12/21/2012

AUTHORS:
KRISTIAN GLOB
METTE LASSEN
OMAR KHAN
DANIEL GROVER

Effect of Methylation on Gene Expression

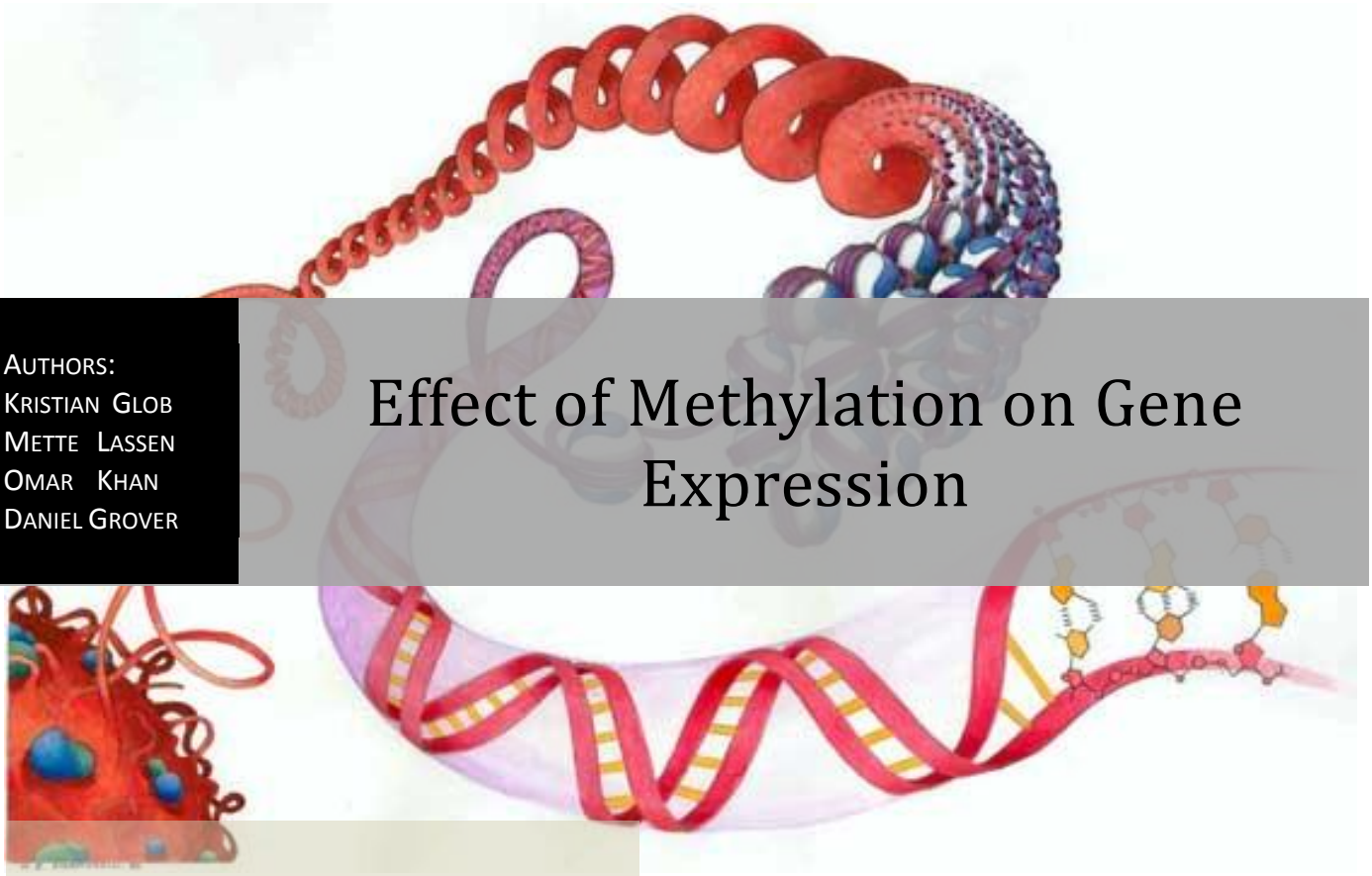


Image from [Bouvior, 2006]

First Semester Project | Supervisor: Christian Theil Have

Abstract

The field of epigenetics, specifically methylation of DNA, plays a crucial role in phenotypical gene expression and cell differentiation. Alterations in the natural pattern of methylation, such as hypermethylation and hypomethylation, can lead to disorders. These alterations can come from environmental influences or can be inherited. There are many biotechnology applications that can determine DNA sequencing patterns and detect methylation. The detection of methylation can help understand the development of diseases such as cancers, and has a possibility of being used as a diagnostic technique in the future.

Index

1 Problem formulation

Previously it was thought that DNA was the final blueprint for the function and expression of cells. However studies have shown a blueprint that is found attached to the DNA structure, the epigenetics of a cell. Etymologically the word epigenetic is derived from the greek word epi (meaning: over, above, outer) and genetic, so that epigenetic can be said to mean on top of or in addition to, genetic [Goldberg, 2007]. The epigenetic pattern of a cell can have an effect on the gene regulation and expression, leading to silencing of genes or over expression of genes. These patterns can be heritable and will continue to be expressed in new cells after cell division.

According to common belief in biology, Darwinism, evolution and phenotypical trait expression can only be changed through generations of being exposed to an environmental factor and natural selection. It has been discovered that epigenetics mechanisms such as DNA methylation can be changed in a single generation. This leads us to the following question:

How can methylation of CpG islands change the phenotypical expression of genes, and what can affect the amount of methylation found?

- Can methylation cause diseases?
- Can methylation be detected?

2 Method

In order to write this project we have spent time reading about the subject, before deciding on a problem formulation, that was interesting and would give us a chance to explore the subject of epigenetics in a deeper sense.

At this point we wrote an outline, containing the chapters and sections we thought we would have to cover, at the same time we wrote a project plan assigning the time we had left, until the hand in deadline, into smaller tasks, so that at any point, we would have an idea about whether our progress was according to plan or not.

During the writing of the project itself, we changed the overall structure of the report several times, and incorporated into the report ideas gained from critique from both our supervisor and from the midterm evaluation.

The main focus of the project is to review and explain the concepts of epigenetics and the areas of gene expression, influenced by it. We have also strived to provide the relevant background knowledge the reader needs in order to understand the more technical sides of the report. Finally we try to put the knowledge we have gained into perspective by covering some of the environmental effects on epigenetics as well as some of its consequences, here exemplified by diseases such as cancers and others.

This report covers some complex topics; therefore we have taken care to make sure all readers, with a basic understanding of biology, will be able to understand the more complicated parts of the report.

The reader is first introduced to the topic of epigenetics in the background section, section 4. In this section the reader is able to develop a general idea of what epigenetics is and where its study came from. A explanation of what DNA is and how it becomes expressed is then provided. Those with a firm grasp of biology, who is already familiar with the working of gene expression and DNA, can safely skip this part and go to section 5.

This is followed by the chapter on methylation and CpG islands, which is the core mechanics that the report focuses on. The influence of the environment on methylation levels is also studied here.

Our perspective section gives the reader some applications and case studies that will help put the knowledge acquired so far into context – allowing it to make more sense. A number of diseases are related to the methylation of DNA, we discuss that in this section. Biotechnological applications are about how the applications used to gain information about methylation on the DNA string.

Instead of a discussion section case studies have been selected in order to give the reader real life examples of how changes in the normal methylation pattern can effect phenotypical expression.

3 Background

3.1 Another look at Lamarckism

In the 19th century Jean-Baptiste Lamarck speculated that the loss of limbs due to amputation or the development of a lean physique from exercise could be passed on to the next generation [Rogers, 2009]. This concept was coined Lamarckism. The theory was created as an attempt to explain unique physical characteristics of some organisms. For example, one of the well-known Lamarckian ideas was that giraffes has such long neck because the past generations had to stretch their necks in order to reach the leaves high on the trees, meaning that the environmental influence of having high trees was able to effect the length of the giraffes necks.

However Lamarck had much opposition to his ideas. In the late 19th century a German biologist by the name of August Weismann decided to disprove his theory. He performed an experiment where he cut of the tails of mice, and then bred the animals. According the Lamarckian theory the offspring's should be born without tails, like their parents had. But Weismann observed that all of the offspring's still had their full length tails. Therefore Lamarckism was considered disproved [Rogers, 2009].

Along with Weismann's theory, Charles Darwin's theory of natural selection also became much more widely accepted than Lamarckism. Charles Darwin suggested that the only way for an organism to become changed is through the slow process of evolution that takes place over many generations.

It became generally accepted that the only way to inherit an acquired characteristic is by a mutation in the DNA of a germ cell [Rogers, 2009]. Environmental factors could affect the future generations, but many generations could pass before these changes where seen. However today scientist are reconsidering parts of Lamarck's theory, the environmental influences may be much more subtle than he thought, but can be just as dramatically influential. Scientists have already identified several acquired characteristics that can be passing on from one generation to the next, even sometimes to multiple generations later. These inheritable changes are recorded on the DNA as small chemical modifications to the DNA and the proteins associated with it. These small modifications became known as epigenetics. Epigenetics has become an area of great interest for scientist when it comes to phenotypical gene expression and the development of diseases.

3.2 Introduction to Epigenetics

Norrbotten, the remote northern county of Sweden, is where the idea that epigenetics is able to change the most basics element of life, DNA, was first proposed. No longer was DNA the final blueprint of the functions of all living things. Somehow the way you live in your current life can affect not only yourself, but also your future generations.

The concept of epigenetics being able to be passed through generations was first proposed by Dr. Lars Olov Bygren, a health specialist, who wanted to know what long-term effect famine and feasting had on generations. Dr. Bygren's study in the 1980's that it was suggested that the epigenetic traits that where acquired during one's lifetime could be passed on to the next generation.

Norrbotten is an isolated area in the northern part of Sweden. On average, only six people live in each square mile of this part of the country. In the 19th century the area was so isolated that if the harvest was bad, people would starve. The harvest each year was unpredictable, one year the people could starve, yet the next they could feast on a surplus. Some of the years of starvation and surplus can be seen in table 1.

Table 1: Years of starvation and surplus of food in Norrbotten in the 19th century

Starvation	Surplus
1800	1801
1812	1822
1821	1828
1836	1844
1856	1863

Dr. Bygren wanted to know what the long-term effects of the famine and feast years might have on the children and next generations of the Norrbotten population. From a random sample of 99 individuals born in the 1905 in the Overkalix perish of Norrbotten [Cloud, 2012]. Dr. Bygren and his two colleagues traced back the parents and grandparents of these individuals using historical

records, and by the help of agricultural records where able to see how much food had been available when the parents and grandparents had been young.

During the time Dr. Bygren was collecting his data, papers in the *Lancet* was published showing a study of pregnant women and their children. They study showed that if a pregnant woman ate poorly, her child would have a much higher chance than the average person of developing a cardiovascular disease once he/she became an adult [Cloud 2012]. This made Dr. Bygren wonder if the mother could affect the child's health even before the pregnancy. Could the way the parents have been living somehow change the traits of their off springs?

In 2001 Dr. Bygren published his results in the Dutch journal *Acta Biotheoretica.* His results showed that the boy's who went from normal eating one season, to gluttony in the next produced sons and grandsons who lived shorter lives. On average these grandsons died six years earlier than grandsons of those who had experienced bad harvest [Cloud, 2012]. Later, once other causes of death and socioeconomic variations were removed, the lifespan of the grandsons of those who starved went all the way up to 32 years greater than the others. Similar results were found on the female side of the families. Somehow a single winter of overeating as a child could create a biological chain of events resulting in one's grandchildren dying decades earlier than their peers [Cloud, 2012].

Dr. Bygren was now sure that the environment of feast and famine had somehow changed the epigenome of the population of Norrbotten. But he did not know how. To get an understanding of how changes to the epigenome can affect the phenotypically expression of DNA, an understanding of what DNA is and how it is expressed must first be present. The next two sections will give the reader an understanding of what DNA is and how it becomes expressed.

3.3 DNA and Genes

Deoxyribonucleic acid, more commonly known as DNA, is the most basic blueprint of all living organisms. All cells in an organism have, mostly, the same set of DNA. DNA becomes translated into usable compounds called RNA (Ribonucleic acid) in the process of transcription, which can then later be translated into macromolecules such as proteins. RNA is similarly build as DNA but instead of having a deoxyribose sugar in the backbone it has a ribose. But even though all

cells have the same basic information in the DNA, it is not all of it that is used in the individual cells.

Genes are segments of DNA which define human traits, such as eye color, and height [Bartoshesky, 2009]. Our DNA, which makes up the roughly 20,000 genes we each have, is often considered to be instruction manuals for the human bodies. However, the genes still need information such as what to do, and where and when to do so [Watters, 2006]. Human brain and liver cells contain identical sets of DNA, yet each is made to express only the proteins which are necessary to keep that organ functioning. This information is not included in the main text of the DNA itself, rather on it, as a collection of chemical switches, known collectively as the epigenome, lying along the span of the DNA double helixes.

3.3.1 Structure and bases

DNA structure is created by modifying the sugar ribose, into deoxyribose which can be seen in figure 1. This is done by replacing one of the OH of the ribose by a hydrogen.

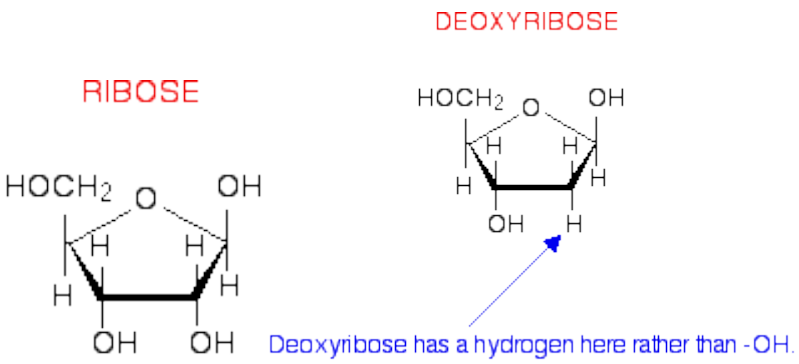


Figure 1: The Ribose and Deoxyribose molecule. The Deoxyribose molecule is used in the backbone of DNA structure, while the ribose sugar is used as the backbone in the RNA molecule. Figure from[Clark, 2011]

Each Carbon atom in the deoxyribose molecule is numbered, which becomes important in figuring out where other compound molecules attach to the sugar. Figure 2 demonstrates how the carbon atoms are numbered, from the carbon atom furthest to the right to then moving down and to the left of the molecule.

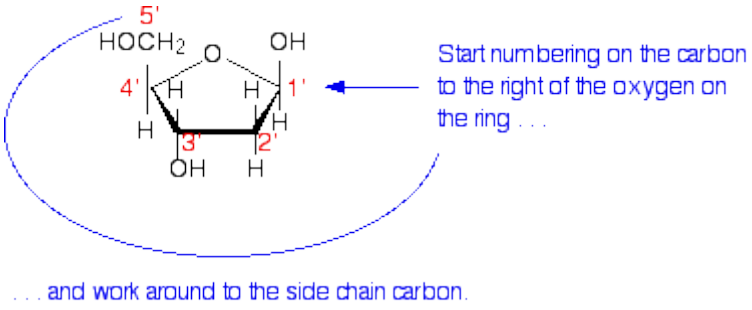


Figure 2: The numbering of the carbon atoms on the

deoxyribose molecule, from 1' to 5'. Figure from [Clark, 2011]

A phosphate group is attached to the 5' carbon atom of the deoxyribose molecule (figure 3) in order to complete the backbone of the DNA structure.

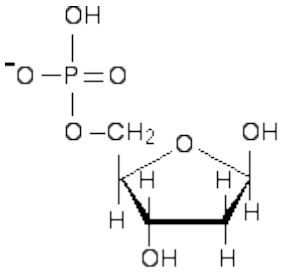


Figure 3: The addition of a phosphate group to the deoxyribose molecule. Figure from [Clark, 2011]

The bases are attached at the 1' Carbon atom (figure 5). In DNA the four different type of nitrogenous bases are: Cytosine (C), Thymine (T), Adenine (A), and Guanine (G). Figure 4 shows the structure of the bases. Cytosine, Thymine, and Uracil (a nitrogenous base in RNA replacing thymine) are categorized in a group called pyrimidines because of their single ring structure while adenine and guanine are categorized as purines due to their double ring structure. By attaching a base to the backbone a nucleotide is created.

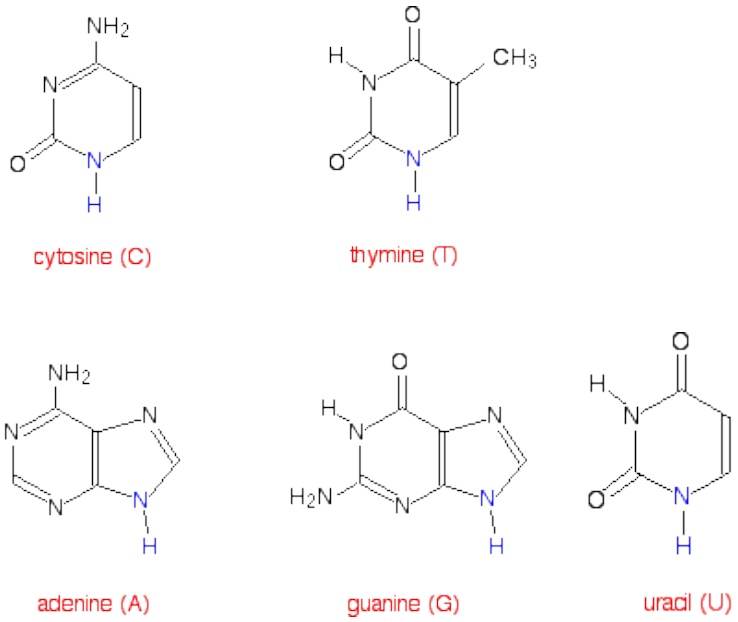


Figure 4: The four different types of bases in DNA. In RNA

thymine becomes replaced with uracil. Purines: A and G. Pyrimidines: C, T, and U .

Figure from[Clark, 2011].

The blue Nitrogen and Hydrogen is where the bases attach at the 1' carbon atom(figure 5). The hydrogen and the -OH group are lost through a condensation reaction when the base is attached to the deoxyribose.

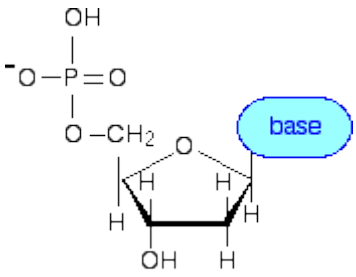


Figure 5: A nucleotide is created by attaching a base at the 1' carbon atom.

Figure from [Clark, 2011].

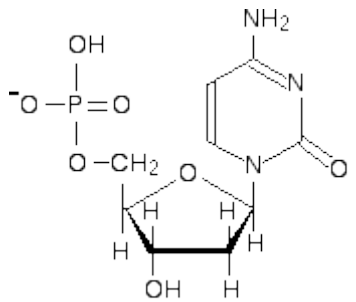


Figure 6: The attachment of a cytosine to a Deoxyribose

through a condensation reaction. Figure from [Clark, 2011]

The DNA strand is a chain of nucleotides. The phosphate group of one nucleotide link with the 3' carbon of another nucleotide (figure 7) through a condensation reaction, losing a water molecule in the process. More nucleotides are added in this formation, elongating the DNA strand.

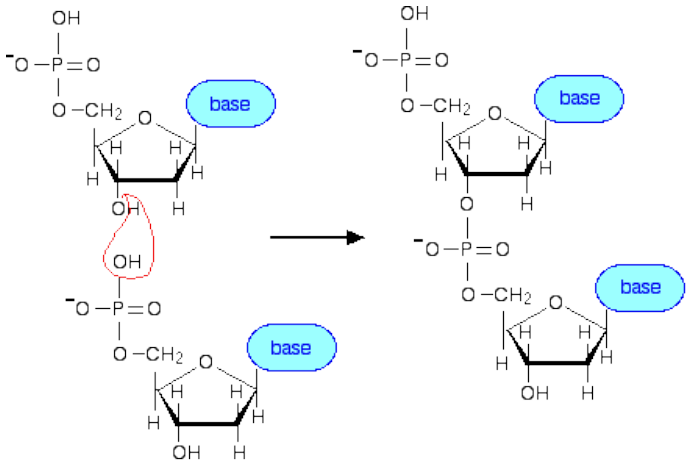


Figure 7: Nucleotides are connected through a condensation reaction

between the 3' carbon of one nucleotide and the phosphate group

of another nucleotide. Figure from[Clark, 2011].

A DNA strand is a double helix, meaning there are two strands of DNA connected together. There are two types of base pairs: Adenine (A) with Thymine (T), and Guanine (G) with Cytosine (C). The Guanine Cytosine base pair has three hydrogen bonds connecting it making it a stronger connection than the Adenine Thymine base pair since it only has two hydrogen bonds (figure 8).

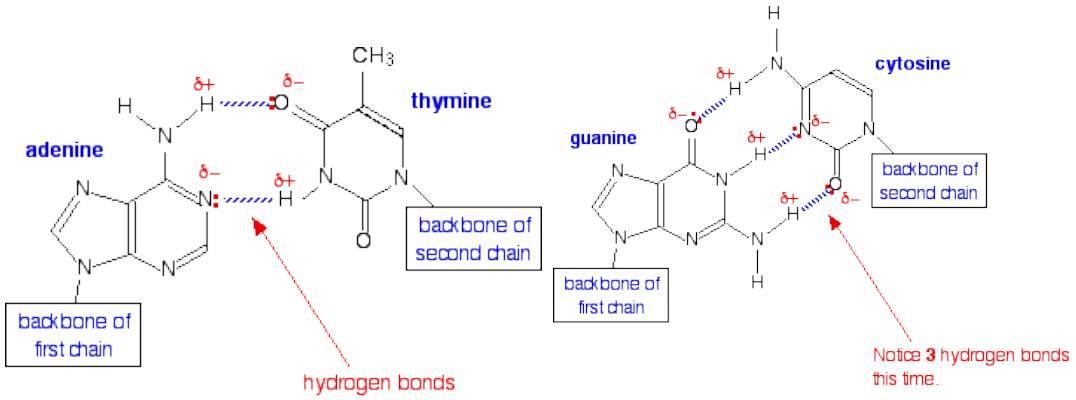


Figure 8: The creation of base pairs, cytosine with guanine and

adenine with thymine, through hydrogen bonds. Figure from [Clark, 2011]

The final structure of the double helix DNA strand appears as shown in figure 9. One strand is 5' to 3', while the other strand is antiparallel running from 3' to 5 '. In figure 9 the pink area is the deoxyribose, the yellow area is the phosphate group, the blue circles are the base pairs with the bonds connecting them. Because of the different size of the base pairs the DNA becomes twisted as can be seen in figure 10.

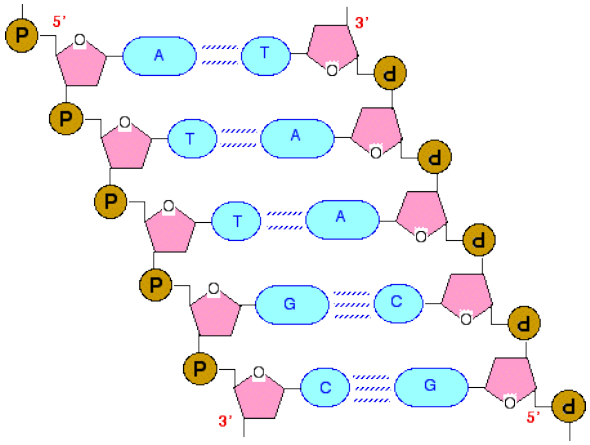


Figure 9: The structure of DNA, pink area is the Deoxyribose, the yellow circles are the phosphate groups, and the blue circles are the base pairs connected with hydrogen bonds. Figure from [Clark, 2012]

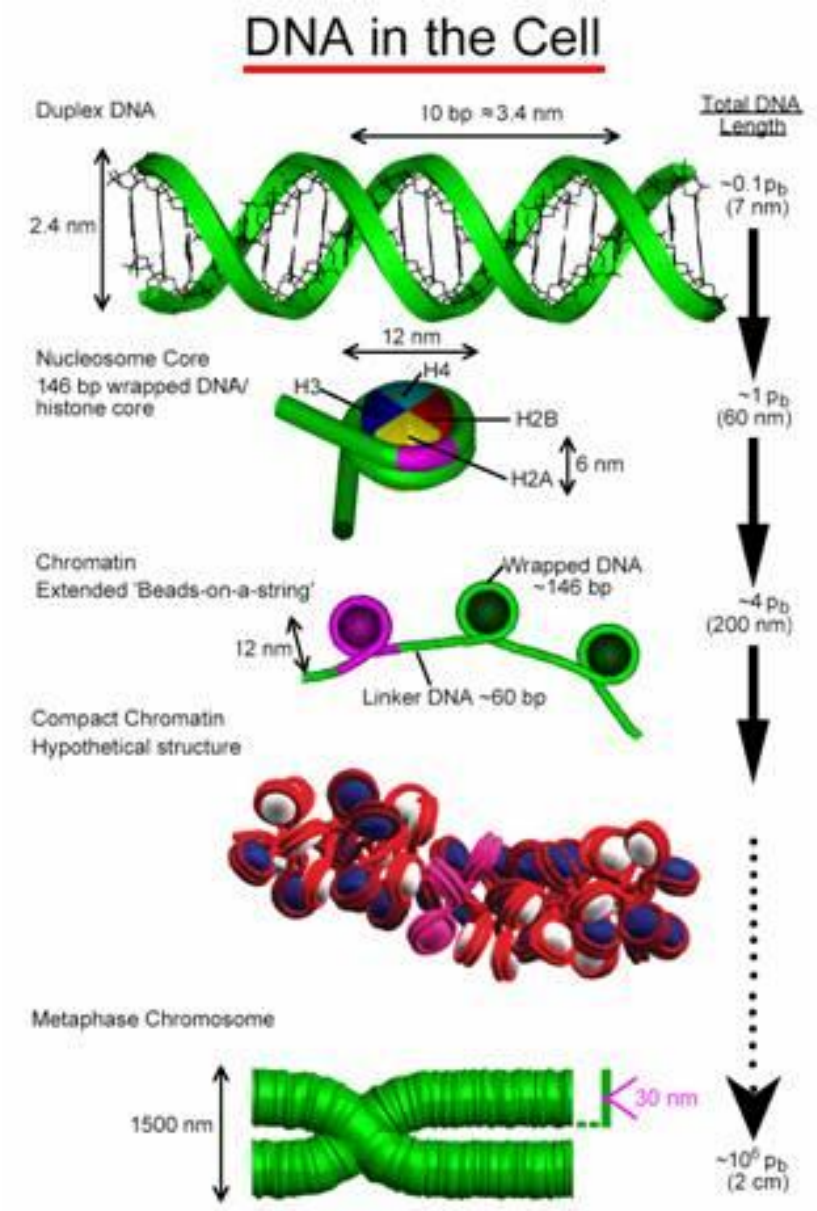


Figure 10: This diagram shows the chromatin being spun around the histones, and this then becoming folded up till it is a chromosome. First step shows a double helix DNA strand. The DNA strand then becomes wrapped around histones (H3, H4, H2B, and H2A are histones) creating a nucleosome in step 2. The wrapping of the DNA around the histones creates a "beads-on-a-string" appearance as shown in step 3. This then becomes tightly packed together (step 4) till it becomes a chromosome (step 5). Figure from [Schlick, 2011]

3.4 Gene expression

All cells of Eukaryotes organism has the same genome, the exact same building blocks. Though each individual cell has the same blueprint, they do not all function the same way. For example, some cells become nerve cells, other muscle cells, or even cells the immune system.

Even though there is such a wide range of possibilities for the cells, they are still made of the exact same DNA. The cells are differentiated by how the genome becomes expressed. The more differentiated a cell is the fewer of the genes on the DNA strand are expressed [Campbell, et al, 2011].

The human genome contains as many as tens of thousands of genes. Out of these thousands of genes from the genome, about 1.5% of bases in DNA code for proteins while others have jobs such as producing RNA, or are simply not transcribed at all. However it is estimated that only about 20% of protein coding genes are expressed at any given time [Campbell, et al, 2011]. The expression of genes can become regulated by the combined effects of DNA and histones modifications of structure.

Histones are the main protein components of chromatin. They act as a spool around which the DNA winds, compacting it into a nucleosome. Nucleosomes come together to form chromatin, as seen in figure 10 step 2. The histone protein has tails sticking out of it that can either be acetylated or unacetylated. A region of DNA with unacetylated histones becomes more compact than a region that has acetylated tails [Campbell, et al, 2011].

DNA wound around a histone, is called chromatin, seen in figure 10 step 2. Chromatin can be packed more or less tight around the histone causing it to be more or less expressed. The more tightly the chromatin is clustered, the harder it is for transcription factors to attach making it difficult to express the genes. If the histone unwinds the DNA, so that it becomes looser, the genes becoming more frequently expressed. Areas that connect the chromatin together (called nuclear lamina or chromosome scaffolding), tend to repress nearby gene promoters [Campbell, et al, 2011 p356].

A promoter is an area or region of DNA where RNA polymerase II can bind with the DNA to begin transcription of a gene [Campbell, et al, p352]. Promoters are known to contain specific DNA sequences and response elements that serve as 'molecular switches' in order to provide a secure primary binding site for RNA polymerase II and proteins known as Transcription factors. These factors tend to recruit RNA polymerase II and have specific activator sequences of complimentary nucleotides that clip to the specific promoters and regulate gene expression. A TATA box is often found in the promoter acting as a binding site for the transcription factors. The TATA box, the sequence 5'- TATAAA - 3' is usually found about 25 bases upstream from the

transcription site. It is considered a core part of the promoter. Promoters are located directly adjacent to genes in the upstream direction. Upstream and downstream are directions relative to the path of the transcription sector of a gene. Downstream is the region towards the 3' end of the DNA strand while upstream is towards the 5' end. [Clancy, 2008].

In order for a gene to become expressed it must go through the process of transcription and translation. Transcription is a 3-part process that involves initiation, elongation, and termination where the DNA becomes translated into a mRNA strand. A regulatory region of the gene, called the enhancer sequence, can help initiate transcription. The enhancer sequence, which can be located far away from the gene, is looped by the help of activator proteins (the red proteins in figure 11) in order to bring it closer. The looping effect can be seen in figure 11. Repressor proteins, the light blue protein in figure 11, do the opposite of the activator proteins and inhibit this looping of the DNA. In addition, transcription factors, which are proteins, bind to particular DNA sequences and recruit RNA polymerase II (the dark blue protein attached to the DNA in figure 11) to the promoter while controlling the output of genetic information from the DNA to mRNA strands. RNA elongation involves the addition of nucleotides to the RNA chain in the 5’ to 3’ direction. Termination of the transcription process occurs when RNA polymerase II transcribes a sequence called the polyadenylation signal sequence [Campbell, p334, 2011]. This sequence codes for a polyadenylation signal (the nucleotide bases AAUAAA) in the pre-mRNA. Downstream from this sequence, the pre-mRNA strand will be cut loose from the RNA Polymerase II [Campbell, p334, 2012] The RNA Polymerase II carries on transcribing up to 1000 or 2000 nucleotides past the point where the 3’ end of the matured mRNA will remain. The actual location for the 3’ is decided during RNA Processing [Winning, 2006].

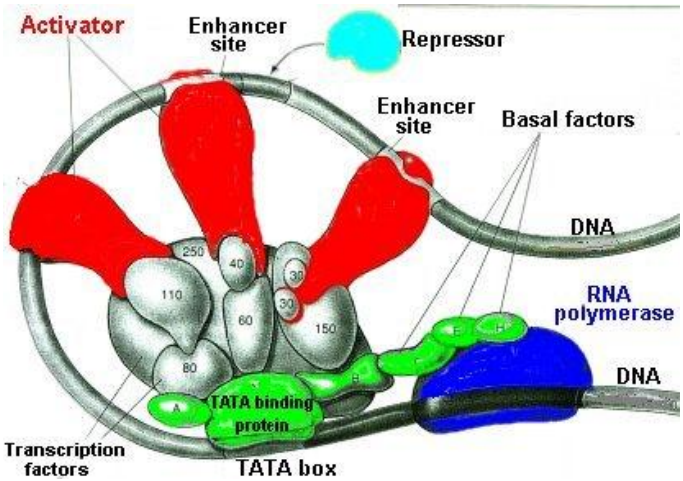


Figure 11: DNA transcription is helped to be initiated by the Enhancer sequence looping the DNA. The looping is done by the activator proteins, while the repressor proteins can prevent it. The DNA structure becomes altered (unwound) so that the RNA polymerase and transcription factors can attach in order to start the process of making a RNA strand copy of the DNA strand. Figure from [Holzer, G]

Once transcription is terminated, the newly formed pre-mRNA strand moves out of the cell nucleus and into the cytoplasm where translation will be completed. However, before it exits the nucleus it must go through RNA processing. During this process the pre-mRNA strand becomes edited. A 5' cap is added to the 5' end of the pre-mRNA and a poly-A tail is added to the 3' end of the pre-mRNA. The 5' cap is a modified form of the nucleotide guanine. A poly-A tail is a chain of adenine, ranging from 50-250 nucleotides that become added on to the pre-mRNA strand by an enzyme. These alterations play three important roles. First they help the mature mRNA strand move out of the nucleus. Secondly they protect the mRNA strand from degradation. And third, they help ribosomes attach to the 5' end of the mRNA strand once it has entered the cytoplasm.

The information of the pre-mRNA strand also becomes edited in a process called RNA splicing/processing. A normal gene in the human DNA is approximately 27,000 nucleotides long, resulting in a pre-mRNA strand the same length. However, 1,200 nucleotides is the size of the average sized protein of 400 amino acids [Campbell, p334, 2011]. To make one amino acid a triplet of nucleotides, called a codon, is needed. A pre-mRNA strand is also made up of noncoding segments of information, called introns. The exons, the coding information of a specific mRNA strand, are spread out between segments of introns. During RNA processing the intron segments become cut out using an enzyme such as snRNP and the splicesome. The exon segments then gets 'glued' back together creating a mRNA strand [Campbell, et al, 2011]. This allows for a larger amount of molecules that can become synthesized during the translation process.

After RNA processing the mRNA strand is able to exit the nucleus into the cytoplasm of the cell where the translation of it will take place. Translation is the process of synthesizing a polypeptide chain by decoding the mRNA into amino acids. The ribosome unit attaches itself to the mRNA strand providing a space for the Transfer RNA (tRNA) to match up with the mRNA. The ribosomal unite consist of two ribosomes, the large subunit and the small subunit. On the large subunit there are 3 sites that the tRNA is attached to, the A site (Aminoacyl-tRNA binding site) the P site (poptidyl-tRNA binding site), and the E site (exit site). The tRNA transfers amino acids found

in the cytoplasm to the translation process and matches the correct amino acid with the correct codon on the mRNA strand. A codon is read from the 5' to the 3; end. Each codon specifies which of the 20 possible amino acids is to be added. However there are 61 types of codons, meaning there are multiple codons coding for the same amino acid. No codon codes for the addition of more than one amino acid. The addition of the amino acid is done by having an anticodon on the tRNA. An anticodon is a triplet with complimentary base-pairs to a specific codon.

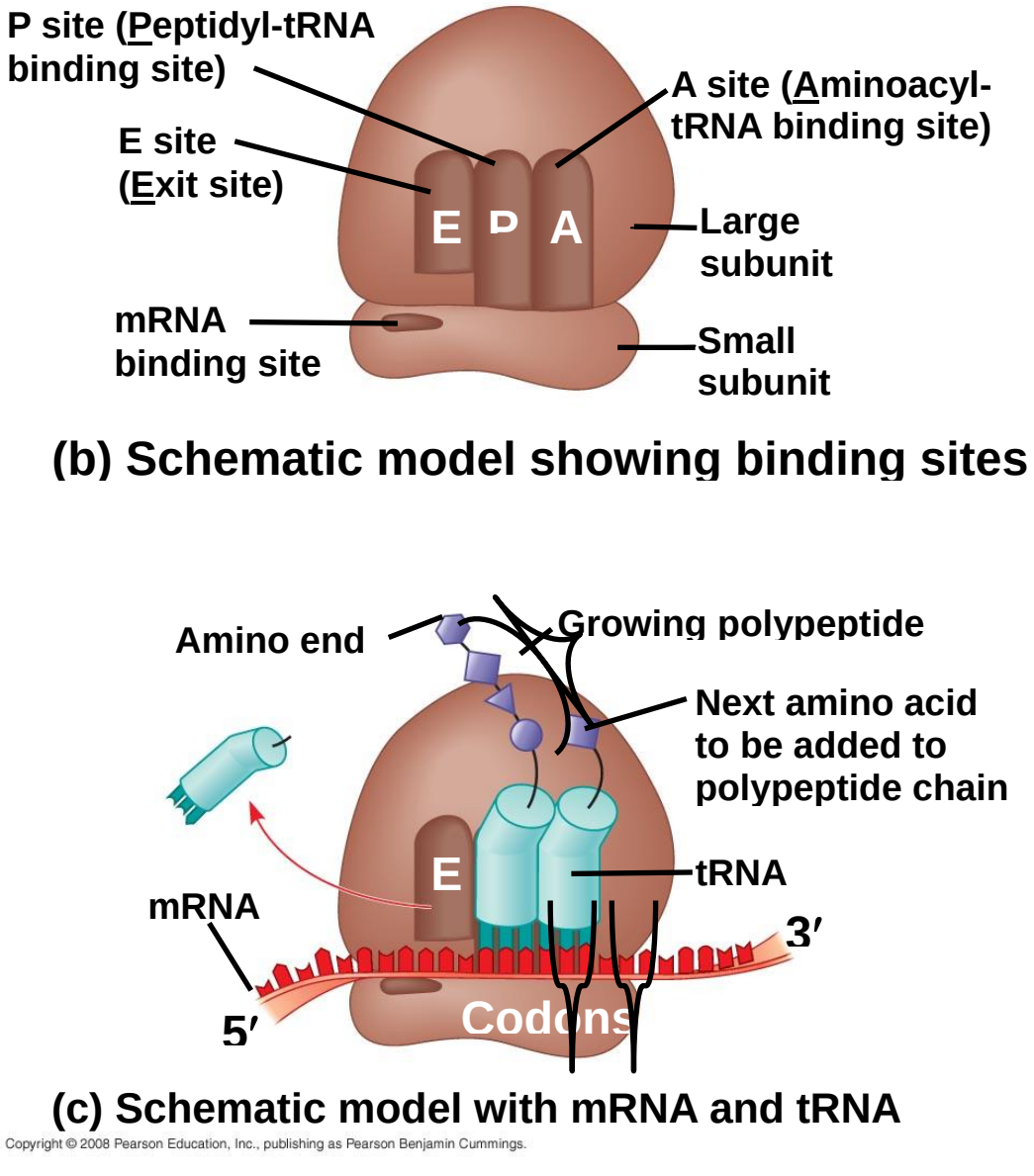


Figure 12: Translation of the mRNA molecule into a polypeptide. The tRNA anticodon matches up with the codon on the mRNA, bringing along the appropriate

amino acid. Figure from [Campbell, et al, p339, 2011]

The AUG, codon coding for Methionine, acts as a start codon (figure 13) for where the translation should begin. Figure 12 shows the translation of the mRNA molecule into a polypeptide. A tRNA first comes in at the A site and is thereafter moved to the P site. The next tRNA comes into the A site where a bond is formed between the amino acid (becoming a polypeptide) on the tRNA in the P site, shifting the polypeptide to the tRNA on the A site. Both the tRNAs are shifted so that the P site tRNA moves to the E site and exit the ribosome, and the A site tRNA moves to the P site, making room for a new tRNA to come in to the now free A site. The process then starts over. This builds a long chain of amino acids, resulting in a polypeptide. A stop codon (UAG, UUA, or UGA) does not code for an amino acid, but is a signal to stop the elongation of the amino acid chain (figure 13). Translation is ended when a release factor takes the place of a tRNA macromolecule, adding a water molecule instead of an amino acid. This reaction releases the polypeptide chain from the tRNA and the ribosomal unit [Campbell, et al, p336-342, 2011].

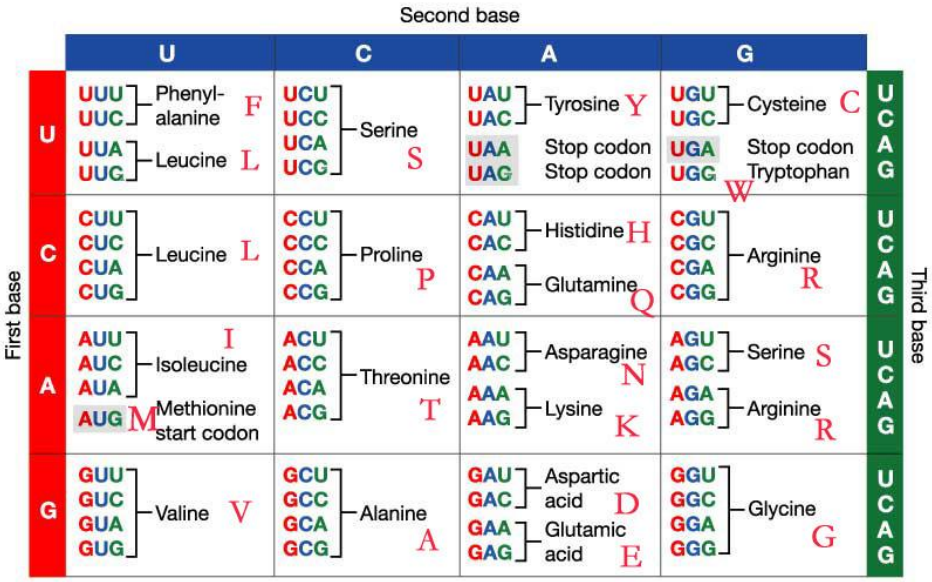


Figure 13: A codon table showing what specific codons code for. Figure from [Harrison, 2012]

The location of where the newly created polypeptide will go depends on which type of ribosome translated it. There are two types of ribosomes in a cell: free ribosome and bound ribosome. Free ribosomes and bound ribosomes are identical and can switch from being one type to the other. The free ribosmes are found in the cytosol of the cell while the bound ribosome are

attached to the endoplasmic reticulum (ER). The proteins that are synthesized by free ribosomes mostly function in the cytosol. Bound ribosomes create proteins of the endomembrane system or proteins that are transported out of the cell.

The expression of genes is very universal. The code for one type of protein is expressed the same way in almost all organisms. This means that it is possible to transfer the genes from one organism to another in order to give the second organism new traits. But since the genome expression is so universal the cells need a way of differentiating. If promoters were the only form of regulation, genes would have the same expression all the time, as it would then be encoded directly in the DNA. An expressed gene would remain expressed and an unexpressed would also remain unexpressed. Cells would be unable to change, become something else, like skin cells, liver cells or neurons, or regulate their function depending on their environment. This is where the epigenetics of the genome comes into play, enabling the cell to differentiate and perform select jobs.

An important part of understanding gene expression and cell differentiation is therefore to understand how promoters themselves can be regulated. Examples of mechanisms that control gene expression are chromatin folding and histone acetylation, but this report will focus on methylation of CpG islands.

Examples of gene regulation includes, E. coli bacteria that are able to digest both glucose and lactose, but prefer to digest glucose, when there is enough of it around. In situations when glucose become scarce, the relevant promoters are regulated, such that the genes for digesting lactose become expressed. They become suppressed again when glucose is available again. [Campbell, et al, 2011 pg 355].

4 Methylation and CpG islands

Joseph Ecker, a Salk Institute biologist and leading epigenetic scientist, offers an explanation of the epigenetic concept: "I can load Windows, if I want, on my Mac. You're going to

have the same chip in there, the same genome, but different software. And the outcome is a different cell type" [Cloud, 2010]. This programming, as Ecker suggest it to be, can be done by a simple unit in organic chemistry, the methyl group. A methyl group is a molecule made up of a carbon atom attached to three hydrogen atoms. This small molecule can attach to DNA, changing the phenotypical expression of genes by turning them off or on, or making its expression higher or lower. The process of adding the methyl group is called DNA methylation. A gene becoming methylated is associated with it becoming less expressed, the more methylation on the gene, the more silenced it will become. Removing methylation from gene can make it more strongly expressed. In mammals nearly all of the DNA methylation is found on cytosine bases [Goldberg, et al, p636, 2007]. Once a cytosine becomes methylated, it will not lose its ability to bond with Guanine [Campbell, et al, 2011]. A methylated cytosine to becomes a 5-Methylcytosine as shown in Figure 14. DNA methylation also tends to cause deacetylation of nearby histones, and thereby tightening the structure of DNA in that area which in itself, tends to make the genes in that area less expressed [Phillips, Ph. D, 2008].

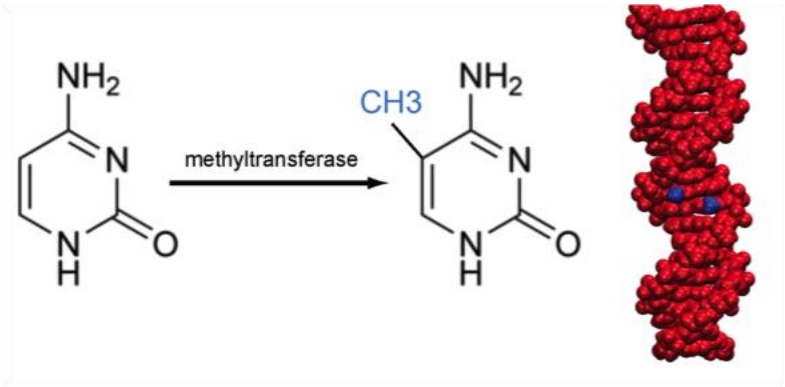


Figure 14: A Cytosine becomes methylated by a methyltransferase. Figure from [Xueqing, 2012]

The methyl group is added to the DNA and maintained by methyltransferase (DNMT). There are two types of methyltranferases, the novo DNMTs and maintenance DNMTs. The methyl groups can be added by de novo or maintenance DNA methyltransferases (DNMTs). Maintenance DNMTs add the methyl group to hemi-methylated DNA during the process of DNA replication. Hemi-methylated DNA is when only one of the two strands of DNA is methylated. This is important for regulating and protecting the DNA. The de novo DNMTs add the methyl groups after DNA replication [Goldberg, et al, p636, 2007].

The amount of methylation found on the DNA can change in response to diet, inherited genetic polymorphisms and exposure to environmental factors. Cells acquire the methyl groups from the organisms diet which is then donated through the folate and methionine pathways [Rodenhiser, Mann, 2006]. Imbalances in ones diet, such as low levels of folate, methionine or seleium, can result in hypomethylation and genetic instability. Hypomethylation is when most of the cytosine is unmethylated, leading to improper gene expression. Hypermethylation is when most of the cytosine has been methylated, resulting in silencing phenotypical traits that should be expressed.

Areas of the genome that have a tendency of being highly methylated are regions known as CpG islands. CpG islands are areas of the genome that have a higher occurrence of cytosine and guanine, more than 50% of the sequence. Their length ranges from 0.5 kb to 5 kb [Sahu, 2012]. Many of the CpG islands are sites of transcription initiation, often found near 5' end of the gene being transcribed [Leng, Zhao, 2009]. Approximately 70% of promoters are associated with CpG islands, meaning that this is the most common type of promoter in the vertebrae genome [Deaton, Bird, 2011]. About 50% of all CpG islands are associated with transcription start sites and promoter regions. However early studies suggest that these promoters located in the CpG islands often lack a TATA box along with other elements found in promoters. The other half of the CpG's are either within or between transcriptional units. These have been termed "orphan" CpG's due to the uncertainty of their significance [Deaton, Bird, 2011].

High concentrations of CpG's can also be found in CpG island shores, near areas of tissue-specific DNA expression. These areas have a slightly lower CpG percentage than the CpG islands. 85% of the CpG's are found in transcriptional silenced regions of the genome, such as transponsons [Fishbein, 133, 2012]. Transponson elements are DNA sequences that are able to change position within the genome. Though CpG islands and CpG island shores are areas of high concentration of cytosine and guanine, it is only 15% of the bases pairs in the genome [Pray, 2008].

The CpG islands usually contain promoters for housekeeping genes, along with a large number of tissue-specific genes and developmental regulatory genes. These regions often have a high chance of becoming methylated. Therefore methylation of CpG islands plays an important role in gene regulation and expression. In 2010 research by Illingworth and Maunakea also shows sections of CpG islands that are located far away from transcription start sites. Their research

suggested that these CpG islands still play a role as promoters [Deaton, Bird, 2011]. This shows an important link between CpG islands and transcription initiation.

Gene expression may be decreased by an increase in methylation. The silencing is thought to be due to inhibition of transcription factors binding by DNA methylation or silenced by methyl-binding domain (MBD) proteins that that calls to it chromatin-modifying activities to methylated DNA. It seems that CpG methylation is not the initiating factor in gene silencing, but helps to lock the silent state. For example, in x-chromosome inactivation in female mammals, the x-linked CpG methylation of the DNA does not happen until after gene silencing [Deaton, Bird, 2011].

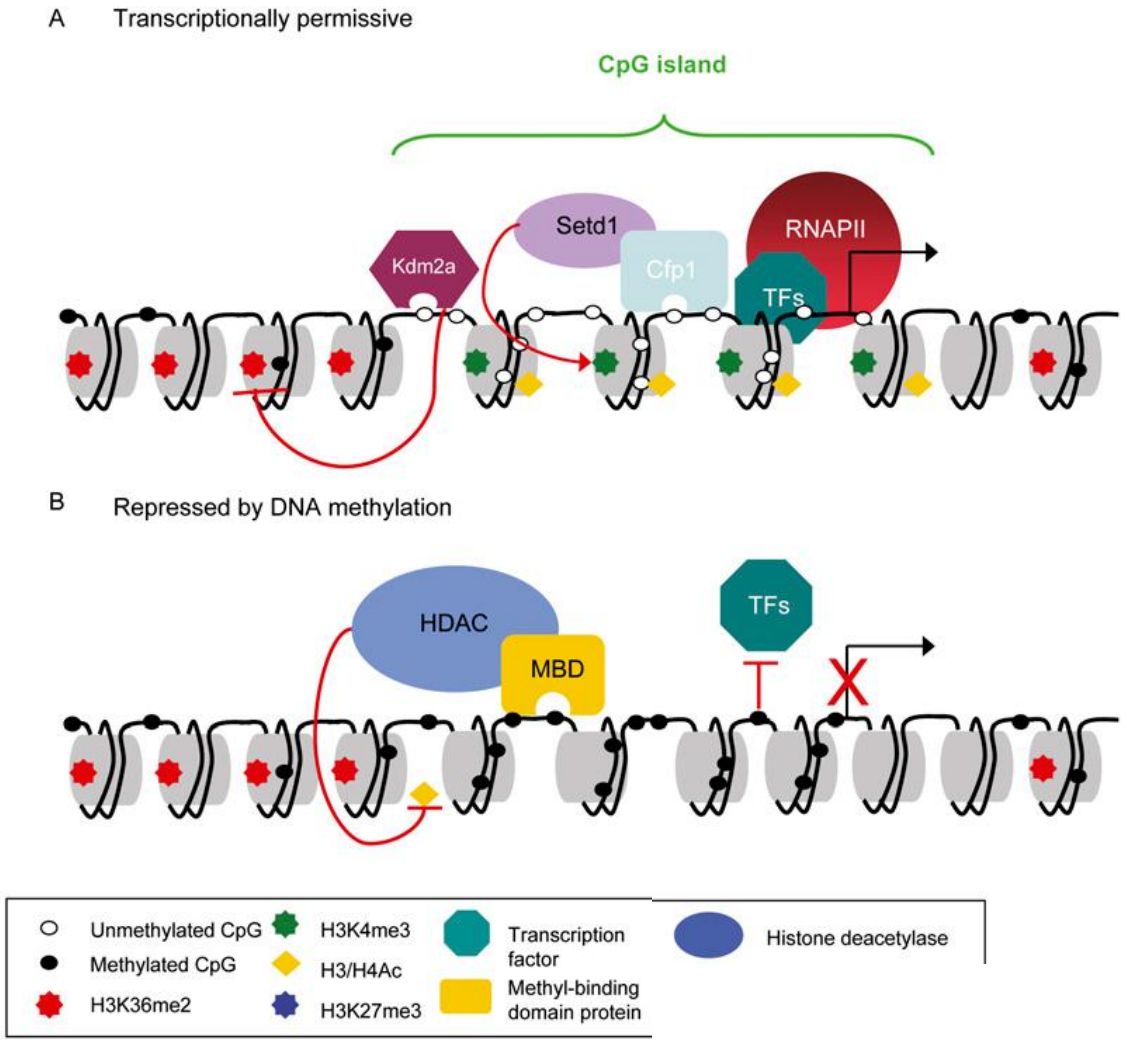


Figure 15: "The chromatin state at CGIs. (A) CGIs usually exist in an unmethylated transcriptionally permissive state. They are marked by histone acetylation (H3/H4Ac) and H3K4me3, which is directed by Cfp1, and show Kdm2a-dependent H3K36me2 depletion. Nucleosome deficiency and constitutive binding of RNAPII may also contribute to this transcriptionally permissive state. (B) DNA methylation is associated with stable long-term silencing of CGI

promoters. This can be mediated by MBD proteins, which recruit corepressor complexes associated with HDAC activity, or may be due to directed inhibition of transcription factor binding by DNA methylation." figure and discription from [Deaton, Bird, 2011]

Figure 15 demonstrates the role of a CpG island in transcription. Part A shows a unmethylated CpG island that will allow transcription to occur (the white circles on the DNA strand shows the unmethylated CpG, while the black shows the CpG has become methylated). Similar to the unmethylated state, this section of the histones are also not acetylated by the H3K36me2 (the red stars). Part B demonstrate how when the section of DNA becomes methylated the Transcription factor is no longer able to bind.

One of the ways that changes to the methylation pattern of CpG islands can occur is environmental factors. The normal methylation pattern of CpG islands to be changed through environmental agents such as metals (e.g. arsenic) and aromatic hydrocarbons (e.g. benzopyrene). Such environmental contaminates can be found in occupational chemicals, fossil fuel emission, contaminated drinking water and cigarette smoke [Rodenhiser, Mann, 2006]. The following section will be discussing studies of environmental effects on methylation.

4.1 From the environment to the genes

People have, up until recently, thought that our genes and DNA were set in stone from conception. Whatever we did to shorten our lives or make us fatter wouldn't affect our genes or DNA. This meant that our children would get a fresh start with their genetic makeup. However, studies have found that for instance, exposure to environmental toxicants such as the pesticide DEET, and dioxin plastics, as well as overeating and smoking cause trans generational alterations in the methylation of genes [Baccarelli, Bollati, 2010]. This leads to a key question facing modern biology: How, and to what extent, does the environment interact with the genome to generate different phenotypic outcomes?

While the effects of the environment on epigenetic changes have been extensively recorded, the actual pathways by which environmental information passes through to the epigenotype is far from understood and represent a field of epigenetics [Aguilera, 2010]. It is hypothesized that the pathways are operated by substrate-dependent activation or inhibition of

epigenetic enzymes. Namely, this could be eating food which is richer in methyl donors for DNA methyltransferases (enzymes which catalyze the transfer of methyl molecules to the DNA) [Brown, 2002] and histone acetyltransferase inhibitors or activators such as the anticonvulsant resveratrol valproate.

One of the first experiments to procure compelling evidence that genes and the environment are not mutually exclusive, but are synergistically linked, was in 1988. A British molecular biologist by the name of John Cairns had an idea for an experiment that would demonstrate that environmental responses can alter the expression of our genes. He proposed that "[w]hen populations of single cells are subject to certain forms of strong selection pressure, variants emerge bearing changes in DNA sequence that bring about an appropriate change in phenotype" [Finanzalarm, 2010]. To experimentally verify his theory, Cairns took E. coli bacteria that did not possess the genes that allow them to produce lactase, the enzyme used to break down lactose into simpler sugars that the bacteria can digest, and placed them in petri dishes where the only food available was lactose. Within a few days, the bacteria had settled in on all of the petri dishes and begun, against all expectations, to eat the lactose. To overcome the environmental conditions, the bacterial DNA adapted. The external environment caused an evolutionary trigger necessary for survival in this very primitive organism mediated by changes in the expression of their gene. [Articles.Mercola, 2011].

Sometimes epigenetic variations happen at random, equivalently to mutations. However, they are usually a response to environmental factors such as the food we eat, our exposure to pollutants, and our social and natural interactions. Medical research in the field of epigenetics is done to discover ways to prevent or reverse epigenetic happenings. It seems that epigenetics has a very strong potential that remains largely untapped, for medical advancement and it is possible epigenetics can help us understand larger-scale illness trends in our society.

Much of epigenetic research is conducted on rats. Long ago it was observed that baby rats deprived of their mother and nest for extensive periods of time exhibit stress related symptoms all their life. On the other hand, if rats that are regularly removed from their comfortable nest and mother for short periods and carefully taken care of by humans they have reduced stress responses compared to their sibling rats. This is attributed to the affection of the mother, who, upon their progeny's return licks them unrelentingly. The mother neglects to lick the infant rat, even treats it like stranger, if they have been set apart for a long time. When mother rats groom their pups by

licking, a sensory stimulation that controls and tempers the stress response affects the rat for life [Francis, 45].

Simultaneously, it was discovered that there are natural variances in the amount different mother rats lick their babies. A research team at McGill University lead by Michael Meaney determined that progeny of good lickers, in their adult age, have additional glucocorticoid receptors (GR) (where cortisol, an anti-stress hormone binds), in various parts of their brains, particularly their hippocampus, than if they had been mothered by a poor licker. Rats with increased levels of glucocorticoid receptors are superior at detecting cortisol, thus they are more capable of recovering from stress overloads [Francis, 44-47].

Seeing as these effects are present in adult offspring, they must have been a result of long-term adjustments in gene regulation. Interestingly, if the baby of a bad licker is put in the care of a good licking mother, its stress response system is more similar to that of its adopted mother than its biological mother. The biological progeny of inferior lickers raised by proficient lickers resemble the true biological pups of good lickers in every way, including the number of GR in the hippocampus. A reversed scenario is equally true. All of these experiments offer evidence of the direct relationship between a mother's care for her babies and the consequent stress levels, as seen in the GR levels, of adults [Francis].

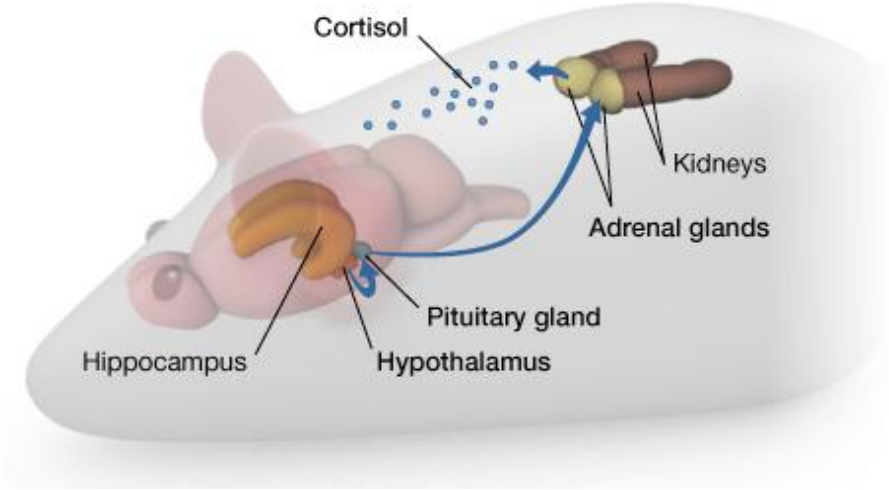


Figure 16: Rats stress response system Figure from [Lick Your Rats, Learn Genetics]

In figure 16, a rat is faced with potential danger. The hypothalamus sends distress signals to the adrenal pituitary glands, passed on to the adrenal glands. The adrenal glands begin to discharge cortisol (and adrenaline, not shown) into the bloodstream. When cells in the hippocampus detect the cortisol, which is now binding to the GR receptors, it communicates with the hypothalamus to end the stress circuit. The

result is the fight or flight response, as the cortisol is used as stored up energy [Lick Your Rats, Learn Genetics].

The change in GR levels persists for life. The cause of the long term changes in GR levels are due to a change in the GR gene itself. Upstream of the GR gene there is a "control panel" to the expression of the gene where the binding site of transcription factor NGFI-A (Nerver Growth Factor Inducible factor A) is located. When NGFI-A binds to the GR gene, it activates, increasing GR gene transcription. Levels of NGFI-A are significantly higher in baby rats who have been well licked than those who have been poorly licked. These differences in NGFI-A levels vanish as the rat develops into an adult, however. Therefore, the difference in NGFI-A quantity must permanently alter the responsiveness of the cortisol receptor gene in the brain by epigenetic means of modifications of the GR gene [Francis, p.45-46].

The differences in parenting in the rat world can be crucial for their survival. The research team at McGill set up open-field tests for rats with the purpose in mind of finding the long-term effects of good and bad parenting on behavior. The rats were put in a large box with the freedom to explore unhindered. Rats raised with bad parenting, portraying anxious behavior stayed close to the walls, timid and afraid, whereas rats with a healthy upbringing were curious and courageous and explored the box. In another test the rats were put into a cage with food. Predictably, the nervous rats would take more time than their counter-parts to feel secure enough before getting to the food, of which they ate less than the less stressed and more confident rats [Edublogs, 2012].

In certain environments certain parenting can mean the difference between life and death. A rat with a neglected upbringing who then displays insecure, anxious and wary behavior would fare better in an environment where food is scarce and danger high. A poorly nurtured rat is more inclined to keep a low profile and act faster to stress responses. The genes controlling stress hormone production were methylated to prepare the rat for danger. A rat that has not been epigenetically prepared by their environment might be too relaxed and unguarded to survive certain environmental situations but would fare better in a socially interactive rat community, for example, where it could get a higher social status and not have to worry about when the next meal is going to happen [Lick Your Rats, Learn Genetics].

However, stress related pathologies can happen in a much earlier phase of one's life. Simply put, what happens in the womb does not stay in the womb. For decades, doctors would administer synthetic cortisol, as another medicinal use of it is to promote maturation of the lungs, to fetuses in danger of being prematurely born, as respiratory failure is a serious risk to prematurely born babies. Doctors and scientists have discovered that there are a multitude of possible life-long stress complications that can be traced back to this treatment. From the fetus' perception, the synthetic glucocorticoid treatments are equivalent to if the mother was stressed. When a pregnant mother is stressed out, she naturally produces more cortisol to compensate for it than she would otherwise. Enough of the cortisol reaches the fetus through the placenta to permanently adjust the stress axis system of the fetus epigenetically to make it more susceptible and responsive to stressful situations. This is known as HPA (Hypothalamic-Pituitary-Adrenal) programming. A mother's stress could originate from a number of situations. To name a few, it might happen due to a breaking/broken marriage, life changes such as moving, or poverty. Fetuses that have been treated with cortisol show hypersensitivity in their stress axes and this can predispose them later to become particularly vulnerable to stress related pathologies such as diabetes and heart diseases [Francis, 41].

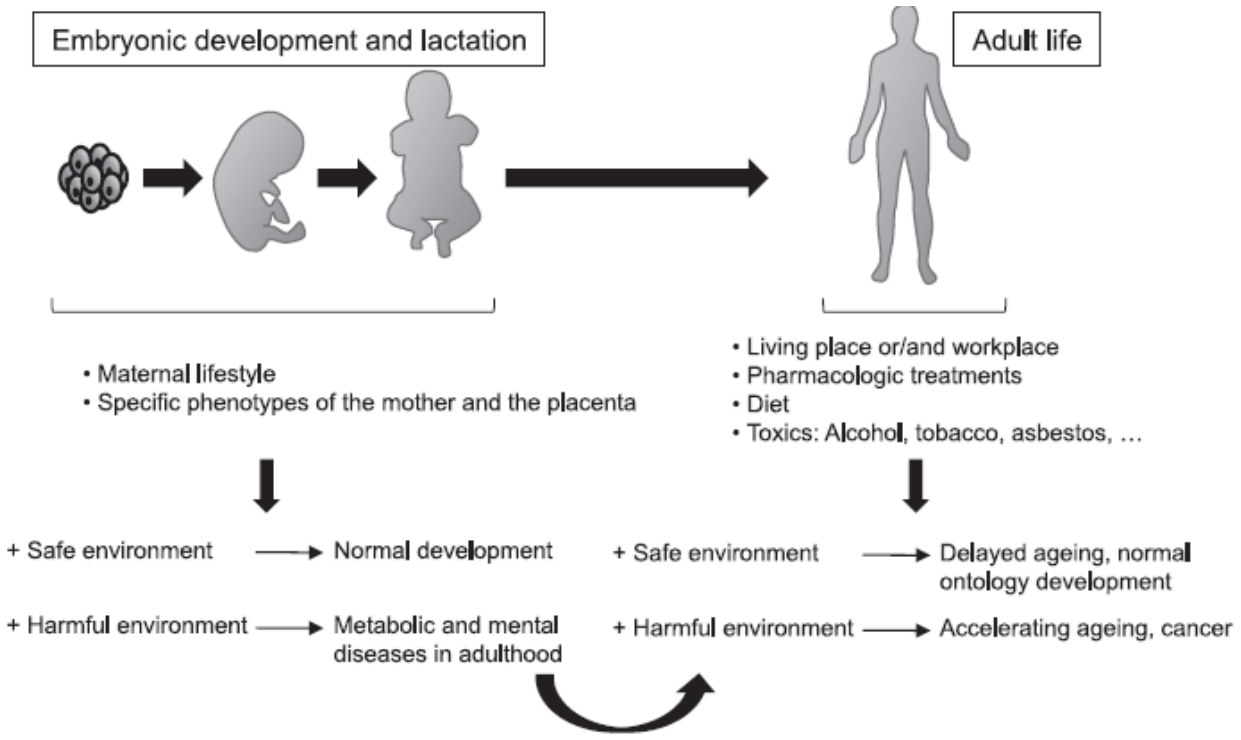


Figure 17: Embryonic development and lactation to adult life. Figure from[Aguilera, 2010]

It is important to distinguish between pre- and post-natal life as there is a large difference in the impact of epigenetic changes might bear on an organism. Specifically, epigenetic changes made during the embryonic stage have a larger impact on someone's epigenetic development because during mitotic division, the epigenetic status of one cell can be transferred over several divisions, resulting in one epigenetic instance that can exponentially impact many more cells. The embryo is therefore very vulnerable to the effects of a single epigenetic change as it can spark a chain reaction of epigenetic alterations [Aguilera, 2010].

Early in development, organisms are more predisposed to instability methylation-wise. In rats, the quality of a mother's care determines the methylation state of the GR gene. Good mothering promotes demethylation and bad mothering leads to methylation of the GR gene. When the GR genes are methylated it decreases the likelihood that the transcription factor NGFI-A can bind to it, causing fewer GR receptors to be produced in the hippocampus. As a result, the stress axis can become overwhelmed, which causes the rats to react anxiously and distressed [Francis].

There is much wisdom to cross over from the old axiom that as a twig bends, so the tree inclines. Changes that occur early in life can have long-lasting ramifications, as can be remarked by the epigenetic differences actuated by basic environmental influences. A walk in the forest will show tress of all shapes and sizes. Some trees start out growing sideways in one direction. Usually, this can be explained by some environmental obstacle, like another tree blocking the sun, or a rock blocking its path. Then it changes course completely, up to 90 degrees at times, toward the sun.

We are not helpless in reversing a negative epigenetic state. Those who are less epigenetically fortunate to start with can change course for the better. Despite that infant-parent interactions lay the groundwork for social behavior; it can be corrected later, as interactions with peers factor similarly into social and emotional development. This is also true with rats. Michael Meaney was able to reverse many of the negative side effects suffered by bad mothering by providing the rats with a socially rich and nourishing environment of same-sexed, healthy rats. With time, his subjects improved their social skills sufficiently for noticeable changes in their stress response. Meaney retested the rats to discover that the NGIA-F gene had been methylated accordingly. Despite their tendency to persist, it seems epigenetic modifications can be reversed [Francis, p. 73-74].

Knowing that epigenetic effects can be changed provides prospective benefits of epigenetic research into epigenetic inheritance. In humans, evidence suggests there is an inheritable epigenetic mark at a locus that promotes a certain type of colon cancer. In addition, smoking cigarettes can prompt the demethylation of metastatic cells of genes in the lungs [Ho, and Schmidlapp] Since it is only recently that scientists have begun to view inheritable diseases from an epigenetic point of view it is likely that with increased funding and research such cases of epigenetic inheritance will become understood and curable in the future.

We inherit more than our genes. The epigenetic code allows the genes a degree of flexibility that goes beyond the fixed ways of our DNA. Among what we can inherit extragentically is our social environment that starts with our progenitors. This allows for a kind of short-term specialization caused by a variety of information to be passed on to offspring without needing to experience the slow special transformation of natural selection or random mutations. Epigenetics is a revolutionary science that enlightens us to the fact that our genetic anatomies are an extension of our environments. We are still confined by certain restrictions of whatever regulatory and normal genes we possess. However, through environmental controls, we now have some ability to consciously guide the direction of our evolution. If environmental factors are managed we could promote certain epigenetic mutations in our genes over others

5 Perspective

5.1 Diseases

Mistakes and defects in the methylation of DNA (and histone modifications) can lead to several diseases and disorders [Rodenhiser, Mann, 2006]. In table 2 the normal functions of some epigenetic mechanisms control is shown, along with what happens when mistakes occur. Table 3 shows us the direct results of such mistakes in methylation of the DNA.

Table 2: Normal cellular functions regulated in part by epigenetics

and molecular abnormalities caused by epigenetic errors. Table from[Rodenhiser, Mann, 2006].

Normal functions	
• Correct organization of chromatin	Controls active and inactive states of embryonic and somatic cells
• Specific DNA methylation and histone modifications	Controls gene- and tissue-specific epigenetic patterns
• Silencing repetitive elements	Ensures that chromatin order and proper gene expression patterns are maintained
• Genomic imprinting	Is essential for development
• X chromosome inactivation	Balances gene expression between males and females
Abnormalities	
• DNA hypermethylation	Results in chromatin condensation and silencing of tumour suppressor and other genes
• DNA hypomethylation	Activates oncogenes, results in chromosomal instability, activates transposons
• Mutations at methylated cytosines	Results in inappropriate gene expression
• Imprinting defects	Results in loss of parental identity

Table 3: Relationship between human diseases and conditions with epigenetics

Disease/Condition	Gene	Biological process
Cancer		
Bladder	Multiple genes	Hypermethylation
Brain (glioma)	RASSF1A	Hypermethylation
Brain (glioblast)	MGMT	Hypermethylation
Breast	BRCA1	Hypermethylation
Breast	Multiple genes	Hypermethylation
Cervix	P16	Hypermethylation
Colon	Multiple genes	Hypermethylation
Colorectal	L1 repeats	Hypomethylation
Esophagus	CDH1	Hypermethylation
Head/neck	P16, MGMT	Hypermethylation
Kidney	TIMP-3	Hypermethylation

Leukemia	P15	Hypermethylation
Liver	Multiple genes	Hypermethylation
Lung	P16, p17	Hypermethylation
Lymphoma	DAPK	Hypermethylation
Myeloma	DAPK	Hypermethylation
Ovary	BRCA1	Hypermethylation
Ovary	Sat2	Hypomethylation
Pancreas	APC	Hypermethylation
Pancreas	Multiple genes	Hypomethylation
Prostate	BRCA2	Hypermethylation
Rhabdomyosarcoma	PAX3	Hypermethylation
Stomach	Cyclin D2	Hypomethylation
Thymus	POMC	Hypomethylation
Urothelial	Satellite DNA	Hypomethylation
Uterus	hMLH1	Hypermethylation
Neurologic		
Schizophrenia	RELN	Hypermethylation
Bipolar disorder	11p?	Unknown
Memory formation	Multiple genes	Hypo-, hypermethylation
Lupus	Retroviral DNA	Hypomethylation
Cardiovascular		
Atherosclerosis	Multiple genes	Hypo-, Hypermethylation
Homocysteinemia	Multiple genes	Hypomethylation
Vascular Endothelium	eNOS	Hypomethylation
Imprinting and pediatric syndromes		
PWS or AS	15q11 - q13	Imprinting
BWS	11p15	Imprinting
SRS	Chromosome 7	Imprinting
UPD14	14q23-q32	Imprinting
PHP, AHO, MAS	20q13.2	Imprinting
Rett syndrome	MECP2	Mutation
ICF syndrome	DNMT3B	Mutation
ATRX	ATRX	Chromatin structure
FraX	Triplet repeat	Silencing
FSHD	3.3 kb repeat	Chromatin structure
Reproductive		
Ovarian teratoma	No paternal genome	Imprinting
CHM	No maternal genome	Imprinting
BiCHM	Maternal genome	Imprinting
Aging	Chromatin	Hypo-,hypermethylation

Note: PWS = Prader–Willi syndrome; AS = Angelman syndrome; BWS = Beckwith–Weidemann syndrome; SRS = Silver–Russell syndrome; UPD14 = uniparental disomy 14;PHP = pseudohypoparathyroidism; AHO = Albright hereditary osteodystrophy; MAS = McCune–Albright syndrome; ICF = immunodeficiency, centromeric instability and facial anomalies; ATRX = a-thalassemia/mental retardation syndrome, X-linked; FraX = Fragile X syndrome; FSHD = facioscapulohumeral muscular dystrophy, CHM = complete hydatidiform mole, BiCHM = familial biparental CHM. [Rodenhiser, Mann, 2006]

5.1.1 Genomic imprinting

Genomic imprinting is a way for the genes to "remember" if they were inherited from the mother or father. This is done so that only the maternally or paternally inherited allele is expressed, and not both of them at the same time. This process is controlled by DNA methylation and histone modification. Developmental and pediatric disorders can develop due to defects (seen in table 3). Prader-Willi and Angelman syndromes are caused by genetic and epigenetic error to the same part of chromosome 15; errors inherited from the father results in Prader-Willi syndrome, and those from the mother results in Angelman syndrome. Beckwith-Wiedemann syndrome is caused by genetic or epigenetic mutations resulting in loss of imprinting on chromosome II [Rodenhiser, Mann, 2006].

5.1.2 Pediatric syndromes

Other than alterations of epigenetic components, mutations can also occur that can affect epigenetic components. Mutations in these areas can be the cause of several syndromes: $DNMT_3B$ in the ICF (immunodeficiency, centromeric instability and facial anomalies) syndrome, $MECP_2$ in Rett syndrome, ATRX in ATR-X syndrome (α-thalassemia/mental retardation syndrome, X linked) and DNA repeats in facioscapulohumeral muscular dystrophy (table 3) [Rodenhiser, Mann, 2006]. For example, in Rett syndrome $MECP_2$ (a specific gene) encodes a protein that binds to methylated DNA. If mutations occur in this area, this protein will cause abnormal gene expression within the first year of life. For girls this means reduced brain growth, loss of developmental milestones and mental disabilities. Severe developmental disabilities is also found in those with ATR-X syndrome due to a lack of the ATRX protein. ATRX is a protein that helps maintain a section of the DNA, keeping it condensed and inactive [Rodenhiser, Mann, 2006]. Alterations of specific genes and chromosomal regions needed for the expression of proper neurologic and physical development can therefore lead to a collection of pediatric syndromes.

5.1.3 Cancer

The complex process of developing cancer is a combination of genetics and epigenetic errors building on one another. This results in cells that invade body parts or develop into metastatic tumor cells. Changing the amount of methylation on specific genes associated with cancer can result in developing a wide range of related diseases (table 3). Hypomethylation can activate oncogenes and creates instability in the chromosome. Hypermethylation sets off tumor suppressor genes. The types of cancers that can develop depends greatly on which genes are affected. For example, hypermethylaton on the promoter on the gene pI6 occurring in a range of degrees (9%-49%), may show up in as many as 15 different cancer types. Hypermethylation of the gene BRCA1 is mostly found in sporadic breast and ovarian cancers (10%-20% of cases). By being able to recognize these changes to the epigenome, scientist are able to diagnose various cancers [Rodenhiser, Mann, 2006].

5.1.4 Aging

Though aging is not a disease, with an older age, often comes diseases associated with it. In studies, increases and decreases in DNA methylation have been seen in the aging process. This suggest that the methylation or demethylation is also age-dependent, and chances are that it is the cause of the development of neurologic disorders, autoimmunity and cancer in elderly people [Rodenhiser, Mann, 2006]. The higher risk of cancer in the elderly could be related to inactivation of cancer associated genes such as tumor suppressors. In other areas of the genome, the levels of methylation of cytosine's decreases as the cells age. This demethylation of these tissues could promote restructuring, instability and rearrangements in the DNA. This increases the risk of neoplasia. Neoplasia is an abnormal new growth of cells that could lead to tumors. Increase in methylation in hypermethylated areas of the genome may also be a precursor event to the increased risk of such cancer with the increase in age such as colon cancer.

5.1.5 Immunity and related disorders

In order for an immune response to take place, epigenetic changes to the cells must take place. Acetylation and methylation of the DNA allows the immune systems cells to carry out a immune response to specific pathogens, and continue to have this recognition over multiple cell generations. Loss of control of these epigenetic controls can result in autoimmune disorders such as Lupus. Lupus is a disease where the immune system becomes hyperactive and attacks healthy cells in the body. An abnormal amount of DNA methylation has been seen in patients with this disease, the T cells showing a decreased extracellular signal-regulated kinase pathway signaling, decreased methyltransferase activity and hypomethylated DNA. Due to the decrease in methylation, leukocyte function-associated factor (LFA_I), and other immunity factors become over expressed [Rodenhiser, Mann, 2006]. Alterations of LFA_I is also found in the development of arthritis, which suggest that alterations to DNA methylation also plays a role in other diseases showing idiopathic autoimmunity [Rodenhiser, Mann, 2006].

5.1.6 Neuropsychiatric disorders

Epigenetic errors have also been shown to play a role in complex adult psychiatric, autistic and neurodegenerative disorders (table 3). Schizophrenia and mood disorders has been associated in many studies with disorder in the DNMT genes. For example, in schizophrenic brains there has been found a selectively over expressed DNMTr gene in gamma-aminobutric acid (GABA) - ergic interneurons. Hypermethylation, repressing the expression of the protein Reelin (which is required for normal neurotransmission, memory formation and synaptic plasticitya) has also been found in patients suffering from schizophrenia, bipolar illnesses and psychosis [Rodenhiser, Mann, 2006].

5.2 Possible treatments

Currently there are not many epigenetic therapies available. Several are in the clinical trial stages, and only few have been approved for specific cancers. Nucleoside analogues, such as azacitidine, are incorporated into replicating DNA. This addition inhibits methylation and is able to reactivate methylated silenced genes. The azacitidine treatment has been found effective in treating myelodysplastic syndrome and leukemias that are caused by hypermethylation (seen in phase I of clinical trials) [Rodenhiser, Mann, 2006]. The results showed that 54% of patients with chronic myelogenous leukemia that was resistant to imatinib (a type of cancer treatment especially used for cancers testing Philadelphia chromosome positive) showed a partial or complete hematologic response. A complete hematologic response is a when blood cell counts return to normal. 46% of the patients had a cytogenetic response (the intensity of the response varying) to 5-aza-2'-deoxycytidine [Rodenhiser, Mann, 2006]. A cytogenetic response is when no Philadelphia (PH) chromosome can be detected in the bone marrow or in the blood.

Another possible treatment currently being tested is the antisense oligonucleotide MG98 that downregulates $DNMT_I$ (the enzyme that adds the methyl group to the DNA). The MG98 treatment is showing promising results in phase I clinical trials in targeting solid tumors and renal cell cancer. Analysis of tumor cells from biopsies of head and neck cancer, that has been treated with MG98, showed demethylation of 2 methylated tumor suppressor genes had occurred. Methylation of an oncogene, a gene that can potentially cause cancer, had also occurred [Rodenhiser, Mann, 2006].

A epigenetic treatment being studied is a small molecules such as valproic acid that downregulated HDACs in order to induce growth arrest and tumor cell death. By combining epigenetic therapies (demethylating agents plus HDAC inhibitors) or epigenetic therapy followed by conventional chemotherapy (or immunotherapy) scientist may be see more effective results since they reactivate silenced genes, including tumor suppressor genes, resensitize drug-resistant cells to standard therapies and at the same time acts to kill cancer cells [Rodenhiser, Mann, 2006]. Future challenges will be to limit the toxic effects the treatments can have on normal cells, and to ensure that the drug effects targets only the tumor cells.

5.3 Biotechnology Applications

Biotechnology is essential when it comes to being able to detect methylation of DNA. Detection of methylation can become a future possible method for detecting disorders such as cancers. The applications used in biotechnology allow scientists to study the sequence of genes, determine how genes are expressed and how methylation can be detected. Scientists these days rely on computational techniques for DNA sequencing such as the Next-Generation sequencing (NGS) or the Sanger sequencing. These sequencing techniques allow quick DNA sequencing of an entire genome and are capable of creating hundreds of gigabytes of DNA into a digital format, thus determining the order of the bases in a strand of DNA. Specific regions in a genome can also be targeted to be sequenced when applying modern day technologies such as the (NGS). DNA cloning plays a vital role where restriction enzymes plus DNA ligase make Recombinant DNA. Gene expressions are measured by DNA microarrays by which various sorts of expression are attained by an array of spots. Bisulfite conversion differentiates methylated and unmethylated sequences, where a double stranded DNA is treated by sodium bisulfite, thus converting unmethylated cytosine to uracil residues while methylated cytosines are not changed.

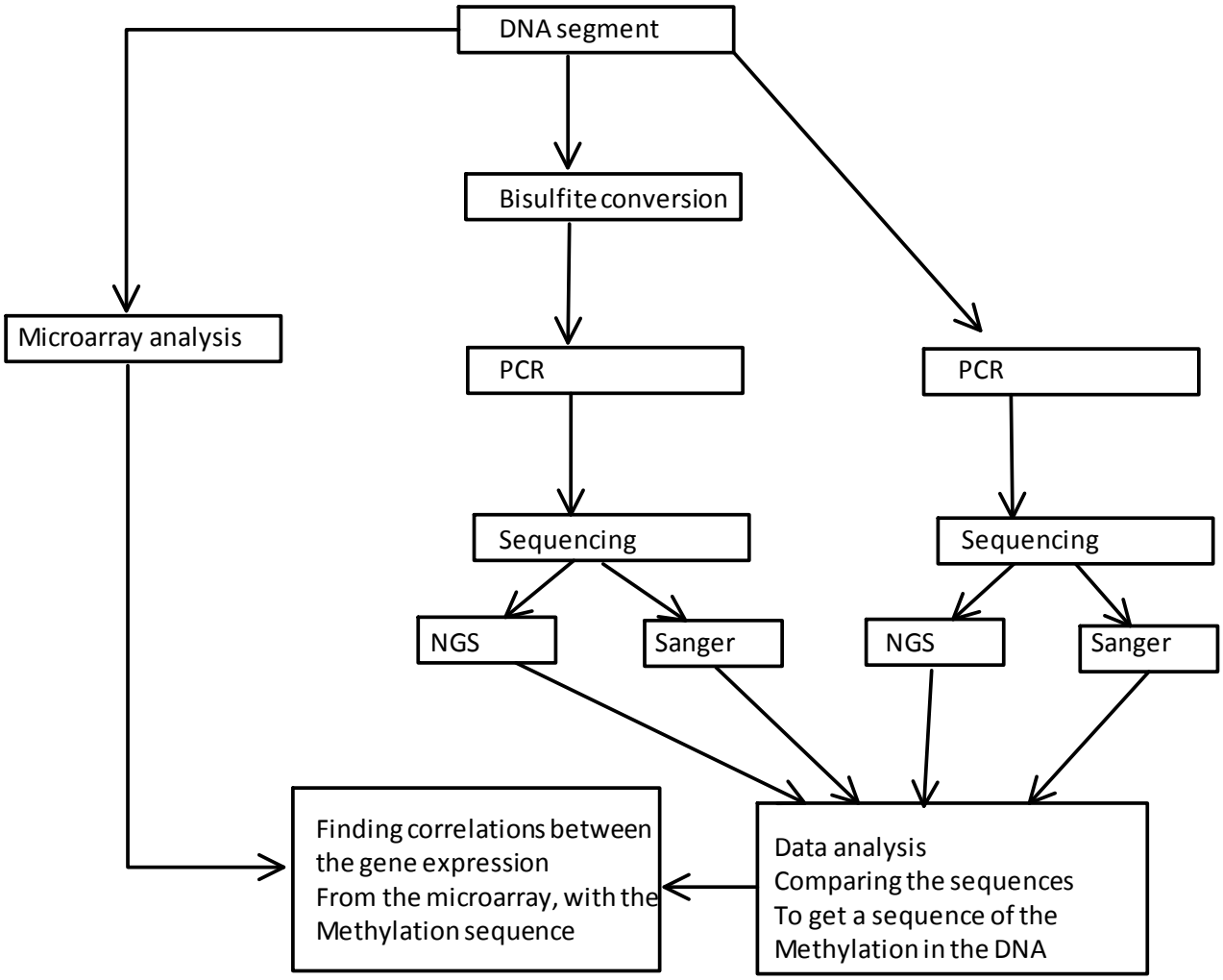


Figure 18: Flow chart of how biotechnological applications to detect methylation

Understanding the flowchart: When sequencing DNA in order to obtain the sequence of methylation, you need to sequence it twice (the middle and right track in fig. 18), once where you just sequence the DNA to get the original sequence - and once where you first treat the DNA with bisulfite, which converts unmethylated cytosines into uracil, following PCR. When you the do analyze the data you can see where the differences are, and that gives you a sequence of the methylation of the DNA. The left track of the flowchart is a microarray analysis, which is a way of collecting data about gene expression. It is then possible to find correlations between the expressed genes and the sequence of methylation.

Figure 18 shows how the biotechnological applications explained in this chapter are used in order to detect methylation.

5.3.1 DNA Cloning

DNA Cloning is a method where multiple copies of an isolated DNA fragment are created. DNA cloning is used in numerous techniques, where it is possible to clone an entire gene fragment or a particular DNA sequence. DNA cloning plays a vital role in various methods such as making recombinant DNA by the help of restriction enzymes and DNA ligase. The Polymerase Chain Reaction (PCR) is an advanced form of DNA cloning where limitless copies of an isolated DNA fragment are created in just a few hours. DNA cloning is being used in various fields, in the field of forensic science to solve crime, viruses, studying cancer cells and in numerous biological studies.

5.3.2 Restriction enzymes

Since the finding of Restriction enzymes or Restriction endonucleases in the late 1960s [Campbell, et al, 396-411], researchers have been able to use these restriction enzymes in the procedure of Gene cloning and Genetic engineering by cutting DNA molecules in specific places.

They have been able to identify hundreds of different kinds of restriction enzymes and have been able to isolate them. The enzymes play a specific role in recognizing a certain short DNA sequence called a restriction site. Both DNA strands are then cut off at exact points within the restriction site.

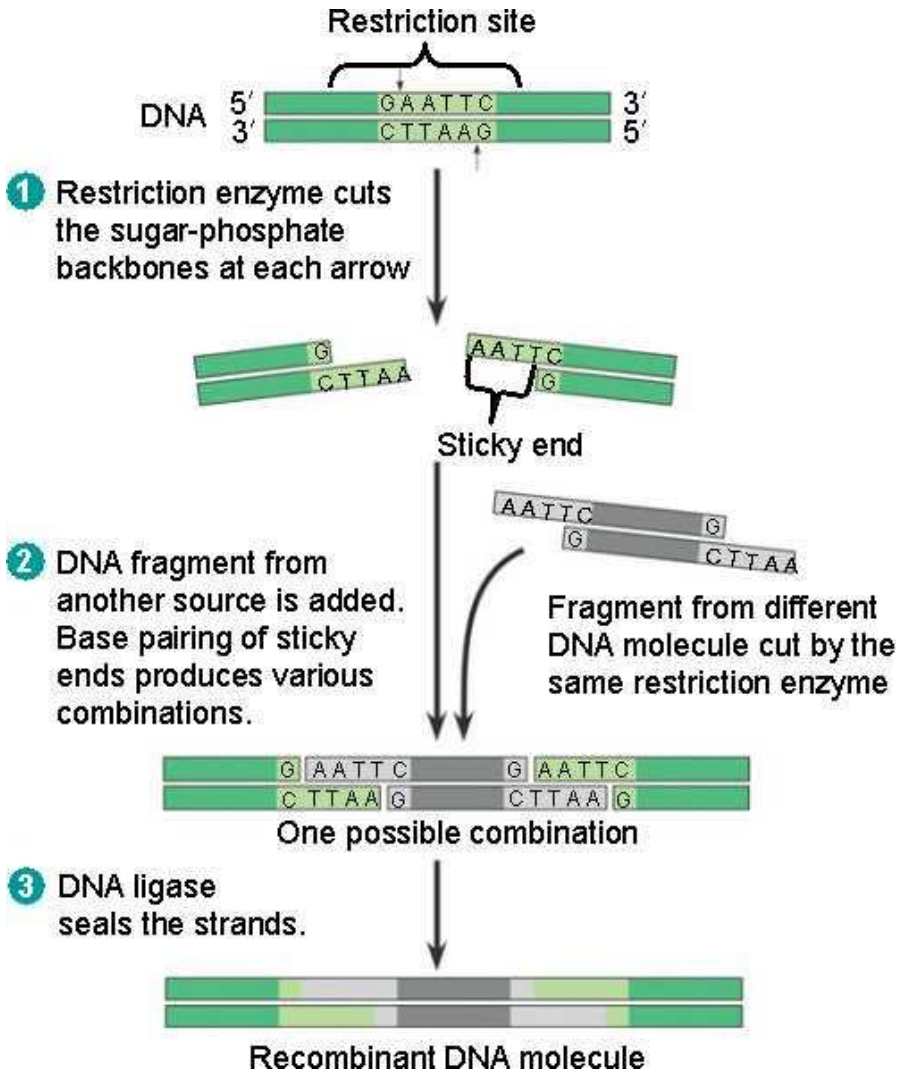


Figure 19: Restriction enzymes. The process of adding a new segment of DNA into a gene using restriction enzymes. Figure from [Campbell, et al 396-411]

Restriction enzymes and DNA ligase are used to make recombinant DNA as shown in the figure 19 [Campbell, et al 396-411]. In this example, a restriction enzyme recognizes a restriction site, containing a six-base pair sequence and thereafter cuts the sugar phosphate backbones in a reproducible way within the sequence, thus producing fragments with sticky ends. A restriction enzyme cuts a long DNA molecule multiple times producing a set of restriction fragments. By cutting different DNA molecules by the same enzyme, fragments are produced. Thereafter an enzyme called DNA ligase seals the strands to make the possible combination permanent and thus creating a final product known as a Recombinant DNA. Genomic templates are

cut off with ‘Methylation-Sensitive Restriction Enzymes’ that recognizes methylation patterns in cytosine residues of CpG sequences.

5.3.3 Polymerase Chain Reaction (PCR)

The Polymerase Chain Reaction also known as (PCR) is a method developed in 1985 where specific target segment inside one or numerous DNA molecules can be rapidly increased [Campbell, et al 396-411]. With the help of mechanization, PCR is able to produce limitless copies of a target segment of DNA in just a few hours and is certainly faster than many other procedures, where the same number of copies takes many days to be made. PCR skips the step of making a library by making a particular DNA fragment that is then inserted into a vector. A vector is used as a vehicle for carrying foreign DNA into a host organism for making copies and isolating the DNA.

The (PCR) method includes a three step cycle where in the first step; a DNA sample is taken and mixed with chemicals, thereafter by heating up the reaction mixture which denaturizes the DNA strands. In other words when the DNA is heated up, it allows the enzymes to break it down. The reaction mixture is then cooled down which allow primers to create hydrogen bonds with ends of target sequence. This step is known as annealing. Finally in the last step which is called extension, DNA polymerase will extend the primers in 5’ → 3’ direction. It adds nucleotides to the 3’ end of every single primer as shown in the figure 20. [Campbell, et al, 396-411]

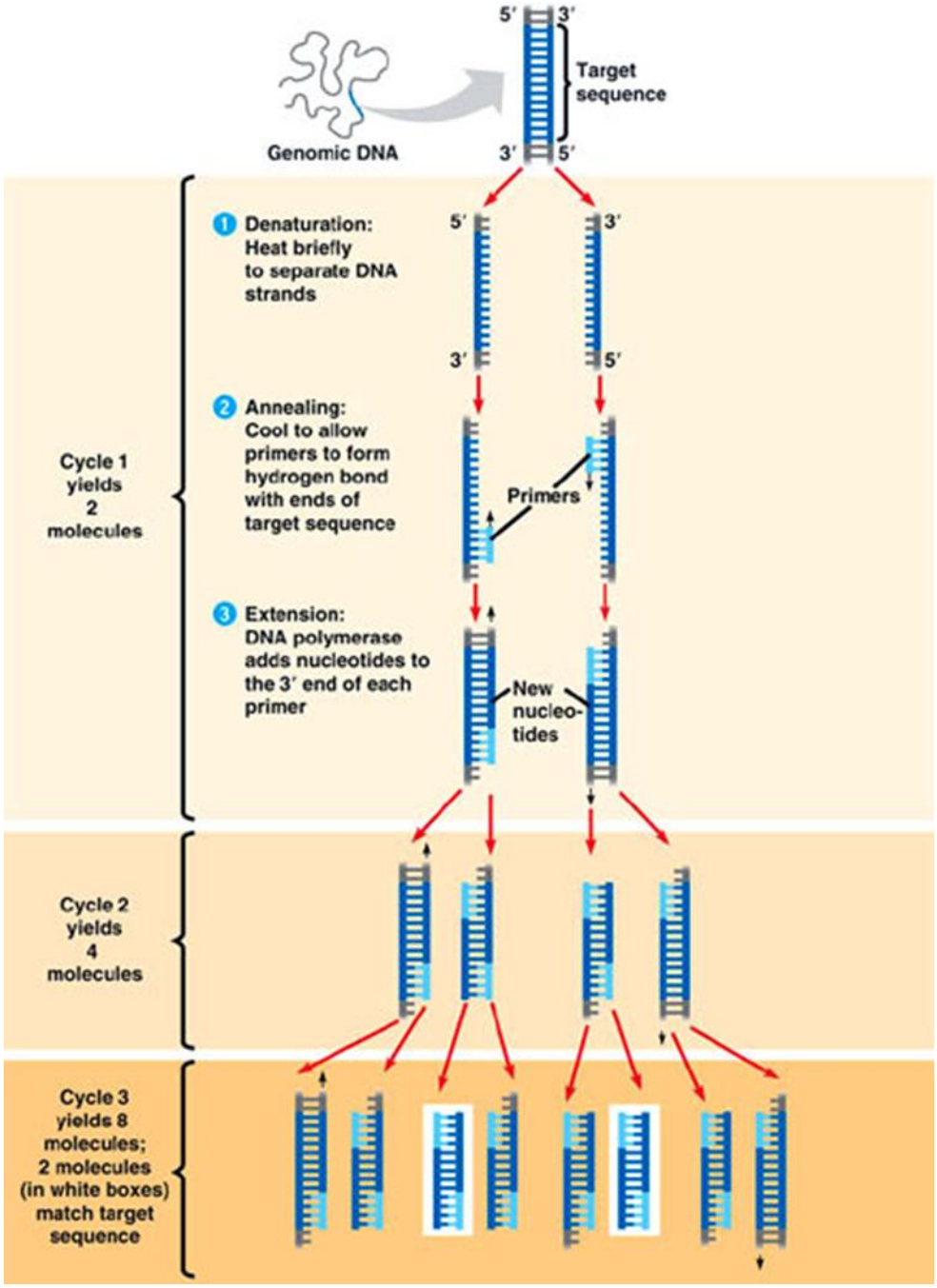


Figure 20: PCR. Figure from [Campbell, et al, 396-411]

Since the discovery of PCR, it has been widely used in criminal investigation cases, where fragments of blood, tissue or semen found at crime scenes has proved to be helpful in solving crime. PCR is also used to piece back together the DNA of an ancient fossil or from a 40,000 year old frozen mammoth. [Campbell, et al, 396-411]

5.3.4 Bisulfite Conversion

There are several different methods of molecular biology that allows examining methylation patterns in the whole genome and in particular genes. By these means we can thus differentiate methylated sequences from unmethylated ones where the genomic templates are cut off with Methylation-Sensitive Restriction Enzymes or by Bisulfite Sequencing.

Bisulfite Conversion is a method which consists of treating double stranded DNA with sodium bisulfite. This converts unmethylated cytosine to uracil residues while methylated cytosines are not changed as shown in figure 21.[Schmacher, Alex.]. Figure 21 shows the chemical reaction of Bisulfite Conversion.

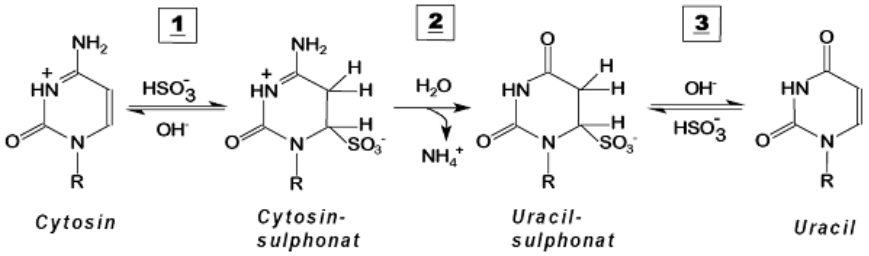


Figure 21: Bisulfite Conversion reaction. Figure from [Schmacher, Alex.]

To differentiate between methylated and unmethylated sequences, a sequencing investigation is carried out where the DNA is first amplified by PCR (Polymerase Chain Reaction) with primers. Uracil matches to adenine and when analyzing the process, adenine matches to thymine in its base paring behavior. By comparing the sequences, the methylation profile of the sample is exposed. Thereby the methylation status of a specific region of DNA is accomplished, as shown in the figure 22.

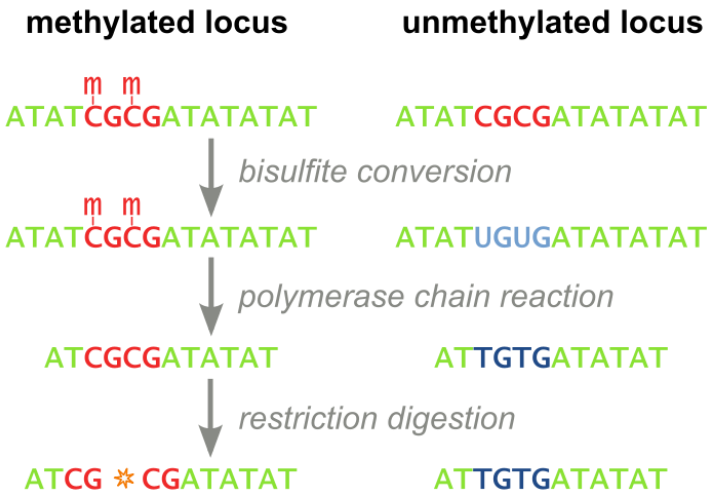


Figure 22: Bisulfite Conversion and the detection of methylated and unmethylated DNA.

Figure from [Hendricksen, 2012]

5.3.5 DNA Sequencing

DNA sequencing is a procedure that locates the exact sequence or order of the nucleotide bases in a DNA molecule. DNA sequencing converts the DNA into a digital format, thus determining the order of the nucleotide bases, which is then used by researchers in various studies such as medical research and forensics.

In DNA sequencing, the DNA is removed from the cell and then it undergoes through a DNA sequencing process where the order of the nucleotide bases are located and identified [Fine, and Wilborn].

Whole Genome Sequencing refers to a technique in which the complete DNA sequence of a living organism's genome is determined at a single time. Whole genome sequencing using 'Capillary Electrophoresis' also known as (CE)-based Sanger technology takes a lot of time

and resources to be completed, whereas compared to the ‘Next-Generation Sequencing’ (NGS) technology, it can be completed just in a few days in a single run.

Over the past few decades DNA sequencing has become essential for almost every branch of biological research. Many sequencing technologies have been introduced but with the arrival of Capillary Electrophoresis also known as (CE)-based Sanger sequencing, researchers are able to obtain genetic information from various organisms.

The Sanger sequencing technique was first developed by Fred Sanger, where he was able to decode the human genome. This is also known as the Dideoxy or Chain Termination technique where a DNA template is copied over and over again to decode the sequence of Adenine, Guanine, Cytosine and Thymine nucleotides in a DNA strand. By adding modified nucleotides which are also called Chain Terminators, the reaction stops copying, thus producing fragments which vary in length and by analyzing these fragments the original sequence can be read.

The Sanger technique, requires preparation of a mixture that contains a single stranded DNA, DNA polymerase, the four deoxyribonucleotides (A, C, G and T), and a single primer. The primer used in Sanger sequencing has a nucleotide sequence matching to the 3' end of the region to be copied so that the DNA polymerase can start the copying process. The mixture prepared earlier, is equally distributed in four test tubes. By inserting a different dideoxyribonucleotide into the growing chain of each tube the copying process carries on, however terminating the strand synthesis. Thereafter the substances of each tube are placed into four lanes of an electrophoresis gel, separating the oligonucleotides by size and nucleotide types. As the shortest oligonucleotide moves at the bottom of the gel, the correct DNA sequence can be read, from the bottom to top order.

The process is widely used in laboratories around the world. However it has limitations. Therefore scientists are not able to attain the information they require, though it is still used as it is capable of sequencing longer fragments. This led to the development of a new technique-‘Next-Generation sequencing’ also known as (NGS). This new technique has a different approach of sequencing and opened up gates for researchers to find many ground breaking discoveries. This new technology also gave researchers and scientists the ability to extract genetic information from biological systems and thus uncovering the full DNA sequence of any organism.

Next-Generation technology is somewhat similar to the Sanger technology where small fragments of DNA are recognized by signals being produced as every single fragment is re-synthesized from a DNA template strand ["Illumina" 1-12]. However the (NGS) technology extends the process of sequencing through limitless reactions in an enormously parallel manner, rather than being limited to just one or a few DNA fragments.

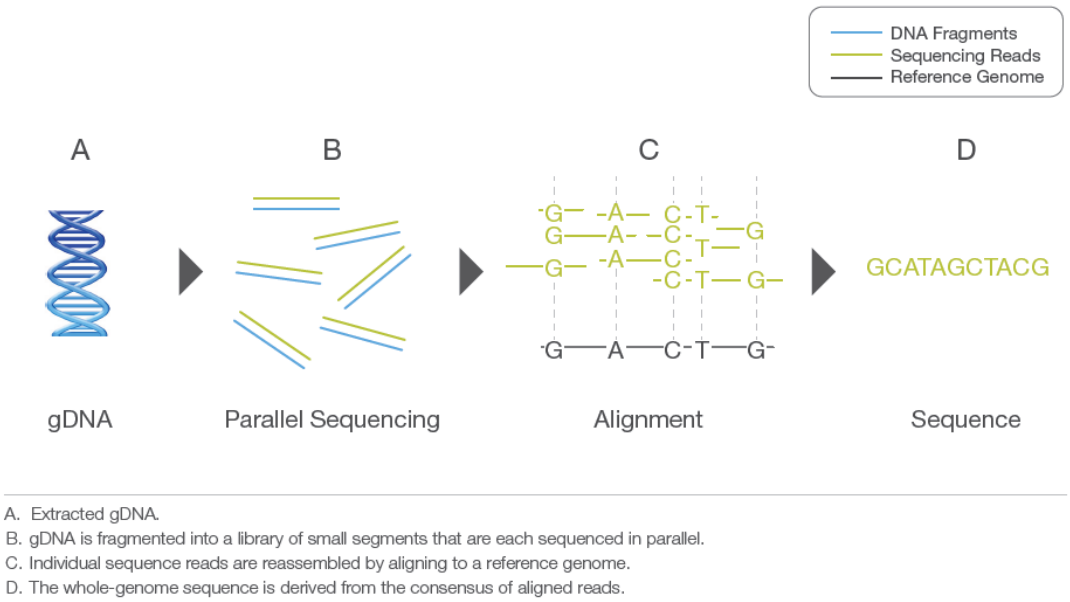


Figure 23: Next-Generation Sequencing. Figure from ["Illumina" 1-12]

‘’The (NGS) technology allows quick sequencing of large stretches of DNA base pairs spanning of entire genomes and is thus capable of creating hundreds of gigabases of DNA (gDNA) samples.’’ Source ["Illumina" 1-12]

This process is half biological and half computational. A single genomic DNA (gDNA) sample is taken and is fragmented into a library of small segments which is then sequenced countless times in a similar manner. New short fragments called reads are reassembled either in the presence or in the absence of a reference genome. When using a reference genome, the sequencing process is repeated all over again as it requires less reads. This process is called Re-sequencing. In the absence of a reference sequence, a process called De-novo sequencing is applied by assembling sequence reads into longer contiguous. DNA sequencing technology does not have the ability to read whole genomes in a single run, therefore small pieces of 20 to 1000 bases are read and these

short fragments are thus called ‘Reads’. Thereafter, the entire sequence of the gDNA sample is exposed by alligned reads as shown in the figure 23 ["Illumina" 1-12].

By using targeted sequencing, researchers are able to target a specific region of interest. Researchers are able to sequence defined regions in a genome. And by these means, they can save time, focus reads and store data, when applying modern day technology such as the Next-Generation Sequencing technology on regions they are most likely to be interested in.

5.3.6 DNA Microarray Assay of Gene Expression Levels

Scientists are now aiming to understand how genes behave, produce and uphold a functioning organism. So far researchers have succeeded in sequencing entire genomes of numerous organisms and are able to study the expression of genes. They are also looking upon groups of genes that are expressed in a coordinated manner and by these means they are able to identify networks of gene expression of an entire genome.

Such genome-wide expression studies are obtained by a technique where mRNAs are isolated from a tissue sample or a specific cell type. Figure 24 illustrates the process of DNA Microarray. By means of reverse transcription, mRNAs are then used as templates to create cDNAs using fluorescently labeled nucleotides. Thereafter the cDNA mixture is applied to a DNA microarray where cDNA hybridizes with DNA fragments representing the genome. It is placed on a microscope slide where copies of DNA fragments of the organism’s genes are set in place in each spot or attached at fixed locations [Campbell, et al 396-411]. The DNA microarray then scans for fluorescence at each spot which represents the gene expressed from the tissue sample by indicating a specific color at that particular spot.

Using this technique, researchers are able to test numerous genes at once to determine which genes are expressed from which particular tissue as well as in many disease conditions and can thus determine the coordination of gene expression.

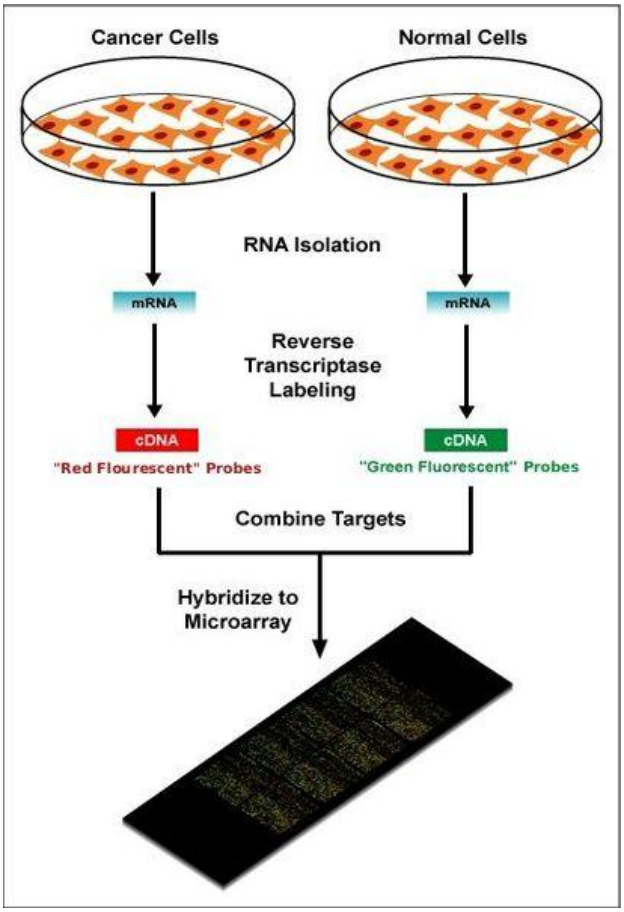


Figure 24: DNA Microarray. Figure from [*Microarray-schema, 2012]*

DNA Microarray consists of small solid supports where numerous or thousands of genes after undergoing sequencing are attached at fixed spots or locations [Microarray, 2007]. The concentration of fluorescence at every single spot measures gene expression in the sample presented, thus indicating that particular spot with colors such as red, green, black or yellow. Normally two different samples are tested in a DNA Microarray. For example, where cell type 1 contains a healthy cell and cell type 2 contains a diseased cell. Thereafter the cDNAs prepared from each sample is labeled, cell type 1 in green and cell type 2 in red. The resulting color at each fluorescence spot will then indicate gene expression levels in both samples. The color green would indicate gene expression in cell type 1 sample, and red in cell type 2. The color yellow indicates expression in both samples and the color black indicates in neither of the two samples [Microarray 2007]. Figure 25 shows the colors that can be seen when using DNA Microarray assay of gene expression levels.

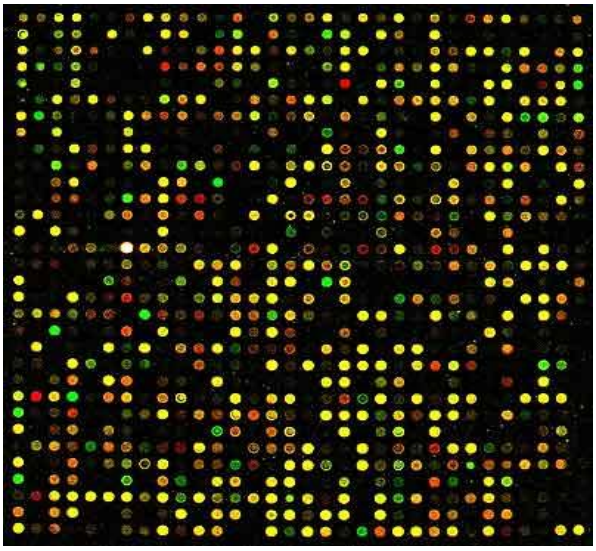


Figure 25: DNA Microarray Assay of Gene Expression levels. Figure from [‘’NTNU Genomics Resource Center’’]

By the help of DNA microarrays, researchers are able to understand diseases such as cancer in a better way and are able to recommend proper diagnostic techniques in the future.

6 Case Studies and Discussion

Knowing that the environment can have a strong effect on development can help people strive for a better life by doing simple things. An example of this is an experiment by Larry Feig on the environmental influence on memory of mice. Feig decided to test if the memory of the mice could be improved by simple environmental influences. A group of mice with genetic memory problems were exposed to an environment rich with toys, exercise and extra attention. An improvement in long-term potentiation (LTP) was found in these mice. LPT is a form of neural transmission that is important to memory formation. The offspring's of the mice from the experiment also showed an improvement in LTP, even though they were not given the extra attention their parents had gotten meaning that the epigenetic traits were passed on to the next generation [Cloud, 2012].

Similarly another study of mice by Dr. Randy Jirtle and Robert Waterland was performed in 2003. The focus of the experiment was a specifically regulated agouti gene that gives

mice a yellow coat and tendency of becoming obese and developing diabetes when the gene is expressed continuously. One group of the pregnant agouti mice were given a diet rich in vitamins B (folic acid and vitamin B12). The B vitamins act as a methyl donor, resulting in the methyl group attaching more often to the agouti gene in utero. The other group of pregnant mice were given a normal diet with no extra vitamin B [Cloud, 2012].

The mothers that expressed the agouti gene and where given the supplement with vitamin B produced healthy brown pups (not expressing the agouti gene). The pups had a healthy weight and did not have a higher risk of developing diabetes [Cloud, 2012]. The DNA had not been changed, but the addition of the methyl groups by the vitamin B silenced the agouti gene.

Both of these mice studies performed, the LTP and the agouti gene, shows that the environment can have an significant effect both for the live currently being lived and for the future generation. Larry Feigs experiment shows that an early introduction to a stimulating environment promotes the production of LTP that is essential in memory formation. Surprisingly this increase in LTP was also found in the next generation of the mice from the study, though they were not given any special treatment. Taking this into consideration, this study can influence the environment that humans should introduce their children into. It suggest that if a child is introduced to a stimulating environment it will likely develop a better memory.

Dr. Randy Jirtle and Robert Waterland study of the agouti gene in mice shows an important concept of how medical treatments could work using methylation to block or activate specific genes. The question with this type of treatment becomes if it is able to affect a specific gene. By increasing the amount of methyl groups available, will the genome wide methylation amount increase, potentially silencing necessary genes? Also is there a difference in the reaction humans and mice have to the methylation? It can be difficult to answer the question of comparing human and mice since it is considered unethical to perform similar experiments on humans. Therefore case studies are the most common type of research done with humans and their epigenome. Dr. Bygren and his study of an isolated region in Sweden, a human case studies, showed results that the way people live can also affect their health and the health of the future generations, similarly as had been seen in mice experiments. Dr. Bygren did future studies within epigenetics in order to understand the extend that the environment could effect phenotypical expression.

Dr. Bygren continued his research on epigenetics and enviromental effects after his study of Norrbotten. He then discovered a paper written in 1996 by Dr. Marcus Pembrey. Dr. Pembrey, a committed Darwinist, speculated that what if the industrial age created new environmental pressures and social changes so extreme that evolution had to occur faster. Pembrey therefore suggested in his article in the Italian journal *Acta Geneticae Medicae et Gemellologiae* that the DNA now had to react within "a few, or moderate number, of generations" [Cloud, 2012]. Pembrey did not know how to test this theory, but suggested that the genome would not have time to change, but the epigenome did.

Bygren contacted Pembrey and together they started to suggest experiments that could test this theory. They would not be able to repeat the study Bygren had done on Norrbotten due to the unethicality of starving some children and over feeding others. Pembrey had for a long time been on the board of the Avon Longitudinal Study of Parents and Children (ALSPAC), a research group that provided a large amount of genetic information extending through generations. Founded by Jean Golding, ALSPAC has followed thousands of young people and their parents since the children were born around the time of 1991 and 1992. A group of 14,024 of pregnant mothers, mostly from the Bristol area of England, was recruited for the study. Every year the parents and children have undergone extensive medical and psychological testing [Cloud, 2012].

Now along with the help of Dr. Golding, Pembrey and Bygren were able to publish a study in 2006 in the European Journal of Human Genetics. The study focused on 14,024 fathers from the ALSPAC database, 166 of them having started to smoke before the age of 11 (just in the beginning of puberty) [Cloud 2012]. In boys, the period around puberty is an vulnerable time for epigenetic changes on the Y chromosome since the sperm is just starting to form.

Their study showed that the sons of these 166 early smokers had a significantly higher body mass indexes than the other boys by the age of 9. This could suggest that that those with a father who started smoking in the early stages of puberty has a higher chance of becoming obese and developing other health problems well into adulthood. These findings could also suggest that the sons of the early smokers could have a shorter life span, similar to the children of the overeaters in Bygren's Norrbotten study. "The coherence between the ALSPAC and Overkalix (Bygren's Norrbotten study) results in terms of the exposure-sensitive periods and sex specificity supports the hypothesis that there is a general mechanism for transmitting information about the ancestral environment down the male line," Pembrey, Bygren, Golding and their colleagues concluded in the

European Journal of Human Genetics paper. Their study suggest that decisions made early in one's own life can result in causing potentially damaging health effects to once children [Cloud, 2012].

7 Conclusion

This report is an introduction into DNA methylation and the cellular mechanisms associated with it such as gene expression. We focus on how DNA is methylated and how it affects phenotypical expression. In addition we have investigated how the environment affects our epigenetic make up, how diseases can be caused by epigenetic changes and how methylation of DNA can be detected.

Methylation of DNA contributes to gene silencing can contribute to a variety of diseases when mistakes or abnormalities occur. Therefore understanding methylation, how it influences gene expression and how it can be detected is important in order to understand and possibly in the future cure diseases as well as understanding what effect the environment have on us and how that could cause us to reconsider some of our behavior.

In recent years scientists have observed the direct relationship between exposure to environmental chemicals and social influences on epigenetics. They examined the environmental effects of the Swedish famine in the early 20^{th} century to prove that epigenetic changes are transmitted transgenerationally. Meanwhile, growing evidence exists that epigenetics can be used to predict which type of environmental exposures could lead to individuals to be more predisposed to develop certain diseases and pathologies.

The ability to detect the specific changes of methylation in the DNA, has been greatly improved in recent years, due to the development of techniques such as Bisulfite conversion combined with gene sequencing and modern day techniques such as Next-Generation sequencing. It is likely that both price and time required for such sequencing techniques to be carried out in the laboratory and will drop further in the future, making methylation detection a potentially important diagnostics tool.

What does all of this mean for the people of today? With the Human Genome Project completed in 2010, scientists have now started on the Human Epigenome Project. The Human Genome Project mapped 25,000 genes. However in the Human Epigenome Project scientist will have to map the 25,000 genes of every type of cell, which are at least 210 types. By creating a map of the epigenome of healthy and diseased cells, this information can be used in developing new epigenetic therapies. As we further our understanding of the field, we will uncover the mechanisms by which epigenetics plays a role in the development of diseases and pathologies.

8 Reference list

Aguilera, Oscar. "Epigenetics in Health and Disease." Epigenetics in Health and Disease. (2010): n. page. Web. 3 Dec. 2012

Baccarelli, A., and V. Bollati. "Environmental epigenetics." *Nature*. N.p., 24 2010. Web. 22 Oct 2012. <http://www.nature.com/hdy/journal/v105/n1/abs/hdy20102a.html>.

Bartoshesky, Louis. "The Basics on Genes." *Kidshealth*. N.p., n.d. Web. 7 Nov 2012. <http://kidshealth.org/teen/your_body/health_basics/genes_genetic_disorders.html

Bouvior, Nicolas. *Epigenetics*. 2006. Science in School, Paris, France. Web. 3 Nov 2012. <http://www.scienceinschool.org/2006/issue2/epigenetics>.

Campbell ,Reece, Urry, Chain, Wasserman, Minorsky, and Jackson. *Campbell Biology*. 9th. United States of America: Pearson, 2011. Print.

Clancy, S. (2008) DNA transcription. Nature Education 1(1)

<http://www.nature.com/scitable/topicpage/dna-transcription-426>.

Clark, Jim. "DNA Structure." *Chem Guide*. Jim Clark, Nov 2011. Web. 7 Oct 2012. <http://www.chemguide.co.uk/organicprops/aminoacids/dna1.html>.

Cloud, John. "Why Your DNA Isn't Your Destiny." *Time Magazine* . 06 2010: n. page. Web. 18 Nov. 2012. <http://www.time.com/time/magazine/article/0,9171,1952313,00.html>.

Deaton, A., and A. Bird. "CSH Press: Genes and Development ." *CSH Press: Genes and Development* . 25 (2011): 1010-1022. Web. 17 Oct. 2012.

D. Goldberg, Aaron, C. David Allis, and Emily Bernstein. (Epigenetics: A Landscape Takes Shape) "Cell." *Cell*. 128. (2007): 635- 638. Print.

File:Microarray-schema.jpg. 2007. Wikipedia Web. 27 Nov 2012. <http://en.wikipedia.org/wiki/File:Microarray-schema.jpg>.

Fine, Miranda, and C. Wilborn, eds. "What is DNA Sequencing?." *wiseGEEK*. N.p.. Web. 27 Nov 2012. <http://www.wisegeek.com/what-is-dna-sequencing.htm>.

Fishbein, James. *Advances in Molecular Toxicology*. 6. Amsterdam: 2012. 133. Print.

Francis, Richard. Epigenetics. 1st ed. New York: W.W. Norton & Co., Inc., 2011. Pages 40, 41 44-47, 73-74. Print.

Genetic Science Learning Center. "Lick Your Rats." Learn.Genetics 3 December 2012 http://learn.genetics.utah.edu/content/epigenetics/rats/

Harrison, . *Putting it all Together*. 2012. Daily Post and HWWeb. 14 Dec 2012. <http://blogs.savcds.org/harrison/category/daily-posts-and-hw/page/11/>.

Hendricksen, Amy. *Bisulfite Conversion, Meet COBRA*. 2012. Promega ConnectionsWeb. 19 Dec 2012. <http://promega.wordpress.com/2012/04/09/bisulfite-conversion-meet-cobra/776px-cobra_workflow-svg/>.

Herman, J., J. Graff, S. Myohane, B. Nelkin, and S. Baylin. "Proceedings of the National Academy of Sciences." *Proceedings of the National Academy of Sciences*. 93. (1996): 9821-9826. Web. 15 Nov. 2012.

Ho, Shuk-Mei Ho, and Jacob Schmidlapp. "Epigenetics: the Interface between Environment and the Genome." *OEHHA*. N.p., 28 2008. Web. 17 Dec 2012. <http://oehha.ca.gov/public_info/lecture/epigenetics.pdf>.

Holzer, G.. "Transcription in eukaryotes ." *Prokaryotic and Eukaryotic Gene Expression*. G. Holzer, n.d. Web. 24 Oct 2012.

<http://www.prism.gatech.edu/~gh19/b1510/genreg.htm>.

"Illumina." *illumina*. February.770-2012-008 (2012): 1-12. Web. 27 Nov. 2012. <http://www.illumina.com/landing/miseq/primer-form.ilmn?scid=2011009PT1>.

Leng, Han, and Zhongming Zhao. "BMC Bioinfomatics." *BMC Bioinfomatics*. 2009, 10:65 Web. 5 Nov. 2012. <http://www.biomedcentral.com/1471-2105/10/65/>.

Meehan, Richard. *Epigenetic Mechanisms in Development and Disease*. Diss. University of Edinburgh, Edinburgh: MRC Human Genetics Unit, Web. <http://www.hgu.mrc.ac.uk/people/r.meehan_researchb.html>.

Mercola, "Epigenetics vs. Determinism." Articles.Mercola. 11 2011. Web. 3 Dec 2012. <http://articles.mercola.com/sites/articles/archive/2012/04/11/epigenetic-vs-determinism.asp&xgt;.

Microarray scanning services. N.d. NTNU Genomics Resource Center Web. 27 Nov 2012. <http://www.microarray.ntnu.no/html/micrarray_scanning_services.html>.

"MICROARRAYS: CHIPPING AWAY AT THE MYSTERIES OF SCIENCE AND MEDICINE." *NCBI*. N.p., 27 2007. Web. 27 Nov 2012. <http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html>.

"News Medical." *News Medical*. n. page. Web. 27 Nov. 2012. <http://www.news-medical.net/health/What-is-Gene-Expression.asp&xgt;.

Panse, Sonal, and Lindsay D., eds. "What is DNA Cloning?." *wiseGEEK*. N.p.. Web. 27 Nov 2012. <http://www.wisegeek.com/what-is-dna-cloning.htm>.

Phillips, Theresa. "Nature Education ." *Nature Education* . 1.1 (2008): n. page. Web. 18 Nov. 2012. <http://www.nature.com/scitable/topicpage/the-role-of-methylation-in-gene-expression-1070>.

Pray, L. "Nature Education ." *Nature Education* . 1.1 (2008): n. page. Web. 17 Oct. 2012. <http://www.nature.com/scitable/topicpage/transposons-the-jumping-genes-518>.

Rodenhiser, David, and Mellissa Mann. "Review." *Review*. 174.3 (2006): 314-348. Web. 6 Nov. 2012.

Rogers, Kara. "The Rebirth of Lamarckism." *Encyclopedia: Britannica Blog*. Encyclopedia Britannica, Inc. 2009.Web. 3 Nov. 2012. <http://www.britannica.com/blogs/2009/02/the-rebirth-of-lamarckism-the-rise-of-epigenetics/>.

Sahu, Saura. *Toxicology and Epigenetics*. Wiley, 2012. Web.

Schmacher, Alex. "Bisulfite Sequencing." *Epigenetics Group*. A. Schumacher & MethyLogiX, n.d. Web. 6 Oct 2012. <http://www.methylogix.com/genetics/bisulfite.shtml.htm>.

Schlick, . "Introduction." *Chromating Folding*. New York University, n.d. Web. 21 Oct 2012. <http://www.biomath.nyu.edu/index/research/chromatin.html>.

"Scitable." *Scitable*. (2012): n. page. Print. <http://www.nature.com/scitable/topicpage/the-order-of-nucleotides-in-a-gene-6525806>.

St. Rosemary Educational Institution. "PCR: Uses, Steps, Purpose." http://schoolworkhelper.net/. St. Rosemary Educational Institution, Last Update: 2012. Web. Retrieved on: Friday 23rd November 2012. <http://schoolworkhelper.net/pcr-uses-steps-purpose/>.

"The Effect of Good Parenting." . Edublogs, 04 Oct 2012. Web. 3 Dec 2012. <http://fifer11.edublogs.org/2012/10/04/the-effect-of-good-parenting/>.

Watters, Ethan. "DNA is not Destiny." *discovermagazine*. N.p., 22 2006. Web. 7 Nov 2012. <http://discovermagazine.com/2006/nov/cover>.

Worthington, Krystal. "Deoxyribonucleis Acid and Related Products." *Worthington Biochemical Corporation* . Worthington Biochemical Corporation, n.d. Web. 21 Oct 2012. <http://www.worthington-biochem.com/DNA/default.html>.

Winning, Robert . "Genes and Transcription." *emunix*. N.p., n.d. Web. 22 Oct 2012. <http://www.emunix.emich.edu/~rwinning/genetics/transcr.htm>.

Xueqing, Zou. "DNA Methylation." *Theoretical and Computational Biophysics Group*. National Insitude of General Medical Science, 15 2012. Web. 13 Nov 2012. <http://www.ks.uiuc.edu/Research/methylation/>.

Zilberman, Daniel, and Steven Henikoff. "Development." *Development*. October 10.doi: 10.1242/dev.001131 (2007): n. page. Web. 27 Nov. 2012. <http://dev.biologists.org/content/134/22/3959.long>.

9 Wordlist

Acetylation: The process of introducing a acetyl group to a compound. Similarly to methylation, it can change the structure of chromatin.

Adenine: One of the four DNA and RNA bases.

Allele: One pair of a gene located on a specific chromosome at a specific site of the genome.

Autoimmune disorders: A disorder where the body's immune system mistakenly attack healthy cells or substances that are normally found in the body. It does this because it mistakenly believes them to be pathogens.

cDNA: Also referred as complementary DNA. And is thus complementary to certain sequences mRNAs (Messenger RNA).

Cell nucleus: Is the core of an eukaryotic cell, contains amongst other things the genetic materiel

Chromatin: Is the proteins and DNA in the cell nucleus

CpG islands: Regions of the DNA with a high percentage of Cytosine–phosphorus-Guanine pairs, 50% or more.

Cytosine: One of the four bases of DNA and RNA. Cytosine often becomes methylated.

Cytoplasm: A watery gel outside the nucleus that protects the cell and helps it keep it's form.

Cytosol: Sort of like cytoplasm, but different

Darwinism: Evolution by natural selection over an extensive generations

De-novo sequencing: It is a form of a primary genetic sequence of an organism, done by assembling sequence reads into longer contiguous in the absence of a reference sequence.

Deoxyribonucleotide: A nucleotide that is constituent of DNA.

Dideoxyribonucleotide: They are chain terminating inhibitors of DNA polymerase, which are used in the Sanger sequencing technique.

DNA: Also known as Deoxyribonucleic acid, it is the basic blueprint of life that codes for everything an organism must be able to do to survive.

DNA ligase: Is an enzyme that joins two DNA strands together, in other words seals two DNA strands together. Used also in the making of Recombinant DNA.

DNA recombinant: A process where two DNA strand are sealed together.

DNA sequencing: Is a range of techniques for finding the sequence of nucleotides in an DNA string.

DNMT genes: DNA methyltransferases, an enzyme that adds the methyl group to the DNA.

Eukaryotes vs. Prokaryotes: Prokaryotes are viruses and bacteria while eukaryotes are all other types of organisms.

Electrophoresis gel: It is used in the Sanger sequencing method, where it separates and analyses DNA and their fragments based on their size.

Endoplasmic reticula: Is part of the cell infrastructure that synthesizes proteins

Enzyme: Are large molecules that functions as catalysts is various metabolic processes

Exon: The coding parts of a specific mRNA strand. Gets separated from the introns during RNA processing.

Extracellular signal-regulated kinase pathway signaling: signal molecules that are involved in the regulation of meiosis, mitosis, and postmitotic functions. Meiosis and mitosis are forms of cell division.

Gene: A section of DNA or RNA that codes for a polypeptide.

Genome: An organisms hereditary information.

Germ Cell: Egg or sperm cells

Guanine: One of the four DNA and RNA bases.

Histone: Is a protein that winds DNA up upon itself, so that the DNA is compacted.

Housekeeping genes: Genes that are constantly expressed because they code for proteins that are always needed in the cell.

Hypermethylated: A region of DNA that has a very high frequency of methylation

Hypomethylated: A region of DNA that has a very low frequency of methylation

Intron: Segments of the pre- mRNA that become cut out using an enzyme such as snRNPs and the splicesome. For the specific polypeptide that is to be made the introns are information that is not needed and therefore cut out.

Lamarckism: A now largely discredited theory that individuals could pass on traits, acquired during their lifetime, to their offspring.

Lupus: Is a disease where the immune system becomes hyperactive and attacks healthy cells in the body

Methylation: The addition of a methyl group to DNA, can be added to the DNA or to the histones the DNA is wound around. The higher percentage of methylation on a gene, the more silenced the gene becomes.

mRNA: Also known as Messenger RNA. It is the RNA that is complementary to a gene on the template strand of the DNA. This RNA strand becomes translated into a macromolecule in the process of translation. Before mRNA goes through RNA processing it is known as pre- mRNA.

Mutation: Is random changes in the gene sequence.

Neoplasia: Abnormal new growth of cells that could lead to tumors.

Nucleotides: Is the basic base pairs that make up DNA.

Oligonucleotide: Single stranded DNA or RNA molecules.

Phenotypical versus Genotypical: phenotypical is an organisms observable traits while genotypical is the genetic makeup of the cell such as whether the trait is recessive or dominant (or mutant).

Polyadenylation signal sequence: Signals for addition of a poly-A tail during RNA processing

Polypeptide: A chain of amino acids linked by peptide bonds. Made during translation of RNA

Promoters: A region of the a gene that initiates its transcription. It is located downstream of the gene.

Protein: Large molecules that are made of one or more polypeptide chains that perform various jobs in the cell.

Reads: DNA sequencing technology does not have the ability to read whole genomes in a single run, therefore small pieces of 20 to 1000 bases are read and these short fragments are thus called Reads.

Recombinant DNA: A process where two DNA strands are sealed together.

Resequencing: Repeating the sequencing all over again. Is much easier as it requires less reads.

Restriction enzyme: Restriction enzyme: Enzymes that play a specific role in recognizing a short DNA sequence called restriction site.

Reverse transcription: Converts RNA to DNA, the opposite of transcription.

Ribosome: A large complex molecular machine that is for example used during translation

RNA Polymerase II: Enzyme found in eukaryotic cells. It initiates the process of transcription.

RNA: Also known as Ribonucleic acid. Single stranded, with a ribose sugar as its backbone. Has the four bases Uracil, Cytosine, Guanine, and Adenine.

snRNPs: A small ribosomal unit involved in the process of RNA processing, making the pre-mRNA into RNA.

spliceosome: A complex of the snRNP that cuts out introns from the pre-mRNA in RNA processing

TATA box: A DNA sequence found in the promoter region, though to be the binding site of transcription factors.

T-Cell: Part of the group known as white blood cells. Plays an essential role in the human immune system.

Thymine: One of the four DNA bases. Is not found in RNA, but is replaced by Uracil.

Transcription: The process of making a mRNA strand by using one side of the double helix DNA as a template. The RNA is complementary to the DNA.

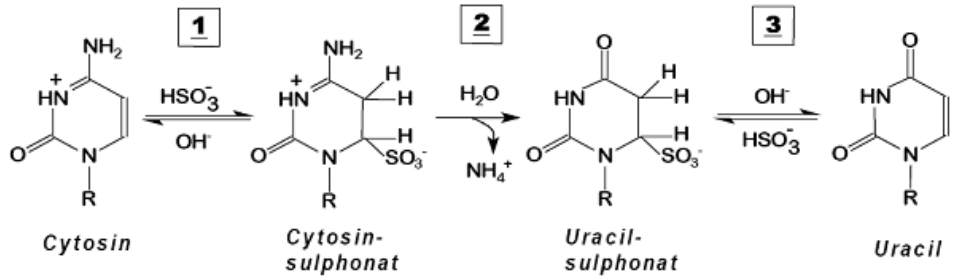
Trithorax group: A group of proteins. Their main job is to maintain gene expression.

Uracil: One of the four RNA bases, is not found in DNA. Replaces the DNA base Thymine.

10 A bit of chemistry

Condensation reaction: A chemical reaction between two compounds where they react to form one compound with a loss of a small compound. In case of loss of water it is called a dehydration reaction.

Bisulfite Conversion:



Hydrogen bond: A weak chemical bond between two molecules where one hydrogen atom from one of the molecules is attracted to an electronegative atom from the other molecule.

Phosphate group: A phosphorus attached to four oxygen atoms. Has a net negative charge.