Allele-specific methylation in the human genome

Emma L. Meaburn\textsuperscript{a}, Leonard C. Schalkwyk\textsuperscript{a} & Jonathan Mill\textsuperscript{b}
\textsuperscript{a} King’s College London; MRC Social; Genetic and Developmental Psychiatry Centre; Institute of Psychiatry, De Crespigny Park; London, UK
\textsuperscript{b} Institute of Psychiatry
Published online: 01 Jan 2010.

To cite this article: Emma L. Meaburn, Leonard C. Schalkwyk & Jonathan Mill (2010) Allele-specific methylation in the human genome, Epigenetics, 5:7, 578-582, DOI: 10.4161/epi.5.7.12960

To link to this article: http://dx.doi.org/10.4161/epi.5.7.12960

PLEASE SCROLL DOWN FOR ARTICLE

It is essential that you check the license status of any given Open and Open Select article to confirm conditions of access and use.
Allele-specific methylation in the human genome
Implications for genetic studies of complex disease

Emma L. Meaburn,1,2 Leonard C. Schalkwyk1 and Jonathan Mill1,*
1King’s College London; MRC Social, Genetic and Developmental Psychiatry Centre; Institute of Psychiatry; London, UK 2Birkbeck; 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.

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. 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, congenital imprinting disorders, and a range of complex chronic disease phenotypes including schizophrenia and bipolar disorder. 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. 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 or chromosomes, recent investigations by several groups 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 methylination quantitative trait loci (mQTL) mapping studies performed in human brain tissue, 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. 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. 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. Although verification is ongoing, our own genome-wide analyses support the computational predications, with non-cis effects accounting for ~10% of detected ASM. 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.

**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. 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. 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 undefined 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...
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, 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. 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. 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 methylation-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 methylene was recently published for two human cell lines, providing detailed location information about the extent and location of methylated loci. 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. It is hoped that single-molecule sequencing-based technologies currently in development will provide the high-resolution quantitative and epilellic information that is required for a comprehensive and sensitive analysis of ASM. 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**