Epigenetics 5:7, 578-582; October 1, 2010; © 2010 Landes Bioscience

Allele-specific methylation in the human genome

Implications for genetic studies of complex disease

Emma L. Meaburn,[1,2] Leonard C. Schalkwyk[1] and Jonathan Mill[1,*]

[1]King's College London; MRC Social, Genetic and Developmental Psychiatry Centre; Institute of Psychiatry; London, UK [2]Birkbeck; University of London; London, UK

Across the genome, outside of a small number of known imprinted genes and regions subject to X-inactivation in females, DNA methylation at CpG dinucleotides is often assumed to be complementary across both alleles in a diploid cell. However, recent findings suggest the reality is more complex, with the discovery that allele-specific methylation (ASM) is a common feature across the human genome. A key observation is that the majority of ASM is associated with genetic variation in cis, although a noticeable proportion is also non-cis in nature and mediated, for example, by parental origin. ASM appears to be both quantitative, characterized by subtle skewing of DNA methylation between alleles, and heterogeneous, varying across tissues and between individuals. These findings have important implications for complex disease genetics; while cis-mediated ASM provides a functional consequence for non-coding genetic variation, heterogeneous and quantitative ASM complicates the identification of disease-associated loci. We propose that non-cis ASM could contribute toward the "missing heritability" of complex diseases, rendering certain loci hemizygous and masking the direct association between genotype and phenotype. We suggest that the interpretation of results from genome-wide association studies can be improved by the incorporation of epi-allelic information and that in order to fully understand the extent and consequence of ASM in the human genome, a comprehensive sequencing-based analysis of allelic methylation patterns across tissues and individuals is required.

Key words: DNA methylation, allele-specific methylation, allele-specific expression, tissue-specific methylation, epigenetics, imprinting, genome-wide association study (GWAS), genetics, complex disease, missing heritability

Submitted: 07/05/10

Accepted: 07/09/10

Previously published online: www.landesbioscience.com/journals/epigenetics/article/12960

DOI: 10.4161/epi.5.7.12960

*Correspondence to: Jonathan Mill; Email: jonathan.mill@kcl.ac.uk

DNA methylation is the best understood and most stable epigenetic modification modulating the transcription of mammalian genomes. The methylation of CpG dinucleotides disrupts the cells' transcriptional machinery by blocking the binding of transcription factors and attracting methyl-binding proteins that initiate chromatin compaction and bring about gene silencing.[1] Because DNA methylation plays a critical role in cellular development and function, aberrant DNA methylation signatures are hypothesized to be involved in diverse human pathologies including cancer,[2] congenital imprinting disorders,[3] and a range of complex chronic disease phenotypes including schizophrenia and bipolar disorder.[4] Elucidating both the genomic patterns of DNA methylation and the factors that determine them thus has important implications for understanding the causes of human health and disease.

Allele-Specific DNA Methylation

Across the majority of the mammalian genome, DNA methylation is assumed to be complementary on both alleles, although there are several classic exceptions where this is known not to be the case and DNA methylation is allele-specific (allele-specific methylation; ASM). First, DNA methylation plays an integral role in regulating the parental-origin-dependent (POD) allele-specific expression (ASE) of imprinted loci. Second, in females, DNA methylation coordinates the random silencing of either the maternally- or paternally-derived X-chromosome to ensure dosage-compensation with males

via the process of X-chromosome inactivation (XCI). A third type of ASM has been reported whereby DNA methylation is determined by DNA sequence in cis and consequently shows Mendelian inheritance patterns.

Recent methylomic studies have uncovered numerous regions of the genome that are characterized by intermediate levels of DNA methylation;[5,6] it is plausible these are largely a result of ASM rather than a consequence of partial methylation across both alleles in the cell population. While early examples of autosomal ASM occurring outside of classical imprinting control regions were largely confined to specific loci[7,8] or chromosomes[9,10] recent investigations by several groups[11-14] have started to yield important insights into the genome-wide nature and prevalence of ASM. These studies suggest that ASM is relatively widespread across the mammalian genome, quantitative rather than qualitative, both cis and POD in nature, and often heterogeneous across tissues and individuals (**Table 1**). In this article we discuss the implications of these findings for maximizing and interpreting statistical association signals that emerge from genome-wide genetic association studies (GWAS), the etiological paradigm currently dominating contemporary biomedical genetic research. We highlight the need for further research into the nature and extent of ASM and suggest that integrating genome-wide surveys of ASM into current GWAS analyses will be a valuable approach for identifying disease-associated genomic loci.

ASM is Often Associated with Genotype In Cis

Genome-wide studies of ASM conclude that cis-effects (i.e., where local genotype is associated with allelic DNA methylation on the same DNA molecule) represent the most prevalent type of ASM. These observations are consistent with the recent methylation quantitative trait loci (mQTL) mapping studies performed in human brain tissue,[15,16] which suggest that a large proportion of differential DNA methylation between unrelated individuals is associated with common cis-acting genetic differences. Widespread genotype-mediated ASM means that allelic methylation is frequently inherited in a Mendelian fashion and provides an epigenetic mechanism by which non-coding sequence variation can have phenotypic effects. This has important implications for the interpretation of results from the current swathe of complex disease GWAS data, where significantly associated alleles are often located a considerable distance from transcribed sequences and have no obvious functional consequence. The integration of epi-allelic data with sequence information will aid in the functional annotation of genetic variation, providing criteria by which to prioritize non-coding disease-associated variants for further study.[17] The publicly available ASM datasets (e.g., http://epigenetics.iop.kcl.ac.uk/ASM) provide an immediate and easily accessible resource for the GWAS community for this purpose, although in the longer term, a comprehensive epigenetic analysis of candidate SNPs and haplotypes resulting from GWAS analyses is warranted.

Non-Cis Mediated ASM is also a Feature of the Genome

Although cis-acting factors can account for the majority of ASM detected, a notable minority is non-cis in nature, presumably due to trans-acting factors, stochastic events or POD effects. Non-cis ASM poses a significant problem for GWAS analyses as it can render loci effectively hemizygous and dilutes or breaks allelic association. It is plausible that such ASM explains a proportion of the "missing heritability" associated with many common human diseases.[18] To date, less than 60 genes have been verified as imprinted in the human genome, although recent computational analyses suggest that the real number may be considerably higher.[19] Although verification is ongoing, our own genome-wide analyses support the computational predications, with non-cis effects accounting for ~10% of detected ASM.[11] These data highlight a limitation of cross-sectional population-based molecular genetic analyses (the prominent design for most current GWAS) in which information about the parental origin of alleles cannot be determined. The utility of being able to track the parental transmission patterns of alleles is exemplified by recent GWAS data in which significant genetic associations have been uncovered, but only when the parental origin of the allele is taken into consideration.[20]

ASM is Quantitative and Heterogeneous in Nature

Genomic imprinting and other classical examples of ASM (e.g., XCI) have been traditionally viewed as all-or-nothing phenomenon with one allele fully methylated and the other unmethylated. Several studies have assessed allelic methylation as a quantitative phenomena and found that many allele-specific effects are in fact relatively subtle, characterized by allelically skewed DNA methylation rather than clear-cut biphasic ASM patterns.[11,14] This is perhaps not surprising; even classically imprinted regions of the genome, believed to be associated with fully monoallelic expression, can show considerable epigenetic heterogeneity.[21] Differences in the degree of methylation between alleles for cis-mediated ASM are important as they suggest that local genotype acts as a "facilitator" of ASM, with as yet undetermined additional factors (most likely trans-acting sequence motifs, stochastic events and environmental factors) determining the absolute pattern of allelic methylation. Again, this has important ramifications for GWAS analyses, as variation in the degree of ASM-skewing across individuals will act to dilute the strength of genetic associations and suggests that effect sizes for disease-associated genetic variants may actually be much larger when epi-allelic variation is taken into consideration.

In addition to the quantitative nature of ASM, several studies have found evidence for tissue (and presumably cellular) heterogeneity in allelic methylation patterns (**Table 1**). While genomic imprinting at several loci is known to be both tissue-specific and developmentally-regulated, the observation that cis-regulated ASM also demonstrates tissue- or cell-type heterogeneity has important implications for genetic studies of complex disease. It suggests, for example, that certain loci may be expressed hemizygously

Table 1. Summary of experimental methods and key findings from large-scale genomic surveys of ASM

Study	Discovery method	Tissue/cell-type	Key findings
Kerkel et al. 2008[12]	MSNP assay[12] using Affymetrix Human Mapping 50K (~50,000 SNPs) and 250K (~250,000 SNPs) arrays.	Peripheral blood leukocytes, bone marrow cells, CD34+Lin- hematopoietic cells, kidney, brain, lung, placental, and buccal cells.	0.16% of queried loci demonstrated clear-cut ASM, with 75% of validated regions associated with genotypoe in cis. Genotype-dependant ASM was associated with ASE for several loci.
Schilling et al 2009[32]	Comparative methyl-CpG immunoprecipitation and tiling arrays, informative for 28Mb of sequence.	Macrophages from two inbred mouse strains (C57BL/6 and BALB/c).	Concluded that ASM is largely attributable to cis-acting polymorphisms.
Zhang et al 2009[9]	Bisulphite conversion, subcloning and sequencing of 190 gene promoter regions on chromosome 21.	Peripheral blood leukocytes, embryo kidney cells, hepatocelluar carcinoma cells, and fibroblasts.	Identified three regions of ASM in leukocytes from a healthy individual.
Schalkwyk et al 2010[11]	Quantitative MSNP assay using Affymetrix Genomewide Human SNP Array 6.0 (~1 million SNPs).	Peripheral blood leukocytes, embryo kidney cells, hepatocelluar carcinoma cells, and fibroblasts.	Identified three regions of ASM in leukocytes from a healthy individual. Demonstrated that genotype-associated ASM is often linearly associated with total mRNA transcript levels
Shoemaker et al 2010[14]	Bisulphite padlock capture and targeted resequencing using the Illumina Genome Analyzer to assess 2,020 CpG islands and targeted regions of the genome.	Peripheral blood leukocytes, fibroblasts, induced pluripotent stem cells, and human embryonic stem cells.	Estimated that between 23-37% of heterozygous SNPs (primarily occurring in CpG positions), are associated with some degree of ASM. The majority of cis-mediated ASM is cell-type specific, with a large amount of between-individual heterogeneity.
Hellman et al 2010[13]	MSNP assay using Affymetrix Genomewide Human SNP Array 5.0 (~500,000 SNPs).	Epstein-Barr virus transformed B-lymphocytes, peripheral blood leukocytes, and post-mortem tissue.	Observed an average genomewide ASM frequency of 20%, which was primarily mediated by cis-effects. Explored the potential mechanisms of genotype-mediated ASM and found a positive relationship between the presence of a CpG at the SNP site and the methylation of close-by CpGs.

in a subset of tissues, and that the use of disease-relevant samples is an important issue for researchers.

ASM-Mediated ASE: Implications for Phenotypic Variation

Like ASM, ASE is common in the human genome and also appears to be largely determined by cis-acting sequence polymorphisms.[22-25] Widespread ASM offers an obvious potential epigenetic mechanism underpinning ASE; there are several examples where allelic methylation levels correlate to allelic or total mRNA expression levels of nearby genes (**Table 1**). These data are limited to only a handful of validated ASM loci and are often inconsistent. Such patterns are expected given that transcription is not a sole consequence of DNA methylation, but is also regulated by other epigenetic processes including histone modification, and is influenced by methylation-independent cis- and trans-acting genetic variation, in addition to environmental factors.

A recent study reported high levels of autosomal ASE occurring stochastically in clonal cell lineages in a process reminiscent of XCI in females.[25] It is likely that this phenomenon is controlled by epigenetic mechanisms such as ASM, with the consequence that loci demonstrating intermediate levels of DNA methylation not caused by POD- or cis-mediated ASM, may still exhibit stochastic patterns of clonally-inherited allelic methylation patterns that directly effect gene expression. Such stochastically-established but clonally-stable ASM/ASE has important implications for research aimed at detecting genetic effects on phenotypic variation, as it renders each individual a unique mosaic for hemizygosity at numerous autosomal loci.

ASM Provides a Biological Mechanism for Genetic and Environmental Effects on Phenotype

One criticism of the GWAS approach is that it investigates genetic effects in isolation; complex phenotypes are now recognized as resulting from interactions between both the genome and the environment (G X E). Several validated

examples of G X E have been identified,[26,27] but these findings are purely statistical in nature and provide few clues about the actual molecular mechanisms operating to mediate susceptibility. Increasing evidence suggests that epigenetic processes can be influenced by a range of external environmental factors including diet, toxins, drugs and stress.[28] The observation that polymorphisms can exert an effect on gene function via epigenetic processes such as ASM occurring in cis, suggests a common pathway behind both genetic and environmental effects and a potential mechanism for G X E. It is pertinent that the regulation of regions characterized by ASM may be particularly vulnerable to environmental influences, especially during embryonic development. Several studies report that inter-individual differences in methylation across differentially methylated regions (DMRs) regulating the monoallelic expression of imprinted domains may be environmentally mediated. Individuals conceived during the Dutch Hunger Winter Famine of 1944–1945, for example, were found to be hypomethylated at the IGF2 DMR on chromosome 11.[29] It is thus plausible that regions characterised by ASM may be environmentally sensitive mediators of disease susceptibility.

Future Directions: Integrating Epi-Allelic Information into GWAS

Studies of ASM to-date have provided a tantalizing insight into the global patterns and common features of ASM in human tissue, but important details have yet to be determined. For instance, many questions remain unanswered about the genomic architecture of ASM (e.g., the true number of loci characterized by allelic methylation differences, how far ASM extends across a region and the presence of functional or positional biases), how distal its effects are on gene expression, and its relationship to other epigenetic processes such as histone modification and non-coding RNA. Furthermore, little is known about the true extent of tissue and cellular heterogeneity, the developmental stability of ASM patterns, the degree of inter-individual variation, what differentiates regions of POD- and cis-mediated ASM, and the specific mechanism(s) by which local genotype influences allelic methylation levels. In order to gain a greater understanding of how genetically-, POD- and stochastically-driven ASM exerts a functional consequence in the cell, a systematic investigation of ASM across tissues and cell types in the context of familial DNA sequence information is required.

Technological advances in methylomic-profiling methodologies mean that it is becoming feasible to map allelic patterns of DNA methylation across the genome at single base-pair resolution; the first reference single-base-resolution map of the methylome was recently published for two human cell lines, providing detailed information about the extent and location of methylated loci.[6] Despite such progress, technical limitations associated with current bisulfite-sequencing methods mean that the ability to determine epi-alleles and epi-haplotypes for all genomic locations is still constrained.[30] It is hoped that single-molecule sequencing-based technologies currently in development will provide the high-resolution quantitative and epi-allelic information that is required for a comprehensive and sensitive analysis of ASM.[31] Ultimately, once normal patterns and sources of inter-individual variation in ASM have been defined, they will contribute greatly to our understanding about the interplay between genetic and epigenetic factors in mediating individual differences in phenotype and disease susceptibility.

References

1. Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci 2006; 31:89-97.
2. Jones PA, Laird PW. Cancer epigenetics comes of age. Nat Genet 1999; 21:163-7.
3. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. Nature 2007; 447:433-40.
4. Mill J, Tang T, Kaminsky Z, Khare T, Yazdanpanah S, Bouchard L, et al. Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. Am J Hum Genet 2008; 82:696-711.
5. Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, et al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat Biotechnol 2009; 27:353-60.
6. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009; 462:315-22.
7. Chandler LA, Ghazi H, Jones PA, Boukamp P, Fusenig NE. Allele-specific methylation of the human c-Ha-ras-1 gene. Cell 1987; 50:711-7.
8. Heijmans BT, Kremer D, Tobi EW, Boomsma DI, Slagboom PE. Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. Hum Mol Gen 2007; 16:547-54.
9. Zhang Y, Rohde C, Tierling S, Jurkowski TP, Bock C, Santacruz D, et al. DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. PLoS Genet 2009; 5:1000438.
10. Yamada Y, Watanabe H, Miura F, Soejima H, Uchiyama M, Iwasaka T, et al. A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. Genome Res 2004; 14:247-66.
11. Schalkwyk LC, Meaburn EL, Smith R, Dempster EL, Jeffries AR, Davies MN, et al. Allelic skewing of DNA methylation is widespread across the genome. Am J Hum Genet 2010; 86:196-212.
12. Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, et al. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. Nat Genet 2008; 40:904-8.
13. Hellman A, Chess A. Extensive sequence-influenced DNA methylation polymorphism in the human genome. Epigenetics Chromatin 2010; 3:11.
14. Shoemaker R, Deng J, Wang W, Zhang K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. Genome Res 2010; 20:883-9.
15. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, et al. Abundant quantitative trait Loci exist for DNA methylation and gene expression in human brain. PLoS Genet 2010; 6:1000952.
16. Zhang D, Cheng L, Badner JA, Chen C, Chen Q, Luo W, et al. Genetic control of individual differences in gene-specific methylation in human brain. Am J Hum Genet 2010; 86:411-9.
17. Tycko B. Mapping allele-specific DNA methylation: a new tool for maximizing information from GWAS. Am J Hum Genet 2010; 86:109-12.
18. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature 2009; 461:747-53.
19. Luedi PP, Dietrich FS, Weidman JR, Bosko JM, Jirtle RL, Hartemink AJ. Computational and experimental identification of novel human imprinted genes. Genome Res 2007; 17:1723-30.
20. Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, et al. Parental origin of sequence variants associated with complex diseases. Nature 2009; 462:868-74.
21. Sakatani T, Wei M, Katoh M, Okita C, Wada D, Mitsuya K, et al. Epigenetic heterogeneity at imprinted loci in normal populations. Biochem Biophys Res Commun 2001; 283:1124-30.
22. Ge B, Pokholok DK, Kwan T, Grundberg E, Morcos L, Verlaan DJ, et al. Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. Nat Genet 2009; 41:1216-22.
23. Palacios R, Gazave E, Goni J, Piedrafita G, Fernando O, Navarro A, et al. Allele-specific gene expression is widespread across the genome and biological processes. PloS One 2009; 4:4150.
24. Serre D, Gurd S, Ge B, Sladek R, Sinnett D, Harmsen E, et al. Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. PLoS Genet 2008; 4:1000006.
25. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A. Widespread monoallelic expression on human autosomes. Science 2007; 318:1136-40.
26. Caspi A, McClay J, Moffitt TE, Mill J, Martin J, Craig IW, et al. Role of genotype in the cycle of violence in maltreated children. Science 2002; 297:851-4.

27. Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science 2003; 301:386-9.
28. Dolinoy DC, Jirtle RL. Environmental epigenomics in human health and disease. Environ Mol Mutagen 2008; 49:4-8.
29. Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc Natl Acad Sci USA 2008; 105: 17046-9.
30. Laird PW. Principles and challenges of genome-wide DNA methylation analysis. Nat Rev Genet 2010; 11:191-203.
31. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 2010; 7:461-5.
32. Shilling E, Chartouni CE, Rehli M, Allele-specific DNA methylation in mouse strains is mainly determined by cis-acting sequences. Genome Res 2009; 11:2028-35