Published in final edited form as:

Biochim Biophys Acta. 2012 July ; 1819(7): 739–742. doi:10.1016/j.bbagr.2012.02.007.

Allele-specific gene expression and epigenetic modifications and their application to understanding inheritance and cancer

Maxwell P Lee

The Laboratory of Population Genetics, Center for Cancer Research, National Cancer Institute, U.S.A

Abstract

Epigenetic information is characterized by its plasticity during development and differentiation as well as its stable transmission during mitotic cell divisions in somatic tissues. This duality contrasts to genetic information, which is essentially static and identical in every cell in an organism with only a few exceptions such as immunoglobulin genes in lymphocytes. Epigenetics is traditionally perceived as a means to regulate gene expression without a change in DNA sequence. This, however, does not exclude a potential role for genetic variations in providing differential backgrounds on which epigenetic modulations and their regulatory consequences are achieved. An effective approach to investigating the interplay between genetic variations and epigenetic variations is through allele-specific analysis of epigenetics and gene expression. Such studies have generated many new insights into functions of genetic variations, mechanisms of gene expression regulation, and the role of mutations and epigenetic alterations in human cancer.

Keywords

allele-specific gene expression; allele-specific methylation; allele-specific chromatin; epigenetic marks; polymorphism; inheritance; cancer

Epigenetics is the study of heritable changes in gene expression that don't involve changes in the DNA sequence. The term was originally coined by Conrad Waddington in 1942 to describe a mechanism by which an organism can be developed from a fertilized egg through the process of differentiation [1]. The term provided a conceptual framework to understand development and cellular differentiation, despite a lack of understanding of their chemical nature at that time. In the past half century, the chemical nature of epigenetics has largely been uncovered. They include chemical modifications in both DNA and chromatin proteins (for review, see [2]). The DNA modification is relatively simple, consisting of adding a methyl group at the 5 position of a cytosine. In mammalian genomes, the frequency of CpG dinucleotide is much lower than the frequency of any other dinucleotide. This is due to selective conversion of 5-methylcytosine to thymine. CpG dinucleotides that are not methylated and therefore protected from the conversion are often found clustered in regions that are usually 400 to 500 bases in length, referred to as CpG islands (CGIs). About half of the CGIs are located within 1 kb of transcription start sites. Modifications of chromatin proteins are far more complex. The H3 and H4 histones undergo extensive post-translational modifications in the tails protruding from the nucleosome. These modifications include methylation, acetylation, phosphorylation, ubiquitination, sumoylation, and ADP-

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ribosylation. The combinations of such modifications are often dubbed the “histone code” to imply that they embody an intrinsic biological meaning [3]. In this review, I will focus on allele-specific gene expression and epigenetic modifications, which are related to two well-known epigenetic phenomena, X chromosome inactivation and genomic imprinting.

X chromosome inactivation is a phenomenon that involves silencing of gene expression from one of the two X chromosomes [4]. X-inactivation ensures that gene expression is quantitatively equal or similar in female and male cells. The choice of the inactivated X chromosome is random and is determined in the early embryonic stage but after tissue progenitors have been established so that mosaicism is observed for this epigenetic feature. X chromosome inactivation involves DNA methylation and histone modifications and these epigenetic modifications are maintained throughout the lifetime of an organism.

Genomic imprinting describes expression of certain genes from only one of the two chromosomes in a cell and the choice of which chromosome is silenced depending on parent-of-origin [5–6]. Imprinted genes express either the allele inherited from the mother, as in the case of the *CDKN1C* gene, or the allele inherited from the father, as with the *IGF2* gene. The underlying basis of genomic imprinting is an epigenetic process involving DNA methylation and histone modifications that silence the expression from one of the two chromosomes. These epigenetic modifications are established during early embryonic development and are maintained throughout the life span of an organism. Abnormal imprinting is the cause of a number of human diseases, including Beckwith-Wiedemann syndrome [7], as well as Angelman syndrome and Prader-Willi syndrome [8], both of which are associated with genetic loci located on 15q11.2. Loss of imprinting at specific loci, notably *IGF2*, also occurs frequently.[9].

Differential expression between the two chromosomes of a cell goes beyond genomic imprinting and X chromosome inactivation. In B lymphocytes, an immunoglobulin gene rearrangement from one chromosome prevents a rearrangement of the same gene from the other chromosome in an individual cell [10–11]. This is referred to as allelic exclusion. Consequently, an individual lymphocyte has a unique sequence of an immunoglobulin protein. Human olfactory systems contain about 1000 olfactory receptor genes. A fourth example of a situation in which an epigenetic mechanism directs expression from only one allele, i.e. mono-allelic expression, is exemplified by protein expression from only a single allele of a single olfactory receptor gene in each sensory neuron [12]. However, in this example, mono-allelic expression is only one of multiple mechanisms that contribute to the random inactivation of one allele.

Allele-specific gene expression

In addition to mono-allelic expression described above, quantitative differences in the degree of gene expression between two chromosomes or two alleles is a widespread phenomenon. We initially made this observation by studying allele-specific gene expression using the Affymetrix SNP arrays and found that allelic variation in expression is common in the human genome [13]. This conclusion was supported subsequently by many studies [14–17]. The difference in gene expression between the two chromosomes or two alleles has been described in the literature using multiple terms, including allele-specific expression (ASE), allelic imbalance (AI) in expression, allelic variation in expression, or mono-allelic expression, to name just a few. We will use ASE in this article. ASE differs from mono-allelic expression of genomic imprinting, X chromosome inactivation, and allelic exclusion in several aspects. First, in ASE the difference is usually in the range of 2~4 fold, reflecting a graded quantitative difference in allelic expression, as opposed to the all-or-none expression differences observed with imprinting and X chromosome inactivation. Second,

ASE affects a large number of genes instead of 75 or so imprinted genes, X-linked genes, immunoglobulin genes, or olfactory receptor genes. Third, ASE is often influenced by genetic polymorphisms near the gene, that is, the decision as to which allele is expressed may be influenced by *cis*-acting polymorphisms.

The association of genetic variations with allelic variation in gene expression is important. This implies that some of the genetic polymorphisms play a regulatory role and exert their effects on gene expression. This is particularly relevant since recent genome wide association studies (GWAS) find that most of the SNPs associated with common diseases do not change protein sequences [18–19]. These GWAS findings include SNPs in intergenic regions, introns, or UTRs as well as synonymous SNPs in coding sequences. Therefore, other mechanisms besides amino acid sequence changes must be invoked to explain the disease association. Specifically, these SNPs can affect phenotypic variations through their impact on gene expression, which may involve several mechanisms including some that operate at the epigenetic level or post-transcription level. An urgent need exists to investigate the underlying mechanisms at a systems level so that a better understanding can be achieved regarding the subtleties of how genetic variations affect phenotypes. To reiterate, this is true especially for those SNPs that don't change amino acid sequences and are postulated to affect gene expression.

ASE is a powerful approach to study the impact of genetic variations on gene expression [20]. It complements the studies of eQTL (expression quantitative trait loci), which quantifies gene expression differences among people with different genotypes. The eQTL combines measurements of genotype and gene expression in a large number of individuals and correlates genotype with gene expression [21]. The approach is versatile. In many cases, genotype data and gene expression are already available through GWAS and other studies. On the other hand, ASE offers some advantage in certain studies. Here are two examples. First, variation in gene expression among different people can be due to a number of causes. Some are real biological factors while others are caused by experimental noise. Allele-specific gene expression provides an effective way to remove technical noise, since the noise often affects both alleles more or less equally and thus should not affect relative expression between the two alleles. Second, transcription regulatory mechanisms such as negative feedback can conceal the effect of variants on the total expression level, but such an effect can be unmasked by means of allele-specific analysis.

Allele-specific DNA methylation and chromatin modifications

Allele-specific DNA methylation (ASM) has also been investigated with high throughput approaches, including SNP arrays and recently Nextgen sequencing technology. The use of a methyl-sensitive restriction enzyme to cut genomic DNA coupled with detection by a SNP array was first described by Yuan et al [22]. Interrogation of DNA that has been cut by such an enzyme will reveal positive hybridization signals only for alleles of SNPs that are methylated; non-methylated alleles are cut and destroyed and therefore yield no signal. Informative, heterozygous loci that exhibit signal for only one of the two alleles suggest the phenomenon of ASM. By analyzing allele-specific DNA methylation, Hellman and Chess showed that methylation in the gene bodies on the active X chromosome (Xa) was higher than those on the inactive X chromosome (Xi) [23]. This contrasts to hypomethylation in promoter regions on Xa. ASM has also been observed on autosomal loci. A recent study by Kerkel et al. identified 16 ASM sites and found 12 of which associating with nearby sequence variants [24]. Such sequence-dependent ASM may help interpret the function of some of non-coding SNPs identified through GWAS.

The combination of chromatin immunoprecipitation (ChIP) and quantitative allelic measurement has been used to study allele-specific RNA polymerase II loading and characterize haplotypes associating with allele-specific transcription of the *LTA* gene [25]. We applied a similar strategy to study allele-specific histone modifications and RNA polymerase II occupancy [26] and found that genetic background can influence global chromatin state mediated by histone modification, the hallmark of the epigenetic phenomena. We also showed that allele-specific histone H3 acetylation segregated with genetic variants.

Inheritance of allele-specific epigenetic marks

An important concept emerging from allele-specific gene expression and epigenetic features is the association of genetic variants with epigenetic marks and gene expression. In other words, epigenetic marks and gene expression is partly determined by genetic variation. Two early studies, one by Chandler et al. [27] and the other by Silva and White [28], reported that differential methylation of alleles, using VNTRs (variable numbers of tandem repeats), on homologous chromosomes occurred in a tissue-dependent manner. Silva and White further demonstrated that allele-specific methylation was transmitted through the germ line [28]. The allele-specific methylation was common, present at 3 out of 10 autosomal loci examined. A recent study by Kerkel et al, described in the previous section, reported sequence-dependent allele-specific methylation (ASM) [24]. Although the inheritance was not investigated in this study, the association between genetic variations and DNA methylation was observed in unrelated individuals. Thus it extends the association between genetic variants and ASM from family to population. Another recent study investigated ASM in a three-generation family using reduced representation bisulfite sequencing (RRBS) method and found that ASM was primarily determined by genotype [29]. In contrast, only a small number of CpG sites showed parent-of-origin dependent ASM, which is a hallmark of genomic imprinting.

ASE has also been shown exhibiting a segregation pattern in families that was consistent with Mendelian inheritance [30]. In this study, six of 13 genes showed ASE; ASE was shown transmitting in the germ line for two genes with enough informative individuals in families for the segregation analysis.

We performed inheritance analysis for seven genes for which we had allele-specific histone H3 acetylation data in a study described in the previous section, and all showed segregation patterns consistent with Mendelian inheritance [26]. In a recent study, allele-specific DNase I hypersensitive and CTCF binding sites were analyzed in two trios (two parents and a child) with deep sequencing technology [31]. 10% of active chromatin sites were found allele-specific and transmitted from parent to child. The results suggest that these chromatin features are also heritable.

Understanding cancer from the perspective of allele-specific analyses

The importance of each allele contributing to cancer risk and tumorigenesis is well recognized, and it is fundamental to our understanding of modern cancer genetics. Sequential inactivation of the two alleles of a tumor suppressor gene in an individual is best illustrated by Knudson two-hit theory [32]. In familial cancer syndrome, the first hit is inherited from parents while the second hit takes place in somatic cells; in sporadic cancer, both hits occur in somatic cells. Despite the importance of allele-specific gene expression and epigenetic marks, direct application of these allele-specific features to cancer research has been scant. A few examples do exist and are briefly described below.

The first example of ASE was provided by Yan et al., who showed that about 50% reduction in gene expression in lymphoblastoid cells from one of the two *APC* alleles and this allele-specific reduction of expression was associated with a pronounced predisposition to hereditary colorectal tumors [33]. In this study, 6 patients from two FAP (familial adenomatous polyposis) families were examined, and no mutation was detected in the *APC* gene. However, using ASE, they found all 6 patients had 2-fold difference in gene expression between the two alleles in lymphoblastoid cells. Furthermore, majority of micro-dissected tumors displayed loss of heterozygosity (LOH), which deleted specifically the allele with higher expression in lymphoblastoid cells. The authors concluded that the reduction of mutant allele expression in germ line cells coupled with the loss of wild type allele by LOH in tumors could lower the expression of APC to below a threshold that is required to suppress intestinal tumorigenesis.

The second example involved ASM. In a study of a three-generation family with the hereditary nonpolyposis colorectal cancer (HNPCC) syndrome, Chan et al. reported that allele-specific hypermethylation of the *MSH2* gene in the germ line cells was associated with the loss of the MSH2 protein in the colorectal adenocarcinomas [34]. Interestingly, a somatic frameshift mutation was detected in *MSH2* but with retention of heterozygosity. The authors concluded that germline transmission of ASM contributed to the first hit while the somatic mutation contributed to the 2nd hit.

The third example also involved ASM as well as *MSH2* and a different family with HNPCC but with an intriguing twist [35]. In this case, the *EPCAM* gene, with its 3' end located 15 kb upstream to the *MSH2* gene, had a germline deletion in the 3' end, and abolished transcription termination. The deletion caused transcription read-through into the downstream gene, *MSH2*, and lead to an increase in DNA methylation in the promoter region of *MSH2*. The deletion co-segregated with the disease and methylation of the *MSH2* promoter. These examples provide a glimpse into how allele-specific gene expression and methylation can bring new insight into understanding the interplay among genetic variation, epigenetics, and cancer.

In conclusion, allele-specific analysis provides a unique perspective to integrate genetic variations with epigenetic information. It enables researchers to identify gene expression that is affected by genetic polymorphisms, to evaluate inheritance of gene expression and epigenetic marks, and to integrate genetic variations with epigenetic modifications, gene expression, and phenotypes. It provides a powerful approach to gaining new insights into understanding gene expression regulation and cancer biology.

Acknowledgments

I like to thank Drs. Mitsutaka Kadota and Barbara Dunn for critical reading of this manuscript. This research was supported by the Intramural Research Program of the NIH and the National Cancer Institute.

References

1. Waddington CH. The epigenotype. *Endeavour*. 1942; 1:18–20.
2. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature*. 2007; 447:433–440. [PubMed: 17522677]
3. Jenuwein T, Allis CD. Translating the histone code. *Science*. 2001; 293:1074–1080. [PubMed: 11498575]
4. Wutz A. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat Rev Genet*. 2011; 12:542–553. [PubMed: 21765457]
5. Barlow DP. Genomic imprinting: a Mammalian epigenetic discovery model. *Annu Rev Genet*. 2011; 45:379–403. [PubMed: 21942369]

6. Feinberg AP. DNA methylation, genomic imprinting and cancer. *Curr Top Microbiol Immunol*. 2000; 249:87–99. [PubMed: 10802940]
7. Choufani S, Shuman C, Weksberg R. Beckwith-Wiedemann syndrome. *Am J Med Genet C Semin Med Genet*. 2010; 154C:343–354. [PubMed: 20803657]
8. Buiting K. Prader-Willi syndrome and Angelman syndrome. *Am J Med Genet C Semin Med Genet*. 2010; 154C:365–376. [PubMed: 20803659]
9. Feinberg AP, Tycko B. The history of cancer epigenetics. *Nat Rev Cancer*. 2004; 4:143–153. [PubMed: 14732866]
10. Cedar H, Bergman Y. Choreography of Ig allelic exclusion. *Curr Opin Immunol*. 2008; 20:308–317. [PubMed: 18400481]
11. Vettermann C, Schlissel MS. Allelic exclusion of immunoglobulin genes: models and mechanisms. *Immunol Rev*. 2010; 237:22–42. [PubMed: 20727027]
12. Chess A, Simon I, Cedar H, Axel R. Allelic inactivation regulates olfactory receptor gene expression. *Cell*. 1994; 78:823–834. [PubMed: 8087849]
13. Lo HS, Wang Z, Hu Y, Yang HH, Gere S, Buetow KH, Lee MP. Allelic variation in gene expression is common in the human genome. *Genome Res*. 2003; 13:1855–1862. [PubMed: 12902379]
14. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A. Widespread monoallelic expression on human autosomes. *Science*. 2007; 318:1136–1140. [PubMed: 18006746]
15. Lin W, Yang HH, Lee MP. Allelic variation in gene expression identified through computational analysis of the dbEST database. *Genomics*. 2005; 86:518–527. [PubMed: 15993562]
16. Pant PV, Tao H, Beilharz EJ, Ballinger DG, Cox DR, Frazer KA. Analysis of allelic differential expression in human white blood cells. *Genome Res*. 2006; 16:331–339. [PubMed: 16467561]
17. Pastinen T, Sladek R, Gurd S, Sammak A, Ge B, Lepage P, Lavergne K, Villeneuve A, Gaudin T, Brandstrom H, Beck A, Verner A, Kingsley J, Harmsen E, Labuda D, Morgan K, Vohl MC, Naumova AK, Sinnott D, Hudson TJ. A survey of genetic and epigenetic variation affecting human gene expression. *Physiol Genomics*. 2004; 16:184–193. [PubMed: 14583597]
18. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009; 106:9362–9367. [PubMed: 19474294]
19. Manolio TA, Collins FS. The HapMap and genome-wide association studies in diagnosis and therapy. *Annu Rev Med*. 2009; 60:443–456. [PubMed: 19630580]
20. Ge B, Pokholok DK, Kwan T, Grundberg E, Morcos L, Verlaan DJ, Le J, Koka V, Lam KC, Gagne V, Dias J, Hoberman R, Montpetit A, Joly MM, Harvey EJ, Sinnott D, Beaulieu P, Hamon R, Graziani A, Dewar K, Harmsen E, Majewski J, Goring HH, Naumova AK, Blanchette M, Gunderson KL, Pastinen T. Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. *Nat Genet*. 2009; 41:1216–1222. [PubMed: 19838192]
21. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. Genetic analysis of genome-wide variation in human gene expression. *Nature*. 2004; 430:743–747. [PubMed: 15269782]
22. Yuan E, Haghighi F, White S, Costa R, McMinn J, Chun K, Minden M, Tycko B. A single nucleotide polymorphism chip-based method for combined genetic and epigenetic profiling: validation in decitabine therapy and tumor/normal comparisons. *Cancer Res*. 2006; 66:3443–3451. [PubMed: 16585166]
23. Hellman A, Chess A. Gene body-specific methylation on the active X chromosome. *Science*. 2007; 315:1141–1143. [PubMed: 17322062]
24. Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, Li K, Murty VV, Schupf N, Vilain E, Morris M, Haghighi F, Tycko B. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet*. 2008; 40:904–908. [PubMed: 18568024]
25. Knight JC, Keating BJ, Rockett KA, Kwiatkowski DP. In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nat Genet*. 2003; 33:469–475. [PubMed: 12627232]

26. Kadota M, Yang HH, Hu N, Wang C, Hu Y, Taylor PR, Buetow KH, Lee MP. Allele-specific chromatin immunoprecipitation studies show genetic influence on chromatin state in human genome. *PLoS Genet.* 2007; 3:e81. [PubMed: 17511522]
27. Chandler LA, Ghazi H, Jones PA, Boukamp P, Fusenig NE. Allele-specific methylation of the human c-Ha-ras-1 gene. *Cell.* 1987; 50:711–717. [PubMed: 2441871]
28. Silva AJ, White R. Inheritance of allelic blueprints for methylation patterns. *Cell.* 1988; 54:145–152. [PubMed: 2898978]
29. Gertz J, Varley KE, Reddy TE, Bowling KM, Pauli F, Parker SL, Kucera KS, Willard HF, Myers RM. Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLoS Genet.* 2011; 7:e1002228. [PubMed: 21852959]
30. Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. Allelic variation in human gene expression. *Science.* 2002; 297:1143. [PubMed: 12183620]
31. McDaniel R, Lee BK, Song L, Liu Z, Boyle AP, Erdos MR, Scott LJ, Morken MA, Kucera KS, Battenhouse A, Keefe D, Collins FS, Willard HF, Lieb JD, Furey TS, Crawford GE, Iyer VR, Birney E. Heritable individual-specific and allele-specific chromatin signatures in humans. *Science.* 2010; 328:235–239. [PubMed: 20299549]
32. Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A.* 1971; 68:820–823. [PubMed: 5279523]
33. Yan H, Dobbie Z, Gruber SB, Markowitz S, Romans K, Giardiello FM, Kinzler KW, Vogelstein B. Small changes in expression affect predisposition to tumorigenesis. *Nat Genet.* 2002; 30:25–26. [PubMed: 11743581]
34. Chan TL, Yuen ST, Kong CK, Chan YW, Chan AS, Ng WF, Tsui WY, Lo MW, Tam WY, Li VS, Leung SY. Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer. *Nat Genet.* 2006; 38:1178–1183. [PubMed: 16951683]
35. Ligtenberg MJ, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, Lee TY, Bodmer D, Hoenselaar E, Hendriks-Cornelissen SJ, Tsui WY, Kong CK, Brunner HG, van Kessel AG, Yuen ST, van Krieken JH, Leung SY, Hoogerbrugge N. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat Genet.* 2009; 41:112–117. [PubMed: 19098912]

Highlights

- I summarize the current research on allele-specific gene expression and epigenetic modifications.
- I discuss how allele-specific features can be applied to study inheritance.
- I provide some examples on how allele-specific analyses can help gain new insights into better understanding of cancer.