Mark Schramm

Genomics and Medicine

12/11/13

Epigenetic Mechanisms and New Technologies for Studying Them

Epigenetics is the study of both heritable and non-heritable changes in gene transcription and expression that are not caused by the DNA sequence, but from a variety of other factors including DNA methylation, histone modification, chromatin remodeling, and transcription factor binding. These factors can be inherited or affected by environmental variables like diet, exercise, and general health and well-being, but there have not been many definitive studies on how these environmental factors affect the epigenome. With the help of Next Generation Sequencing, more progress is being made to study the epigenome, and more can be learned about how epigenetics is heritable, how it is affected by environmental variables, and its impact on DNA expression and phenotypes likes traits or disease.

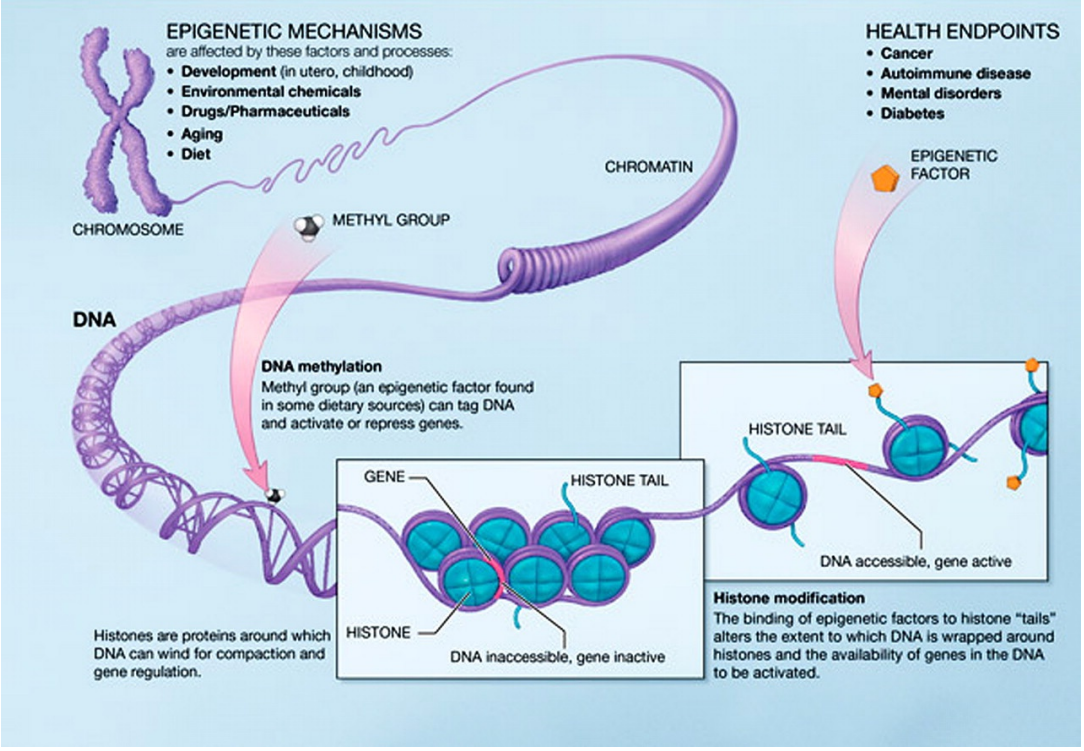
(2)Epigenetic mechanisms do not alter DNA sequencing but can regulate DNA expression that produces different cell types, which give rise to the variety of tissues within the body. Epigenetics is also essential in processes like cell-division and embryogenesis, and epigenetic abnormalities in this process have been linked to disease. Furthermore, many complex human diseases and phenotypes are not very strongly linked to particular genetic variants in (genome-wide association studies) GWAS studies, meaning that epigenetic mechanisms could potentially provide explanations for heritability of diseases. Studies of monozygotic but disease-discordant twins is a very powerful way to study epigenetic heritability that occurs in the germline during meiosis. “However, it is challenging to determine which epigenetic events are germline heritable. Comparing the epigenetic profile between the ‘disease epigenome’ such as a cancer epigenome, with the epigenome from constitutional DNA from the same individual as a reference, is required for distinguishing between germline and somatic epigenetic events” (Ku et. al,

Introduction). So studies of the epigenome are also important in tracking the causes of diseases, such as various forms of complex cancer. Because the significant advances in studying the epigenome are relatively recent, our current understanding is still limited relative to what is know about the genome. In this paper I will discuss at a broad level what is currently understood about epigenetic mechanisms, how it is thought to link to phenotypes and disease, and then talk about new approaches and technologies that are being used to study epigenetics, particularly those introduced by Next Generation Sequencing.

The term 'epigenome', which we learned in lecture means 'on top of' or 'in addition to' the genome, encapsulates a multitude of biochemical mechanisms that control DNA expression. The most studied and understood epigenetic mechanism is DNA methylation, which consists of the methylation of cytosine in DNA. This is "the addition of a methyl group at the carbon 5 position of cytosine through DNA methyltransferase(DNMT)"(Ku et. al, DNA methylation and histone modifications). Methylation suppresses transcription in regions around CpG islands. In fact, most CpG sites in the genome are methylated, although CpG islands in the promoter regions of human genes are generally not methylated. Hypermethylation of the genome has been associated with cancer because it leads to genomic instability (Ku et. al, DNA methylation and histone modifications).

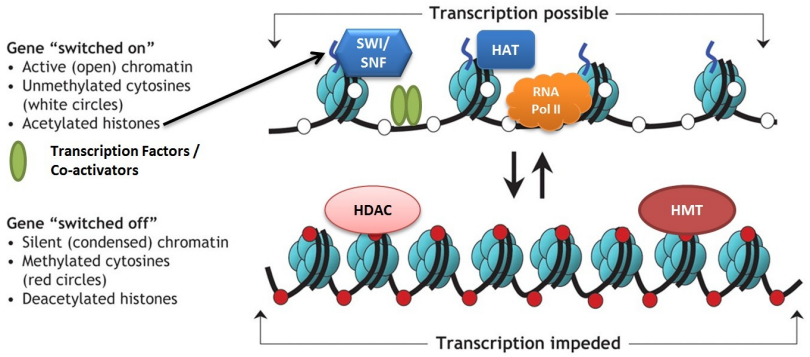
Another important epigenetic mechanism that we also discussed in class is histone modification. Each nucleosome in the chromatin consists of two copies of the four core histones which are wrapped around by 146 bp of DNA. Modification of these histones can cause the chromatin that is wrapped around it to condense or unwind, which has an effect on transcription of this DNA. "The N-terminal tails of histone polypeptides can be modified by over 100 different post-translational modifications including methylation, acetylation, phosphorylation, and ubiquitination" (Ku et. al, DNA methylation and histone modifications). Because there are all of these different reactions, which depend

on a variety of different chemical compounds, that can modify the histones and change chromatin structure, there are many environmental variables which involve these compounds that can affect the epigenome. One example is the methylation of histone H3 lysine 4 (H3K4) and H3 lysine 36 associated with transcription activation, but the methylation of H3 lysine 9 (H3K9), H3 lysine 27 (H3K27), and H4 lysine 20 (H4K20) being correlated with repression of transcription (Ku et. al, DNA methylation and histone modifications). Other mechanisms, such as histone acetylation, discussed in class, have been shown affect transcription regulation. This diagram, taken from the Journal of Medical Genetics (url http://jmg.bmj.com/content/48/11/721.long), demonstrates how epigenetic factors control chromatin coiling.



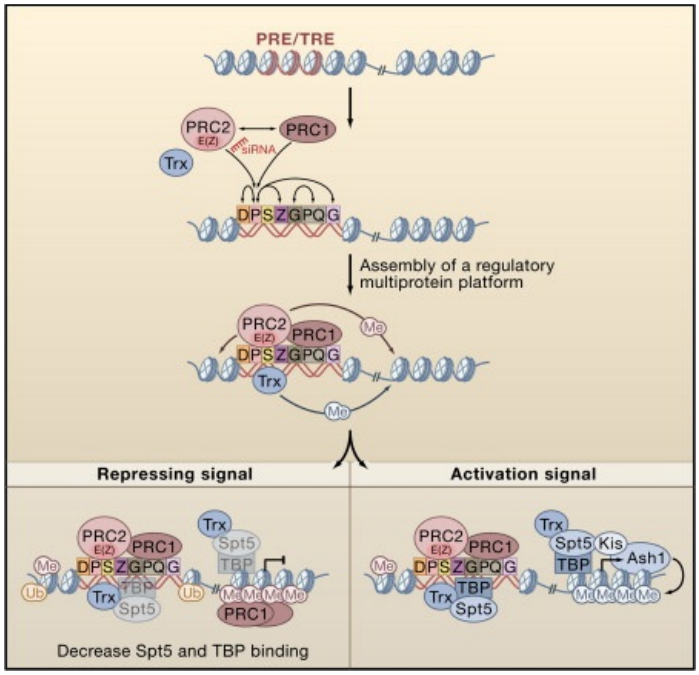
Cytosine methylation, and histone modification through a variety of histone polypeptide post-translational modifications such as acetylation and methylation are currently the most studied mechanisms among the translational regulatory processes that fall under epigenetics, but there are still other biochemical mechanisms. One such mechanism is ATP-Dependent Chromatin-Remodeling Complexes. These complexes

use energy from ATP hydrolysis to locally disrupt or alter the association of histones with DNA. These are different from the earlier discussed histone modifiers which “regulate transcriptional activity of the genes by determining the level of acetylation of the amin-terminal domains of nucleosomal histones associated with them”(Vignali et. al, introduction). All ATP based chromatin structure remodeling proteins are derived from the SNF2 protein family, and the particular proteins are “classified into two main groups, the SWI2/SNF2 group and the imitation SWI (ISWI) group”(Vignali et. al, Classification and Subunit Composition). SNF2 proteins are also involved in DNA repair (Ryan, Owen-Hughes, Snf2-family proteins: chromatin remodelers for any occasion). Here is a diagram, taken from wikipedia (url <http://en.wikipedia.org/wiki/File:Luong_LD_SA_F2.jpg>, which illustrates how ATP-based chromatin remodeling complexes SWI and SNF work with histone acetyltransferase(HAT), histone deacetylase(HDAC), and histone methyltransferase(HMT).



Yet another epigenetic mechanism is the polycomb and trithorax group proteins. Just as ATP dependent complexes and enzymes like HAT and HMT, the polycomb group (PcG) and trithorax group (trxG) silence or activate gene expression by binding to specific regions of DNA and direct posttranslational modification of histones. Polycomb repressive complex 1 and polycomb repressive complex 2(PCR1 and PCR2), and a third complex PhoRC, are classifications of different PcG and trxG complexes which carry out different functions. These different proteins have been seen in *Drosophila* to interact and bind to specific sites on DNA and regulate transcription. Binding of PcG

proteins have been associated with the absence of RNA polymerase 2 which suggests that PcG bound genes are being silenced, also “PcG proteins bind preferentially to genes encoding transcription factors” (Schuettengruber, Genome-Wide Distribution and Biological Functions of PcG Proteins), suggesting that PcG proteins affect transcription pathways. This image, from sciencedirect (url http://www.sciencedirect.com/science/article/pii/S0092867407001900),



illustrates how PcG and Trx proteins bind to specific sites along the DNA (PRE and TRE sites), and can act to repress or activate gene transcription.

All of these mechanisms are now better understood because of new tools and sequencing made sequencing technology. Next generation sequencing has developed new methods to account for the epigenome when sequencing DNA and to analyze the epigenome. DNA methylation has and continues to be the main focus of the tools, but now other mechanisms are also being analyzed. Older methods to recognize DNA methylation could only be used on a small subset of methylation sites, or could only detect regions containing relatively high levels of cytosine methylation(CpG islands) (Ku et. al, Traditional methods in studying epigenomics). To enable genome-wide analysis of

methylation and histone modifications, microarrays were used, but this technique required that the user have prior knowledge of areas of interest. Microarray-based techniques, used in conventional ChIP-chip (chromatin immunoprecipitated chip) studies, only detected precipitated DNA fragments associated with a particular histone modification if that region of the genome was covered by probes (Ku et. al, Traditional methods in studying epigenomics). However, new sequencing based approaches capture data as the genome is actually sequenced, so in theory all areas of the genome are covered depending on sequencing depth and coverage. “The current high density microarrays High density microarrays such as the Infinium Human Methylation 450 BeadChip allow researchers to investigate >450 000 CpG sites out of the approximately 28 million CpG sites in the human genome (<0.02%).There is also ascertainment bias in selecting these >450 000 CpG sites to be interrogated. This type of microarray is widely accepted as a genome-wide tool for DNA methylation yet it is still restricted to preselected CpG sites. This incomplete interrogation of DNA methylation patterns in the human genome will reduce the power of discoveries. For example, methylation in the non-CG context, revealed through sequencing of the entire DNA methylome, would be overlooked when using microarrays”(Ku et. al, Traditional methods in studying epigeneomics). Current genome-wide mapping of histone modifications and other DNA-protein interactions are heavily dependent on the precision and breadth of genome mapping provided by next generation sequencing. However, while the high throughput of next generation sequencing have allowed for an enormous amount of data to be collected (94% of all cytosines in the genome covered), new biological insights have been made to improve upon the conventional approach of bisulfite conversion to differentiate methylated from unmethylated cytosines.

While bisulfite treatment of DNA is the most widely used, there are several weaknesses with the technique. This method and also the method of methylated DNA immunoprecipitation with anti-5mC antibodies cannot detect 5hmC or methyladenine. The importance of 5hmC is unknown, but it is found in promoter and intragenic regions, but rarely in non-gene regions. It has been positively correlated with gene expression levels, and perhaps its epigenetic significance will become more obvious when more precise sequencing methods are applied. Another weakness of bisulfite treatment is that it damages DNA and actually fragments it, hindering the study of multiple methylcytosines at adjacent loci due to fragmentation (Ku et. al, New opportunities from TGS technologies). This conventional sequencing method, used by Illumina, could soon be challenged by newer sequencing methods with advantages.

One method, considered a part of third generation sequencing, was developed by Pacific Biosciences and is called single-molecule, real-time (SMRT) sequencing. SMRT sequencing does not required bisulfite treatment of DNA prior to DNA sequencing. Bisulfite treatment of DNA “deaminates cytosine residues to uracil which are subsequently read out as thymine,” but cytosines that are methylated are converted with much lower efficiency by the bisulfite, and thus remain cytosine (Clark, et. al, background). The chemically treated strand is then compared to an untreated strand, and methylated regions are discovered in this manner. SMRT sequencing offers a much more direct way to identify methylated cytosines, and also produces more detailed data about the modification that it encounters. SMRT sequencing involves monitoring the progress of a DNA polymerase as it makes a copy of a strand of DNA. When the polymerase “encounters a modified base on a template strand, its rate of progression

changes in a characteristic way relative to an unmodified template with the same sequence context" (Clark et. al, Background). This is measured by the time between fluorescent pulses that indicate nucleotide incorporation. Not only is this data indicative of the existence of a modification, but provides a data on a range of different modifications because the time between pulses, called the interpulse duration (IPD), may vary. This is seen in the results as different modifications such as 5mC (5-methylated cytosines), 5hmC (5-hydroximethylated cytosines), 5fC, or 5caC have different average IPD's, which are also called kinetic signatures (Clark et. al, Results). The size of the chemical structure of a particular modification is correlated with the magnitude of the kinetic signature. This offers a particular advantage over bisulfite treatment which cannot distinguish between 5mC and 5hmC. SMRT sequencing detection of 5mC is further enhanced by converting each 5mC to 5caC. There is a lot of potential in this technique because it offers a generalized way to detect and identify all sequence modifications along the genome. It also has an intuitive appeal because it is making detection based on how long it takes DNA polymerase to transcribe a particular modified region, and producing data that is based in time, and primed for statistical analysis.

Another approach for detecting DNA methylation that does not have the disadvantages of bisulfite treatment is the use of solid-state nanopores to sequence DNA. This technique involves transporting a double stranded DNA through a solid state nanopore and using "ionic current spectroscopy to electrically interrogate individual DNA molecules with the sensitivity to discern subtle structural motifs...The translocation of the methylated DNA - MBD-1x complex through a solid-state nanopore induced approximately a 3-fold increase in the measured blockage current relative to unmethylated DNA. The binding of a single MBD-1x protein to a methylated DNA fragment was suffi-

cient for differentiation with high fidelity, thereby enabling single CpG dinucleotide sensitivity" (Shim et. al, Introduction). The advantage of this technology is that it is very fast at processing DNA and identifying methylation sites based on changes in current.

These "Third generation sequencing" techniques provide sequencing methods that can detect methylation and other modifications as the DNA is being sequenced. By allowing more precise and efficient analysis these techniques can provide new insights into where modifications occur, so that they can be correlated with effects on transcription and phenotypic variants and disease. There are a huge number of different types of modifications that can occur and cause biochemical and structural changes to chromatin. A more nuanced understanding of what these mechanisms are, where they occur, and what the effects of the mechanisms are will be key to understanding how epigenetic abnormalities give rise to disease like cancer. It will allow a better understanding of epigenetic inheritance. Finally, It will provide a deeper understanding of how environmental variables affect us at the level of gene transcription, and hopefully drastically increase the power of personalized medicine by allowing people to make informed decisions not only based on knowledge of their genome, but of their epigenome.

Bibliography

1) Schuettengruber, Bernd, Daniel Chourrout, Michel Vervoot, Benjamin Lablanc, and Giacamo Cavalli. "Cell - Genome Regulation by Polycomb and Trithorax Proteins." Cell - Genome Regulation by Polycomb and Trithorax Proteins. N.p., 23 Feb. 2007. Web. 11 Dec. 2013.

2) Ryan, Daniel P., and Tom Owen-Hughes. "Snf2-family Proteins: Chromatin Remodellers for Any Occasion." Snf2-family Proteins: Chromatin Remodellers for Any Occasion. N.p., 5 Oct. 2011. Web. 11 Dec. 2013.

3) Vignalli, Marissa, Ahmed H. Hassan, Kristen E. Neely, and Jerry L. Workman. "American Society for MicrobiologyMolecular and Cellular Biology." ATP-Dependent Chromatin-Remodeling Complexes. American Society for Microbiology, Mar. 2000. Web. 11 Dec. 2013.

4) Shim, Jiwook, Gwendolyn I. Humphreys, Bala M. Venkatesan, Jan M. Munz, Xueqing Zou, Chaitanye Sathe, Klaus Schulten, Farhad Kosari, Ann M. Nardulli, George Vasmatzis, and Rashid Bashir. "Detection and Quantification of Methylation in DNA Using Solid-State Nanopores." Nature.com. Nature Publishing Group, 15 Nov. 2012. Web. 11 Dec. 2013.

5) Clark, Tyson A., Xingyu Lu, Khai Luong, Qing Dai, Matthew Boitano, Stephen W. Turner, Chuan He, and Jonas Korlach. "Enhanced 5-methylcytosine Detection in Single-molecule, Real-time Sequencing via Tet1 Oxidation." BMC Biology. BMC Biology, 22 Jan. 2013. Web. 11 Dec. 2013.

6) Ku, Chee S., Nasheen Naidoo, Mengchu Wu, and Richie Soong. "Studying the Epigenome Using next Generation Sequencing." Studying the Epigenome Using next Generation Sequencing. Journal of Medical Genetics, 8 Aug. 2011. Web. 11 Dec. 2013.