the RBC, the type of globin produced is highly regulated by heterochromatin formation.

C. Epigenetic mechanisms and regulation of gene expression

Epigenetic mechanisms pass information regarding gene expression from parent to progeny cells. These include modification of DNA or regulatory gene proteins.

A differentiated cell expresses a set of genes that is characteristic of that cell type. Certain genes are turned off and others are turned on. When the cell divides, the same gene expression pattern is found in the progeny cell as in the parental cell. Several mechanisms (listed on the left) are used to pass this information from one generation to the next.
1. **Positive feedback loops (self-activation, auto-activation)**
Positive feedback loops are a simple and common mechanism for transmitting information to progeny cells. In this example, we see that when gene A is expressed, it produces protein A. Protein A is a transcription factor and there is a regulatory element for protein A in the regulatory region of gene A. Therefore, protein A feeds back and promotes its own expression. When the cell divides, protein A is present in the cytoplasm of the cell and will continue to feedback on its own gene even in the absence of the initial transient signal that leads to its initial expression. This autofeedback mechanism will continue for generation to generation as protein A will continue to regulate its own expression in the progeny cells.

**Figure 7-39.** A positive feedback loop creates cell memory.

a. **Example: Transcription factor expression patterns.**
An example of positive feedback loops is in transcription factor expression. The placenta is an asymmetrical organ. The cells in one region do not get the same signals as cells in other regions. This allows the production of a variety of cell types. In the figure below, we can see that the right side of the placenta sends out a signal for the cell to produce transcription factor 1 (TF1). The left side does not send out this signal. Thus, TF1 will only be expressed by the cell on the right side and not by a cell on the left side. The placenta then sends out a signal on both sides for cells to express either TF2 or TF3. Some cells will express only TF2, only TF3, TF1+TF2 or TF1+TF3. Next, the placenta signals for expression of TF4 or TF5. At the end, we have generated 14 different cells types expressing different combinations of transcription factors and therefore, different combinations of proteins.

**Figure 7-33.** The importance of combinatorial gene control for development.
2. **DNA methylation**

DNA methylation is another epigenetic mechanism influencing gene expression. This is not the same as histone methylation where lysines and arginine amino acids are methylated. This is methylation of cytosine in the DNA. Different enzymes are involved in this methylation. The methylation occurs on cytosines that are followed by a guanine. This is a mini palindrome of CG. Methylation of a promoter or enhancer region generally inhibits gene expression. If a C on one strand is methylated, the C on the complementary strand is methylated.

DNA methylation does not disrupt hydrogen bonding so it will not disrupt the double helix, but it can block transcription factor binding to regulatory elements. Methylation interferes with the binding interaction and can interfere with RNA polymerase binding at the transcriptional start site.

Thus, DNA methylation in the promoter or enhancer region of a gene can inhibit transcription and silence genes.

**Figure 7-43.** Formation of 5-methyl cytosine occurs by methylation of a cytosine base in the DNA double

When a cell divides, only one of the strands of DNA has methylation marks as shown in this figure. The newly prepared DNA strand is unmethylated. A maintenance methylase is used to methylate the corresponding C on the new DNA strand.

This is an interesting process. The maintenance methylase recognizes sites of hemimethylation where the C on one strand is methylated but the corresponding C on the other strand is not. Thus, the methylation marks are passed from parent to progeny cells.

**Figure 7-44.** How DNA methylation patterns are faithfully inherited.
a. **DNA methylation and silencing gene transcription**

DNA methylation upstream of the transcriptional start site tends to block transcription. At the top of the figure is a gene that is turned on. If the transcription factors promoting transcription were lost, the gene would be off, but would be leaking meaning that there is low level, baseline gene expression from the basal promoters (repressed). To completely turn off the gene, the upstream DNA can be methylated. This will block transcription factor binding, polymerase binding and the region will attract inhibitory proteins, HDACs, and chromatin remodeling complexes to promote heterochromatin formation and prevent all expression of the gene.

b. **CpG islands inhibit methylation of essential gene regulatory regions.**

About 40-50% of our genes are essential and complete silencing of these genes would be detrimental to the cell. Essential genes tend to have concentrations of CpG in the promoter and enhancer region called CpG islands that protect the gene from methylation.
It seems counterintuitive. Why put a concentration of cytosines in an area that cannot be methylated without detrimental effects? High densities of CpG are resistant to methylases. CpG islands resist methylation while less dense, less concentrated areas of CpG are methylated to inhibit transcription.

**Figure 7–46 The CG islands surrounding the promoter in three mammalian housekeeping genes.** The yellow boxes show the extent of each island. As for most genes in mammals, the exons (dark red) are very short relative to the introns (light red).

This can be confusing when you get to the lectures on cancer where you will discuss the methylation of CpG islands of tumor suppressor genes. CpG islands in normal cells resist methylation, but they can be methylated, and an example is tumor suppressor genes in cancer cells.

Re-drawn from *Nature* 2002 21(35) 5427

In a normal cell (top of figure), the CpG island is unmethylated. Activators, histone acetyltransferases and basal transcriptional machinery protect the CpG island and help ensure continued expression of this tumor suppressor gene.

In tumor cells (bottom of figure), the tumor suppressor gene is turned off by methylation of the CpG island. Transcriptional repressors, HDACs, DNA methyltransferases, and methyl-binding proteins shut down the CpG and prevent expression of this tumor suppressor gene.
3. Histone acetylation and deacetylation

Modified from Clement and Almouzni, Nat Struc Biol. 2015, 22(8):587

During DNA replication, the histones in the parental DNA strand are removed, recycled, and randomly inserted into one of the two newly formed DNA strands. Only half of the histones are from the parental cell and are appropriately acetylated. The other half of the histones are new and unacetylated. A maintenance acetylase is used to place acetylation marks in the same pattern as in the parental cell.

D. Imprinting

Genomic imprinting can be a difficult topic. These are some important facts about imprinting.

Normally, two functional copies of a gene are inherited from our parents. Because we have two copies, if one is mutated, the other can usually compensate, and the individual has no obvious symptoms. They are carriers of the mutation.

When the sperm and egg are made, the methylation marks are erased. It is unclear how they are removed. This is true for most genes, but there are a small number of genes (estimated from 5 to 300)
that are methylated in the sperm and the egg. These genes are turned off in a parent specific manner. The genes that are affected in the egg and the sperm are different. Certain genes are always turned off in the mother and other genes are always turned off in the father. These genes are called imprinted.

After fertilization, the genes retain their methylation marks and are not expressed. Imprinted genes are susceptible targets for pathologies because one copy of the gene is already silenced. The genes are considered functionally haploid meaning that a mutation in the one copy of the gene that is expressed can have detrimental effects.

Terminology:
In this example, gene A is said to be imprinted in both the male and female. It is imprinted off in the male (silenced) but imprinted on in the female.

Figure 7-48. Imprinting in the mouse.

This figure highlights the concept of imprinting. There is a male mouse and a female mouse. Each have two copies of gene A, but the paternal copy is silenced and the maternal copy is expressed. The egg and the sperm are shown where methylation has been erased in the female, but both copies are methylated in the male (silenced).

When the animals mate, and the egg is fertilized, the offspring will inherit one copy of gene A from the mother and one from the father. Regardless of whether the offspring inherits the orange or the yellow gene A from the father, it will not be expressed because it is imprinted off. What is important is whether the offspring gets the yellow or the orange from the mother as it is the only functional copy of the gene the offspring will receive.
1. **Example of passage of a mutated gene through imprinting.**

   This modified figure illustrates a potential problem with imprinting. Here, the mother has inherited a mutated gene from her father, but the gene is imprinted off (the gene is methylated in males so the copy inherited from her father is silenced). She is phenotypically normal, but has the potential to pass on a mutated gene.

   Her eggs will carry either the normal gene or the mutated gene. The methylation has been erased during meiosis (imprinted on), and the mutated gene can now be transmitted to her offspring. In the sperm, the copy of the gene is imprinted off. Regardless of which copy of the gene is present in the sperm, it is silenced and will not be expressed.

   Because the male copy of the gene is silenced, the female copy will be expressed. 50% of offspring will have the mutated gene and 50% will have the normal gene. Offspring with the mutated gene will express the disease.