Gene Expression and Regulation (Laura Hoopes)

How does a gene, which consists of a string (sequence) of DNA hidden in a cell's nucleus, know when it should express itself? How does this gene cause the production of a string (sequence) of amino acids called a protein? How do different types of cells know which types of proteins they must manufacture? The answers to such questions lie in the study of gene expression. Thus, this topic begins by showing how a quiet, well-guarded sequence of DNA is expressed to make RNA, and how the messenger RNA is translated from nucleic acid coding to protein coding to form a protein. Along the way, we examine the nature of the genetic code, how the elements of code were predicted, and how the actual codons were determined.

The first topic is the **regulation of genes**. Genes can't control an organism on their own; rather, they must interact with and respond to the organism's environment. Some are **constitutive genes**, or always "on," regardless of environmental conditions. Such genes are among the most important elements of a cell's genome, and they control the ability of DNA to replicate, express itself, and repair itself. These genes also control protein synthesis and much of an organism's central metabolism. In contrast**, regulated genes** are needed only occasionally — but how do these genes get turned "on" and "off"? What specific molecules control when they are expressed?

It turns out that the regulation of such genes differs between prokaryotes and eukaryotes. For prokaryotes, most regulatory proteins are negative and therefore turn genes off. Here, the cells rely on protein–small molecule binding, in which a ligand or small molecule signals the state of the cell and whether gene expression is needed. The repressor or activator protein binds near its regulatory target: the gene. Some regulatory proteins must have a ligand attached to them to be able to bind, whereas others are unable to bind when attached to a ligand. In prokaryotes, most regulatory proteins are specific to one gene, although there are a few proteins that act more widely. For instance, some repressors bind near the start of mRNA production for an entire operon, or cluster of co-regulated genes. Furthermore, some repressors have a fine-tuning system known as attenuation, which uses mRNA structure to stop both transcription and translation depending on the concentration of an operon's end-product enzymes. (In eukaryotes, there is no exact equivalent of attenuation, because transcription occurs in the nucleus and translation occurs in the cytoplasm, making this sort of coordinated effect impossible.) Yet another layer of prokaryotic regulation affects the structure of RNA polymerase, which turns on large groups of genes. Here, the sigma factor of RNA polymerase changes several times to produce heat- and desiccation-resistant spores.

For eukaryotes, cell-cell differences are determined by expression of different sets of genes. For instance, an undifferentiated fertilized egg looks and acts quite different from a skin cell, a neuron, or a muscle cell because of differences in the genes each cell expresses. A cancer cell acts different from a normal cell for the same reason: It expresses different genes. (Using microarray analysis, scientists can use such differences to assist in diagnosis and selection of appropriate cancer treatment.) Interestingly, in eukaryotes, the default state of gene expression is "off" rather than "on," as in prokaryotes. Why it is the case? The secret lies in chromatin, or the complex of DNA and histone proteins found within the cellular nucleus. The histones are among the most evolutionarily conserved proteins known; they are vital for the well-being of eukaryotes and brook little change. When a specific gene is tightly bound with histone, that gene is "off." But how, then, do eukaryotic genes manage to escape this silencing? This is where the histone code comes into play. This code includes modifications of the histones' positively charged amino acids to create some domains in which DNA is more open and others in which it is very tightly bound up. DNA methylation is one mechanism that appears to be coordinated with histone modifications, particularly those that lead to silencing of gene expression. Small noncoding RNAs such as RNAi can also be involved in the regulatory processes that form "silent" chromatin. On the other hand, when the tails of histone molecules are acetylated at specific locations, these

molecules have less interaction with DNA, thereby leaving it more open. The regulation of the opening of such domains is a hot topic in research. For instance, researchers now know that complexes of proteins called chromatin remodeling complexes use ATP to repackage DNA in more open configurations. Scientists have also determined that it is possible for cells to maintain the same histone code and DNA methylation patterns through many cell divisions. This persistence without reliance on base pairing is called epigenetics, and there is abundant evidence that epigenetic changes cause many human diseases.

In order for transcription to occur, the area around a prospective transcription zone needs to be unwound. This is a complex process requiring the coordination of histone modifications, transcription factor binding and other chromatin remodeling activities. Once the DNA is open, specific DNA sequences are then accessible for specific proteins to bind. Many of these proteins are **activators**, while others are **repressors**; in eukaryotes, all such proteins are often called **transcription factors (TFs).** Each TF has a specific DNA binding domain that recognizes a 6-10 base-pair motif in the DNA, as well as an effector domain. In the test tube, scientists can find a footprint of a TF if that protein binds to its matching motif in a piece of DNA. They can also see whether TF binding slows the migration of DNA in gel electrophoresis.

For an activating TF, the effector domain recruits RNA polymerase II, the eukaryotic mRNA-producing polymerase, to begin transcription of the corresponding gene. Some activating TFs even turn on multiple genes at once. All TFs bind at the promoters just upstream of eukaryotic genes, similar to bacterial regulatory proteins. However, they also bind at regions called enhancers, which can be oriented forward or backwards and located upstream or downstream or even in the introns of a gene, and still activate gene expression. Because many genes are co-regulated, studying gene expression across the whole genome via microarrays or massively parallel sequencing allows investigators to see which groups of genes are co-regulated during differentiation, cancer, and other states and processes.

Most eukaryotes also make use of small non-coding RNAs to regulate gene expression. For example, the enzyme Dicer finds double-stranded regions of RNA and cuts out short pieces that can serve in a regulatory role.

Imprinting is yet another process involved in eukaryotic gene regulation; this process involves the silencing of one of the two alleles of a gene for a cell's entire life span. Imprinting affects a minority of genes, but several important growth regulators are included. For some genes, the maternal copy is always silenced, while for different genes, the paternal copy is always silenced. The epigenetic marks placed on these genes during egg or sperm formation are faithfully copied into each subsequent cell, thereby affecting these genes throughout the life of the organism.

Still another mechanism that causes some genes to be silenced for an organism's entire lifetime is **X inactivation**. In female mammals, for instance, one of the two copies of the X chromosome is shut off and compacted greatly.

Repressor protein: A protein that binds to an operator, blocking transcription of an operon and the enzymes for which the operon codes.

Operon: A unit of genetic material that functions in a coordinated manner by means of an operator, a promoter, and one or more structural genes that are transcribed together.

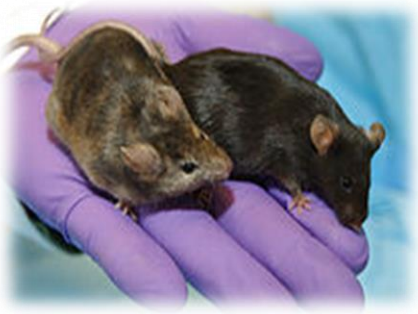
Chromatin: A complex of nucleic acids and proteins, primarily histones, in the cell nucleus that stains readily with basic dyes and condenses to form chromosomes during cell division.

RNA interference (RNAi): It is also called post transcriptional gene silencing (PTGs), is a biological process in which RNA molecules inhibit gene expression, typically by causing the destruction of specific mRNA molecules.

Ligand: A molecule that binds to another molecule, used especially to refer to a small molecule that binds specifically to a larger molecule.

Gene knockout: A gene knockout (KO) is a genetic technique in which one of an organism's genes is made inoperative or non-functional ("knocked out" of the organism's genome). Also known as knockout organisms or simply knockouts, they are used in learning about a gene that has been sequenced, but which has an unknown or incompletely known function. Researchers draw inferences from the difference between the knockout organism and normal individuals.

The term also refers to the process of creating such an organism, as in "knocking out" a gene. The technique is essentially the opposite of a "gene knock-in". Knocking out two genes simultaneously in an organism is known as a double knockout (DKO). Similarly the terms triple knockout (TKO) and quadruple knockouts (QKO) are used to describe three or four knocked out genes, respectively.



A laboratory mouse in which a gene affecting hair growth has been knocked out (left), is shown next to a normal lab mouse

These female flies opposite have a mean lifespan difference of nearly 40% as a result of a single gene knocked out in the insulin/insulin-like signaling pathway. The one on the left lives longer; the gene knockout results in dwarfism as well as longevity.

Method

Knockout is accomplished through a combination of techniques, beginning in the test tube with a plasmid, a bacterial artificial chromosome or other DNA construct, and proceeding to cell culture. Individual cells are genetically transfected with the DNA construct. Often the goal is to create a transgenic animal that has the altered gene. If so, embryonic stem cells are genetically transformed and inserted into early embryos. Resulting animals with the genetic change in their germ-line cells can then often pass the gene knockout to future generations. To create knockout moss, transfection of protoplasts is the preferred method. Such transformed *Physcomitrella*-protoplasts directly regenerate into fertile moss plants. Already eight weeks after transfection the plants can be screened for gene targeting via PCR.

The construct is engineered to recombine with the target gene, which is accomplished by incorporating sequences from the gene itself into the construct. Recombination then occurs in the region of that sequence within the gene, resulting in the insertion of a foreign sequence to disrupt the gene. With its sequence interrupted, the altered gene in most cases will be translated into a nonfunctional protein, if it is translated at all.

A knockout mouse (left) that is a model of obesity, compared with a normal mouse.

A conditional knockout allows gene deletion in a tissue or time specific manner. This is done by introducing short sequences called loxP sites around the gene. These sequences will be introduced into the germ-line via the same mechanism as a knock-out. This germ-line can then be crossed to another germ line containing Cre-recombinase which is a viral enzyme that can recognize these sequences, recombines them and deletes the gene flanked by these sites.

Because the desired type of DNA recombination is a rare event in the case of most cells and most constructs, the foreign sequence chosen for insertion usually includes a reporter. This enables easy selection of cells or individuals in which knockout was successful. Sometimes the DNA construct inserts into a chromosome without the desired homologous recombination with the target gene. To eliminate such cells, the DNA construct often contains a second region of DNA that allows such cells to be identified and discarded.

In diploid organisms (such as insects), which contain two alleles for most genes, and may as well contain several related genes that collaborate in the same role, additional rounds of transformation and selection are performed until every targeted gene is knocked out. Selective breeding may be required to produce homozygous knockout animals.

Knock-in is similar to knock-out, but instead it replaces a gene with another instead of deleting it.

Knock-in in molecular biology, refers to a genetic engineering method that involves the insertion of a protein coding cDNA sequence at a particular loucs in an organism's chromosome. Typically, this is done in mice since the technology for this process is more refined, and because mouse embryonic stem cells are easily manipulated. The difference between knock-in technology and transgenic technology is that a knock-in involves a gene inserted into a specific locus, and is a "targeted" insertion.

A common use of knock-in technology is for the creation of disease models. It is a technique by which scientific investigators may study the function of the regulatory machinery (e.g. promoters) that governs the expression of the natural gene being replaced. This is accomplished by observing the new phenotype of the organism in question. The BACs (Bacterial artificial chromosomes) BACs and YACs (Yeast artificial chromosomes) are used in this case so that large fragments can be transferred.

Use of transgenic technology:

- Knockouts are primarily used to understand the role of a specific gene or DNA region by comparing the knockout organism to a wild-type with a similar genetic background.
- Knockouts organisms are also used as screening tools in the development of drugs, to target specific biological processes or deficiencies by using a specific knockout, or to understand the mechanism of action of a drug by using a library of knockout organisms spanning the entire genome, such as in yeast *Saccharomyces cerevisiae*.

Epigenetics: In biology, and specifically genetics, epigenetics is the study of changes in gene expression or cellular phenotype, caused by mechanisms other than changes in the underlying DNA sequence – hence the name *epi-* (Greek: *επί-* over, above, outer) -*genetics*, some of which are heritable.

It refers to functionally relevant modifications to the genome that do not involve a change in the nucleotide sequence. Examples of such modifications are **DNA methylation and histone modification**, both of which serve to regulate gene expression without altering the underlying DNA sequence. Gene expression can be controlled through the action of repressor proteins that attach to silencer regions of the DNA. These changes may remain through cell divisions for the remainder of the cell's life and may also last for multiple generations. However, there is no change in the underlying DNA sequence of the organism; instead, non-genetic factors cause the organism's genes to behave (or "express themselves") differently.

There are objections to the use of the term epigenetic to describe chemical modification of histone, since it remains unclear whether or not histone modifications are heritable.

One example of epigenetic changes in eukaryotic biology is the process of **cellular differentiation**. During morphogenesis, totipotent (ability to generate an organism) stem cells become the various pluripotent cells (a cell capable of differentiating into most cell types found in an organism but not capable of forming a functional organism) of the embryo, which in turn become fully differentiated cells. In other words, a single fertilized egg cell – the zygote – changes into the many cell types including neurons, muscle cells, epithelium, endothelium of blood vessels, etc. as it continues to divide. It does so by activating some genes while inhibiting others.

Gene silencing: is a general term describing epigenetic processes of gene regulation. The term gene silencing is generally used to describe the "switching off" of a gene by a mechanism other than genetic modification. That is, a gene which would be expressed ("turned on") under normal circumstances is switched off by machinery in the cell. Gene silencing occurs when RNA is unable to make a protein during translation.

Genes are regulated at either the transcriptional or post-transcriptional level.

Transcriptional gene silencing is the result of histone modifications, creating an environment of hetero-chromatin around a gene that makes it inaccessible to transcriptional machinery (RNA polymerase, transcription factors, etc.).

Post-transcriptional gene silencing is the result of mRNA of a particular gene being destroyed or blocked. The destruction of the mRNA prevents translation to form an active gene product (in most cases, a protein). The blocking of the gene occurs through the activity of silencers, which bind to repressor regions. A common mechanism of post-transcriptional gene silencing is RNAi.

Both transcriptional and post-transcriptional gene silencing are used to regulate endogenous genes. Mechanisms of gene silencing also protect the organism's genome from transposons and viruses. Gene silencing thus may be part of an ancient immune system protecting from such infectious DNA elements.

Genes may be silenced by DNA methylation during meiosis, as in the filamentous fungus *Neurospora crassa.*