



# Crosstalk of Genetic Variants, Allele-Specific DNA Methylation, and Environmental Factors for Complex Disease Risk

Huishan Wang<sup>1,2</sup>, Dan Lou<sup>1</sup> and Zhibin Wang<sup>1,3\*</sup>

<sup>1</sup> Laboratory of Human Environmental Epigenome, Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States, <sup>2</sup> Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong, <sup>3</sup> State Key Laboratory of Biocatalysis and Enzyme Engineering, School of Life Sciences, Hubei University, Wuhan, China

## OPEN ACCESS

### Edited by:

Feng C. Zhou,  
Indiana University Bloomington,  
United States

### Reviewed by:

Amy C. Lossie,  
National Institutes of Health (NIH),  
United States  
Cuncong Zhong,  
University of Kansas, United States

### \*Correspondence:

Zhibin Wang  
zwang47@jhu.edu

### Specialty section:

This article was submitted to  
Behavioral and Psychiatric Genetics,  
a section of the journal  
Frontiers in Genetics

**Received:** 08 March 2018

**Accepted:** 12 December 2018

**Published:** 09 January 2019

### Citation:

Wang H, Lou D and Wang Z (2019)  
Crosstalk of Genetic Variants,  
Allele-Specific DNA Methylation, and  
Environmental Factors for Complex  
Disease Risk. *Front. Genet.* 9:695.  
doi: 10.3389/fgene.2018.00695

Over the past decades, genome-wide association studies (GWAS) have identified thousands of phenotype-associated DNA sequence variants for potential explanations of inter-individual phenotypic differences and disease susceptibility. However, it remains a challenge for translating the associations into causative mechanisms for complex diseases, partially due to the involved variants in the noncoding regions and the inconvenience of functional studies in human population samples. So far, accumulating evidence has suggested a complex crosstalk among genetic variants, allele-specific binding of transcription factors (ABTF), and allele-specific DNA methylation patterns (ASM), as well as environmental factors for disease risk. This review aims to summarize the current studies regarding the interactions of the aforementioned factors with a focus on epigenetic insights. We present two scenarios of single nucleotide polymorphisms (SNPs) in coding regions and non-coding regions for disease risk, via potentially impacting epigenetic patterns. While a SNP in a coding region may confer disease risk via altering protein functions, a SNP in non-coding region may cause diseases, via SNP-altering ABTF, ASM, and allele-specific gene expression (ASE). The allelic increases or decreases of gene expression are key for disease risk during development. Such ASE can be achieved via either a “SNP-introduced ABTF to ASM” or a “SNP-introduced ASM to ABTF.” Together with our additional in-depth review on insulator CTCF, we are convinced to propose a working model that the small effect of a SNP acts through altered ABTF and/or ASM, for ASE and eventual disease outcome (named as a “SNP intensifier” model). In summary, the significance of complex crosstalk among genetic factors, epigenetic patterns, and environmental factors requires further investigations for disease susceptibility.

**Keywords:** allele-specific DNA methylation (ASM), single nucleotide polymorphisms (SNPs), allele-specific gene expression (ASE), genetic variants, regional “autosomal chromosome inactivation (ACI)”, quantitative trait locus (QTL), allele-specific binding of transcription factors (ABTFs), SNP intensifier model

## INTRODUCTION

Genetic variants identified from genome-wide association studies (GWAS) promise to uncover the understanding of inter-individual differences in phenotypes and the risk of complex diseases. One hypothesis is that the susceptibility of an individual to a complex disease is due to the interaction of genetic variants with environmental factors acting through epigenetic mechanisms. Therefore, understanding the complex crosstalk among genetic variation, environmental exposure, and epigenetic patterns is essential for unraveling the etiology of common disease. This review aims to illustrate potential mechanisms for the crosstalk of genetic factors, epigenetic patterns, and allelic binding of transcriptional factors (ABTF), as well as the crosstalk with environmental exposure for disease susceptibility. We begin the review with a basic and brief, but hopefully necessary introduction on the fundamental insights of epigenetic mechanisms (DNA methylation and histone modifications) in the regulation of gene transcription, because both DNA methylation and histone modifications are linked with genetic variants and environmental exposure for disease or complex traits in many burgeoning reports (detailed below). We then describe the crosstalk among different factors (including genetic variation, epigenetic variation, gene expression, and environmental factors) for complex disease risk. We complement the aforementioned broad topics with an in-depth summary on the crosstalk among insulator CTCF, genetic variation, and epigenetic variation in disease or complex traits. This subtopic is currently of high interest in the biomedical arena because of CTCF's pleiotropic roles in biology. CTCF helps shape the high-order genome organization (e.g., "4D nucleosome" program), acts as a tumor suppressor (Kemp et al., 2014), and plays an important role in complex traits. To better envision the underlying mechanisms, we selectively discuss the most relevant investigations regardless of the disease investigated. That is, we did not focus on one type of complex diseases, because a relevant functional study may not be available if we restrain our references to one disease. Hence, with such selection, we may unintentionally overlook many important reports within a given disease. In cases where the expected or hypothesized mechanism may not have been thoroughly addressed, we try to provide a diagram to aid our illustration. We will start our review with brief information on epigenetic background.

## EPIGENETIC MECHANISMS INVOLVED IN TRANSCRIPTIONAL REGULATION

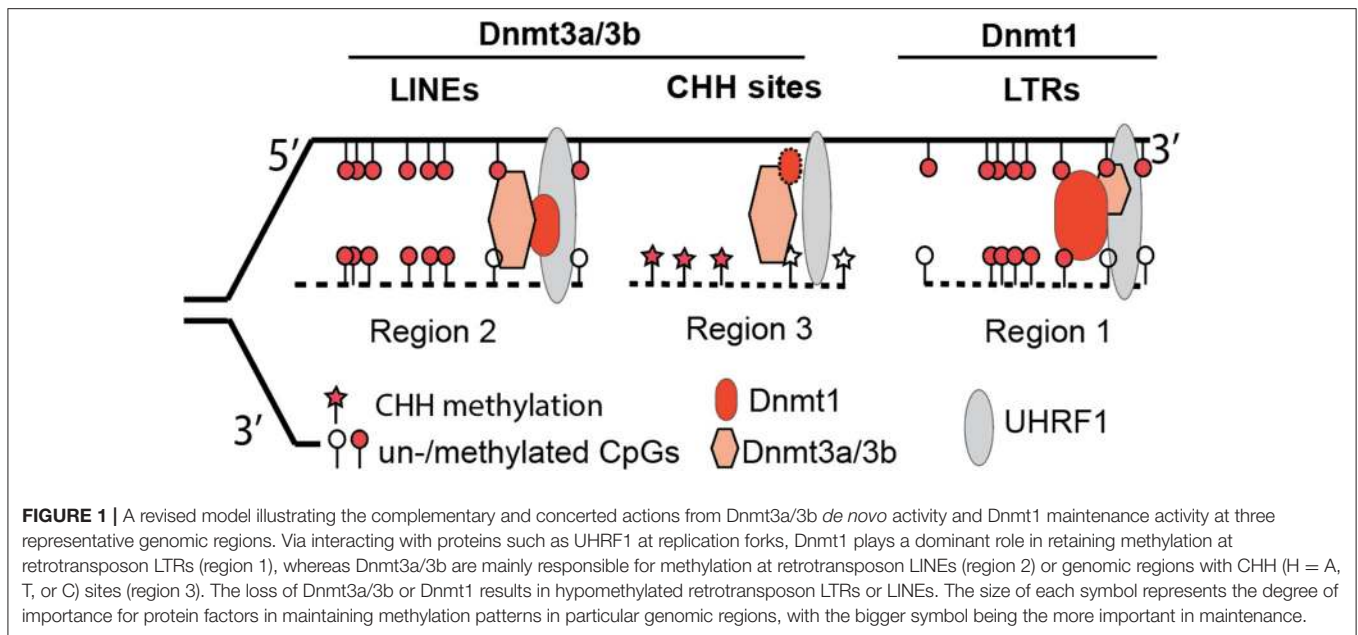
### DNA Methylation Patterns: Establishment, Maintenance, and Functional Roles

DNA methylation patterns are established and maintained by three DNA methyltransferases (Dnmts), namely Dnmt1, Dnmt3a, and Dnmt3b, in both the mouse and human genomes (Bestor, 2000). In a simplified "two-step" model, it is the role of Dnmt3a and Dnmt3b, which initiate methylation of globally hypomethylated genomic DNA (after implantation) to establish methylation patterns during early embryonic development

(Okano et al., 1999). Afterwards, via coupling with DNA replication machinery, Dnmt1 faithfully copies methylation information from the parental strands to the daughter strands during DNA replication (Jones and Liang, 2009; Jurkowska et al., 2011). Because of this copying role, Dnmt1 is referred to as the "maintenance methyltransferase," whereas Dnmt3a and Dnmt3b are called the "de novo methyltransferases." In support of this model, Dnmt3a and Dnmt3b have high activity toward unmethylated DNA, while Dnmt1 shows low activity toward unmethylated DNA and prefers hemi-methylated DNA (Jeltsch, 2006; Jurkowska et al., 2011).

However, with the recent availability of base-resolution methylome data for Dnmt knockout cells, this "two-step" model seems oversimplified (Li et al., 2015). Our data demonstrate that in the absence of Dnmt1, Dnmt3a and Dnmt3b can still maintain symmetrical methylation of CpG dinucleotides (where a cytosine nucleotide is followed by a guanine nucleotide linearly along 5' to 3' direction) to some degree. In addition, the presence of Dnmt1 alone allows cells to maintain the methylation levels of retrotransposon long terminal repeats (LTRs) in Dnmt double knockout cells (*Dnmt3a<sup>-/-</sup>/3b<sup>-/-</sup>*), at levels comparable to wild type cells. In *Dnmt1<sup>-/-</sup>*, the presence of only two *de novo* methyltransferases (Dnmt3a and Dnmt3b) still enables embryonic stem cells to retain methylation at retrotransposon long interspersed nuclear elements (LINEs). Therefore, Dnmt1 is required for methylation of LTRs, whereas Dnmt3a and Dnmt3b are necessary for LINE methylation (Figure 1) (Li et al., 2015). Collectively, it is more accurate to state that the concerted actions of three Dnmt enzymes are required for the maintenance and establishment of DNA methylation patterns in the mouse genome. While at the last stage of publishing this review, a new report suggests that Dnmt1 does have *de novo* methyltransferase activity (Li et al., 2018).

The unexpected division of labor between maintenance Dnmt1 and *de novo* Dnmt3a and Dnmt3b for retrotransposons is informative for deeper mechanistic insights in future investigations. In early epidemiological studies, the methylation levels of LINEs or Alu sequences in the human/mouse genome were assumed to represent global DNA methylation changes. Though the assumption was largely untested (Nelson et al., 2011), LINE or Alu methylation has been used as a biomarker for environmental exposure or a diseased state. For example, in a recent cohort study that examined the association between DNA methylation of pre-diagnostic leukocyte and gastric cancer risk, it was reported that the latter was inversely associated with Alu methylation. Intriguingly, LINE methylation was not associated with gastric cancer risk (Gao et al., 2012). The association of Alu but not LINE methylation with gastric cancer risk is in line with other distinct mechanisms in control of different classes of retrotransposons (Li et al., 2015). Presumably, the mechanism for Alu methylation was affected in gastric cancer, but the mechanism for LINE methylation was not. In an effort to assess Alu (LINE not investigated) methylation from peripheral blood DNA of healthy donors and patients with alcohol use disorders, Kim and colleagues found that Alu methylation levels are significantly higher in the latter than the former (Kim et al., 2016). The underlying mechanism for the aforementioned



associations, however, remains to be determined. In the agouti viable yellow ( $A^y$ ) mouse studies, methylation of CpG sites within a LTR retrotransposon is vulnerable (hypomethylated) to environmental insults, including Bisphenol A (BPA), folate, and alcohol (Dolinoy, 2008; Kaminen-Ahola et al., 2010). Because LTR's methylation depends on Dnmt1 (Li et al., 2015), it may imply that Dnmt1-dependent methylation activity is more sensitive to environmental exposure. It is also worth mentioning that Dnmt1 is highly expressed in most somatic tissues, whereas Dnmt3a and Dnmt3b are not. In summary, depending on exposure-alteration of Dnmt1 or Dnmt3a/3b, the LINE or LTR methylation level should be selected accordingly, in order to represent global DNA methylation changes in cells.

Methylation of CpG sites can occur in the transcribed regions of genes (either silenced or expressed) (Lister et al., 2009), suggesting that the alteration of DNA methylation patterns may not affect the expression of a large number of genes. Indeed, recent investigations demonstrate that global DNA hypomethylation did not alter the transcriptome as drastically as previously expected (Blattler et al., 2014; Li et al., 2015). In contrast to methylation changes in a global fashion in *Dnmt*-knockout cells, environmental exposure including BPA exposure seems to change methylation patterns only at specific genomic loci (Dolinoy et al., 2007; Kundakovic et al., 2015). Therefore, if the environmental exposure-altered methylation has an important impact on gene transcription, we expect that the CpG sites bearing the altered methylation must play critical roles in gene regulation, such as CpG sites within enhancers or binding motif of critical transcriptional factors (TF) or the imprinting control regions (ICRs).

## Histone Modification for Permissive or Inhibitive Transcription

Another layer of epigenetic mechanisms for the regulation of gene expression is posttranslational modifications (PTMs) of

histone tails. To date, there are a few hundred distinct histone tail modifications, including histone acetylation, methylation, and phosphorylation (Strahl and Allis, 2000; Tan et al., 2011). The former two have been known for decades to possess roles in gene transcription, and several representative histone acetylation and methylation marks have been actively pursuing by the community for many years (Allfrey et al., 1964; Barski et al., 2007; Wang et al., 2008). For example, the ENCODE project selects eight histone marks out of many potential histone acetylations and methylations: H3K4me1 (H3 lysine 4 monomethylation), H3K4me2, H3K4me3, H3K9ac (H3 lysine 9 acetylation), H3K27ac, H3K36me3, H3K9me3, and H3K27me3 (Yue et al., 2014). The former six are active histone marks in association with expressed genes, whereas the latter two are repressive marks in association with silent genes. Except H3K36me3 at gene bodies, the remaining seven histone marks can be enriched at promoters and enhancers (Dai and Wang, 2014).

Each histone mark demonstrates useful and maybe distinct information, but they can also share overlapping information (Wang et al., 2008). Seven marks can be present at promoters and/or enhancers, but each mark may better serve a particular application. H3K4me3, H3K9ac, and H3K27ac are enriched at promoters or transcription start sites of active genes (Wang et al., 2008). Similarly, H3K4me1, H3K9ac, and H3K27ac are enriched at enhancers, with H3K27ac particularly enriched at active enhancers (Creighton et al., 2010). One active mark (out of the former five) and one inactive mark (either H3K9me3 or H3K27me3) can be used to predict bivalent promoters (Bernstein et al., 2006; Roh et al., 2006). Among potential combinations, H3K27me3 and H3K4me3 are popularly chosen for testing promoter bivalency. In addition to these popularly analyzed histone acetylations and methylations, many more new histone marks have been recently identified (Tan et al., 2011) and their roles in gene transcription and biological pathways

are less clear (Goudarzi et al., 2016). Recently, even one of the well-characterized histone acetylation marks, H3K27ac, was shown to have an unexpected suppressing role. In contrast to the transcriptional activation of H3K27ac at enhancers and promoters, age-related up-regulated genes contain hyper H3K27ac in gene bodies, acting to suppress the overexpression of inflammaging genes (Cheng et al., 2018). In addition, the expression changes of these age-related genes can be predicted by gene body H3K27ac level. It seems that histone marks have the potential to reflect the aging stage or disease conditions.

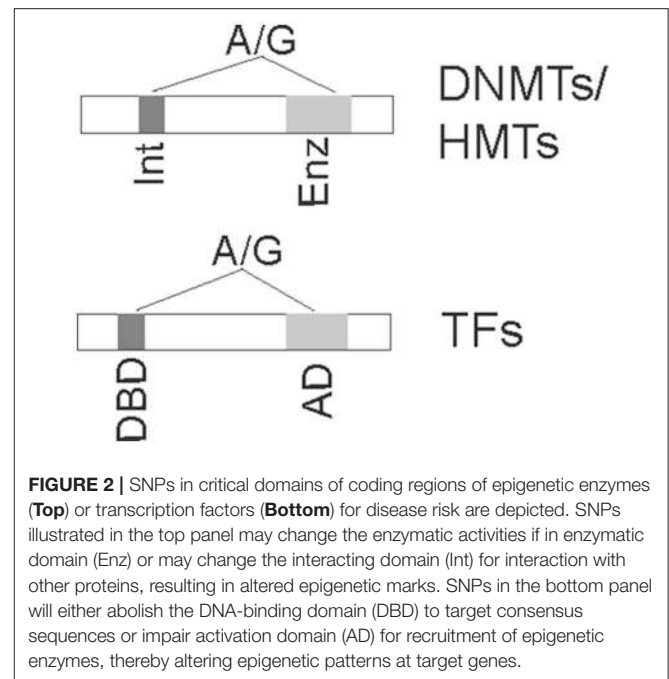
## INTERACTIONS OF GENETIC VARIANTS AND EPIGENETIC PATTERNS FOR COMPLEX DISEASE RISK

A common SNP is defined as a single base change in a DNA sequence that occurs among a significant proportion ( $\geq 1\%$ ) of a population (Lockwood et al., 2014). SNPs may reside in coding regions or non-coding regions. Contrary to the common belief, it is estimated that about 90% of GWAS-associated genetic variants reside in non-coding regions (Welter et al., 2014; Farh et al., 2015). While variants in coding regions may confer disease risk through altered protein sequences and, therefore, altered protein functions, variants in non-coding regions (usually enhancers) contribute to disease susceptibility through changing gene transcription and non-coding RNAs (Hrdlickova et al., 2014). These genetic variants might engage in crosstalk with sequence-specific transcription factors and epigenetic patterns (including DNA methylation) (McVicker et al., 2013), thereby impacting the transcription of genes locally or remotely.

### Genetic Variants in Coding Regions for Disease Susceptibility via Altering Epigenetic Patterns

There are at least two scenarios, including SNPs within epigenetic enzymes and SNPs in TFs (Figure 2), that lead to the altered epigenetic patterns for increased disease risk. SNPs within epigenetic enzymes (including DNMTs and histone modifying enzymes) may potentially alter their functions, thereby subsequently changing epigenetic patterns in a genome-wide fashion. SNPs in TFs may impact the function of TFs' binding to DNA or its recruitment of epigenetic enzymes, thereby changing epigenetic patterns indirectly (Khandanpour et al., 2012).

SNPs in DNA methyltransferases have been reported in association with complex diseases. For example, polymorphisms in three DNMTs with methylation activity (DNMT1, DNMT3A, and DNMT3B) and in one DNMT without methylation activity (DNMT3L) are associated with an increased risk of schizophrenia (Saradalekshmi et al., 2014). Though many SNPs reside inside non-coding regions, at least one of them (rs2228611, within DNMT1 exon) has been found to be significantly associated with schizophrenia at genotypic and allelic levels in a South Indian population. SNPs in DNMT1 (exonic rs16999593) and DNMT3A (intronic rs1550117) may contribute to the gastric cancer risk, according to a recent meta-analysis (Li et al., 2016). Polymorphism rs1550117 of DNMT3A has been shown in



association with the late-onset Alzheimer's disease. Specifically, patients with an AA genotype showed a 2.08-fold risk when compared to patients with a GG genotype (Ling et al., 2016). When investigating the imprinting disorder Beckwith-Wiedemann syndrome, Dagar and colleagues screened variants within the DNMT1 coding region and identified three patients (out of 53 examined) who contained three rare missense variants: rs138841970: C>T, rs150331990: A>G, and rs757460628: G>A; encoding NP\_001124295 p.Arg136Cys, p.His118Arg, and p.Arg1223His, respectively (Dagar et al., 2018). Using the DNMT1 binding as a surrogate for DNMT1 enzymatic activity, GFP-tagged DNMT1 fusion proteins with site-directed mutation show a reduced binding affinity (40-70%) of variants compared to that of the wild-type DNMT1. While it would be more informative to validate the expected DNA demethylation with bisulfite sequencing in three patients, the report at least provides a reasonable support for an association between variants in DNMT1 and increased disease risk (Dagar et al., 2018).

SNPs in histone-modifying enzymes and/or cofactors are also associated with complex disease susceptibility. GWAS from the GABRIEL Consortium (a multidisciplinary study of genetic and environmental causes of asthma) identified significant SNPs within histone-modifying enzymes in asthmatic patients, including histone deacetylases (HDAC4, HDAC7, HDAC9) and H3 lysine 36 demethylases (KDM4C, KDM2A) (Moffatt et al., 2010; Kidd et al., 2016). A focused summary of SNPs and mutations within histone lysine methyltransferases (KMTs) and histone lysine demethylases (KDMs) (Van Rechem and Whetstone, 2014) listed a few SNPs in coding regions of KMTs and KDMs in association with diseases. In contrast, the less frequent mutations or the deletion of KMTs and KDMs seem to be associated with more diseases. Although the variations in coding regions are expected to affect protein stability,

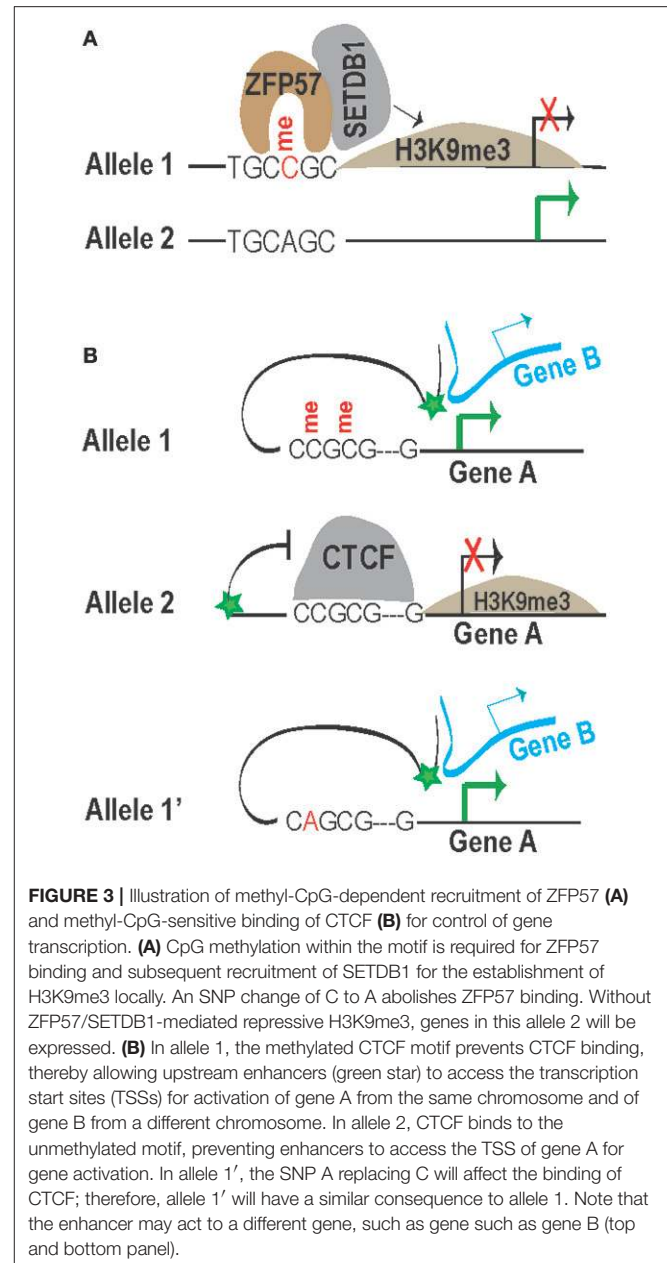
folding, ligand-binding, and/or post-translational modification, it remains unclear how these SNPs impact the function of DNMTs and histone-modifying enzymes.

SNPs within coding regions of DNA-binding TFs can also potentially impact epigenetic patterns (Figure 2). These variants have an indirect effect, compared to the direct effect of SNP-altering epigenetic enzymes. The human growth factor independence 1 (GFI1), a DNA-binding transcription repressor, is important for hematopoietic stem cell and B and T cell differentiation. About 3 to 7% of white subjects contain GFI136N, which impairs binding to target genes such as *HOXA9*, resulting in elevated histone H3K4me2 signals and thereby promoting gene transcription. Compared with the GFI136S genotype, people with GFI136N have an increased risk (60%) for acute myeloid leukemia (Khandanpour et al., 2012). In one early cited report, identified common SNPs that have a potential impact on DNA binding of zinc finger TFs are unlikely to alter gene transcription in trans (Lockwood et al., 2014). However, Lockwood et al.'s results also suggest that large-scale analyses might offer a different conclusion, as in the case of GFI1.

### Genetic Variants in Non-coding Regions for Complex Disease Risk via Crosstalk With DNA Methylation Status and/or TF Binding for Shaping Chromatin Structure

As aforementioned, about 90% of GWAS-identified genetic variants reside within the non-coding regions (Farh et al., 2015). SNPs in non-coding regions may impact gene function via several mechanisms. Intronic SNPs may affect splicing and/or mRNA stability, and studies have shown that an intronic SNP (rs910083-C) within *DNMT3B* is associated with an increased risk of nicotine dependence and squamous cell lung carcinoma. Though the mechanism is yet unclear, this SNP is associated with hypermethylation of about 252 bp upstream of the *DNMT3B* gene (Hancock et al., 2017). SNPs affecting the TFs' binding sites can predispose disease susceptibility by impacting the associated TFs' regulatory pathways (Figure 3), partially because SNPs at conserved residues within the consensus sequences for TFs can decrease or even abolish the binding of these TFs or because SNPs generate a new binding site for TFs. For example, the SNP A>C within the *Hcn2* locus for an allele-specific metal responsive element of metal regulatory transcription factor 1 (MTF1) (Martos et al., 2017). Hence, SNPs in this category could alter merely the DNA sequence, or, if they alter or generate CpG dinucleotides, they could simultaneously alter both the consensus sequence and its DNA methylation.

Accumulating evidence suggests that GWAS-identified and disease risk-associated genetic variants may largely confer their impacts by changing the expression of neighboring genes (Kilpinen et al., 2013; McVicker et al., 2013; Soldner et al., 2016; Allen et al., 2017; Gallagher et al., 2017; Li X. et al., 2017). Such changes may act through variant-disrupted motifs that eliminate or introduce the TF binding as described above. In our opinion, two TFs—ZFP57 and CTCF—are particularly intriguing to work with because of their known potential to impact chromatin in a global fashion (Ong and Corces, 2014; Riso et al., 2016). Both



**FIGURE 3** | Illustration of methyl-CpG-dependent recruitment of ZFP57 (A) and methyl-CpG-sensitive binding of CTCF (B) for control of gene transcription. (A) CpG methylation within the motif is required for ZFP57 binding and subsequent recruitment of SETDB1 for the establishment of H3K9me3 locally. An SNP change of C to A abolishes ZFP57 binding. Without ZFP57/SETDB1-mediated repressive H3K9me3, genes in this allele 2 will be expressed. (B) In allele 1, the methylated CTCF motif prevents CTCF binding, thereby allowing upstream enhancers (green star) to access the transcription start sites (TSSs) for activation of gene A from the same chromosome and of gene B from a different chromosome. In allele 2, CTCF binds to the unmethylated motif, preventing enhancers to access the TSS of gene A for gene activation. In allele 1', the SNP A replacing C will affect the binding of CTCF; therefore, allele 1' will have a similar consequence to allele 1. Note that the enhancer may act to a different gene, such as gene such as gene B (top and bottom panel).

are DNA-binding TFs, and polymorphisms at different residues within their binding motifs seem to carry different weight with TF binding. In addition, these two TFs represent two distinct classes of TFs in terms of the impact of methylation status on their recruitments. That is, methylation is required for the binding of ZFP57 (Quenneville et al., 2011), while methylation prevents the binding of CTCF (Bell and Felsenfeld, 2000).

### Methylation-Sensitive CTCF

#### CTCF biology

The insulator CTCF is an intriguing TF with potential roles for altering chromatin in a genome-wide fashion (Phillips and Corces, 2009; Ong and Corces, 2014). CTCF may also

serve as a pioneer TF to shape local chromatin for recruiting additional TFs to their target sites (Tehranchi et al., 2016). In contrast to ZFP57, which is dependent upon methylation for binding, hypomethylation of the consensus sequence CCGCGNGGNGGCAG is critical for CTCF binding (Bell and Felsenfeld, 2000; Kim et al., 2007; Cuddapah et al., 2009). One essential role of CTCF is to block the spread of chromatin structure to the opposite side of a CTCF binding site (barrier) or to block the enhancer activity toward a promoter (Cuddapah et al., 2009; Phillips and Corces, 2009). The prototypical role of insulation has been revealed by investigation of CTCF at the imprinted *H19/Igf2* locus (Bell and Felsenfeld, 2000). *H19* and *Igf2* are located on the opposite side of CTCF-bound regions and there are a few enhancers downstream of the *H19* gene. CTCF binding motifs show allelic methylation, thereby dictating allelic CTCF binding. CTCF cannot bind to the methylated motifs of the paternal allele, allowing enhancers to access the paternal *Igf2* gene for activation. In contrast, CTCF binds to the unmethylated motifs of the maternal allele, thereby blocking enhancer access of *Igf2*, but allowing access of *H19* for stimulation. Experimental mutations to abolish CTCF binding affect imprinted expression within the locus (Singh et al., 2012).

CTCF's insulation is not limited to imprinted loci and is expected to act in a three dimensional fashion, presumably because enhancers can loop with multiple loci to regulate genes on the same or even different chromosome(s) (Spilianakis et al., 2005). **Figure 3B** presents a simplified cartoon to show the enhancer-blocking activity of CTCF in crosstalk with SNPs. In Allele 1, allelically methylated CpG sites prevent CTCF binding, enabling the enhancer to interact with gene A from the same chromosome and gene B from a different chromosome for stimulation. In Allele 2, the unmethylated motif allows CTCF insulation, thereby blocking the enhancer's access to stimulate gene A and, hence, keeping gene A silenced. A SNP change from C to A will potentially alter the conserved residues (illustrated in Allele 1') and perhaps also the methylation status, preventing, or enabling CTCF insulation. Depending on the "conserveness" affected, the SNPs or mutations may reduce or even completely abolish CTCF binding to its motifs (Li W. et al., 2017), eventually affecting its insulation effect.

### CTCF, genetic variation, and gene expression changes

Having briefly summarized CTCF insulation, we then did an in-depth examination of genetic variants within CTCF-binding motifs and their crosstalk with ABTF and ASMs in terms of impact on human health. Given the diverse functions of CTCF in transcription, imprinting, and X-chromosome inactivation (Phillips and Corces, 2009; Ong and Corces, 2014), it is not surprising that investigations begin to reveal the associations of these genetic variants with diseases or complex traits (**Table 1**) by impacting the allelic binding of CTCF. Among many reports (**Table 1**), several thorough investigations have presented their detailed insights of CTCF's allelic binding through long-range chromatin interactions that affect gene expression and, therefore, disease risks. Like ZFP57, CTCF's insulation or regulation of

long-range chromatin interaction is not limited to ICRs of imprinted loci, but also applies to regular genomic regions (exemplified below).

For example, severe human influenza was reported in association with SNP rs34481144 A/G in the 5' UTR of a regular *IFITM3* gene (not imprinted) that encodes antiviral protein IFITM3 for inhibition of viral entry (Allen et al., 2017). IFITM3 in memory CD8<sup>+</sup> T cells promotes the adaptive immunity for antiviral resistance. Functional studies have revealed that the risk allele A decreases the binding of interferon regulatory factor 3 (IRF3) but increases the binding of CTCF. The latter is expected to impact the expression of *IFITM3*-neighboring genes via its insulation activity. Authors also demonstrate that increased methylation at rs3448114 blocks the binding of CTCF, thereby enhancing the expression of *IFITM3*. These investigations exemplify our proposed "SNP intensifier" model; a tiny difference at one residue (A/G, with the former disrupting a CpG site) is intensified via changing allelic DNA methylation, thereby affecting CTCF affinity and altering the expression of neighboring genes for disease risk (Allen et al., 2017). While this line of study has been limited to *in vitro* assays and would be much improved with *in vivo* confirmation of CRISPR/cas9-altered SNP A/G, the presented results are agreeable and consistent with the model presented later in section Conclusion, Challenges, and Perspectives.

Another example is the investigation of birth weight of babies conceived through assisted reproductive technology (ART) that was related to the imprinted locus. In ART, fresh embryo transfer-derived newborns are associated with low birth weight, while newborns derived from frozen embryo transfer are associated with increased birth weight (Marjonen et al., 2018). The underlying mechanism remains largely unknown. ART is expected to increase imprinting defects (Eroglu and Layman, 2012); therefore, the *IGF2/H19* locus, with its role in normal placental and embryonic growth, was selected for characterization. As described above, the *IGF2/H19* locus is regulated by an ICR that bears seven CTCF-binding motifs. In placentas from women who have used ART, a SNP rs10732516 A/G within the sixth binding motif (CCGCGc/tGGNGGCAG or complementary strand CTGCCNCCa/gCGCGG) of CTCF results in allele-specific demethylation on the paternal allele of rs10732516 paternal A/maternal G genotype, but not on the paternal G/maternal A genotype (Marjonen et al., 2018). The SNP allele A would interrupt the first CpG site of the complementary strand sequence CTGCCNCCa/gCGCGG and is associated with hypomethylation. It seems that the first CpG site might carry more weight in controlling DNA methylation status than do the following two CpG sites. These investigations also lack the expression status of *H19* and *IGF2* among genotypes (A/A, A/G, and G/G) for further mechanistic analysis. Fresh embryo transfer-derived newborns with the G/G genotype have shown an increased birth weight and larger head circumference when compared to the parameters of the A/A genotype (Marjonen et al., 2018).

The same group above has also examined SNP rs10732516 at the *IGF2/H19* locus for the consequences of prenatal alcohol

**TABLE 1** | The crosstalk among CTCF, genetic variants, epigenetic variation in complex diseases or traits. (N/A, not available).

Complex traits or diseases	Loci	Mechanism			Key results and/or model suggested	References
		Genetic variants	Epigenetic variation	ABTF		
Asthma	ORMDL3	Yes	Yes	Yes	A SNP (rs4065275) in an enhancer within the 1st intron of <i>ORMDL3</i> promotes CTCF binding, whereas another SNP (rs12936231) downstream of the enhancer impairs CTCF binding. SNPs therefore alters three-dimensional organization in the asthma-risk allele to facilitate the expression of <i>ORMDL3</i> , which inhibits IL-2 production.	Schmiedel et al., 2016; Bérubé et al., 2017
Birth weight	H19/IGF2	Yes	Yes	N/A	A SNP rs10732516 A/G within the sixth binding motif of CTCF within the ICR results in allele-specific demethylation in paternal allele of rs10732516 paternal A/maternal G genotype. Allelic binding of CTCF is expected.	Marjonen et al., 2018
Cerebellum weight (CW)	H19/IGF2	Yes	Yes	Yes	DNA methylation at CTCF-binding site 3 within ICR explains ~25% of the CW variation; Genetic variation of the ICR in strong association with CW in a parental-origin dependent fashion.	Pidsley et al., 2012
Dementia	TMEM106B	Yes	Yes	Yes	The risk allele of rs1990620 increases the recruitment of CTCF, thereby leading to haplotype-specific effects on three-dimensional chromatin interactions and thus increased <i>TMEM106B</i> expression. The latter increases cytotoxicity for risk of neurodegeneration.	Gallagher et al., 2017
Influenza	IFITM3	Yes	Yes	Yes	A SNP (rs34481144) in the 5' UTR of antiviral <i>IFITM3</i> gene renders the risk allele with lower TF IRF3 binding but higher CTCF binding, thereby altering expression correlations among <i>IFITM3</i> -neighboring genes. The risk allele also disrupts a CpG site that is under differential methylation in CD8+ T cell subsets.	Allen et al., 2017
Lung cancer	DAGLA	Yes	N/A	N/A	A SNP within CTCF binding site inside an intron of <i>DAGLA</i> was significantly associated with increased risk of lung cancer	Dai et al., 2015
Lynch syndrome	MLH1	Yes			A SNP rs143969848 (G>A) within CTCF motif, part of an enhancer and also upstream of <i>MLH1</i> transcription start site, disrupts enhancing activity and <i>MLH1</i> expression.	Liu et al., 2018
Mental illness	3p22 (TRANK1)	Yes		Yes	The risk allele of SNP rs9834970 shows lower baseline expression of <i>TRANK1</i> that may further alter genes important for neurodevelopment/differentiation. While the role of rs9834970 unknown, a nearby SNP rs906482 alters CTCF binding and the allele with increased CTCF binding is the risk allele of rs9834970.	Jiang et al., 2018
Osteoporosis	SOST	Yes		Yes	Four SNPs within the locus of <i>SOST</i> (negative regulator of bone formation and positive regulator of bone resorption). Among them, the SNP rs1230399 shows FOXA1 binding activity, resulting a T allele-specific activation; the SNP rs1107748 renders C allele transcriptional enhancer activity through a CTCF binding site; Variant rs75901553 C > T abolishes the binding site of miR-98-5p that is negative responsive to parathyroid hormone.	Ye et al., 2018
Osteoporosis	1p36.12 (LINC00339)	Yes		Yes	A SNP rs6426749 functions as a distal allele-specific enhancer stimulating the expression of a lncRNA	Chen et al., 2018
Type 2 diabetes	TF binding Sites	Yes	N/A	N/A	SNPs within motifs of CTCF, EP300, FOXA1/2, HNF4A, and TCF7L2 are associated with T2D from computational analyses	Cheng et al., 2017

exposure, which is known to affect development of the fetal nervous system and to restrict fetal head growth (Treit et al., 2016). Investigators found that alcohol exposure decreases the hypermethylation of the paternal allele of rs10732516 paternal A/maternal G genotype in placentas (Marjonen et al., 2017).

Additional reports listed in **Table 1**, including rs1990620 within the *TMEM106B* locus in association with neurodegeneration, support the “SNP intensifier” working

model (presented later) (Gallagher et al., 2017). Lastly, CTCF is linked with intellectual disability (Bastaki et al., 2017; Hori et al., 2017). Altogether, a change of one residue (for example, C to A in **Figure 3**) may abolish or introduce the allelic recruitment of CTCF (and its associated insulation or long-range chromatin interactions), thereby causing a change from biallelic to monoallelic (or vice versa) expression of genes, and thus the copy numbers of the mRNA of related genes are altered.

### Additional methylation-sensitive TFs

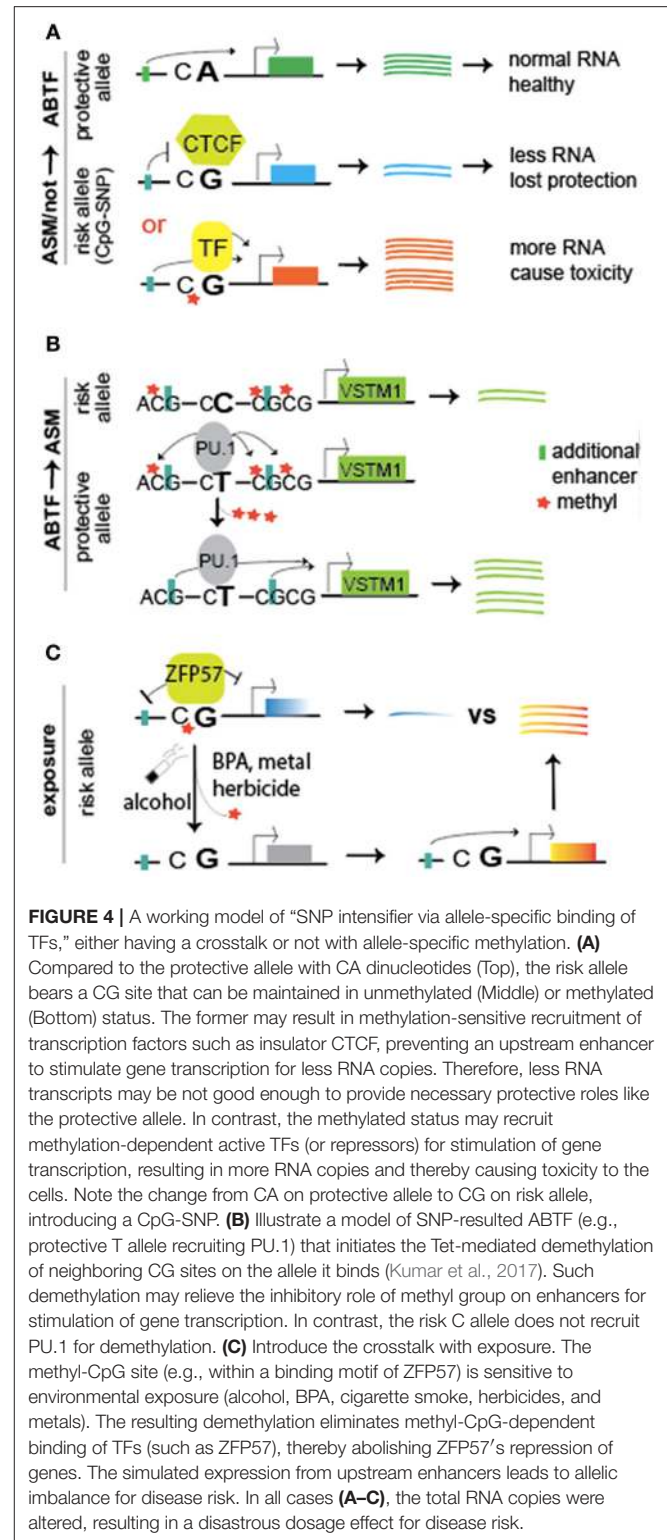
Many additional TFs—including bHLH, bZIP, and ETS family members—are also inhibited by methyl-CpG (Yin et al., 2017). Therefore, any SNPs that change these TFs' consensus motifs and/or result in methyl-CpG (e.g., "AG" becoming methylation-prone "CG") have the potential to alter the binding of these TFs and, subsequently, to alter these TFs-controlled genes. Future investigations in these similar TFs will certainly be promising.

### Methylation-Dependent ZFP57

The DNA-binding transcription factor ZFP57 only binds to the methylated TGCCGC hexanucleotide and, subsequently, recruits cofactor KAP1 and DNMTs, as well as SETDB1 (Quenneville et al., 2011). The recruited SETDB1, as a histone-modifying enzyme, is expected to establish H3K9me3 heterochromatin mark around ZFP57-targeted sites (Figure 3A) (Anvar et al., 2016). In addition, polymorphisms at CpG dinucleotides will not only change the consensus residue, but also change the methylation status (namely CpG-SNP), which may impact TF binding more severely. The A replacing the C in Allele 2 will change the consensus residue and methylation status (Figure 3A). Indeed, when AT replaced the fourth and fifth residues CG, the resulting oligonucleotide (mut1) was shown to lose both the consensus sequence and its methylation status. When TGC was changed to GAG, the resulting oligonucleotide (mut2) only altered consensus sequences, but still maintained the methyl-CpG. Intriguingly, the mut1 probe did not compete as well as the mut2 probe, at least in electrophoresis shift assays (Quenneville et al., 2011), suggesting that the methylation status may be more critical than the consensus sequence for ZFP57 binding.

While the association of methylated TGCCGC in H3K9me3 patterns was initially investigated in imprinting control regions (ICRs) of imprinted loci (Quenneville et al., 2011), regular genomic regions are expected to have similar ZFP57-associated H3K9me3. In other words, so long as genomic loci contain the TGCCGC sequences, these loci are potentially subjected to ZFP57/SETDB1-regulated H3K9me3. Therefore, SNPs at ZFP57 motifs are expected to play a role beyond imprinted genes (Anvar et al., 2016). Because ZFP57-mediated H3K9me3 signal distribution seems to have a relatively long range, any alterations of TGCCGC sequences could potentially affect the transcription of multiple genes within the region.

The effect of SNPs within the ZFP57 motifs on disease risk remains to be determined in human populations. Our literature search was relatively futile in this aspect; however, a few reports indeed linked methylation QTLs to expression changes of ZFP57 in disease/traits, including post-traumatic stress disorder (Rutten et al., 2018), metabolic trait (Volkov et al., 2016), and psychosis (Rivollier et al., 2017). Compared to the extensive studies of SNPs within CTCF binding motifs for complex traits/diseases (which burgeoned only in the past 2 to 3 years) (Table 1), we expect more upcoming reports to show the effect of SNPs within ZFP57 binding motifs on disease risk because these SNPs seem to abnormally alter allelic expression of genes (Anvar et al., 2016). From F1 hybrid ES cells between C57BL/6 and Cast/EiJ, genetic variants with disruption of ZFP57 consensus



motif and methylation status are linked to monoallelic expression of neighboring genes (Strogantsev et al., 2015). Note: Allelic increases or decreases of gene expression are key for CTCF-involved disease risk (Figure 4).



In summary, this section presents the mechanistic insights of SNP-resulted allele-specific methylation that facilitates the allelic binding of TFs such as ZFP57 (Figure 4A, bottom panel) and of SNP-resulted motifs (unmethylated) for insulator CTCF (Figure 4A, middle panel). In both cases, ABTFs play a key role in allelic regulation of gene transcript. In the following section, we will review SNP-resulted ABTF, thereby leading to the subsequent ASMs and potential ASE (Figure 4B).

## GENETIC VARIANTS IN SHAPING ALLELE-SPECIFIC DNA METHYLATIONS (ASMS) AND POTENTIAL ALLELE-SPECIFIC GENE EXPRESSION (ASE) FOR DISEASE RISK

### Unexpected ASMs and ASE for Autosomal Genes in the Human or Mouse Genome

It is worth presenting genetic variants in shaping ASMs (separately from the above section Interactions of Genetic Variants and Epigenetic Patterns for Complex Disease Risk). In this section, we envision the underlying mechanism for ASMs and ASE. Traditionally, both ASM and ASE are considered as the phenomena of genomic imprinting and X chromosome inactivation (XCI). However, more recent reports have started to unveil unexpected allelic asymmetries of many non-imprinted autosomal genes in the mammalian genome (Klengel et al., 2013; Izzi et al., 2016).

Increasing evidence shows that both the human and mouse genomes contain hundreds or even thousands of ASMs (Cheung et al., 2017). Originally aimed to identify unknown imprinted genes, one study identified many unexpected ASMs from genomic loci outside of known imprinted regions in several human tissues (Kerkel et al., 2008). Further investigations demonstrate that these ASMs are tissue-specific and individual-specific, suggesting genetic background in affecting ASMs (Yang et al., 2010). Looking at two mouse strains (C57BL/6 and BALB/c) and their F1 hybrid offspring, investigators focused on 181 large genomic intervals with an approach based on methyl-CpG immunoprecipitation and locus-wide tiling arrays, identifying several hundreds of differentially methylated regions and strain-specific methylation patterns controlled by *cis*-acting polymorphisms (Schilling et al., 2009). Regarding the identification of unknown imprinted loci in the mouse genome, our group has developed a computational approach based on the feature of monoallelic hyper- or hypomethylation at ICRs to scan the base-resolution DNA methylomes from the mouse ES cell J1 line (Li et al., 2015; Martos et al., 2017) that identified more than 2,000 regions showing bimodal methylation patterns. These regions are potentially associated with monoallelic methylation. In the subsequent validation, we generated four independent hybrid ES lines (from a reciprocal cross between 129S1/SvImJ and Cast/EiJ or between C57BL/6NJ and Cast/EiJ) and confirmed the ASM of the *Hcn2/Polrmt* locus—that is, within this locus, the Cast allele with a SNP C is always hypomethylated (i.e., independent of parental origin), whereas the 129 allele or the C57 allele with a SNP A is always hypermethylated. Intriguingly, the

SNP C renders the Cast allele a new motif for metal regulator transcription factor MTF1 (Martos et al., 2017).

### Do Non-imprinted ASMs Control ASE?

The extent to which the non-imprinted ASMs control allelic gene expression, remains to be defined. Because imprinted ASMs control ASE, one would expect the similar control of ASMs for the allelic expression of genes on autosomes. Indeed, reports support that ASMs are at least one of the factors to cause allelic imbalance of gene expression.

1. **Genetic variants-resulting demethylation of one allele for ASM and ASE.** This class of variant regulation is initially featured by SNP-facilitated ABTF, which initiates allelic demethylation for ASM and ASE (Figure 4B). In other words, ABTF leads to ASM. One example is an atopic dermatitis-associated SNP rs612529 T/C in the promoter of *VSTM1* that encodes SIRT1. It turns out that the protective T allele facilitates the recruitment of transcription factors YY1 and PU.1 (Kumar et al., 2017). YY1 can either activate or repress gene transcription, depending on the context in which it binds (Gordon et al., 2006), whereas PU.1 seems to demethylate its target genes through Tet2 demethylation (de la Rica et al., 2013). Indeed, the neighboring CpG sites of the PU.1 binding site on the protective T allele, as compared to those of the risk C allele, are hypomethylated in monocytes, thereby leading to allelic upregulation of gene transcripts. The risk C allele does not recruit PU.1, resulting in the low expression of SIRT1 in monocytes. The latter leads to a higher risk for manifestation of an inflammatory skin disease (Kumar et al., 2017). Another example is the allele-specific demethylation of long-range enhancers of the *FKBP5* (FK506 binding protein 5) gene, which increases the susceptibility of developing stress-related psychiatric disorders in adulthood (Klengel et al., 2013). *FKBP5* is an important regulator of the glucocorticoid receptor complex that is involved in the stress hormone system. Exposure to early childhood abuse is associated with demethylation of the enhancer on the risk allele (with the AA genotype), bringing the enhancer through long-range interaction to transcription machinery for allelically increased transcript of *FKBP5*. The latter results in a long-term malfunction of the stress hormone system and a genome-wide impact on the function of immune cells (Klengel et al., 2013). The third example is growth differentiation factor 5 (*GDF5*), a ligand of the TGF-beta superfamily of proteins, essential for normal skeletal development. In the 5' UTR of the *GDF5* gene, a C-to-T SNP rs144383 is a risk factor for osteoarthritis of the knee. Methylation of a highly conserved CpG site (4 bp upstream of rs144383 and part of SP1 and SP3 motif) affects allele-specific binding of repressor SP1 and SP3, thereby attenuating the repressive effect of SP1 and SP3 proteins and resulting in allelic expression of *GDF5* (Reynard et al., 2014). This ABTF of SP1 and SP3 may exemplify how the above mentioned MTF1 predisposed the hypomethylation of Cast allele within the *Hcn2* locus in our studies. Presumably, this MTF1 predisposes the Cast allele for

Tet proteins for demethylation; however, the extent to which the ABTF predisposes allelic hypomethylation remains to be determined (Martos et al., 2017). Working on chromosome 21 in leukocytes from healthy individuals, investigations also show that the genetic variation among individuals affects ASMs, thereby leading to ASE (Zhang et al., 2009). In acute lymphoblastic leukemia, ASE for 470 SNPs within 400 genes was detected. The level of ASE varies from a 1.4-fold overexpression of one allele to strictly monoallelic expression. Further investigation suggests that ASE is associated with promoter CpG site methylation (Milani et al., 2009).

2. **Genetic variants-resulting methylation of one allele for ASM and ASE.** In contrast to the demethylation of one allele for ASMs mentioned above, genetic variant-related methylation of one allele can also lead to ASMs and potential ASE. For example, the SNP rs12041331 has been linked to cardiovascular disease and platelet reactivity (Izzi et al., 2016). The major G allele of rs12041331 (leading to a CpG-SNP within the intron of the *PEAR1* locus) is linked to a higher transcript level than the minor A allele in endothelial cells and platelets. The resulting CpG site of GG carriers is fully methylated in leukocytes. Intriguingly, this methylated CpG site of the G allele recruits more nuclear proteins than does the unmethylated A allele. However, the authors did not characterize these nuclear proteins (Izzi et al., 2016); otherwise, their insights would be clearer. Based on our summary above, it is reasonable to expect a similar binding of TFs like ZFP57 to the methylated CpG site. That is, methylation-dependent recruitment of c-Jun and/or ATF3 (resulted due to CpG-SNP on the G allele) may understand the higher *PEAR1* expression and risk. More recent reports linking ASMs to disease or complex traits include the SNP rs174537-regulated ASM at the *FADS* gene locus for long-chain polyunsaturated fatty acid biosynthesis (Rahbar et al., 2018) and ASMs of susceptibility genes for inflammatory bowel disease (Chiba et al., 2018).

## Is Dosage Effect the Key for Disease Risk?

Imprinting disorders have been linked to abnormal biallelic expression or biallelic silencing of imprinted genes in contrast to monoallelic expression (Bartolomei, 2009; Weksberg, 2010). The dosage effect is certainly key for imprinting disorders. As for non-imprinted loci with ASMs and/or ASE, we expect a similar role. Allelic imbalance—a hallmark of cancer—has been known to contribute to cancers for many years. For example, monoallelic expression of cancer-related genes, including *TP53* and *IDH1*, seems to be in association with tumor aggressiveness and progression (Walker et al., 2012). The ASE of *BRCA1* and, to a lesser extent, of *BRCA2*, contributes to an increased risk for breast cancer (Chen et al., 2008). ASE is also observed in acute lymphoblastic leukemia, as described above (Milani et al., 2009). Our thorough examinations including insulator CTCF-involved diseases/traits (examples in **Table 1**) and other TFs-mediated traits/diseases (described above) demonstrate the causative increase or decrease of gene transcripts of one allele.

In the authors' opinion, this is a dosage effect, like imprinting disorders for complex traits or diseases.

## VULNERABILITY OF GENETIC VARIANT-INFLUENCED ASMS AND/OR ASE IN RESPONSE TO ENVIRONMENTAL FACTORS FOR DISEASE RISK

The extent to which these genetic variants-influenced ASMs and ASE are vulnerable to environmental factors remains to be determined. However, we may learn from previous investigations on imprinted ASMs, as both imprinted ASMs and genetic variants-influenced (or non-imprinted) ASMs share the feature of mono-allellicity. Since there is no backup from the complementary allele, the mono-allellicity of methylation, presumably, is the cause for vulnerability. It is, therefore, reasonable to expect that ASMs (both imprinted and non-imprinted) are vulnerable to exposure.

The imprinted ASMs at ICRs, the key for imprinted gene expression, are known to be vulnerable to environmental exposures, including BPA and cadmium, as they can alter the allelic expression of imprinted genes and, consequently, increase disease susceptibility (Heijmans et al., 2008; Susiarjo et al., 2013; Van de Pette et al., 2017; Cowley et al., 2018). Indeed, in response to cadmium exposure, ASMs at ICRs do show a higher sensitivity to cadmium when compared to other loci in newborn cord blood and maternal blood (Cowley et al., 2018). As for alcohol use disorder, prenatal alcohol exposure alters one ASM of the *H19/Igf2* locus, thereby causing about a 1.5-fold decrease of *Igf2* transcripts (Downing et al., 2011). The ASMs of imprinted loci, including *Dio3*, and *H19/Igf2*, are susceptible to fetal alcohol exposure, thereby changing the allelic expression of *Dio3* and *Igf2* (Haycock and Ramsay, 2009; Tunc-Ozcan et al., 2016, 2018). Lastly, calorie restriction can also alter imprinted *Igf2* expression in a sex-dependent manner. In rats, moderate calorie restriction during gestational days 8 through 21 of Sprague-Dawley dams (F0) increases the adult hippocampal *Igf2* transcripts in F1 females, and in these F1 females-produced (in cross with naïve male Brown Norway) F2 offspring (Harper et al., 2014). Although it was not fully investigated as to whether calorie restriction altered ASM within the *H19/Igf2* locus for this intergenerational increase of *Igf2* in female offspring, similar treatment with maternal vitamin D depletion has suggested changes in DNA methylation (Xue et al., 2016).

With the established vulnerability of imprinted ASMs to exposure, the non-imprinted ASMs are expected to behave similarly. As exemplified above, the risk allele (AA carrier) of the *FKBP5* locus is indeed vulnerable to child abuse exposure (Klengel et al., 2013). With social or mental stress, the oxytocin receptor gene (*OXTR*) SNP rs53576 (G-A) is expected to be associated with social behavior. Investigations demonstrate that prenatal mental stress exposure is linked to child autistic traits, but not related to *OXTR* methylation across the rs53575 G allele homozygous children or A allele holder (Rijlaarsdam et al., 2017). Though not clear about the allelic sensitivity to

mental stress from the investigation, the *OXTR* methylation levels were positively in association with social problems for the G allele homozygous children (but not the A allele carriers). Using five human cell types for 50 treatments, Luca and colleagues identified 1,455 genes with ASE and 215 genes with gene-by-environment (GxE) interactions (Moyerbrailean et al., 2016). More importantly, exposure-perturbed genes showed a 7-fold increased odds of being reported in GWAS. Almost half of 215 genes showing GxE interactions are associated with complex traits, as revealed by GWAS. These results are consistent with the idea that genes with ASE are vulnerable to exposure. Additional evidence regarding alcohol use disorder also supports this idea. In rats, drinking alcohol affects the ASE of about 300 genes, as reported recently from Zhou and colleagues (Lo et al., 2018). This affection is expected through the crosstalk among genetic variants, epigenetic patterns, and alcohol exposure.

## CONCLUSION, CHALLENGES, AND PERSPECTIVES

Accumulating evidence supports that the GWAS-identified variants in non-coding regions are implicated in complex human diseases or traits, an implication that involves an extensive crosstalk among SNPs, ASMs, and ABTFs for impacting ASE (Figures 4A,B), as well as environmental exposure-altering ASMs and ABTFs (Figure 4C), thereby rendering a person more susceptible to diverse diseases. Exposure to environmental factors (bisphenol A, metal, herbicides), alcohol, and cigarette smoke has been demonstrated to affect chromatin structure (DNA methylation and histone modifications), thereby potentially altering ASMs and ABTFs (Cowley et al., 2018; Meehan et al., 2018; Pathak and Feil, 2018; Strakovsky and Schantz, 2018; Zhu et al., 2018). With more investigations of the crosstalk of SNP-TFs-allelic expression changes and of SNP-insulator CTCF-allelic expression changes, which have burgeoned only in the past 2 years (Table 1), it becomes clearer that the signal of a small change at one residue (i.e., SNP) is intensified by introducing or abolishing allelic DNA methylation, thereby impacting the allelic recruitment or abolishment of TFs. The allele-specificity of TFs eventually increases or decreases transcript copies from one allele, and the resulting changes of dosage seem to be the key for disease risk (Figure 4). We thus coin this working model as the “SNP intensifier.”

A future challenge is to provide more evidence to elucidate this mechanism of variants contributing to human disease susceptibility and inter-individual phenotypic differences. In regard to the elucidation, several roadblocks must first be cleared and a better standard must be established, including the quick characterization of all GWAS-associated genetic variants in affecting epigenetic patterns and gene transcription, and the conclusive determination of each variant's role in complex human disease susceptibility. More specifically, we should seek to identify all SNPs-resulted ASMs or methylation QTLs in a cost-effective and high throughput manner, characterize the chronic effect of slightly increased RNA transcripts of a gene for

disease risk, and integrate layers of information from variants, ASMs/ASE, and exposure for deeper mechanistic insights.

The community certainly needs an integrated system for biological analyses of data from layers of genetic variants, epigenetic patterns, and expression QTL (eQTL). Intriguingly, a summary data-based Mendelian Randomization (SMR) method was recently developed (Zhu et al., 2016). The method of SMR borrows the concept of Mendelian Randomization (MR) analysis, which uses a genetic variant (e.g., a SNP) as an instrumental variable to assess the putative causative effect of an exposure (e.g., gene expression level) on an outcome (e.g., diseased phenotype). Because the variance in a phenotype explained by a single genetic variant, or the expression change of a single gene, is likely to be very small, a limitation of MR analysis is the requirement of an extremely large sample size. To overcome this limitation, Zhu et al. used the summary-level data (for instance, effect sizes or test statistics) available from the very large-scale GWAS and eQTL studies (Zhu et al., 2016). With SMR analyses, investigators integrated summary-level GWAS data on up to 339,224 individuals and eQTL data on 5,311 individuals for the identification of 126 genes in association with five human complex traits. Out of these 126 genes, *TRAF1* and *ANKRD55* were found to be associated with rheumatoid arthritis, and *SNX19* and *NMRAL1* were found to be associated with schizophrenia. Using the SMR approach, an independent group identified DNA methylation sites in association with GWAS-identified variants for multiple complex traits out of more than 40 traits examined (Hannon et al., 2017) and showed the potential role of genetic variants in the *RNASSET2* locus in association with eQTL and mQTL (methylation quantitative trait locus) for Crohn's disease. Similarly, the group that first reported the SMR method also integrated the summary-level data of mQTL with that of GWAS and eQTL (Wu et al., 2018) and revealed pleiotropic associations between 7,858 DNA methylation sites and 2,733 genes, which can be regarded as a map of the methylome to the transcriptome. Further analyses identified 149 DNA methylation sites and 66 genes showing pleiotropic associations with 12 complex traits. Wu et al. hypothesize a mechanism whereby a genetic variant impacts complex disease by way of genetic modulation of transcription through DNA methylation.

Characterization of genetic variant-influenced ASMs remains a challenge as well. Not surprisingly, these ASMs show tissue specificity (Do et al., 2016; Marzi et al., 2016). Therefore, the cost-effective identification of all ASMs in different tissues is needed. With tissue specificity, methylation status of CpGs within a given ASM can be changed during cell development, and differentiation. For example, our data demonstrate that HCN2 ASM seems to exist in embryonic stem cells, but that CpG sites of unmethylated allele were methylated in differentiated neural progenitor cells and neurons (Martos et al., 2017). In addition, the characterization of vulnerability of each ASM, from different tissues to different environmental stresses, is also critical for understanding the complex human diseases. Lastly, the current method to analyze DNA methylation for large cohort studies relies on an array-based approach that is limited to low resolution, and

there is an urgent need for an alternative sequencing-based method for efficient screening of hundreds or thousands of samples.

Due to the accompanied ethnic issues of human samples, animal studies are required to solve the mystery of the detailed mechanistic insights. For example, animal studies are needed for exploration of the effect of chronic and even subtle changes of one allele. F1 hybrid mice, between two strains of rat or mouse, have many more SNPs (for example, up to 25 million SNPs between mouse strains 129 and Cast) for insights (Schilling et al., 2009; Lossie et al., 2014; Lo et al., 2016; Martos et al., 2017). The interactions between genetic variants and disease susceptibility may be tested via different crossing strategies in the Diversity Outbred mouse population and the Collaborative Cross inbred strains (French et al., 2018), another hot topic that will not be described in detail in this review.

While the mechanistic insights gained through the extensive hard work above have brought the community closer to our ultimate goal of finding opportunities for the intervention and prevention of human diseases, at least one investigation has shown encouraging results in this regard. Administration of thyroxin (T4) or metformin to neonatal rats after fetal alcohol exposure was shown to reverse the expression changes of *Dio3* and *Igf2* and to alleviate the fear memory deficit that was triggered by fetal alcohol exposure (Tunc-Ozcan et al., 2018). This reverse encourages the exploration of drugs for potential intervention.

## REFERENCES

- Allen, E. K., Randolph, A. G., Bhangale, T., Dogra, P., Ohlson, M., Oshansky, C. M., et al. (2017). SNP-mediated disruption of CTCF binding at the IFITM3 promoter is associated with risk of severe influenza in humans. *Nat. Med.* 23, 975–983. doi: 10.1038/nm.4370
- Allfrey, V. G., Faulkner, R., and Mirsky, A. E. (1964). Acetylation and methylation of histones and their possible role in the regulation of Rna synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 51, 786–794. doi: 10.1073/pnas.51.5.786
- Anvar, Z., Cammisà, M., Riso, V., Baglivo, I., Kukreja, H., Sparago, A., et al. (2016). ZFP57 recognizes multiple and closely spaced sequence motif variants to maintain repressive epigenetic marks in mouse embryonic stem cells. *Nucleic Acids Res.* 44, 1118–1132. doi: 10.1093/nar/gkv1059
- Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., et al. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837. doi: 10.1016/j.cell.2007.05.009
- Bartolomei, M. S. (2009). Genomic imprinting: employing and avoiding epigenetic processes. *Genes Dev.* 23, 2124–2133. doi: 10.1101/gad.1841409
- Bastaki, F., Nair, P., Mohamed, M., Malik, E. M., Helmi, M., Al-Ali, M. T., et al. (2017). Identification of a novel CTCF mutation responsible for syndromic intellectual disability - a case report. *BMC Med. Genet.* 18:68. doi: 10.1186/s12881-017-0429-0
- Bell, A. C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 405, 482–485. doi: 10.1038/35013100
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315–326. doi: 10.1016/j.cell.2006.02.041
- Bérubé, S., Al Tuwaijri, A., Kohan-Ghadr, H. R., Elzein, S., Farias, R., Berube, J., et al. (2017). Role of DNA methylation in expression control of the IKZF3-GSDMA region in human epithelial cells. *PLoS ONE* 12:e0172707. doi: 10.1371/journal.pone.0172707
- Bestor, T. H. (2000). The DNA methyltransferases of mammals. *Hum. Mol. Genet.* 9, 2395–2402. doi: 10.1093/hmg/9.16.2395
- Blattler, A., Yao, L., Witt, H., Guo, Y., Nicolet, C. M., Berman, B. P., et al. (2014). Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biol.* 15:469. doi: 10.1186/s13059-014-0469-0
- Chen, X., Weaver, J., Bove, B. A., Vanderveer, L. A., Weil, S. C., Miron, A., et al. (2008). Allelic imbalance in BRCA1 and BRCA2 gene expression is associated with an increased breast cancer risk. *Hum. Mol. Genet.* 17, 1336–1348. doi: 10.1093/hmg/ddn022
- Chen, X. F., Zhu, D. L., Yang, M., Hu, W. X., Duan, Y. Y., Lu, B. J., et al. (2018). An osteoporosis risk SNP at 1p36.12 acts as an allele-specific enhancer to modulate LINC00339 expression via long-range loop formation. *Am. J. Hum. Genet.* 102, 776–793. doi: 10.1016/j.ajhg.2018.03.001
- Cheng, H., Xuan, H., Green, C. D., Han, Y., Sun, N., Shen, H., et al. (2018). Repression of human and mouse brain inflammaging transcriptome by broad gene-body histone hyperacetylation. *Proc. Natl. Acad. Sci. U.S.A.* 115, 7611–7616. doi: 10.1073/pnas.1800656115
- Cheng, M., Liu, X., Yang, M., Han, L., Xu, A., and Huang, Q. (2017). Computational analyses of type 2 diabetes-associated loci identified by genome-wide association studies. *J. Diabetes* 9, 362–377. doi: 10.1111/1753-0407.12421
- Cheung, W. A., Shao, X., Morin, A., Siroux, V., Kwan, T., Ge, B., et al. (2017). Functional variation in allelic methylation underscores a strong genetic contribution and reveals novel epigenetic alterations in the human epigenome. *Genome Biol.* 18:50. doi: 10.1186/s13059-017-1173-7
- Chiba, H., Kakuta, Y., Kinouchi, Y., Kawai, Y., Watanabe, K., Nagao, M., et al. (2018). Allele-specific DNA methylation of disease susceptibility genes in Japanese patients with inflammatory bowel disease. *PLoS ONE* 13:e0194036. doi: 10.1371/journal.pone.0194036
- Cowley, M., Skaar, D. A., Jima, D. D., Maguire, R. L., Hudson, K. M., Park, S. S., et al. (2018). Effects of cadmium exposure on dna methylation at imprinting control regions and genome-wide in mothers and newborn children. *Environ. Health Perspect.* 126:037003. doi: 10.1289/EHP2085

## AUTHOR CONTRIBUTIONS

HW, DL, and ZW prepared the manuscript. HW and DL contribute significantly in terms of literature reading, summarizing, and presenting.

## ACKNOWLEDGMENTS

The laboratory of ZW at Johns Hopkins was supported by U.S. National Institutes of Health (R01ES25761, U01ES026721 opportunity fund, and R21ES028351) and Johns Hopkins Catalyst Award. ZW also thanks Dr. Jian Yang at the University of Queensland for the critical reading of the manuscript, and sincerely apologizes to fellow investigators whose important contributions were not directly referenced due to space and reference limitations.

- Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21931–21936. doi: 10.1073/pnas.1016071107
- Cuddapah, S., Jothi, R., Schones, D. E., Roh, T. Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* 19, 24–32. doi: 10.1101/gr.082800.108
- Dagar, V., Hutchison, W., Muscat, A., Krishnan, A., Hoke, D., Buckle, A., et al. (2018). Genetic variation affecting DNA methylation and the human imprinting disorder, Beckwith-Wiedemann syndrome. *Clin. Epigenetics* 10:114. doi: 10.1186/s13148-018-0546-4
- Dai, H., and Wang, Z. (2014). Histone modification patterns and their responses to environment. *Curr. Envir. Health Rep.* 1, 11–21. doi: 10.1007/s40572-013-0008-2
- Dai, J., Zhu, M., Wang, C., Shen, W., Zhou, W., Sun, J., et al. (2015). Systematical analyses of variants in CTCF-binding sites identified a novel lung cancer susceptibility locus among Chinese population. *Sci. Rep.* 5:7833. doi: 10.1038/srep07833
- de la Rica, L., Rodríguez-Ubreva, J., García, M., Islam, A. B., Urquiza, J. M., Hernando, H., et al. (2013). PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. *Genome Biol.* 14:R99. doi: 10.1186/gb-2013-14-9-r99
- Do, C., Lang, C. F., Lin, J., Darbary, H., Krupska, I., Gaba, A., et al. (2016). Mechanisms and disease associations of haplotype-dependent allele-specific DNA methylation. *Am. J. Hum. Genet.* 98, 934–955. doi: 10.1016/j.ajhg.2016.03.027
- Dolinoy, D. C. (2008). The agouti mouse model: an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome. *Nutr. Rev.* 66 (Suppl. 1), S7–S11. doi: 10.1111/j.1753-4887.2008.00056.x
- Dolinoy, D. C., Huang, D., and Jirtle, R. L. (2007). Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13056–13061. doi: 10.1073/pnas.0703739104
- Downing, C., Johnson, T. E., Larson, C., Leakey, T. I., Siegfried, R. N., Rafferty, T. M., et al. (2011). Subtle decreases in DNA methylation and gene expression at the mouse Igf2 locus following prenatal alcohol exposure: effects of a methyl-supplemented diet. *Alcohol* 45, 65–71. doi: 10.1016/j.alcohol.2010.07.006
- Eroglu, A., and Layman, L. C. (2012). Role of ART in imprinting disorders. *Semin. Reprod. Med.* 30, 92–104. doi: 10.1055/s-0032-1307417
- Farh, K. K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518, 337–343. doi: 10.1038/nature13835
- French, J. E., Gatti, D. M., Morgan, D. L., Kissling, G. E., Shockley, K. R., Knudsen, G. A., et al. (2018). Erratum: “diversity outbred mice identify population-based exposure thresholds and genetic factors that influence benzene-induced genotoxicity”. *Environ. Health Perspect.* 126:069003. doi: 10.1289/EHP3950
- Gallagher, M. D., Posavi, M., Huang, P., Unger, T. L., Berlyand, Y., Gruenewald, A. L., et al. (2017). A dementia-associated risk variant near TMEM106B alters chromatin architecture and gene expression. *Am. J. Hum. Genet.* 101, 643–663. doi: 10.1016/j.ajhg.2017.09.004
- Gao, Y., Baccarelli, A., Shu, X. O., Ji, B. T., Yu, K., Tarantini, L., et al. (2012). Blood leukocyte Alu and LINE-1 methylation and gastric cancer risk in the Shanghai Women’s Health Study. *Br. J. Cancer* 106, 585–591. doi: 10.1038/bjc.2011.562
- Gordon, S., Akopyan, G., Garban, H., and Bonavida, B. (2006). Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene* 25, 1125–1142. doi: 10.1038/sj.onc.1209080
- Goudarzi, A., Zhang, D., Huang, H., Barral, S., Kwon, O. K., Qi, S., et al. (2016). Dynamic competing histone H4 K5K8 acetylation and butyrylation are hallmarks of highly active gene promoters. *Mol. Cell* 62, 169–180. doi: 10.1016/j.molcel.2016.03.014
- Hancock, D. B., Guo, Y., Reginsson, G. W., Gaddis, N. C., Lutz, S. M., Sherva, R., et al. (2017). Genome-wide association study across European and African American ancestries identifies a SNP in DNMT3B contributing to nicotine dependence. *Mol. Psychiatry* 23, 1–9. doi: 10.1038/mp.2017.193
- Hannon, E., Weedon, M., Bray, N., O’Donovan, M., and Mill, J. (2017). Pleiotropic effects of trait-associated genetic variation on dna methylation: utility for refining GWAS loci. *Am. J. Hum. Genet.* 100, 954–959. doi: 10.1016/j.ajhg.2017.04.013
- Harper, K. M., Tunc-Ozcan, E., Graf, E. N., Herzing, L. B., and Redei, E. E. (2014). Intergenerational and parent of origin effects of maternal calorie restriction on Igf2 expression in the adult rat hippocampus. *Psychoneuroendocrinology* 45, 187–191. doi: 10.1016/j.psyneuen.2014.04.002
- Haycock, P. C., and Ramsay, M. (2009). Exposure of mouse embryos to ethanol during preimplantation development: effect on DNA methylation in the h19 imprinting control region. *Biol. Reprod.* 81, 618–627. doi: 10.1095/biolreprod.108.074682
- Heijmans, B. T., Tobi, E. W., Stein, A. D., Putter, H., Blauw, G. J., Susser, E. S., et al. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17046–17049. doi: 10.1073/pnas.0806560105
- Hori, I., Kawamura, R., Nakabayashi, K., Watanabe, H., Higashimoto, K., Tomikawa, J., et al. (2017). CTCF deletion syndrome: clinical features and epigenetic delineation. *J. Med. Genet.* 54, 836–842. doi: 10.1136/jmedgenet-2017-104854
- Hrdlickova, B., de Almeida, R. C., Borek, Z., and Withoff, S. (2014). Genetic variation in the non-coding genome: involvement of micro-RNAs and long non-coding RNAs in disease. *Biochim. Biophys. Acta* 1842, 1910–1922. doi: 10.1016/j.bbdis.2014.03.011
- Izzi, B., Pistoni, M., Cludts, K., Akkor, P., Lambrechts, D., Verfaillie, C., et al. (2016). Allele-specific DNA methylation reinforces PEAR1 enhancer activity. *Blood* 128, 1003–1012. doi: 10.1182/blood-2015-11-682153
- Jeltsch, A. (2006). On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. *Epigenetics* 1, 63–66. doi: 10.4161/epi.1.2.2767
- Jiang, X., Detera-Wadleigh, S. D., Akula, N., Mallon, B. S., Hou, L., Xiao, T., et al. (2018). Sodium valproate rescues expression of TRANK1 in iPSC-derived neural cells that carry a genetic variant associated with serious mental illness. *Mol. Psychiatry* doi: 10.1038/s41380-018-0207-1. [Epub ahead of print].
- Jones, P. A., and Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. *Nat. Rev. Genet.* 10, 805–811. doi: 10.1038/nrg2651
- Jurkowska, R. Z., Jurkowski, T. P., and Jeltsch, A. (2011). Structure and function of mammalian DNA methyltransferases. *ChemBiochem* 12, 206–222. doi: 10.1002/cbic.201000195
- Kaminen-Ahola, N., Ahola, A., Maga, M., Mallitt, K. A., Fahey, P., Cox, T. C., et al. (2010). Maternal ethanol consumption alters the epigenotype and the phenotype of offspring in a mouse model. *PLoS Genet.* 6:e1000811. doi: 10.1371/journal.pgen.1000811
- Kemp, C. J., Moore, J. M., Moser, R., Bernard, B., Teater, M., Smith, L. E., et al. (2014). CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. *Cell Rep.* 7, 1020–1029. doi: 10.1016/j.celrep.2014.04.004
- Kerkel, K., Spadola, A., Yuan, E., Kosek, J., Jiang, L., Hod, E., et al. (2008). Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat. Genet.* 40, 904–908. doi: 10.1038/ng.174
- Khandanpour, C., Krongold, J., Schütte, J., Bouwman, F., Vassen, L., Gaudreau, M. C., et al. (2012). The human GF1136N variant induces epigenetic changes at the Hoxa9 locus and accelerates K-RAS driven myeloproliferative disorder in mice. *Blood* 120, 4006–4017. doi: 10.1182/blood-2011-02-334722
- Kidd, C. D., Thompson, P. J., Barrett, L., and Baltic, S. (2016). Histone modifications and Asthma. The interface of the epigenetic and genetic landscapes. *Am. J. Respir. Cell Mol. Biol.* 54, 3–12. doi: 10.1165/rcmb.2015-0050TR
- Kilpinen, H., Waszak, S. M., Gschwind, A. R., Raghav, S. K., Witwicki, R. M., Orioli, A., et al. (2013). Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* 342, 744–747. doi: 10.1126/science.1242463
- Kim, D. S., Kim, Y. H., Lee, W. K., Na, Y. K., and Hong, H. S. (2016). Effect of alcohol consumption on peripheral blood Alu methylation in Korean men. *Biomarkers* 21, 243–248. doi: 10.3109/1354750X.2015.1134661
- Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., et al. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128, 1231–1245. doi: 10.1016/j.cell.2006.12.048

- Klengel, T., Mehta, D., Anacker, C., Rex-Haffner, M., Pruessner, J. C., Pariante, C. M., et al. (2013). Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. *Nat. Neurosci.* 16, 33–41. doi: 10.1038/nn.3275
- Kumar, D., Puan, K. J., Andiappan, A. K., Lee, B., Westerlaken, G. H., Haase, D., et al. (2017). A functional SNP associated with atopic dermatitis controls cell type-specific methylation of the VSTM1 gene locus. *Genome Med.* 9:18. doi: 10.1186/s13073-017-0404-6
- Kundakovic, M., Gudsnuk, K., Herbstman, J. B., Tang, D., Perera, F. P., and Champagne, F. A. (2015). DNA methylation of BDNF as a biomarker of early-life adversity. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6807–6813. doi: 10.1073/pnas.1408355111
- Li, H., Li, W., Liu, S., Zong, S., Wang, W., Ren, J., et al. (2016). DNMT1, DNMT3A and DNMT3B polymorphisms associated with gastric cancer risk: a systematic review and meta-analysis. *EBioMed.* 13, 125–131. doi: 10.1016/j.ebiom.2016.10.028
- Li, W., Shang, L., Huang, K., Li, J., Wang, Z., and Yao, H. (2017). Identification of critical base pairs required for CTCF binding in motif M1 and M2. *Protein Cell* 8, 544–549. doi: 10.1007/s13238-017-0387-5
- Li, X., Kim, Y., Tsang, E. K., Davis, J. R., Damani, F. N., Chiang, C., et al. (2017). The impact of rare variation on gene expression across tissues. *Nature* 550, 239–243. doi: 10.1038/nature24267
- Li, Y., Zhang, Z., Chen, J., Liu, W., Lai, W., Liu, B., et al. (2018). Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature* 564, 136–140. doi: 10.1038/s41586-018-0751-5
- Li, Z., Dai, H., Martos, S. N., Xu, B., Gao, Y., Li, T., et al. (2015). Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells. *Genome Biol.* 16:115. doi: 10.1186/s13059-015-0685-2
- Ling, C., Fangyu, D., Wanhua, H., Kelong, C., Zhimin, W., Yuting, Z., et al. (2016). DNMT3A rs1550117 Polymorphism is associated with late-onset alzheimer's disease in a Chinese population. *Am. J. Alzheimers. Dis. Other Demen.* 31, 278–281. doi: 10.1177/1533317515603688
- Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315–322. doi: 10.1038/nature08514
- Liu, Q., Thoms, J. A. I., Nunez, A. C., Huang, Y., Knezevic, K., Packham, D., et al. (2018). Disruption of a–35 kb enhancer impairs CTCF binding and MLH1 expression in colorectal cells. *Clin. Cancer Res.* 24, 4602–4611. doi: 10.1158/1078-0432.CCR-17-3678
- Lo, C. L., Lossie, A. C., Liang, T., Liu, Y., Xuei, X., Lumeng, L., et al. (2016). High resolution genomic scans reveal genetic architecture controlling alcohol preference in bidirectionally selected rat model. *PLoS Genet.* 12:e1006178. doi: 10.1371/journal.pgen.1006178
- Lo, C. L., Lumeng, L., Bell, R. L., Liang, T., Lossie, A. C., Muir, W. M., et al. (2018). CIS-acting allele-specific expression differences induced by alcohol and impacted by sex as well as parental genotype of origin. *Alcohol. Clin. Exp. Res.* 42, 1444–1453. doi: 10.1111/acer.13776
- Lockwood, S. H., Guan, A., Yu, A. S., Zhang, C., Zykovich, A., Korf, I., et al. (2014). The functional significance of common polymorphisms in zinc finger transcription factors. *G3* 4, 1647–1655. doi: 10.1534/g3.114.012195
- Lossie, A. C., Muir, W. M., Lo, C. L., Timm, F., Liu, Y., Gray, W., et al. (2014). Implications of genomic signatures in the differential vulnerability to fetal alcohol exposure in C57BL/6 and DBA/2 mice. *Front. Genet.* 5:173. doi: 10.3389/fgene.2014.00173
- Marjonen, H., Auvinen, P., Kahila, H., Tšuiiko, O., Kōks, S., Tiirats, A., et al. (2018). rs10732516 polymorphism at the IGF2/H19 locus associates with genotype-specific effects on placental DNA methylation and birth weight of newborns conceived by assisted reproductive technology. *Clin. Epigenetics* 10:80. doi: 10.1186/s13148-018-0511-2
- Marjonen, H., Kahila, H., and Kaminen-Ahola, N. (2017). rs10732516 polymorphism at the IGF2/H19 locus associates with a genotype-specific trend in placental DNA methylation and head circumference of prenatally alcohol-exposed newborns. *Human Reprod. Open* 2017, 1–12. doi: 10.1093/hropen/hox014
- Martos, S. N., Li, T., Ramos, R. B., Dai, H., Xu, J. C., et al. (2017). Two approaches reveal a new paradigm of 'switchable or genetics-influenced allele-specific DNA methylation' with potential in human disease. *Cell Discov.* 3:17038. doi: 10.1038/celldisc.2017.38
- Marzi, S. J., Meaburn, E. L., Dempster, E. L., Lunnon, K., Paya-Cano, J. L., Smith, R. G., et al. (2016). Tissue-specific patterns of allelically-skewed DNA methylation. *Epigenetics* 11, 24–35. doi: 10.1080/15592294.2015.1127479
- McVicker, G., van de Geijn, B., Degner, J. F., Cain, C. E., Banovich, N. E., Raj, A., et al. (2013). Identification of genetic variants that affect histone modifications in human cells. *Science* 342, 747–749. doi: 10.1126/science.1242429
- Meehan, R. R., Thomson, J. P., Lentini, A., Nestor, C. E., and Pennings, S. (2018). DNA methylation as a genomic marker of exposure to chemical and environmental agents. *Curr. Opin. Chem. Biol.* 45, 48–56. doi: 10.1016/j.cbpa.2018.02.006
- Milani, L., Lundmark, A., Nordlund, J., Kiialainen, A., Flaegstad, T., Jonmundsson, G., et al. (2009). Allele-specific gene expression patterns in primary leukemic cells reveal regulation of gene expression by CpG site methylation. *Genome Res.* 19, 1–11. doi: 10.1101/gr.083931.108
- Moffatt, M. F., Gut, I. G., Demenais, F., Strachan, D. P., Bouzigon, E., Heath, S., et al. (2010). A large-scale, consortium-based genomewide association study of asthma. *N. Engl. J. Med.* 363, 1211–1221. doi: 10.1056/NEJMoa0906312
- Moyerbrailean, G. A., Richards, A. L., Kurtz, D., Kalita, C. A., Davis, G. O., Harvey, C. T., et al. (2016). High-throughput allele-specific expression across 250 environmental conditions. *Genome Res.* 26, 1627–1638. doi: 10.1101/gr.209759.116
- Nelson, H. H., Marsit, C. J., and Kelsey, K. T. (2011). Global methylation in exposure biology and translational medical science. *Environ. Health Perspect.* 119, 1528–1533. doi: 10.1289/ehp.1103423
- Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247–257. doi: 10.1016/S0092-8674(00)81656-6
- Ong, C. T., and Corces, V. G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* 15, 234–246. doi: 10.1038/nrg3663
- Pathak, R., and Feil, R. (2018). Environmental effects on chromatin repression at imprinted genes and endogenous retroviruses. *Curr. Opin. Chem. Biol.* 45, 139–147. doi: 10.1016/j.cbpa.2018.04.015
- Phillips, J. E., and Corces, V. G. (2009). CTCF: master weaver of the genome. *Cell* 137, 1194–1211. doi: 10.1016/j.cell.2009.06.001
- Pidsley, R., Dempster, E., Troakes, C., Al-Sarraj, S., and Mill, J. (2012). Epigenetic and genetic variation at the IGF2/H19 imprinting control region on 11p15.5 is associated with cerebellum weight. *Epigenetics* 7, 155–163. doi: 10.4161/epi.7.2.18910
- Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., et al. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell* 44, 361–372. doi: 10.1016/j.molcel.2011.08.032
- Rahbar, E., Waits, C. M. K., Kirby, E. H. Jr., Miller, L. R., Ainsworth, H. C., Cui, T., et al. (2018). Allele-specific methylation in the FADS genomic region in DNA from human saliva, CD4+ cells, and total leukocytes. *Clin. Epigenetics* 10:46. doi: 10.1186/s13148-018-0480-5
- Reynard, L. N., Bui, C., Syddall, C. M., and Loughlin, J. (2014). CpG methylation regulates allelic expression of GDF5 by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis susceptibility SNP rs143383. *Hum. Genet.* 133, 1059–1073. doi: 10.1007/s00439-014-1447-z
- Rijlaarsdam, J., van IJzendoorn, M. H., Verhulst, F. C., Jaddoe, V. W., Felix, J. F., Tiemeier, H., et al. (2017). Prenatal stress exposure, oxytocin receptor gene (OXTR) methylation, and child autistic traits: the moderating role of OXTR rs53576 genotype. *Autism Res.* 10, 430–438. doi: 10.1002/aur.1681
- Riso, V., Cammisà, M., Kukreja, H., Anvar, Z., Verde, G., Sparago, A., et al. (2016). ZFP57 maintains the parent-of-origin-specific expression of the imprinted genes and differentially affects non-imprinted targets in mouse embryonic stem cells. *Nucleic Acids Res.* 44, 8165–8178. doi: 10.1093/nar/gkw505
- Rivollier, F., Chaumette, B., Bendjema, N., Chayet, M., Millet, B., Jaafari, N., et al. (2017). Methylomic changes in individuals with psychosis, prenatally exposed to endocrine disrupting compounds: lessons from diethylstilbestrol. *PLoS ONE* 12:e0174783. doi: 10.1371/journal.pone.0174783
- Roh, T. Y., Cuddapah, S., Cui, K., and Zhao, K. (2006). The genomic landscape of histone modifications in human T cells. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15782–15787. doi: 10.1073/pnas.0607617103

- Rutten, B. P. F., Vermetten, E., Vinkers, C. H., Ursini, G., Daskalakis, N. P., Pishva, E., et al. (2018). Longitudinal analyses of the DNA methylome in deployed military servicemen identify susceptibility loci for post-traumatic stress disorder. *Mol. Psychiatry* 23, 1145–1156. doi: 10.1038/mp.2017.120
- Saradalekshmi, K. R., Neetha, N. V., Sathyan, S., Nair, I. V., Nair, C. M., and Banerjee, M. (2014). DNA methyltransferase (DNMT) gene polymorphisms could be a primary event in epigenetic susceptibility to schizophrenia. *PLoS ONE* 9:e98182. doi: 10.1371/journal.pone.0098182
- Schilling, E., El Chartouni, C., and Rehli, M. (2009). Allele-specific DNA methylation in mouse strains is mainly determined by cis-acting sequences. *Genome Res.* 19, 2028–2035. doi: 10.1101/gr.095562.109
- Schmiedel, B. J., Seumois, G., Samaniego-Castruita, D., Cayford, J., Schulten, V., Chavez, L., et al. (2016). 17q21 asthma-risk variants switch CTCF binding and regulate IL-2 production by T cells. *Nat. Commun.* 7:13426. doi: 10.1038/ncomms13426
- Singh, P., Lee, D. H., and Szabó, P. E. (2012). More than insulator: multiple roles of CTCF at the H19-Igf2 imprinted domain. *Front. Genet.* 3:214. doi: 10.3389/fgene.2012.00214
- Soldner, F., Stelzer, Y., Shivalila, C. S., Abraham, B. J., Latourelle, J. C., Barrasa, M. I., et al. (2016). Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. *Nature* 533, 95–99. doi: 10.1038/nature17939
- Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R., and Flavell, R. A. (2005). Interchromosomal associations between alternatively expressed loci. *Nature* 435, 637–645. doi: 10.1038/nature03574
- Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. *Nature* 403, 41–45. doi: 10.1038/47412
- Strakovsky, R. S., and Schantz, S. L. (2018). Impacts of bisphenol A (BPA) and phthalate exposures on epigenetic outcomes in the human placenta. *Environ. Epigenet.* 4:dvy022. doi: 10.1093/eep/dvy022
- Strogantsev, R., Krueger, F., Yamazawa, K., Shi, H., Gould, P., Goldman-Roberts, M., et al. (2015). Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biol.* 16:112. doi: 10.1186/s13059-015-0672-7
- Susiarjo, M., Sasson, I., Mesáros, C., and Bartolomei, M. S. (2013). Bisphenol A exposure disrupts genomic imprinting in the mouse. *PLoS Genet.* 9:e1003401. doi: 10.1371/journal.pgen.1003401
- Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146, 1016–1028. doi: 10.1016/j.cell.2011.08.008
- Tehranchi, A. K., Myrthil, M., Martin, T., Hie, B. L., Golan, D., and Fraser, H. B. (2016). Pooled ChIP-seq links variation in transcription factor binding to complex disease risk. *Cell* 165, 730–741. doi: 10.1016/j.cell.2016.03.041
- Treit, S., Zhou, D., Chudley, A. E., Andrew, G., Rasmussen, C., Nikkel, S. M., et al. (2016). Relationships between head circumference, brain volume and cognition in children with prenatal alcohol exposure. *PLoS ONE* 11:e0150370. doi: 10.1371/journal.pone.0150370
- Tunc-Ozcan, E., Harper, K. M., Graf, E. N., and Redei, E. E. (2016). Thyroxine administration prevents matrilineal intergenerational consequences of *in utero* ethanol exposure in rats. *Horm. Behav.* 82, 1–10. doi: 10.1016/j.yhbeh.2016.04.002
- Tunc-Ozcan, E., Wert, S. L., Lim, P. H., Ferreira, A., and Redei, E. E. (2018). Hippocampus-dependent memory and allele-specific gene expression in adult offspring of alcohol-consuming dams after neonatal treatment with thyroxin or metformin. *Mol. Psychiatry* 23, 1643–1651. doi: 10.1038/mp.2017.129
- Van de Pette, M., Abbas, A., Feytout, A., McNamara, G., Bruno, L., To, W. K., et al. (2017). Visualizing changes in Cdkn1c expression links early-life adversity to imprint mis-regulation in adults. *Cell Rep.* 18, 1090–1099. doi: 10.1016/j.celrep.2017.01.010
- Van Rechem, C., and Whetstone, J. R. (2014). Examining the impact of gene variants on histone lysine methylation. *Biochim. Biophys. Acta* 1839, 1463–1476. doi: 10.1016/j.bbagr.2014.05.014
- Volkov, P., Olsson, A. H., Gillberg, L., Jørgensen, S. W., Brøns, C., Eriksson, K. F., et al. (2016). A genome-wide mQTL analysis in human adipose tissue identifies genetic variants associated with DNA methylation, gene expression and metabolic traits. *PLoS ONE* 11:e0157776. doi: 10.1371/journal.pone.0157776
- Walker, E. J., Zhang, C., Castelo-Branco, P., Hawkins, C., Wilson, W., Zhukova, N., et al. (2012). Monoallelic expression determines oncogenic progression and outcome in benign and malignant brain tumors. *Cancer Res.* 72, 636–644. doi: 10.1158/0008-5472.CAN-11-2266
- Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddapah, S., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* 40, 897–903. doi: 10.1038/ng.154
- Weksberg, R. (2010). Imprinted genes and human disease. *Am. J. Med. Genet. C Semin. Med. Genet.* 154C, 317–320. doi: 10.1002/ajmg.c.30268
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42, D1001–D1006. doi: 10.1093/nar/gkt1229
- Wu, Y., Zeng, J., Zhang, F., Zhu, Z., Qi, T., Zheng, Z., et al. (2018). Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat. Commun.* 9:918. doi: 10.1038/s41467-018-03371-0
- Xue, J., Schoenrock, S. A., Valdar, W., Tarantino, L. M., and Ideraabdullah, F. Y. (2016). Maternal vitamin D depletion alters DNA methylation at imprinted loci in multiple generations. *Clin. Epigenetics* 8:107. doi: 10.1186/s13148-016-0276-4
- Yang, H. H., Hu, N., Wang, C., Ding, T., Dunn, B. K., Goldstein, A. M., et al. (2010). Influence of genetic background and tissue types on global DNA methylation patterns. *PLoS ONE* 5:e9355. doi: 10.1371/journal.pone.0009355
- Ye, W., Wang, Y., Mei, B., Hou, S., Liu, X., Wu, G., et al. (2018). Computational and functional characterization of four SNPs in the SOST locus associated with osteoporosis. *Bone* 108, 132–144. doi: 10.1016/j.bone.2018.01.001
- Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* 356:eaaj2239. doi: 10.1126/science.aaj2239
- Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515, 355–364. doi: 10.1038/nature13992
- Zhang, Y., Rohde, C., Reinhardt, R., Voelcker-Rehage, C., and Jeltsch, A. (2009). Non-imprinted allele-specific DNA methylation on human autosomes. *Genome Biol.* 10:R138. doi: 10.1186/gb-2009-10-12-r138
- Zhu, Y., Li, Y., Lou, D., Gao, Y., Yu, J., Kong, D., et al. (2018). Sodium arsenite exposure inhibits histone acetyltransferase p300 for attenuating H3K27ac at enhancers in mouse embryonic fibroblast cells. *Toxicol. Appl. Pharmacol.* 357, 70–79. doi: 10.1016/j.taap.2018.08.011
- Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 48, 481–487. doi: 10.1038/ng.3538

**Conflict of Interest Statement:** ZW is a consultant of Shandong Bio-focus Gene-tech Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Wang, Lou and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.