Epigenetics: Gene Regulation from Conception throughout Life

By Orlando Jack Miller

A paratrooper (11th Airborne Division) during the Korean War, Orlando Jack Miller received B.S. and M.D. degrees at Yale, as well as completing an obstetrics and gynecology residency. He began genetic research at University College London and was part of the team that first discovered that there were two extra chromosomes both in persons with Down syndrome and those with Klinefelter syndrome.

Jack was a Professor of Obstetrics and Gynecology and of Human Genetics and Development at Columbia University, leaving after 26 years to found a department of Molecular Biology and Genetics at Wayne State University in Detroit, Michigan. He published more than 200 peer-reviewed papers, many in such leading journals as *Science, Nature*, and *Cell*.

He served on the editorial boards of nine scientific journals, was the second President of the American Board of Medical Genetics, and was a member of the first Genome Study Section of the National Institutes of Health.

Jack is a past president of the Blue Ridge Torch Club of Leesburg, Virginia. "Epigenetics" was originally presented to this Blue Ridge Torch Club on October 18, 2016.

Epigenetics is one of the most exciting areas in genetics. The Greek prefix "epi-" means roughly "above, over, or around," and "epigenetics" looks at the processes and contexts that govern or regulate the behavior of genes. Its primary goal is to find out how genes are turned on or off, from conception throughout life. For eighty years, geneticists focused on identifying genes and the location of genes on specific chromosomes. With only a few exceptions—that genes for hemoglobin are turned on only in red blood cells, genes for insulin are turned on only in islet cells in the pancreas, etc.—almost nothing was learned about how genes are regulated in any type of cell. Since no gene can turn itself on or off, or regulate the level of its activity, some other part of our genetic heritage, or genome, must do this.

Our genome is transmitted to each of us from our parents via a single sperm and single egg. The sperm nucleus contains a unique set of the twenty-two human non-sex chromosomes and an X or a Y chromosome; the egg nucleus contains another unique set of the twenty-two non-sex chromosomes and an X chromosome. Each individual thus receives two complete nuclear genomes, each containing a unique set of our 22,000 or so protein-coding genes. In addition, mitochondria in the egg cytoplasm contain thirty-seven genes that are essential for energy production within cells.

Virtually every cell in the human body contains a copy of each chromosome present in the fertilized egg from which it arose. How does the six feet of DNA of a cell fit into the nucleus of each of our trillions of cells? The answer is that chromosomal DNA undergoes a multi-step process of compaction. In the first step, about 200 base pairs of the extremely thin DNA thread are wrapped around a cluster of eight tiny histone proteins to form a nucleosome (for a helpful multi-color diagram of a nucleosome, see the Wikipedia article on "Nucleosome".)

Most of our DNA undergoes several additional rounds of compaction, producing a much sturdier complex of DNA, RNA, and protein: a chromatin thread. However, the enzymes necessary for synthesizing RNA and DNA cannot reach the DNA in such highly compacted chromatin. Thus, only genes in naked DNA and the DNA that is wrapped around a nucleosome can be turned on, not genes that have undergone further

compaction. Compaction was, for many years, the only known mechanism of gene inactivation.

* * *

The twenty-three human chromosome pairs contain a total of over three billion base pairs of DNA. Genetic information is stored in DNA using a triplet code that is identical in virtually all forms of life. (That is why almost any type of life form can become our food!) The genomes (i.e., the whole of the genetic material) of tens of thousands of humans have been completely sequenced. This research revealed that the 22,000 genes that code for our proteins are contained in less than 10% of our DNA. This tiny fraction of our DNA directs the synthesis of the thousands of messenger RNAs that direct the synthesis of our approximately 70,000 different proteins.

Research during the last forty years has shown that some segments of the other 90% of our DNA direct the synthesis of thousands of long and short non-protein-coding RNAs that turn genes on and off. They regulate gene activity from embryonic development throughout life (Carey 191).

All the genes in sperm are inactive (turned off), and so are most of the genes in an egg cell. Fertilization of the egg cell by a sperm triggers the activation of several specific genes. After three successive cell divisions, the embryo consists of eight identical "stem" cells. As cell divisions continue, different combinations of genes are turned on in different lineages, leading to the development of several hundred different types of cells. Virtually every cell in the body has a complete copy of the two sets of chromosomes that were present in the fertilized egg, but each type of cell has a unique combination of active genes!

How are genes activated?

A gene is usually activated by the addition of an acetyl group (CH3,CO—six atoms) to the amino acid lysine in the histones (alkaline protein found in cell nuclei) that DNA is wrapped around. This process is mediated by the enzyme histone acetylase. The eight histones in a nucleosome (the basic unit of DNA packaging) contain more than 50 sites where an acetyl group can be attached. This is enough attachment-sites to provide more than 4,000 unique combinations. If each of the many different types of cells in our bodies has a different combination of acetyl-containing sites in its genes, this could produce a unique set of active genes in each type of cell. This may be how a fertilized egg develops into such a complex organism!

How are genes inactivated?

Genes that have been activated often have to be turned off in some types of cell. The first gene inactivation method to be discovered was the methylation of cytosine, one of the four nucleotide bases (adenine, guanine, and thymine are the other three) that DNA is made of. This inactivation requires only the addition of a methyl group (CH3: four atoms) to the DNA base cytosine. My research group at Columbia showed that highly compacted DNA, which is genetically inert, is heavily methylated (Miller et al., 1974). Seven years later, Weintraub et al. (1981) showed that the DNA of inactive hemoglobin genes is methylated. At the same time, my research group showed that excess copies of the gene for ribosomal RNA synthesis are methylated (Tantravahi et al., 1981). DNA cytosine methylation is by far the most common mechanism for inactivating genes.

The ARH1 tumor suppressor gene is inactivated in nearly half of all breast and ovarian tumors.

DNA methylation of a gene that is normally turned on in a particular type of cell has now been shown to cause hundreds of different diseases, including many types of cancers. For example, the ARH1 tumor suppressor gene is methylated, and thus inactivated, in nearly half of all breast and ovarian tumors. Methylation of any of the thirty other known tumor suppressor genes can cause renal, colon, rectal, and other cancers. (Carey 217).

Another epigenetic mechanism for gene inactivation is called imprinting, which inactivates a specific gene on only one chromosome of a particular pair. An inactivating mutation of the homologous (same) gene on the non-imprinted chromosome leaves no active copy of the gene and causes a serious disorder, such as the Angelman syndrome, in which the individual has severe mental retardation, a small brain, and minimal speech. More than 50 other imprinted human genes have been discovered (Carey 135).

Such discoveries suggest the immense potential of this field. A major goal of epigenetics is to identify causes of developmental disorders, cancers, etc. We now know that many disorders arise from blocking histone acetylation. For example, inactivation of the histone acetylase gene causes the Rubinstein-Taybi syndrome (mental retardation, heart defects, and broad hands; Carey 257). Blocking histone acetylation can cause early development of rheumatoid arthritis, an autoimmune disease. Blocking the histone deacetylase SIRT6 gene leads to breakdown of the telomeres at the ends of chromosomes, causing premature aging (Carey 277). Conversely, activation of any of the 100 known cancer-causing oncogenes by histone acetylation can cause cancer (Sharma et al., 2010).

A fourth epigenetic mechanism of gene regulation is the formation of DNA loops. Chromosomes can be seen under the microscope during the process of cell division, but are invisible in most cells of the body. However, it is sometimes possible to show that two genes that are normally far apart have been brought close together within the cell nucleus by the formation of a DNA loop. This occurs when one "insulator protein" binds to two DNA sites that are about 900,000 base pairs apart on a chromosome, forming a DNA loop. Genes within the loop are blocked from interacting with DNA elements outside the loop.

42% of the human genome is now viral in origin.

At least one of the many cancer-causing oncogenes in the human genome is kept inactive by the formation of a DNA loop. However, if the loop is broken by methylation of DNA at the binding site on the insulator protein, the oncogene is activated, causing cancer (Flavahan et al., 2016).

Carolyn Brown et al. discovered a fifth type of epigenetic mechanism in 1991. They showed that the normal inactivation of one of the two X chromosomes in human females requires coating one X chromosome with copies of the long non-protein-coding RNA product of the Xist (X-inactive specific transcript) gene on that X chromosome. Loop formation has recently been shown to speed up the coating process (Engreitz et al., 2016).

Many environmental agents can modify our genome. X-rays, radioactivity, and various chemicals have long been known to cause mutations, but viruses are the major cause. In order to reproduce, viruses have to insert a DNA copy of themselves into their host's genome, where it directs the synthesis of many copies of the virus. One or more copies of the virus may remain in the person's DNA and may be reactivated many years later, producing shingles in the case of the chickenpox virus. Such processes have gone on for millions of years, and 42% of the human genome is now viral in origin, though much of it has been adapted for many epigenetic uses by its human hosts (Carey 127).

Another epigenetic mechanism is based on the methylation of adenine, one of the four nucleotide bases in RNA. T.P. Wu and colleagues found that antibodies to methylated adenine bind primarily to viral DNA sequences in mammalian DNA, especially the more recent invaders (Wu et al., 2016).

Many people are concerned that the growing number of chemicals used for widespread medical and agricultural purposes may affect future generations. Evidence for this is limited, but so is our knowledge of the human genome and of the effects of new chemicals or social environments. Here are two examples. In one study, the fungicide Vinclozolin reduced the fertility of several male rats. Four generations of their male offspring were said to have had reduced fertility despite no further

exposure to Vinclozolin (Carey 113). A different group of scientists, including a recent Nobel laureate, showed that chronic stress in people reduces telomerase activity and leads to shortened telomeres at the ends of chromosomes, shortening life spans by up to ten years (Epel et al., 2004). Our social environment affects our genomes in profound ways!

WORKS CITED

Brown, C.J. et al. "The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus." *Cell* 71:527-542, 1992.

Carey, N. *The Epigenetics Revolution: How Modern Biology is Rewriting Our Understanding of Genetics, Disease, and Inheritance.* Columbia University Press, 2012.

Engreitz, J.M. "The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome." *Science.* 2013 Aug 16; 341(6147). doi:10.1126/science.1237973

Epel, E.S. "Accelerated telomere shortening in response to life stress." *Proceedings of the National Academy of Sciences of the United States of America.* 2004 December 7; 101 (49) 17312-17315. doi. 10.1073/pnas.0407162101

Flavahan, W.A. et al. "Insulator dysfunction and oncogene activation in IDH mutant gliomas." *Nature.* 2016 Jan 7; 529(7584):110-4. 2016 doi:10.1038/Nature

Miller, O.J. et al.: "5-Methylcytosine localized in mammalian constitutive heterochromatin." *Nature.* 1974; 251: 636-637.

Sharma, S., Kelly, T.K., Jones, P.A. "Epigenetics in cancer." *Carcinogenesis.* 2010 Jan; 31(1): 27-36. Carcinogenesis–2010–Sharma-27-36-Scribd

Tantravahi, U. et al. "Amplified ribosomal RNA genes in a rat hepatoma cell line are enriched in 5-methylcytosine." *Proceedings of the National Academy of Sciences of the United States of America.* 1981. Jan. 1; 78 (1): 489-493, 1981

Weintraub, H.et al. "Alpha-globin-gene switching during the development of chicken embryos: expression and chromosome structure." *Cell* 24:333-344, 1981

Wu, T.P. et al. "DNA methylation of N6-adenine in mammalian embryonic stem cells. *Nature*. 2016 April 21; 532 (7599): 329-33. doi:10.1038/nature 17640.

Torch Clubs Online

In addition to the great resource IATC provides at Torch.org, many clubs have their own websites, where you can get timely information about upcoming events and more.

Albany, NY	albanytorchclub.org
Central PA	acuri.net/torchclub
Chambersburg, PA	aliveinthevalley.org/default-1.html
Durham-Chapel Hill, NC	durhamchapelhilltorch.org
Erie, PA	erietorchclub.com
Fox Valley, WI	torchfoxvalley.org
Frederick, MD	fredtorch.org/home.html
Hagerstown, MD	hagertorch.org
Montgomery County, VA	montgomerytorch.org
Rochester, NY	torchclubrny.wix.com/torch-c
Schenectady, NY	schenectadytorchclub.org
St. Catharines, ON	vaxxine.com/torch

Where are you?

Does your club have a website, but you don't see it listed here? Please contact IATC to make sure we have all your club's information up to date.