Published in final edited form as:

Behav Sci Law. 2015 October ; 33(5): 675–690. doi:10.1002/bsl.2202.

A Review of Epigenetic Markers of Tobacco and Alcohol Consumption

Robert Philibert, M.D., Ph.D.^{†,*} and Cheryl Erwin, J.D., Ph.D.[‡]

[†]Behavioral Diagnostics Inc., Iowa City, IA, U.S.A

[‡]Departments of Medical Education and Psychiatry, Texas Tech University Health Sciences Center, Lubbock, TX, U.S.A

INTRODUCTION

The past three decades of scientific and biotechnical progress have revolutionized our conceptualization of healthcare. For a reasonable fee, tech savvy consumers can choose to have their whole genome sequenced. Then, depending on the level of complexity needed, these consumers can have the resulting data interpreted and stored for future use. In considering the impact of this technology on the healthcare market, it is important to realize that the science that led to this revolution developed not overnight but incrementally.

Perhaps the sentinel event for genetic revolution was the publication in 1975 of a new method for DNA sequencing by future Nobel laureate Fred Sanger and associates (Sanger & Coulson, 1975). Sanger sequencing, as the technique is known, was the first scalable method of determining DNA sequence. When this technique was combined with other molecular methods such as plasmid cloning, early molecular biologists were soon able to determine the sequence for entire genes. Over the next 15 years, the sequencing process was industrialized by companies such as Perkin Elmer (who was later purchased by Applied Biosystems) while the processing of the resulting data was accelerated by government-funded DNA alignment computer algorithms capable of assembling large stretches of DNA sequence. The success of this integrated academic industrial effort in determining the biological basis of select genetic disorders set the basis for subsequent announcement of what would be termed the “Human Genome Project” in 1990 (Watson, 1990). In an event that perhaps foreshadowed the current controversies with regard to these technologies, the success of those efforts by both an academic and a commercial consortium were announced in June 2000 in a carefully scripted announcement in the White House hosted by President Bill Clinton (Federal News Service, 2000). Since that time, the pace of genetic sequencing technologies has only accelerated. The first, rather rudimentary, whole genome sequence took the efforts of an entire generation of scientists and hundreds of millions of dollars. Now, for less than \$3000, a high-quality whole genome sequence can be generated in less than a day by a single well-trained technician.

*Correspondence to: Robert Philibert, Department of Psychiatry, University of Iowa Rm 2-126 MEB, 500 Newton Road, Iowa City, IA, 52242, USA. robert-philibert@uiowa.edu.

In response to these developments, a large number of federal and state laws to regulate the performance and usage of these technologies have been developed. Perhaps the best-known of these regulations is the Genetic Information Nondiscrimination Act (GINA) (Genetics Information Non-Discrimination Act, 2008). The implementation of the network of this and other regulations has largely quelled the fears of the general public with respect to possible misuse of these technologies.

However, over the past 10 years, biologically oriented health scientists have increasingly been interested in epigenetics, now regarded as the new frontier of human molecular biology. This review provides an introduction for the reader to the fundamentals of the burgeoning field of epigenetics beginning with an epidemiological perspective and ending with the latest in molecular approaches. Next, the authors will review how epigenetic tools are helping scientists extend our current genetically oriented conceptualization to a more holistic understanding of the biology of human disease. Then, some of the latest findings from some of these studies in particular are reviewed, focusing on the development of biomarkers for potentially preventable common complex illnesses. The authors will then conclude by discussing the possible uses of epigenetic information by medical, civil and governmental audiences.

EPIDEMIOLOGICAL EVIDENCE OF EPIGENESIS

Perhaps one of the most important points about epigenetics is that our current notions of medical illness are rooted in centuries of observations and the observations are not always what they seem. For example, a re-visitation to the early experiments of Mendel suggests that the results may have been “optimized” to convey his points (Novitski, 2004). Similarly, while the data from several decades of twin studies have not changed, our conclusions from the analyses of these data have undergone a remarkable reshaping in light of subsequent discoveries.

For geneticists, twin studies are the paradigm of choice when seeking to understand the etiology of medical illnesses. Over the past 70 years, tens of thousands of analyses of the contribution of genetic and environmental factors to the etiology of medical illnesses and medically related traits have been published. The fundamentals of twin studies are easy to understand. Twin studies compare and contrast the frequency of a given trait (e.g. obesity) in identical and fraternal twins. The results are analyzed using a deceptively simple analytical framework, $1 = G + E + G \times E$, where “1” represents the summed fractional observations of the heritable and environmental contributions to a given phenotype (i.e. they always must add up to 100%), G represents the contribution of heritable genetic factors, E represents the contribution of environmental factors and $G \times E$ represents the environmentally contextual effects of genetic factors. Whereas the terms G and E are readily understandable to almost all, $G \times E$ effects are the changes in the way genes express themselves in response to environmental exposures. Fortunately, some of these $G \times E$ effects can be readily observed. For example, phenylketonuria (PKU) is a somewhat rare syndrome, which is caused by receiving a dysfunctional phenylalanine hydrolase (PAH) gene from both parents (i.e., autosomal recessive inheritance) (de Groot, Hoeksma, Blau, Reijngoud, & van Spronsen, 2010). As result of having two dysfunctional copies of the PAH gene, patients are unable to

metabolize phenylalanine and may develop a syndrome of progressive cognitive impairment if they ingest phenylalanine. Fortunately, PKU is very rare in the United States because almost all individuals with the disorder are identified by newborn screening and placed on a phenylalanine free diet. These individuals do not develop significant clinical difficulties while those with an identical genotype who do ingest phenylalanine (commonly found in chocolate) develop the disorder. In other words, the effects of the two mutant PAHs (or G) are dependent on the presence of an environmental (E) dietary factor (phenylalanine). Somewhat surprisingly, the impact of these $G \times E$ effects on medical illness is more common than first appreciated, with the $G \times E$ effects having significant roles in the etiology of diverse disorders such as cirrhosis (e.g., alcoholic cirrhosis) and chronic obstructive pulmonary disease (COPD, with the E being cigarette smoke) being readily apparent (Bataller, North, & Brenner, 2003; Pillai et al., 2009).

The relative contributions of G, E and $G \times E$ effects are commonly illustrated using bar graphs. Figure 1 illustrates the etiology of several representative medical disorders in today's environment as shown by twin studies. Not surprisingly, the presentation of Huntington's chorea, which is caused by a trinucleotide repeat expansion in the first exon of the huntingtin gene, is almost strictly secondary to heritable genetic factors.

In contrast, type 2 diabetes (T2DM) and tobacco dependence (i.e., smoking) have only a 40% heritable component (Lehtovirta et al., 2010; Rose, Broms, Korhonen, Dick, & Kaprio, 2009; Tsuang, Bar, Harley, & Lyons, 2001). Remarkably, although the occurrence of all kinds of cancer (herein lumped together) is mediated through genetic mechanisms, heritable components only represent 15% of a given individual's vulnerability to developing cancer (Ahlbom et al., 1997).

While these findings are generally accepted, the interpretation of the data is where "the devil in the details" lies. For example, it is important to stipulate that the observed frequencies for any illness are dependent on the environment. For example, 2000 years ago, very few people developed Alzheimer's disease. It is not that the genetic risk factors were not present in the population. Simply, people did not live long enough to experience the illness. In today's society, where individuals can expect to live until their mid to late 70s, it is readily observable.

However, for common medical disorders such as T2DM, an additional level of complexity is indicated by the fact that the presence of other common disorders, each of which has its separate G, E and $G \times E$ components, are risk factors for T2DM. For example, the presence of either smoking or obesity increases the risk for T2DM by 50% (Cameron et al., 2008; Willi, Bodenmann, Ghali, Faris, & Cornuz, 2007). Interestingly, this does not seem to have any effect on the diagnosis or treatment of T2DM. Whether or not the patient smokes (or is obese), the same tests are used to diagnose (hemoglobin A1C) and treat (e.g., insulin and metformin) the T2DM. This observation is the first important element in our understanding of epigenetic factors in common illnesses. In other words, the biological expression of T2DM is similar regardless of the distal G, E and $G \times E$ contributions. To the cells involved, the disorder looks identical even though the factors that contributed to the favorable conditions for its development may differ.

A second observation is no less important—particularly for the smoking related medical disorders such as coronary artery disease (CAD), T2DM and hypertension. Those who smoke may have a vulnerability to illness that has a heritable (genetic) component for CAD, T2DM or hypertension, yet once the patient quits smoking, the likelihood of having/developing the disease phenotype starts to remit (Athiros, Katsiki, Doumas, Karagiannis, & Mikhailidis, 2013; Critchley & Capewell, 2003). This is important because the heritable genetic factors cannot change—yet the biology of these illnesses, which is manifested at the cellular level, changes. To understand the importance of this statement, we need to understand how cell biology is regulated at the molecular level.

GENE ACTIVITY CONTROLS CELLULAR BEHAVIOR

The complexity of the biological symphony that has led to the development of each human being can be appreciated by a simple experiment. First, touch your head and then touch your knee. These body parts are very distinct yet the cells in the each of these anatomical parts have the same genetic code. The reason for their anatomical difference is straightforward. Just as software controls the output of the “hardwired” portions of the computer, epigenetic programming or “code” moderates the output of the cellular “hard wiring” found in the genetic code. Indeed, at all 23,000 genes in the human genome in every cell of the body, epigenetic programming is essential for moderating the process of transcribing the DNA into messenger RNA (mRNA), which is then translated by ribosomes into the proteins that constitute the bulk of our bodies.

A good example of how the epigenetic code moderates the output of the genome can be found by examining the regulation of the serotonin transporter (see Figure 2). Serotonin is a neurotransmitter that modulates key behaviors such as mood, appetite and sleep. Like all neurotransmitters, once it is released from the neuron into the synaptic cleft (the space between two linked neurons) and transmits its signal to an adjacent neuron, its action has to be terminated. The serotonin transporter (SLC6A4 or 5HTT) is commonly known to most Americans as a primary site of action for antidepressants. Selective serotonin re-uptake inhibitors (SSRIs) such as fluoxetine (Prozac®) or sertraline (Zoloft®) accomplish their effects in part by blocking the re-uptake of the serotonin back into the neuron (Kroeze, Zhou, & Homberg, 2012). The gene coding for this “off switch” for serotonergic neurotransmission stretches nearly 25,000 DNA base pairs and contains 14 exons (including a variably spliced or included tripartite exon 1)—which are the portions of the gene transcribed into mRNA by an assemblage of proteins referred to as RNA polymerase (Philibert, Madan, Andersen, Cadoret, Packer, & Sandhu, 2007). You could think of the gene as the blueprint used by them RNA to make different parts of the final protein. The start site (i.e., transcription start site or TSS) for this RNA polymerase copying (the process of making the different mRNA parts) is at the start of exon 1A, while the stop site for synthesis of mRNA by RNA polymerase is at the end of exon 14. However, not all the mRNA that is produced by RNA polymerase is translated into protein by the ribosomes (another part of the cell that is a worker used in building proteins). For example, with respect to the serotonin transporter, the genetic code that triggers protein translation by the ribosomes is found in the middle of exon 2 while the stop translation sequence is found in the middle of exon 14. The portions of the exons that are not translated into protein are referred to as untranslated

regions (UTRs). These UTRs are critical in regulating the amount of protein that will be eventually produced. Whether coding or noncoding, each of these exons is essential for normal function of the gene. If important parts of the blueprint do not get copied the final protein could be unusable or not produced in sufficient quantities.

Like the rest of the genome, the serotonin transporter contains a significant amount of genetic variation. Many of these variants appear to be functional. In fact, perhaps the best studied variant in the human genome (referred to as the 5HTTLPR variable nucleotide repeat, VNTR), is found 1400 base pairs upstream of the promoter or start site of the gene. The promoter variant comes in two forms—long (14 copies of a 22 base pair repeat) or short (12 copies of a 22 base pair repeat). Laboratory and brain imaging studies show that individuals with two copies of the short variant produce less of the serotonin transporter mRNA than those with two long copies of the 5HTTLPR VNTR in those cells in which the protein is produced (Heinz et al., 2000; Philibert et al., 2008).

However, the serotonin transporter is only produced in a fraction of the cells in the human body. In particular, it is found in high concentrations in neurons that are connected to the raphe nucleus in the human brain stem, where it serves to moderate neurotransmission related to sleep, mood and appetitive behavior (Berger, Gray, & Roth, 2009). In addition, lower levels of the serotonin transporter are found in an odd collection of cells including certain types of lymphocyte and bone cell, where scientists believe that the protein may serve as a cellular sensor of stress.

How does the body make sure that the serotonin transporter is produced only in those cells that need it? The answer can be found in the concept of the “CpG island” (placement is indicated by the gray box covering exon 1A in Figure 2 and Figure 3, see later). The human genome consists of approximately three billion base pairs—cytosines (C), guanines (G), thymines (T) and adenines (A). The guanines bind to cytosines, forming three hydrogen bonds, while thymines bind to adenines, forming two hydrogen bonds. What is not commonly appreciated is that these nucleotide bases do not occur with equal frequencies, and at even the smallest levels of organization their distribution is tightly regulated in all organisms (Suzuki & Bird, 2008). In humans, Gs or Cs of the human genome are found at approximately 41% of all positions, while As or Ts are found at 59% of all the positions. Even at the dinucleotide level, the order is not random. Because there are four different nucleotides there are 16 possible dinucleotide pairs (e.g., CC or AT). If the distribution of dinucleotide pairs were random, we would expect that about 1 in 20 of each of the dinucleotide pairs in the genomes would be cytosine phosphoguanine (CpG) dinucleotide pairs. Instead, what we find is that fewer than one in 90 dinucleotide pairs are CpGs. In turn, even these CpG pairs are not randomly distributed but tend to occur in areas referred to as CpG islands. Understanding the reasons for this distribution is critical to understanding the regulation of cellular functioning.

The reason for this distribution is deeply rooted within evolution. Our current understanding of the origin of life is that cells with nuclei (eukaryotes) originated from the fusion of two bacteria several billion years ago (Rivera & Lake, 2004). These first simple eukaryotic organisms, like their bacterial precursors and all their subsequent descendants, were

susceptible to viral attacks against the genome, such as the viral attack posed by HIV. In order to combat these viruses, eukaryotes evolved a mechanism called DNA methylation, which silenced the viruses that had successfully incorporated themselves into the genome. Over the course of hundreds of millions of years of evolution, these viruses have continued to attempt to incorporate themselves into our genomes. Indeed, at the current time, between 60 and 70% of the genome is believed to be derived as a consequence of viral incursions.

Nature has taken advantage of some of these events to increase the evolutionary fitness of organisms. In fact, all mammals use the same DNA methylation machinery that is used to silence that viral signature of our evolution to regulate the activity of genes (Suzuki & Bird, 2008). The method through which this occurs is illustrated in Figures 3 and 4. As noted above, the distribution of CpG dinucleotide pairs is not random. Instead, CpG pairs tend to be found in areas of high concentration referred to as CpG islands. Two-thirds of all genes in the human genome have these CpG islands associated with their first exon or promoter. As it turns out, it is the association of these islands with gene promoters that gives human cells an exquisite ability to regulate gene expression.

Human gene promoters can be conceptualized as the 300–400 base pairs of DNA sequence immediately surrounding the TSS of the gene. Unlike bacteria, the sequence and structure of human gene promoters can vary widely. However, the serotonin transporter (illustrated in Figure 3) is fairly typical. Immediately upstream of the TSS there is a sequence motif referred to as a TATA box (bold and boxed). The transcription start site of the gene is demarcated by the start of exon 1A, which is underlined. The promoter associated CpG island, which is approximately 800 base pairs in length and contains 81 CpG residues, completely envelops exon 1A.

By modifying the cytosine residue found in the CpG dinucleotide pairs, cells can turn on or turn off transcription at the gene. How does this work? Similar to their ancient ancestors, human cells turn off regions of DNA by adding methyl groups (one carbon and three hydrogen atoms) to the cytosine of CpG dinucleotide pairs. In brief, enzymes referred to as DNA methyltransferases (DNMTs) take a methyl group from S-adenosyl methionine then transfer it to the cytosine of a CpG residue (Suzuki & Bird, 2008). When a sufficient number of the CpG residues in a region are methylated, a synergistic interaction of other epigenetic modifications made to the histone protein scaffold on which the DNA strand is suspended in the nucleus results. This addition of methyl groups, which are hydrophobic (literally water fearing), and the removal of acetyl groups, which are hydrophilic (water loving), causes the molecular complex of DNA and histone proteins to become less soluble in the nuclear cytosol. As a result, the DNA quite literally tightly winds itself around histone proteins in a shape shifting change process referred to as nucleosome formation and condensation. As a result, the DNA is physically much less accessible to RNA polymerase for gene transcription. Conversely, the process of methylation can be reversed by the TET family of proteins, which through a series of processes remove the methylation (Pastor, Aravind, & Rao, 2013). This process, which is coordinated with other enzymes that modify the closely aggregated histone proteins, results in greater access of RNA polymerase to the DNA for gene transcription.

USING EPIGENETICS TO ASSESS SMOKING STATUS

We can take advantage of changes in DNA methylation to measure the health of cells and their exposure to the environment. To illustrate this, we will focus our discussion on the DNA methylation changes associated with smoking. However, the principles outlined will apply to many other illnesses induced by harmful substances as well.

Smoking is the largest preventable cause of morbidity and mortality in the United States (Centers for Disease Control, 2002). However, no one actually dies of cigarette intoxication. Rather, they suffer and die as a consequence of the processes initiated by smoking—to be precise the induction of other disease processes such as diabetes and heart disease. The interesting thing about these smoking associated disease processes is that they are reversible and that they are accompanied by changes in cellular behavior. For example, with respect to coronary artery disease, smoking induces both the proliferation and degeneration of the cells that both constitute and line the blood vessels of the heart (Pittilo, 2000; Villablanca, 1998). As a result, the lumen of the vessel is narrowed and it becomes more susceptible to blockade. These changes in cellular behavior are all directed by changes in the cell's epigenetic signature (Breitling, Salzmann, Rothenbacher, Burwinkel, & Brenner, 2012). Likewise, with respect to COPD, the inhalation of smoke recruits a type of white blood cell called monocytes from the blood stream to reside in the lung. Once there and in the presence of continued smoking, these cells metamorphosize into cells that destroy local lung tissue and release inflammatory chemicals into the surrounding lung tissue. As this process continues, the ability to absorb oxygen from air decreases and the natural elasticity of the lung diminishes, resulting in the syndrome of shortness of breath and inability to breathe deeply known as COPD. Once again, these changes in activity of monocytes are moderated at the cellular level by changes in DNA methylation (Kremens et al., 2012).

Over the past 10 years, building on the success of the Human Genome Project, scientists have made rapid advancements in our ability to quantify these epigenetic changes. Since the primary sequence of the human genome was already known and the tools, such as high throughput sequencing machines and microarrays, were already developed, the next steps were rather straightforward with respect to the more accessible forms of epigenetic modifications. In particular, this is very true with respect to the portion of the epigenetic signature that represents the methylation status at all the human genome's 27 million CpG dinucleotide pairs, which is collectively referred to as the "methylome."

The rapid definition of the human methylome is in large part due to the invention of the technique of bisulfite conversion by Frommer and colleagues in 1992. This simple treatment of DNA with sodium bisulfite in the presence of sulfuric acid results in the deamination (removal of an amine or nitrogen group) from cytosine, but not methylated cytosine (see Figure 5). Since uracil can be amplified by DNA polymerases by substituting thymine in place of uracil, by first amplifying bisulfite treated DNA and then measuring the amount of conversion of cytosine to thymine, we can infer the methylation status of any cytosine residue in the human genome (see Figure 5). Thanks to the development of high throughput sequencing and hybridization arrays, scientists can routinely measure the degree of DNA methylation at any position in the human genome.

These techniques have been a boon for investigators seeking to understand the ways in which smoking cigarettes cause the myriad of disorders with which they are associated. Building on work begun nearly 10 years earlier, in 2012, our group published the first genome wide study of the effects of smoking on DNA methylation (Monick et al., 2012). In particular, we demonstrated that smoking changes the DNA methylation signature of the aryl hydrocarbon receptor repressor (AHRR). Since AHRR is the key regulator of the metabolic system that degrades polycyclic aromatic hydrocarbons and dioxins, and tobacco smoke is far and away the largest source of these environmental toxins, the finding had strong biological support (Nguyen & Bradfield, 2007). Since publication of these initial findings, these changes of AHRR methylation in response to smoking have become the most highly replicated findings in biological psychiatry. Over a dozen studies have exactly replicated the findings that this locus is the most highly sensitive locus in the human genome (see Table 1) (Allione et al., 2015; Besingi & Johansson, 2013; Dogan et al., 2014; Elliott et al., 2014; Guida et al., 2015; Harlid, Xu, Panduri, Sandler, & Taylor, 2014; Joubert et al., 2012; Philibert, Beach, & Brody, 2012; Philibert, Beach, Li, & Brody, 2013; Shenker et al., 2012; Teschendorff et al., 2015; Tsaprouni et al., 2014; Zaghlool et al., 2015; Zeilinger et al., 2013).

Many of these studies have described the molecular signatures of the mechanisms through which smoking causes coronary artery disease, COPD and stroke (Breitling et al., 2012; Dogan et al., 2014; Kremens et al., 2012). In particular, many of the studies have highlighted the relationship between smoking and inflammation. As total smoke consumption increases, there is a strong positive correlation with the amount of inflammatory factors produced by a variety of cells in the body. This is important in understanding the relationship between smoking and the myriad of disorders with which smoking is associated, because the vast majority of these disorders have strong inflammatory components to their etiology. This is particularly evident with respect to methylation of a coagulation factor gene referred to as F2RL3 (Breitling et al., 2012). Smoking markedly alters gene methylation at this locus, and increased methylation at this locus is a strong predictor of heart attacks.

The robustness of these findings and the industrialization of molecular techniques now allow methylation analyses to be translated into the clinical realm. During the initial exploration phase of the human methylome, only large laboratories with access to advanced machinery could participate. Now that the relationship of the methylome to smoking is relatively well understood, clinical scientists can focus on key areas predictive of clinical status. In direct contrast to the scientific findings with respect to the genetics of smoking in which hundreds of small effect loci are implicated and whose effects tend to be additive, the epigenetic response of the human methylome can be summarized by the assessment of only one or two CpG residues. This permits techniques such as the quantitative polymerase chain reaction (qPCR) to be used in assessments of clinical status. This inexpensive, rapid approach allows scientists to summarize the average methylation of millions of cells at a given locus, such as cg05575921, in a matter of hours. Figure 6 illustrates the relationship of DNA methylation at cg05575921 as assessed by a qPCR assay being developed for the clinical market as “Smoke SignatureTM” under the auspices of federal Small Business Innovation Research (SBIR) funding. As the arrow indicates, in a group of 61 adult subjects ascertained as part of a

recent commercial study, a cutoff value of approximately 0.89 (89%) methylation at the cg05575921 successfully distinguished smokers from non-smokers (Philibert et al., 2015).

OTHER APPLICATIONS OF EPIGENETICS IN THE DEVELOPMENT PIPELINE

Smoking and smoking related comorbidities are not the only types of substance or condition that can be assessed via epigenetic status. For example, with respect to other substances, we published a genome wide study demonstrating that alcohol consumption also could be assessed via epigenetic methods (Philibert et al., 2014a). Funding for the commercial translation of this set of findings via the SBIR mechanism is currently pending. In addition, we and others are attempting to develop epigenetic methods to assess consumption of other substances of abuse such as opioids, amphetamines and cannabis. In contrast to the challenges of developing tests for alcohol or smoking consumption, there are substantive scientific hurdles that need to be overcome before these tests are ready for commercial translation.

Perhaps the most important lesson that can be learned from our experiences with the development of epigenetic markers for smoking is that any medical condition that is associated with changes in cellular behavior can be potentially detected using epigenetic techniques if one can get access to the cells which are affected. For example, with respect to smoking, the collection of relevant biomaterial is easy because cells whose metabolism is disturbed via smoking can be collected via standard blood draws, finger sticks or saliva. Since most common medical disorders such as diabetes, hypertension and heart disease affect white cell metabolism, and their DNA is easy to obtain via blood or saliva (because they migrate into the mouth via the gums and the salivary glands), a number of investigators have demonstrated that DNA methylation changes are associated with each one of these disorders as well (Bell et al., 2010; Kulkarni, Chavan-Gautam, Mehendale, Yadav, & Joshi, 2011; Toperoff et al., 2012). In fact, the major question in the field of epigenetics is whether peripheral blood epigenetic signatures can be used to detect medical conditions before they become clinically manifest, thus giving physicians a chance to intervene and prevent illness.

THE CHALLENGES PRESENTED BY EPIGENETIC TOOLS

As with any new technology, there is always the potential for misuse. For example, common tools such as hammers are essential for the construction of our homes and offices. Nevertheless, each year people are harmed or killed with hammers. Although no one is likely to be physically harmed from epigenetic assessments, the potential for misuse is still there. While many individuals have noted that the standard high throughput methylation array used by scientists around the world included probes capable of determining genotype, the discovery that these arrays could be used simultaneously to assess both substance use status and provide an absolute DNA fingerprint capable of distinguishing anyone in the world came as a shock to many in the field (Philibert et al., 2014b). Previously, it was not generally held that these arrays contained personally identifying information. Now, it is clear to all concerned that these arrays can be used to provide high resolution genotypic fingerprints and assessments of substance use or other limited medical disorders. This is a

potential issue because these arrays, such as the Illumina 450K, have been used over 100,000 times and some of the resulting data are easily accessible via the Internet.

This will not pose a significant issue in the vast majority of cases, because the requisite genetic information to link any individual to the data in these arrays is not generally available. However, in this day and age of computer hacking, it is not unreasonable to believe that medical records containing the requisite genetic information could be stolen from either hospitals or commercial firms providing genotyping services. Under this scenario, it would be relatively easy to scan the Internet to identify possible matches, although currently only limited medical information can be extracted from these arrays. A more likely abuse of the information would not require internet matching. Employers, insurers and others with an interest in the future health of an individual are under no present legal constraints that would prohibit asking for epigenetic information that could reveal potential health problems in the future. This information could be used to discriminate against an individual who has not yet developed an illness, and may never develop the health condition. The potential problem of employers and others misusing personal information to discriminate has been extensively documented in the literature on genetic discrimination (Erwin et al., 2010; Rothstein & Anderlik, 2001). Like genetic information, epigenetic information reveals individually identifiable traits.

Because of the potential for misuse of an individual's information, these epigenetic techniques should be viewed as tools that can be used constructively or destructively. With respect to constructive purposes, we believe that the smoking detection technologies have great potential to aid in the prevention and treatment of tobacco use disorders. For example, if this technology can be successfully employed in the pediatric setting, the early detection of smoking in adolescents, followed by family-based psychoeducational interventions, may forestall the development of a variety of substance use related externalizing disorders (Beach, Gerrard, Gibbons, Brody, & Philibert, 2015). For adults, the potential use of this technology to guide evidence based treatment of smoking is already being explored in a Phase I SBIR project funded by the National Institutes of Health (R43DA037620).

The potential for abuse is also evident. Under some circumstances, disclosure of smoking or other health-related status that can conceivably be obtained via epigenetics (including alcohol consumption, diabetes and potentially even data such as Alzheimer's disease) to someone by a third party could be used to deny the rights, privileges or opportunities to which one would otherwise be entitled. Hence, there is a substantial need for discussion by the medical legal community on how epigenetic data should be handled.

Epigenetics additionally raises questions of social justice concerning differential exposure to environmental toxins that may lead to epigenetic changes that predispose those exposed to the development of disease (Lanphear, 2015). Justice requires that we provide all individuals in society with a fair equality of opportunity (Rawls, 1971). Epigenetic justice may require that we provide our citizens with a safe environment free of substances that may damage the epigenome. Because we can now identify and measure the epigenetic impact of exposure to substances, we can identify environmental substances that may have adverse health impacts on those individuals who are unable to avoid them.

CONCLUSION

In summary, like its genetics predecessor, the field of epigenetics is rapidly advancing and is leading to the development of new diagnostic tools for the prevention and treatment of common human illnesses. However, as with any new development, these new tools can be used both constructively and destructively. We anticipate here challenges to our understanding of the ways in which epigenetic information may be used in society and in the clinical care of patients. The most adept implementation of these technologies will be fostered by dialogue between all the potential stakeholders, including healthcare providers, patients and experts in the ethical, legal and social issues presented by these technologies in our society.

Acknowledgments

The use of DNA methylation to assess alcohol use status is covered by pending property claims. The use of DNA methylation to assess smoking status is covered by U.S. patent 8,637,652 and other pending claims. Dr. Philibert is a potential royalty recipient on these intellectual right claims. Dr. Philibert is an officer and stockholder of Behavioral Diagnostics Inc. (www.bdmethylation.com). This production of this article was supported by NIH grants R01DA037648 and R43DA037620. Dr. Erwin is a consultant for Behavioral Diagnostics.

References

- Ahlbom A, Lichtenstein P, Malmström H, Feychting M, Pedersen NL, Hemminki K. Cancer in twins: Genetic and nongenetic familial risk factors. *Journal of the National Cancer Institute*. 1997; 89(4): 287–293. DOI: 10.1093/jnci/89.4.287 [PubMed: 9048832]
- Allione A, Marcon F, Fiorito G, Guarrera S, Siniscalchi E, Zijno A, Crebelli R. Novel epigenetic changes unveiled by monozygotic twins discordant for smoking habits. *PLoS ONE*. 2015; 10(6):e0128265.doi: 10.1371/journal.pone.0128265 [PubMed: 26043106]
- Athyros VG, Katsiki N, Doumas M, Karagiannis A, Mikhailidis DP. Effect of tobacco smoking and smoking cessation on plasma lipoproteins and associated major cardiovascular risk factors: A narrative review. *Current Medical Research and Opinion*. 2013; 29(10):1263–1274. DOI: 10.1185/03007995.2013.827566 [PubMed: 23879722]
- Battaller R, North KE, Brenner DA. Genetic polymorphisms and the progression of liver fibrosis: A critical appraisal. *Hepatology*. 2003; 37(3):493–503. DOI: 10.1053/jhep.2003.50127 [PubMed: 12601343]
- Beach SRH, Gerrard M, Gibbons FX, Brody G, Philibert R. A role for epigenetics in broadening the scope of pediatric care in the prevention of adolescent smoking. *Epigenetics Diagnosis and Therapy*. 2015; 1(1):1–7.
- Bell CG, Finer S, Lindgren CM, Wilson GA, Rakyan VK, Teschendorff AE, ... Hitman GA. Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. *PLoS ONE*. 2010; 5(11):e14040.doi: 10.1371/journal.pone.0014040 [PubMed: 21124985]
- Berger M, Gray JA, Roth BL. The expanded biology of serotonin. *Annual Review of Medicine*. 2009; 60:355–366.
- Besingi W, Johansson Å. Smoke related DNA methylation changes in the etiology of human disease. *Human Molecular Genetics*. 2013; doi: 10.1093/hmg/ddt621
- Breitling LP, Salzmann K, Rothenbacher D, Burwinkel B, Brenner H. Smoking, F2RL3 methylation, and prognosis in stable coronary heart disease. *European Heart Journal*. 2012; doi: 10.1093/eurheartj/ehs091
- Cameron AJ, Boyko EJ, Sicree RA, Zimmet PZ, Soderberg S, Alberti KGM, ... Shaw JE. Central obesity as a precursor to the metabolic syndrome in the AusDiab study and Mauritius. *Obesity*. 2008; 16(12):2707–2716. DOI: 10.1038/oby.2008.412 [PubMed: 18820650]

- Centers for Disease Control. Annual smoking-attributable mortality, years of potential life lost, and economic costs—United States, 1995–1999. *Morbidity and Mortality Weekly Report*. 2002; 51:300–303. [PubMed: 12002168]
- Critchley JA, Capewell S. Mortality risk reduction associated with smoking cessation in patients with coronary heart disease: A systematic review. *Journal of the American Medical Association*. 2003; 290(1):86–97. [PubMed: 12837716]
- de Groot MJ, Hoeksma M, Blau N, Reijngoud DJ, van Spronsen FJ. Pathogenesis of cognitive dysfunction in phenylketonuria: Review of hypotheses. *Molecular Genetics and Metabolism*. 2010; 99(Suppl 0):S86–S89. DOI: 10.1016/j.ymgme.2009.10.016 [PubMed: 20123477]
- Dogan MV, Shields B, Cutrona C, Gao L, Gibbons FX, Simons R, Philibert RA. The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. *BMC Genomics*. 2014; 15:151.doi: 10.1186/1471-2164-15-151 [PubMed: 24559495]
- Elliott H, Tillin T, McArdle W, Ho K, Duggirala A, Frayling T, Davey Smith G. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clinical Epigenetics*. 2014; 6(1):4. [PubMed: 24485148]
- Erwin C, Williams JK, Juhl AR, Mengeling M, Mills JA, Bombard Y. I-RESPOND-HD Investigators of the Huntington Study Group. Perception, experience, and response to genetic discrimination in Huntington disease: The international RESPOND-HD study. *American Journal of Medical Genetics B: Neuropsychiatric Genetics*. 2010; 153B(5):1081–1093. DOI: 10.1002/ajmg.b.31079 [PubMed: 20468061]
- Federal News Service. Special White House Briefing. Washington, DC: Author; 2000.
- Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, ... Molloy PL. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences*. 1992; 89(5):1827–1831. DOI: 10.1073/pnas.89.5.1827
- Genetics Information Non-Discrimination Act, Pub. L. No. 122, 881 Stat. 110–233 (2008).
- Guida F, Sandanger TM, Castagné R, Campanella G, Polidoro S, Palli D, Chadeau-Hyam M. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Human Molecular Genetics*. 2015; doi: 10.1093/hmg/ddu751
- Harlid S, Xu Z, Panduri V, Sandler DP, Taylor JA. CpG sites associated with cigarette smoking: Analysis of epigenome-wide data from the Sister Study. *Environment Health Perspectives*. 2014; 122(7):673–678.
- Heinz A, Jones DW, Mazzanti C, Goldman D, Ragan P, Hommer D, Linnoila M, Weinberger DR. A relationship between serotonin transporter genotype and in vivo protein expression and alcohol neurotoxicity. *Biological Psychiatry*. 2000; 47(7):643–649. DOI: 10.1016/S0006-3223(99)00171-7 [PubMed: 10745057]
- Joubert BR, Håberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, London SJ. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environmental Health Perspectives*. 2012; 120(10):1425–1431. [PubMed: 22851337]
- Kremens K, Powers L, Gerke A, Hassan I, Sears R, Philibert R, Monick M. Smoking induced changes in CpG DNA methylation at gene promoter regions target inflammation pathways. *American Journal of Respiratory and Critical Care Medicine*. 2012; 185:A2668.
- Kroeze Y, Zhou H, Homberg JR. The genetics of selective serotonin reuptake inhibitors. *Pharmacology and Therapeutics*. 2012; 136(3):375–400. DOI: 10.1016/j.pharmthera.2012.08.015 [PubMed: 22944042]
- Kulkarni A, Chavan-Gautam P, Mehendale S, Yadav H, Joshi S. Global DNA methylation patterns in placenta and its association with maternal hypertension in pre-eclampsia. *DNA and Cell Biology*. 2011; 30(2):79–84. DOI: 10.1089/dna.2010.1084 [PubMed: 21043832]
- Lanphear BP. The impact of toxins on the developing brain. *Annual Review of Public Health*. 2015; 36(1):211–230. DOI: 10.1146/annurev-publhealth-031912-114413
- Lehtovirta M, Pietiläinen KH, Levälähti E, Heikkilä K, Groop L, Silventoinen K, ... Koskenvuo M. Evidence that BMI and type 2 diabetes share only a minor fraction of genetic variance: a followup study of 23,585 monozygotic and dizygotic twins from the Finnish Twin Cohort Study. *Diabetologia*. 2010; 53(7):1314–1321. DOI: 10.1007/s00125-010-1746-4 [PubMed: 20401462]

- Monick MM, Beach SR, Plume J, Sears R, Gerrard M, Brody GH, Philibert RA. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. *American Journal of Medical Genetics B: Neuropsychiatric Genetics*. 2012; 159B(2):141–151. DOI: 10.1002/ajmg.b.32021 [PubMed: 22232023]
- Nguyen LP, Bradfield CA. The search for endogenous activators of the aryl hydrocarbon receptor. *Chemical Research in Toxicology*. 2007; 21(1):102–116. DOI: 10.1021/tx7001965 [PubMed: 18076143]
- Novitski E. On Fisher's criticism of Mendel's results with the garden pea. *Genetics*. 2004; 166(3): 1133–1136. DOI: 10.1534/genetics.166.3.1133 [PubMed: 15082533]
- Pastor WA, Aravind L, Rao A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature Reviews of Molecular and Cellular Biology*. 2013; 14(6):341–356. DOI: 10.1038/nrm3589 [PubMed: 23698584]
- Philibert RA, Beach SR, Brody GH. Demethylation of the aryl hydrocarbon receptor repressor as a biomarker for nascent smokers. *Epigenetics*. 2012; 7(11):1331–1338. [PubMed: 23070629]
- Philibert R, Beach SR, Li KM, Brody G. Changes in DNA methylation at the aryl hydrocarbon receptor repressor may be a new biomarker for smoking. *Clinical Epigenetics*. 2013; 5:19–26. DOI: 10.1186/1868-7083-5-19 [PubMed: 24120260]
- Philibert R, Hollenbeck N, Andersen E, Osborn T, Gerrard M, Gibbons R, Wang K. A quantitative epigenetic approach for the assessment of cigarette consumption. *Frontiers in Psychology*. 2015; : 6.doi: 10.3389/fpsyg.2015.00656 [PubMed: 25759672]
- Philibert R, Madan A, Andersen A, Cadoret R, Packer H, Sandhu H. Serotonin transporter mRNA levels are associated with the methylation of an upstream CpG island. *American Journal of Medical Genetics*. 2007; 144B(1):101–105. DOI: 10.1002/ajmg.b.30414 [PubMed: 16958039]
- Philibert R, Penaluna B, White T, Shires S, Gunter TD, Liesveld J, Osborn T. A pilot examination of the genome-wide DNA methylation signatures of subjects entering and exiting short-term alcohol dependence treatment programs. *Epigenetics*. 2014a; 9(9):1–7. [PubMed: 24739672]
- Philibert RA, Sandhu H, Hollenbeck N, Gunter T, Adams W, Madan A. The relationship of 5HTT (SLC6A4) methylation and genotype on mRNA expression and liability to major depression and alcohol dependence in subjects from the Iowa Adoption Studies. *American Journal of Medical Genetics B, Neuropsychiatric Genetics*. 2008; 147B(5):543–549. DOI: 10.1002/ajmg.b.30657 [PubMed: 17987668]
- Philibert R, Terry N, Erwin C, Philibert W, Beach SRH, Brody G. Methylation array data can simultaneously identify individuals and convey protected health information: An unrecognized ethical concern. *Clinical Epigenetics*. 2014b;6. [PubMed: 24678997]
- Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Goldstein DB. A genome-wide association study in chronic obstructive pulmonary disease (COPD): Identification of two major susceptibility loci. *PLoS Genetics*. 2009; 5(3):e1000421. [PubMed: 19300482]
- Pittilo M. Cigarette smoking, endothelial injury and cardiovascular disease. *International Journal of Experimental Pathology*. 2000; 81(4):219–230. [PubMed: 10971743]
- Rawls, J. A theory of justice. Cambridge, MA: Belknap Press of Harvard University Press; 1971.
- Rivera MC, Lake JA. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature*. 2004; 431(7005):152–155. DOI: http://www.nature.com/nature/journal/v431/n7005/supinfo/nature02848_S1.html. [PubMed: 15356622]
- Rose, RJ.; Broms, U.; Korhonen, T.; Dick, DM.; Kaprio, J. Genetics of smoking behavior. In: Kim, YK., editor. *Handbook of behavior genetics*. New York: Springer; 2009. p. 411–432.
- Rothstein MA, Anderlik MR. What is genetic discrimination, and when and how can it be prevented? *Genetic Medicine*. 2001; 3(5):354–358.
- Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*. 1975; 94(3):441–448. DOI: 10.1016/0022-2836(75)90213-2 [PubMed: 1100841]
- Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, ... Flanagan JM. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Human Molecular Genetics*. 2012; doi: 10.1093/hmg/dd5488

- Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*. 2008; 9(6):465–476. DOI: 10.1038/nrg2341
- Teschendorff AE, Yang Z, Wong A, Pipinikas CP, Jiao Y, Jones A, ... Widschwendter M. Correlation of smoking-associated DNA methylation changes in buccal cells with DNA methylation changes in epithelial cancer. *JAMA Oncology*. 2015; doi: 10.1001/jamaoncol.2015.1053
- Toperoff G, Aran D, Kark JD, Rosenberg M, Dubnikov T, Nissan B, Hellman A. Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. *Human Molecular Genetics*. 2012; 21(2):371–383. DOI: 10.1093/hmg/ddr472 [PubMed: 21994764]
- Tsaprouni LG, Yang TP, Bell J, Dick KJ, Kanoni S, Nisbet J, Deloukas P. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics*. 2014; 9(10):1382–1396. DOI: 10.4161/15592294.2014.969637 [PubMed: 25424692]
- Tsuang MT, Bar JL, Harley RM, Lyons MJ. The Harvard Twin Study of Substance Abuse: What we have learned. *Harvard Review of Psychiatry*. 2001; 9(6):267–279. [PubMed: 11600486]
- Villablanca AC. Nicotine stimulates DNA synthesis and proliferation in vascular endothelial cells in vitro. *Journal of Applied Physiology*. 1998; 84(6):2089–2098. [PubMed: 9609804]
- Watson J. The human genome project: past, present, and future. *Science*. 1990; 248(4951):44–49. DOI: 10.1126/science.2181665 [PubMed: 2181665]
- Willi C, Bodenmann P, Ghali WA, Faris PD, Cornuz J. Active smoking and the risk of type 2 diabetes: A systematic review and meta-analysis. *Journal of the American Medical Association*. 2007; 298(22):2654–2664. DOI: 10.1001/jama.298.22.2654 [PubMed: 18073361]
- Zaghlool S, Al-Shafai M, Al Muftah W, Kumar P, Falchi M, Suhre K. Association of DNA methylation with age, gender, and smoking in an Arab population. *Clinical Epigenetics*. 2015; 7(1):6. [PubMed: 25663950]
- Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Illig T. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS ONE*. 2013; 8(5):e63812.doi: 10.1371/journal.pone.0063812 [PubMed: 23691101]

Figure 1.
The heritability of selected common complex medical disorders.

Genetic and Epigenetic Regulation of the Serotonin Transporter

Figure 2.
The structure of the serotonin transporter gene (SL6A4 or 5HTT).

CAAGGGCCTCTAAGCTGAGCTCACATCCCAGCCGGTCAGTCAGATAAACG
 CATGGGTATCGAGTACTGCTAGGTCCCAGGAAGAAAGAGAGAGCAGCTTT
CGGGATGGGACGATGGGAGGTGTCCGAGGTCAAGAGAAAGCGGCAGCA
 GCAGACCCCTGTGTGC**CGTCCTGTGGGCGCGGGCGGCAGGGGAGGCGCA**
 CACCTGCTCCTTTGTGCAGCCTCCCCCTCC**CGCAAAGTTAAAGAGCAGG**
 AAAGTCAGGATTCC**CTCGCTCGGCCCTGCCCTGCGGCTGCTCCGCGCTCC**
GCTCCTCCCTGCGAGCGTGTGTGTGTGTGGGGTCCCTCCCTCCTGGC
 TCTGGGGT**CGGGCGCGCACCCCGCCCCGTAGCGCGGCCCTCCCTGGCGA**
CGCGAACCCCATCCAGCGGGAGCGCGGAGCCGCGGC CGCGGGGAAGCATT
 AAGTTTATT**CGCCTCAAAGTGACGCAAAATTCTTCAAGAGCTCTTTGGC**
GGCGGCTATCTAGAGATCAGACCATGTGAGGGCCCGCGGGTACAAA**TACG**
GC CGCGCGCGCGCCCCCTCGCACAGCCAGCGCGCGCGGGTGCCTCGAGGG
CGCGAGGCCAGCCCGCCTGCCAGCCCGGGACCAGCCTCCC CGCGCAGCC
TGGCAGGTGGGTCCGCTTTTCTCTCCGCTCGAACCCACGTTTCTTTCC
 AGACCTTCTTCCC**CGCCTCGGGGAGGGGATAGAACCGCTGCGCCCCACC**
GCCCTGCGAGGAGGCGAGGAGGTGCATGCGCCCCAGCGGTGGGCGCGGA
 TCCTGCCCCT**CGCCCTCCACGCTCAGCAAGAGCCAGAGCTGAAGCTGAC**
CGGCCAGAGTGGGAGACGAGGAACGTGGAGTGCTCGAAGTGGGCGGGCGT
 AGGGGGCTCCTTTGTCTATTGTTGCAGGGGCTTTGCAACCTCTTGAGGAC
 CCGTGAAAGACCCCTTAAAGAGGGTCTTGAGCAAAGTGCTGTCCTGCGC

Figure 3.

The nucleotide sequence of the 799 base pair CpG island that regulates serotonin transporter transcription. The island consists of 83 CpG residues (bold and red) and completely envelops exon 1A (underlined). The six base pair motif indicating a TATA box (indicated by boxed text) is immediately upstream of exon 1A.

Figure 4.

A depiction of the relationship of DNA methylation to chromosome condensation. In brief, in response to either internal and external cellular cues, enzymes referred to as DNA methyltransferases (DNMTs) transfer a methyl group from a methyl donor such as folate or methionine to the cytosine residue of a CpG dinucleotide pair. When sufficient numbers of CpG residues in a given area have been methylated and the histone scaffold that is associated with the DNA region has been deacetylated by histone deacetylase (HDAC), the region is condensed into a tightly coiled, transcriptionally inactive chromatin conformation referred to as heterochromatin.

Figure 5.
Bisulfite conversion.

Figure 6.
Using DNA methylation to determine smoking status.

Table 1

Results of attempts to replicate the original findings (Monick et al., 2012) using independent populations with respect to smoking and methylation status at cg05575921

Author	Significance			Population
	Year	Rank (of 485 557 probes)	level	
Philibert	2012	1st	3×10^{-7}	adolescents
Joubert	2012	1st	8×10^{-33}	newborns
Shenker	2012	1st	2×10^{-15}	adults
Philibert	2013	1st	2×10^{-3}	young adults
Zeilinger	2013	1st	3×10^{-182}	adults
Dogan	2014	2nd	6×10^{-19}	adults
Besingi	2014	1st	7×10^{-70}	adults
Elliot	2014	1st	6×10^{-59}	adults
Tsaprouni	2014	1st	9×10^{-69}	adults
Harlid	2014	2nd	2×10^{-2}	adults
Guida	2015	1st	1×10^{-106}	adults
Zaghlool	2015	1st	7×10^{-7}	adults
Allione	2015	1st	N.A.	adults
Teschendorf	2015	1st	8×10^{-60}	adults