Published in final edited form as:

Am J Med Genet B Neuropsychiatr Genet. 2017 September ; 174(6): 589–594. doi:10.1002/ajmg.b.32569.

Optimizing the Chances of Success in the Search for Epigenetic Biomarkers: Embracing Genetic Variation

Robert Philibert, MD PhD^{1,2} and Stephen J Glatt, PhD³

¹Behavioral Diagnostics, Coralville, Iowa 52241

²Department of Psychiatry, University of Iowa, Iowa City, IA 52242

³Psychiatric Genetic Epidemiology & Neurobiology Laboratory (PsychGENe Lab), Department of Psychiatry and Behavioral Sciences, SUNY Upstate Medical University, 750 East Adams Street, Syracuse, NY 13210

Abstract

The emphasis on clinical translation in biomedical research continues to grow. This focus has been particularly notable in those investigators using epigenetic approaches to decipher the biology of complex behavioral disorders. As a result of these efforts, reproducible findings for several disorders, such as smoking, have been generated, giving rise to hopes that biomarkers for other behavioral illnesses would be forthcoming. Unfortunately, that biomedical cornucopia has not yet materialized. In this editorial, we review progress to date and discuss barriers to generating epigenetic biomarkers for complex behavioral disorders. We highlight the need to incorporate information on genetic variation and develop more powerful bioinformatics tools in order to optimize the likelihood of success. We emphasize that searches should focus on clearly defined, readily distinguishable behavioral constructs and suggest that some well-intentioned methods, such as correction for cellular heterogeneity, may actually impede the identification of clinically relevant biomarkers in peripheral blood. Finally, we describe how the understanding created by the development of these biomarkers may lead to more valid animal models of neuropsychiatric illness. We conclude that the prospects for epigenetic biomarkers for complex disorders are bright, but emphasize that the journey to the clinical implementation of these findings will be a slow, iterative process.

Contrary to popular belief, the concept of “epigenetics” is not a new one. Although controversial, some have argued that Lamarckism touched upon, or at least foreshadowed, many concepts which are now being refined throughout the field. Hence, for this and many other reasons, it is probably both accurate and charitable to assert that our current foundation for the understanding of behavioral genetics has been contributed by a large number of individuals over an extended period of time.

This more general statement is also true for the more restricted area of substance use epigenetics. For example, with respect to alcohol intake, a correlation between alcohol-

associated alterations in folate cycle enzymes and increased risk of colon cancer had been suspected as early as 1999 by Slattery and colleagues while a link between changes in p16 DNA methylation in lung cancer, which is overwhelmingly associated with smoking, was reported in 2002 by Belinsky and associates.^{1,2} Over the subsequent several years, a steady drumbeat of studies examining the relationship between changes in DNA methylation status with various forms of substance use-related diseases, generally cancer, was published.³⁻⁶ Sadly, despite the interest of the scientific community, at these earlier times, a number of factors precluded more rapid progress with respect to substance use.

In these prior years, the primary barrier was the inadequacy of techniques to quantify methylation. The earliest studies used methyl sensitive restriction enzymes and hybridization techniques to quantify relative methylation.⁷ The subsequent coupling of bisulfite conversion and other laboratory techniques to increasingly sophisticated nucleic acid sequencing technologies lead to increasingly more sophisticated understandings of the methylome.^{8,9} Still, this progress in measurement approaches was incremental. For example, using one of these earlier technologies, matrix-assisted laser desorption ionization time of flight mass spectroscopy (MALDI-TOF MS),¹⁰ in this journal, we published the first reproducible findings of smoking associated methylation changes at a specific gene, *monoamine oxidase A*. Over the subsequent years, the use of MALDI-TOF MS based approaches for single-locus assessments of methylation was complemented by introduction of other single-locus technologies including pyrosequencing.¹¹ Finally, methylation array technologies advanced to the point that genome-wide studies could be conducted¹² which allowed investigators to systematically assess the epigenetic response to smoking. Although each of the technologies used to assess DNA methylation status had and continues to have its limitations,¹³ it is irrefutable that over the past decade the precision and scope of methylation assessment tools are no longer the primary barriers to the elucidation of a better understanding of behavioral genetics.

A second barrier to progress was and continues to be an inadequate conceptual framework for the molecular mechanisms through which changes in the environment alter the “epigenetic code.” Part of this failing is quite understandable: like the genome, the epigenome is extraordinarily complex. For example, much to the chagrin of investigators, including ourselves, hoping for concise mechanisms with clear boundaries (e.g., a small change in methylation status at a single CpG residue results in both altered transcription and behavior), it has become evident that the individual components of the “epigenome” including the “methylome” and the “histone code” are extraordinarily complex at the individual level and interactive at the systems level. What is more, the patterns of these structures and interactions are both chronologically and anatomically contextual. Finally, these modifiable factors interact with the underlying genetic variation in a tissue-specific and chronologically contextual manner to regulate gene expression. Therefore, although regrettable, it is rather understandable to realize that a major impediment to our understanding of substance use epigenetics was a lack of appreciation for the complexity of the epigenome.

Not all of these failures in efforts to identify epigenetic biomarkers were secondary to the complexity of systems biology. Perhaps the avoidable foible has been the failure to fully

embrace the simplicity of the term “biomarker”. Simply defined a biomarker is “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention”.¹⁴ The value of such a measurement is dependent on the sensitivity and specificity of the characteristic for the given condition. Whereas the predictive value of the biomarker can be achieved by a better understanding of what drives changes in that characteristic, it is not necessary to understand the mechanism of action for the biomarker to successfully employ it. For example, for millennia, we have used patient temperature (i.e., fever) as an indicator of illness. The exact mechanism(s) through which infections cause elevated temperatures are still being fully delineated. Nevertheless, the value of this biomarker as a measurement of patient status is indicated by its usage as a core component of every physical exam.

This point is particularly relevant for those of us conducting genome-wide epigenetic studies in order to identify biomarkers for substance use. In this regard, numerous genome-wide and single-locus studies have shown that peripheral blood DNA methylation at cg05575921, a well-studied CpG residue in the aryl hydrocarbon receptor repressor (*AHRR*), predicts smoking status.^{15–17} To a certain extent, the mechanism through which this reliability is achieved is irrelevant as long as cg05575921’s methylation status reliably predicts smoking status. Analyzing cellular composition of the blood could conceivably tell us whether changes in cellular composition are associated with the change in average DNA methylation at cg05575921 and lead to improvements in how the assay is constructed. But whether or not such differences are present would not change the reliability of cg05575921 methylation status as a predictor of smoking status as currently assessed.

Over the past several years there has been an increasing demand by many to use corrections for cellular heterogeneity in the primary analysis of genome-wide studies attempting to identify biomarkers.¹⁸ Although well intentioned, these demands can also be counterproductive because the degrees of freedom consumed by these correction techniques often substantially erode, if not eradicate, statistical power to detect meaningful results. Furthermore, an increasing number of studies have shown that these approaches do not improve the results.^{19–21} As a consequence, we agree with Bauer and colleagues¹⁹ and recommend that those conducting genome-wide methylation studies in the hopes of identifying biomarkers for a given trait report results both with and without adjustments for cellular heterogeneity. This approach allows the possibility that blood sample cell composition itself may have biomarker utility. Regardless, powerful biomarkers can be identified and the cause of advancing healthcare aided even if we are totally ignorant of the mechanism leading to the change (e.g., whether an epigenomic biomarker is detected because of or despite a disease-associated change in cell composition)..

A less than perfect appreciation of the term “biomarker” has not been the only conceptual obstacle. These failings also extend to our clinical classifications of substance use. Perhaps the clearest example of this was the failure to appreciate the difference between state and trait. The vast majority of early studies, including several of our own, utilized lifetime categorical diagnoses as the principal dependent variable in analyses.^{22–24} Unfortunately, this type of approach resulted in the inclusion of individuals who used substances at an

earlier point in their life, but may have become abstinent for as many as 10 to 20 years. Since subsequent analyses have clearly shown that methylation signatures associated with substance use consumption may revert, sometimes rather rapidly, to baseline values,^{15,19,25} this lack of understanding resulted in the inclusion of individuals whose methylation signatures may have more resembled the “unaffecteds” in the “affected” category. Subsequent appreciation of the pliability of the methylome has led to improvements in both the chronological and quantitative consumption phenotyping of subjects used in epigenetic studies.

These improvements in phenotyping are drastically needed. In this issue, Allan Andersen details the extent of unreliable self-report in adult populations and the improvement that biomarkers, in this case both ELISA and DNA methylation assessments, can bring to subject classification.²⁶ Alarming, he shows nearly half of the subjects from a high-risk cohort with clear biochemical evidence of smoking unreliaibly report their smoking status. Since misclassification has a drastic impact on power,²⁷ ironically noted with regard to smoking status as early as 1965,²⁸, these findings may give pause to those seeking to generate diagnostic biomarker panels for major depression and general anxiety disorders, each of which have low inter-rater reliability.

However, a more disabling bias in the conceptual framework was a lack of appreciation of the protean effects of substance use on the methylome. Smoking and drinking are the 1st and 3rd leading causes of preventable morbidity and mortality in the United States.²⁹ Without a doubt, the excessive consumption of these substances is mediated through central nervous system (CNS) processes. Rightfully so, as a result, many investigators have focused their attention on the epigenetic properties of areas of the CNS that subserve reward behaviors. However, the assertion that exposure of human tissues to these environmental toxins could alter methylation signatures in non-CNS peripheral tissues independent of active co-morbid disease processes such as cancer or substance dependence met steady resistance in earlier years. Sadly, part of this resistance was attributable to the continuing stigma against those with behavioral illnesses.³⁰ Over the past several years, thanks to the strong educational efforts by numerous individuals and entities, this barrier has diminished. However, a close corollary, a lack of comprehension of the chronologically contextual nature of shared and unshared effects of substance use exposure amongst tissues from the same individual, still is a limitation. To a small extent, some of this limitation has been alleviated by investigators who have included assessments of multiple tissues from the same individual in their studies. However, a more comprehensive set of studies, such as that which could be generated from resources such as the GTEx Consortium,³¹ should markedly accelerate the pace of the acquisition of this crucial knowledge.

Taking Advantage of an Interactive View of the Epigenome

Just as the epigenetic signature of a cell or tissue is environmentally or developmentally contextual, the signature is also genetically contextual. For studies of humans, this has been both a blessing and a curse. For our conduct of earlier studies of the effects of smoking on *MAOA* methylation in which we showed the degree of demethylation in response to smoking was dependent on *MAOA* VNTR genotype, it was much more the latter than the

former for several reasons. Obviously, the loss of degrees of freedom secondary to the need to correct for incorporation of genotype into our regression analyses diminished the power of the subsequent calculations. But the second reason was a little more surprising; we assumed that because we had genotyped the well characterized VNTR system originally described by Sabol and colleagues,³² that we had identified the most important genetic modifier. The subsequent demonstration of another more extensive VNTR system immediately upstream whose status is more tightly linked to *MAOA* promoter methylation status quickly dispelled that belief.³³ Despite that discovery, the observation that the diminished methylation at *MAOA* is associated with the allele that is more active in standard *in vitro* assays suggests that even this understanding is incomplete and that attempts to generate an understanding of interactions between the methylome and genome for some loci will be challenging.

In this issue, we take the next step in this journey using the resources of the Framingham Heart Study which includes complete genome-wide genetic and epigenetic data for a large number of subjects. Using these resources and an analysis strategy that plumbs both the direct (no knowledge of genetic status) and genetically contextual effects of smoking on the methylome, we demonstrate that the effects of smoking on the epigenome extend beyond our wildest expectations.³⁴ Specifically, we identify over 4000 genes whose methylation status was affected by smoking and show that the SNPs modifying the strength of many of those changes map to loci known to be associated with disease. In essence, by removing the masking effects of *cis* and *trans* SNP interaction effects, we show that even that number is a vast underestimate of the extent of genetic remodeling, and that it appears the methylomic signatures of the vast majority of genes in the genome are somehow modified in association with smoking.

However, this first of what is likely to be many integrated genetic-epigenetic genome-wide studies is just the beginning of the journey of understanding. One part of the next portion of the journey may be to better understand which genetic variation at a given locus identified by the tag SNP is interacting with the methylation status at a given CpG site. Although there are likely a variety of methods, one approach that has been successfully implemented is illustrated by the work of Vockley and colleagues who used chromatin immunoprecipitation (ChIP) to create a library of glucocorticoid (GC) binding sites (GBS) to identify those local GBS directly coupled to transcription as well more distal GBS sites not directly coupled to transcription but having epigenetic features of enhancer elements for glucocorticoid stimulated expression.³⁵ When required, the functionality of individual regulatory elements then can be explored by searching for eQTL in consensus motifs contained at these sites or by site-directed mutagenesis of critical sequence using CRISPR as recently demonstrated by Tewhey and colleagues.³⁶ The net result will be a better understanding of the portion of the genomes that regulate GC signaling.

These molecular approaches may come in handy for those defining the epigenomics of post-traumatic stress disorder (PTSD). Dysregulation of GC signaling is a core feature of the pathophysiology of PTSD. Unfortunately, despite the wealth of analyses devoted to this subject, clear reproducible findings have been slow to emerge. In many ways, this is not surprising and may be secondary to the strong degree of clinical heterogeneity known to

exist for this anxiety spectrum disorder. However, to overcome the challenges posed by this heterogeneity, several tools for managing large datasets must emerge. In this issue, Alicia Smith and colleague describe one tool, a DNA methylation pipeline, for handling the increasing number of arrays that are informative for PTSD.³⁷ When coupled to molecular techniques such as those described above which prioritize certain regions of the genome, the increased power of additional arrays combined with a reduction in the complexity of the genome may lead to the first reproducible epigenetic findings.

However, trauma is not the only environmental variable whose role in fine-scale sculpting of the epigenome will benefit from creation of large-scale data pipelines such as those described by Smith and colleagues. In this issue, Steve Beach and associates describe the strong effects that nutrient levels such as folate may have on altering methylation levels.³⁸ In this case, the Beach collaboration uses methylation levels of methylene tetrahydrofolate reductase (*MTHFR*), a key enzyme in the methyl donor cycle, as a proxy for folate levels. They show that methylation of the CpG island covering the first exon moderates the changing methylation at cg05575921 in response to smoking. Because folate levels co-segregate with a number of other environmental factors such as obesity and diet,³⁹ it is likely that a large-scale meta-analysis of methylation data that is also informative for key dietary parameters will be necessary to dissect the relationship between nutrient levels and the efficiency of the cellular epigenetic apparatus. Given the complexity of the methylation and de-methylation pathways, it is likely that these relationships will be complicated as well.

To a large extent, these “complicated” relationships will be moderated by the microbiome. An unnerving discovery for many was the realization that our health and perhaps our existence is dependent on a symbiotic relationship with a rich variety of micro-organisms.⁴⁰ Over the past several years, a number of investigators have shown the important role this flora play in moderating vulnerability to more conventional medical disorders such as diabetes and obesity. In this issue, Abdolmeleky and colleagues peek behind the door and discuss the growing evidence of effects that the microbiome has on behavioral health.⁴¹ Over the next several decades, it is very likely that continuing advances in this area of research will lead to a fuller and more holistic understanding of the relationship of inflammation on all aspects of human health—including behavior—as well as insights to mechanisms for scalable preventive interventions.

Still A Need for Animal Models

No matter what is found with respect to peripheral biomarkers, it is clear that animal models will be needed to complete our understanding of CNS disease pathogenesis and point the way to better prevention and therapeutics. This is particularly clear for smoking. For example, the smoking biomarker described by Allan Andersen may find a significant role in the assessment and treatment of those with tobacco use disorder. However, by itself, the cg05575921 only measures biological processes in peripheral white blood cells. It does not assess the biological processes in the brain that moderate the initiation and maintenance of the smoking behavior.

To an extent, we can learn quite a bit about the biology of these central processes through clinical examinations of smokers. However, to fully understand the mechanisms through which the interplay of genetic variation with key environmental variables drives these and other addictions, animal models will be absolutely critical. But even using mice, these studies will not be easy to conduct. First, it will be necessary to fully understand the role of the epigenome in shaping the murine CNS. Then, it will be necessary to be able to introduce, in a controlled manner, human genetic variation that may affect addictive processes. Currently, the only genetic risk factors generally accepted with respect to smoking are found in the CHRNA5 region of chromosome 15.⁴² However, if even a fraction of the significant *cis* and *trans* interactions in whole blood DNA noted by Dogan and colleagues with respect to tobacco smoke exposure (some of which involve loci in the nicotinic receptor cluster on chromosome) are evident in the CNS as well and moderate vulnerability to smoking, modeling the important interplay in tobacco use disorders may be a substantial endeavor.

In the final article in this issue, Javidar and colleagues illustrate one pathway through which this modeling process for each putative genetic or methylation variant can be accomplished with respect to schizophrenia.⁴³ Specifically, they describe approaches that feature the CRISPR-Cas family of nucleases to either introduce human specific genetic variation associated with schizophrenia that may affect the local epigenetic signature or recruit chromatin modifiers to key neurodevelopmental gene in mice. Because, unlike in humans, we can repeatedly assess the effects of these modifications throughout the murine lifecycle and can use a broad array of tools, such as tissue- or cell-specific gene expression drivers, to pinpoint a particular area of the brain, the possibility exists that we may finally begin to understand through which mechanism environmental variability moderates the effects of specific risk variants on the expression of the development of psychosis. If so, that could be a pivotal point in the development of personalized medicine for those at risk for the major psychiatric disorders.

In summary, in many ways, our journey to a fuller understanding of the epigenome has faced many of the same hurdles faced by pioneers in the genome project, namely the need for better tools and more exact diagnoses. At the same time, the challenges posed by the interactive nature of the various layers of the epigenome, such as the methylome and histone code, within themselves, with the genome as well as with extracellular signals, are new and unexpected complications. To overcome these barriers and develop better treatments and interventions, investigators will need expertise in any of the number of unique knowledge domains, ranging from psychology and microbiology to high performance computing and animal models, that can be best acquired through collaborative networks. Nevertheless, the current progress to date in isolating the low-hanging fruit in the methylome, such as markers for tobacco smoke exposure, is promising and suggests the likelihood that biomarkers for more complex disorders, such as coronary heart disease, are in the offing. Processes that facilitate collaborations between clinicians, bench scientists and bioinformaticians could accelerate the pace of those discoveries and lead to significant advancements in the prevention, diagnosis and treatment of a wide variety of medical disorders.

Acknowledgments

Dr. Philibert is supported by U.S. NIH grants R44 CA213507, R44DA041014, R44AA022041 and R01DA037648. He is also the Chief Executive Officer of Behavioral Diagnostics and inventor on a number of granted and pending patent applications with respect to both alcohol and tobacco consumption, and a variety of other medical disorders. Dr. Glatt is supported by U.S. NIH grants R01MH101519 and R01AG054002.

References

- Slattery ML, Potter JD, Samowitz W, Schaffer D, Leppert M. Methylenetetrahydrofolate Reductase, Diet, and Risk of Colon Cancer. *Cancer Epidemiology Biomarkers & Prevention*. 1999; 8:513–518.
- Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, Grimes MJ, Harms HJ, Tellez CS, Smith TM, Moots PP, Lechner JF, Stidley CA, Crowell RE. Aberrant Promoter Methylation in Bronchial Epithelium and Sputum from Current and Former Smokers. *Cancer Res*. 2002; 62:2370–2377. [PubMed: 11956099]
- Bonsch D, Lenz B, Reulbach U, Kornhuber J, Bleich S. Homocysteine associated genomic DNA hypermethylation in patients with chronic alcoholism. *J Neural Transm*. 2004; 111:1611–1616. [PubMed: 15565495]
- Chang HW, Ling GS, Wei WI, Yuen AP. Smoking and drinking can induce p15 methylation in the upper aerodigestive tract of healthy individuals and patients with head and neck squamous cell carcinoma. *Cancer*. 2004; 101:125–132. [PubMed: 15221997]
- Jiang YH, Bressler J, Beaudet AL. Epigenetics and human disease. *Annu Rev Genomics Hum Genet*. 2004; 5:479–510. [PubMed: 15485357]
- Marutha Ravindran CR, Ticku MK. Changes in methylation pattern of NMDA receptor NR2B gene in cortical neurons after chronic ethanol treatment in mice. *Brain Res Mol Brain Res*. 2004; 121:19–27. [PubMed: 14969733]
- Singer-Sam J, LeBon JM, Tanguay RL, Riggs AD. A quantitative HpaII-PCR assay to measure methylation of DNA from a small number of cells. *Nucleic Acids Res*. 1990; 18:687. [PubMed: 1689825]
- Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL. 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; 89:1827–1831.
- Thomassin H, Oakeley EJ, Grange T. Identification of 5-methylcytosine in complex genomes. *Methods*. 1999; 19:465–475. [PubMed: 10579942]
- Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, Cantor CR, Field JK, van den Boom D. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci USA*. 2005; 102:15785–15790. [PubMed: 16243968]
- Tost J, Gut IG. DNA methylation analysis by pyrosequencing. *Nat Protocols*. 2007; 2:2265–2275. [PubMed: 17853883]
- Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan JB, Shen R. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011; 98:288–295. [PubMed: 21839163]
- Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, Jager N, Gnirke A, Stunnenberg HG, Meissner A. Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotechnol*. 2010; 28:1106–1114. 3066564. [PubMed: 20852634]
- Colburn W, DeGruttola V, DeMets D, Downing G, Hoth D, Oates J, Peck C, Schooley R, Spilker B, Woodcock J. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Biomarkers Definitions Working Group. *Clinical Pharmacol & Therapeutics*. 2001; 69:89–95.
- Philibert R, Hollenbeck N, Andersen E, McElroy S, Wilson S, Vercande K, Beach SR, Osborn T, Gerrard M, Gibbons FX. Reversion of AHRR Demethylation Is a Quantitative Biomarker of Smoking Cessation. *Frontiers in Psychiatry*. 2016; 7

16. Philibert R, Hollenbeck N, Andersen E, Osborn T, Gerrard M, Gibbons R, Wang KA. Quantitative Epigenetic Approach for the Assessment of Cigarette Consumption. *Front Psychol.* 2015; 6
17. Shenker NS, Ueland PM, Polidoro S, van Veldhoven K, Ricceri F, Brown R, Flanagan JM, Vineis P. DNA methylation as a long-term biomarker of exposure to tobacco smoke. *Epidemiology.* 2013; 24:712–716. [PubMed: 23867811]
18. Houseman, EA. DNA Methylation and Cell-Type Distribution. In: Teschendorff, AE., editor. *Computational and Statistical Epigenomics.* Vol. 7. Springer; Netherlands: 2015. p. 35–50.
19. Bauer M, Fink B, Thürmann L, Eszlinger M, Herberth G, Lehmann I. Tobacco smoking differently influences cell types of the innate and adaptive immune system—indications from CpG site methylation. *Clin Epigenetics.* 2016; 8:83.
20. Dogan MV, Shields B, Cutrona C, Gao L, Gibbons FX, Simons R, Monick M, Brody G, Tan K, Philibert R. The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. *BMC Genomics.* 2014; 15:151. [PubMed: 24559495]
21. Joubert, Bonnie R., Felix, Janine F., Yousefi, P., Bakulski, Kelly M., Just, Allan C., Breton, C., Reese, SE., Markunas, Christina A., Richmond, Rebecca C., Xu, CJ., Küpers, Leanne K., Oh, Sam S., Hoyo, C., Gruzieva, O., Söderhäll, C., Salas, Lucas A., Baiz, N., Zhang, H., Lepeule, J., Ruiz, C., Ligthart, S., Wang, T., Taylor, Jack A., Duijts, L., Sharp, Gemma C., Jankipersadsing, Soesma A., Nilsen, Roy M., Vaez, A., Fallin, MD., Hu, D., Litonjua, Augusto A., Fuemmeler, Bernard F., Huen, K., Kere, J., Kull, I., Munthe-Kaas, Monica C., Gehring, U., Bustamante, M., Saurel-Coubizolles, Marie J., Quraishi, Bilal M., Ren, J., Tost, J., Gonzalez, Juan R., Peters, Marjolein J., Håberg, Siri E., Xu, Z., van Meurs, Joyce B., Gaunt, Tom R., Kerkhof, M., Corpeleijn, E., Feinberg, Andrew P., Eng, C., Baccarelli, Andrea A., Benjamin Neelon, Sara E., Bradman, A., Merid, Simon K., Bergström, A., Herceg, Z., Hernandez-Vargas, H., Brunekreef, B., Pinart, M., Heude, B., Ewart, S., Yao, J., Lemonnier, N., Franco, Oscar H., Wu, Michael C., Hofman, A., McArdle, W., Van der Vlies, P., Falahi, F., Gillman, Matthew W., Barcellos, Lisa F., Kumar, A., Wickman, M., Guerra, S., Charles, MA., Holloway, J., Auffray, C., Tiemeier, Henning W., Smith, George D., Postma, D., Hivert, MF., Eskenazi, B., Vrijheid, M., Arshad, H., Antó, Josep M., Dehghan, A., Karmaus, W., Annesi-Maesano, I., Sunyer, J., Ghantous, A., Pershagen, G., Holland, N., Murphy, Susan K., DeMeo, Dawn L., Burchard, Esteban G., Ladd-Acosta, C., Snieder, H., Nystad, W., Koppelman, Gerard H., Relton, Caroline L., Jaddoe, Vincent WV., Wilcox, A., Melén, E., London, Stephanie J. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *The American Journal of Human Genetics.* 2016; 98:680–696. [PubMed: 27040690]
22. Philibert RA, Gunter TD, Beach SR, Brody GH, Madan A. MAOA methylation is associated with nicotine and alcohol dependence in women. *Am J Med Genet.* 2008; 147B:565–570. [PubMed: 18454435]
23. Bleich S, Lenz B, Ziegenbein M, Beutler S, Frieling H, Kornhuber J, Bonsch D. Epigenetic DNA Hypermethylation of the HEP Gene Promoter Induces Down-regulation of Its mRNA Expression in Patients With Alcohol Dependence. *Alcoholism: Clinical and Experimental Research.* 2006; 30:587–591.
24. Hillemecher T, Frieling H, Hartl T, Wilhelm J, Kornhuber J, Bleich S. Promoter specific methylation of the dopamine transporter gene is altered in alcohol dependence and associated with craving. *J Psychiatr Res.* 2009; 43:388–392. [PubMed: 18504048]
25. Brückmann C, Di Santo A, Karle KN, Batra A, Nieratschker V. Validation of differential GDAP1 DNA methylation in alcohol dependence and its potential function as a biomarker for disease severity and therapy outcome. *Epigenetics.* 2016:00–00.
26. Andersen A, Philibert R, Gibbons FX, Simons RL, Long J. Accuracy and Utility of an Epigenetic Biomarker for Smoking in Populations with Varying Rates of False Self-Report. *American Journal of Medical Genetics, Part B.* ((In Press)).
27. Silverberg MS, Daly MJ, Moskovitz DN, Rioux JD, McLeod RS, Cohen Z, Greenberg GR, Hudson TJ, Siminovitch KA, Steinhart AH. Diagnostic misclassification reduces the ability to detect linkage in inflammatory bowel disease genetic studies. *Gut.* 2001; 49:773–776. [PubMed: 11709510]

28. Mote VL, Anderson RL. An Investigation of the Effect of Misclassification on the Properties of Chi Square Tests in the Analysis of Categorical Data. *Biometrika*. 1965; 52:95–109. [PubMed: 14341284]
29. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. *JAMA*. 2004; 291:1238–1245. [PubMed: 15010446]
30. Ledford H. Medical Research: If Depression were Cancer. *Nature*. 2014; 515
31. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, Compton CC, DeLuca DS, Peter-Demchok J, Gelfand ET, Guan P, Korzeniewski GE, Lockhart NC, Rabiner CA, Rao AK, Robinson KL, Roche NV, Sawyer SJ, Segrè AV, Shive CE, Smith AM, Sobin LH, Undale AH, Valentino KM, Vaught J, Young TR, Moore HM. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation and Biobanking*. 2015; 13:311–319. [PubMed: 26484571]
32. Sabol SZ, Hu S, Hamer D. A functional polymorphism in the monoamine oxidase A gene promoter. *Hum Genet*. 1998; 103:273–279. [PubMed: 9799080]
33. Philibert RA, Wernett P, Plume J, Packer H, Brody GH, Beach SR. Gene environment interactions with a novel variable Monoamine Oxidase A transcriptional enhancer are associated with antisocial personality disorder. *Biol Psychol*. 2011; 87:366–371. 3134149. [PubMed: 21554924]
34. Dogan MV, Beach SRH, Philibert R. Genetically Contextual Effects of Smoking on Genome Wide DNA Methylation. *American Journal of Medical Genetics, Part B*. ((In Press)).
35. Vockley, Christopher M., D'Ippolito, Anthony M., McDowell, Ian C., Majoros, William H., Safi, A., Song, L., Crawford, Gregory E., Reddy, Timothy E. Direct GR Binding Sites Potentiate Clusters of TF Binding across the Human Genome. *Cell*. 2016; 166:1269–1281.e1219. [PubMed: 27565349]
36. Tewhey R, Kotliar D, Park Daniel S, Liu B, Winnicki S, Reilly Steven K, Andersen Kristian G, Mikkelsen Tarjei S, Lander Eric S, Schaffner Stephen F, Sabeti Pardis C. Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. *Cell*. 2016; 165:1519–1529. [PubMed: 27259153]
37. Ratanatharathorn A, Youssef NA, PGC PTSD Epigenetics Workgroup. Epigenome-wide Association of PTSD from Heterogeneous Cohorts with a Common Multi-Site Analysis Pipeline. *American Journal of Medical Genetics, Part B*. ((In Press)).
38. Beach SRH, Kit-Lei MK, Ong ML, Brody GH, Dogan MV, Philibert R. MTHFR Methylation Moderates the Impact of Smoking on DNA Methylation at AHRR. *American Journal of Medical Genetics, Part B*. ((In Press)).
39. Bird JK, Ronnenberg AG, Choi SW, Du F, Mason JB, Liu Z. Obesity Is Associated with Increased Red Blood Cell Folate Despite Lower Dietary Intakes and Serum Concentrations. *The Journal of Nutrition*. 2015; 145:79–86. [PubMed: 25527662]
40. Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI. Human nutrition, the gut microbiome and the immune system. *Nature*. 2011; 474:327–336. [PubMed: 21677749]
41. Abdolmaleky H, Alam R, Zhou J-R. Microbiome, inflammation, epigenetic alterations and mental diseases. *American Journal of Medical Genetics, Part B*. ((In Press)).
42. Bierut LJ, Stitzel JA, Wang JC, Hinrichs AL, Grucza RA, Xuei X, Saccone NL, Saccone SF, Bertelsen S, Fox L, Horton WJ, Breslau N, Budde J, Cloninger CR, Dick DM, Foroud T, Hatsukami D, Hesselbrock V, Johnson EO, Kramer J, Kuperman S, Madden PAF, Mayo K, Nurnberger J Jr, Pomerleau O, Porjesz B, Reyes O, Schuckit M, Swan G, Tischfield JA, Edenberg HJ, Rice JP, Goate AM. Variants in Nicotinic Receptors Risk for Nicotine Dependence. *Am J Psychiatry*. 2008; 165:1163–1171. [PubMed: 18519524]
43. Javidfar B, Park R, Kassim B, Bicks L, Akbarian S. The Epigenomics of Schizophrenia, in the Mouse. *American Journal of Medical Genetics, Part B*. (in press).