Methods 102 (2016) 50–56

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Intersection of genetics and epigenetics in monozygotic twin genomes

Kwoneel Kim, Kibaick Lee, Hyoeun Bang, Jeong Yeon Kim, Jung Kyoon Choi *

Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea

ARTICLE INFO

Article history:
Received 21 August 2015
Accepted 18 October 2015
Available online 5 November 2015

ABSTRACT

As a final function of various epigenetic mechanisms, chromatin regulation is a transcription control process that especially demonstrates active interaction with genetic elements. Thus, chromatin structure has become a principal focus in recent genomics researches that strive to characterize regulatory functions of DNA variants related to diseases or other traits. Although researchers have been focusing on DNA methylation when studying monozygotic (MZ) twins, a great model in epigenetics research, interactions between genetics and epigenetics in chromatin level are expected to be an imperative research trend in the future. In this review, we discuss how the genome, epigenome, and transcriptome of MZ twins can be studied in an integrative manner from this perspective.

Contents

1. Identical twins, the same but different

Epigenetics often refers to mechanisms that affect gene expression and cellular phenotype through changes that do not alter the DNA sequence. Although pure instances of heritable epigenetic change are rare, the term typically encompasses changes in gene expression at the molecular level in response to environmental cues, even if these changes are ultimately underpinned by DNA sequence. For example, variations in the level of DNA methylation have been observed following exposure certain chemicals [1], but these changes are likely due to protective systems encoded in the DNA. Here, we refer to such modifications and the mechanisms that read and write them as epigenetic. The semantics of the word itself reflect a general problem in disentangling cause and effect of these modifications. Studying monozygotic (MZ) twins, which are by definition genetically identical, has long been a gold standard for separating the epigenetic from genetic. We focus here on how MZ twins can also be used for integrating epigenetics with genetics.

As alluded to above, much of what we call epigenetic is actually dependent on genetic variation, especially in noncoding regions, which can alter transcriptional processes via epigenetic mechanisms, such as DNA methylation, chromatin remodeling, and small RNA regulation. In cases such as this, studying the epigenetic mechanism can facilitate our understanding of the genetic mechanisms that affect specific phenotypes. At the forefront of this type of work are large-scale studies integrating genomics and epigenomics, a trend that has not yet been widely implemented in MZ twin research, but which offers great promise for moving towards a holistic view of phenotypic diversity.

Another case in which epigenetics, and particularly epigenetics in MZ twins, can inform us about genetics is somatic changes to DNA. Somatic mutations arise throughout the life of the organism,

* Corresponding author.
E-mail address: jungkyoon@kaist.ac.kr (J.K. Choi).

http://dx.doi.org/10.1016/j.ymeth.2015.10.020
1046-2023/

and therefore can be pursued with the MZ twin model. Thus, transcriptional regulation, which is achieved in large part through epigenetics, can be affected both by genetic variation, or polymorphisms, but also by acquired genetic changes, or somatic mutations. As the former will be identical in MZ twins while the latter may diverge, this model offers unique opportunities for understanding how these two classes of mutation impact epigenetic processes.

Yet another level of variation exists in the form of external stimuli, responses to which depend on the interaction between genetic factors and epigenetic mechanisms. For example, although twins who share a certain genetic factor may not show differences when exposed to particular environmental stimuli, twins who share other genetic factors may exhibit discrepancy in their responses to the same stimuli, despite identical genetic makeup. In this manner, genetic factors may contribute to differential disease susceptibility between identical twins.

It is this last observation that is the most compelling: the discordance between MZ twins. Such instances, which we review here, offer a window into the workings of genetics and epigenetics. We discuss twin-based, new approaches to disentangle the affects of nature and nurture and dissect the complex interplay between genetics and epigenetics at a molecular level. MZ twin models, when coupled with recently emerging research methodologies, tools, and resources in genomics and epigenomics, may open up new research avenues at the interface between genetics and epigenetics.

2. Integration of genome, epigenome, and transcriptome at the chromatin level

One of the most important discoveries in genetics in the recent years is that a majority of DNA variations associated with particular traits reside in genomic regions that are outside of protein-coding regions in what was once referred to as junk DNA. Now it is widely accepted that these noncoding regions contain genetic instructions that control the expression of specific genes. Noncoding regions account for >95% of the human genome, and therefore many trait-associated DNA variants may alter regulatory elements affecting transcriptional processes rather than the sequence of the protein itself. These observations have now been systematically catalogued by ENCODE, the Encyclopedia Of DNA Elements [2–6]. As of 2012, it was reported that the vast majority (>80%) of the human genome participates in regulatory function [5], although there has been some debate about claim [7]. It is notable that most of the supporting ENCODE data were based on chromatin accessibility.

Chromatin structure regulates access to the DNA for a wide spectrum of DNA binding proteins to regulate transcription, DNA repair, recombination, and replication [8]. As such, the profiling of open chromatin and histone modifications has been used to identify the genomic locations of various regulatory regions including promoters, enhancers, insulators, silencers, etc. [9–12]. Specific combinations of histone modifications can dictate the increase of decrease of gene expression by modulating the chromatin accessibility of transcription factors (TFs). Chromatin immunoprecipitation followed by genome sequencing (ChIP-seq) has been widely used to profile histone modifications that mark active, inactive, or poised promoters or enhancers [13]. Chromatin accessibility itself can be directly measured via next-generation sequencing by taking advantage of the fact that accessible chromatin is hypersensitive to digestion by DNase I. Similarly, the DNA binding sites of TFs can be extensively profiled based on the distribution of sequencing tags derived from DNase I hypersensitive sites (DHSs) [14,15]. The FAIRE-seq (formaldehyde-assisted isolation of regulatory elements) assay has also been used to capture accessible chromatin regions in the genome [10,16–20].

Given the well-established biological mechanisms and a large volume of relevant data, chromatin structure and the histone modifications that modulate it have been a focal point in studies aimed at a broader understanding of gene regulatory mechanisms. The strength of the connection between chromatin and genetic variation was demonstrated in 2010 when it was shown that linked chromatin accessibility patterns and underlying genetic polymorphisms constitute heritable features [21]. Association mapping of DHSs was used to understand the genetic basis of chromatin regulation for transcription control [22]. A similar attempt was made based on the genetic linkage of FAIRE signals [20]. Importantly, disease-associated regulatory variations identified through genome-wide association studies (GWASs) are concentrated in regulatory DNA marked by DHSs [23]. This study also identified distant gene targets for hundreds of variant-containing DHSs that may explain phenotype associations. Histone modifications also have implications for the interpretation of mechanisms for disease-associated regulatory variations. Disease variants frequently coincide with enhancer elements marked with particular histone modifications specific to a relevant cell type [24]. Large clusters of enhancers called super-enhancers were identified in a number of human cell and tissue types based on histone modification profiles, and it was found that disease-associated variation is especially enriched in the super-enhancers of disease-relevant cell types [25].

These findings spurred the development of a genetic and epigenetic fine-mapping method to identify causal variants in linkage disequilibrium with tag SNPs detected in GWASs [26]. In the context of autoimmune diseases, the predicted causal variants tend to occur near binding sites for critical regulators of immune function, but only 10–20% directly alter recognizable transcription factor binding motifs. In other words, we cannot label trait-associated genetic variants as causative factors simply because they are located in a region of accessible chromatin. Causality can be tested by examining whether chromatin accessibility mechanistically changes as DNA sequence changes. At heterozygous sites, the ratio of the reads from each allele is supposed to be close to 1:1 when sequencing a diploid genome. But if a particular variant changes the chromatin structure, the allele ratio generated from DHS/FAIRE sequencing or histone modification ChIP-seq deviates from 1:1 [27] (Fig. 1). A computational model was recently developed to detect this deviation [28].

A trio of recent reports [29–31] demonstrated that allele-specific TF binding, which occurs mainly through TF motif disruption by DNA variants, underlies the allelic imbalance in chromatin accessibility. This illustrates how genetics drives epigenetics [32]: DNA variants influence the epigenetic layer of transcriptional regulation by altering the sequence-specific activity of TFs. Remarkably, all three studies found that many of the DNA variants that led to allelic imbalance in TF binding sites were not associated with gene expression variation. This may reflect presence of non-consequential regulatory variations or mechanisms that compensate for the consequences of functional variations. In any case, this highlights the need to directly examine RNA expression. RNA-seq also can be used to interrogate allelic effects when RNA reads overlap a site that provides a heterozygote call. Allele-specific expression (ASE) refers to a phenomenon whereby transcriptional activity at the different alleles of a gene in a diploid genome can differ considerably [27] (Fig. 1). Genome-wide ASE was investigated in human, mice and cell lines [33–40]. ASE can be used to identify causal variants associated with diseases and more importantly, to further characterize them particularly in terms of the function of target genes [40,41].

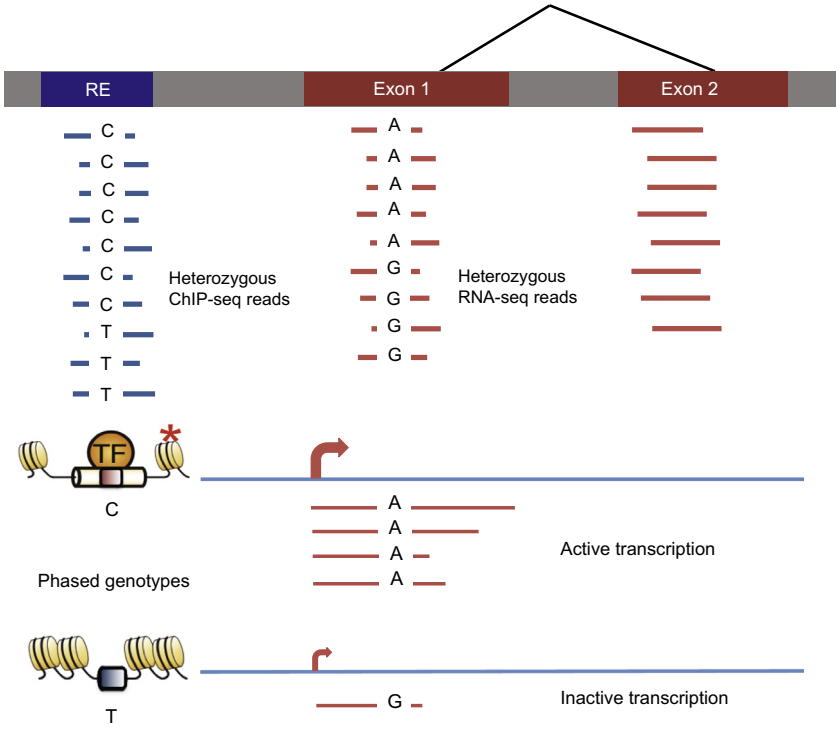

Fig. 1. Schematic showing allelic imbalance and allele-specific expression (ASE). The activity of the regulatory element (RE) is estimated based on the number of histone modification ChIP-seq reads. The expression level of the target gene controlled by the RE is measured in the number of RNA-seq reads. The deviation of the allelic ratio from 1:1 at a heterozygous site infers that the RE activity and gene expression level are different at two chromosomes. Allelic imbalance in RE proves its functionality such as disease-related variant affecting TF binding. The target gene that a RE affects can be predicted by analyzing phased genotypes. In the example above, the C allele at the RE facilitates activator binding with increased activation histone marks, leading to an amplified ASE of RNA that harbors the A allele from the same chromosome.

3. Linking genomic, epigenomic, transcriptomic, and phenotypic variations using MZ twins

The previous section described how genomic and epigenomic approaches are recently being employed to investigate the connections between variations in DNA, chromatin, RNA, and phenotype. MZ twins can shed new lights on these multi-layer interactions among variations at different levels. Especially, the twins discordant for a particular trait despite identical genetic makeup can offer a unique perspective.

A bottleneck when describing the usefulness of MZ twins in this aspect is the scarcity of genome-wide twin chromatin studies. A majority of twin studies examined DNA methylation, which is thought of as a primary epigenetic mechanism that could explain non-genetic discordance between MZ twins [42–47]. Although DNA methylation does function as an intermediary of genetic factors associated with particular traits or phenotypes [48], the cis effect of the underlying DNA sequences on DNA methylation is not as distinct as on chromatin structure. Only those mutations that arise at CpG sites can affect DNA methylation. Moreover, compared to a point mutation that disrupts sequence-specific binding sites, a methylation change at a single CpG site may exert weaker effects on chromatin structure and transcription activity. Furthermore, chromatin structure is central to epigenetic regulation because its modulation is typically the sum of various signals, including DNA methylation [49–53]. Hence, chromatin studies in twins can be as valuable as DNA methylation studies.

A genome-wide discordance in chromatin accessibility between MZ twins was recently shown for the first time based on in-depth FAIRE-seq for 36 pairs of MZ twins discordant for a particular trait [54]. However, allelic imbalance in FAIRE-seq and its associated ASE were not investigated in this work. In another study, twin samples were used in H3K4me3 ChIP-seq, but only for a technical purpose [55]. RNA-seq was used to understand the phenotypic discordance of MZ twins [56,57], and one of these studies also examined genome-wide H3K4me3 and DHS profiles to characterize the chromatin environment of differential gene expression [56]. However, the contribution of genetic factors was not investigated in these studies. According to a study of ASE [58], the extent of ASE is highly similar between MZ twin siblings compared to among unrelated individuals. This implies that ASE is strongly dependent on genetic factors, most likely residing in associated regulatory elements, the functionality of which can be examined based on allelic imbalance in DHS/FAIRE/ChIP sequences. Therefore, many more studies examining genetic variations in regulatory regions together with those in transcripts are needed in twin research.

Chromatin architecture research is playing an instrumental part in explaining regulatory roles of noncoding variants discovered in GWASs [23–26]. Finding functional variants and the corresponding affected chromatin site(s) and target gene(s) is made possible through allelic imbalance and ASE (Fig. 1). If a chromatin region affected by a causal variant that is associated with a particular disease shows a significant difference between discordant MZ twins, the chromatin site can be thought of as a causal locus that affects the phenotype of the twin siblings. This analysis is important because simply identifying distinct chromatin regions in bulk cannot differentiate cause and effect. The expected results of this type of analysis are illustrated in Fig. 2. GWAS fine mapping can be performed by identifying imbalanced variants in DHS/FAIRE/ChIP sequencing that are in linkage disequilibrium with reporter SNPs. Associated ASE can also be identified from RNA-seq data. If the identified chromatin regions and transcripts exhibit disparity in MZ twins discordant for the same disease associated with the GWAS variants, they can be regarded as causal molecular events underlying non-genetic phenotypic discordance of the twins.

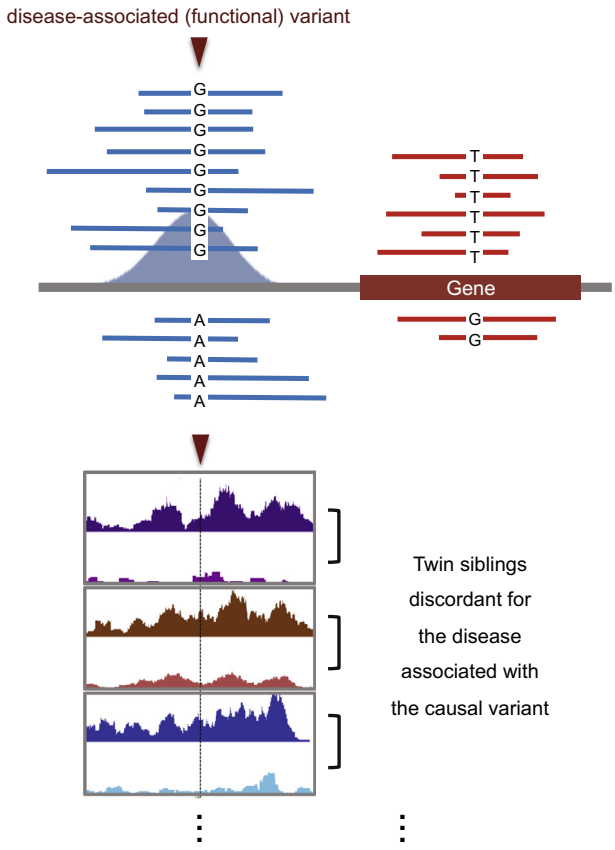

Fig. 2. Illustration of intrapair MZ twin chromatin variation at the locus associated with a GWAS SNP. Shown above are the ASE of a gene represented by the T and G alleles on RNA-seq reads, and allelic imbalance between the G and A alleles on DHS/FAIRE/ChIP sequencing reads. The chromatin-associated imbalanced variant is in linkage disequilibrium with a GWAS tag (reporter) SNP. Therefore, differential chromatin accessibility at this locus in the MZ twins may play a causal role in their discordance for a specific disease, unlike a majority of other chromatin regions that most likely are secondary consequences of or irrelevant to their phenotypic discordance.

To test this idea, we performed a case study (unpublished work) based on RNA-seq and histone modification (H3K4me1, H3K4me3, H3K27me3, and H3K27ac) ChIP-seq data in lymphoblastoid cell lines (LCLs) from 24 different human subjects [29–31]. We first identified imbalanced variants, that is, heterozygous variants whose allelic ratio was >15% and <85% from >8 sequence reads, and the binomial *P* value <0.05 [27,41], from the RNA-seq and ChIP-seq data separately. On the other hand, we collected 45,473 single nucleotide polymorphisms (SNPs) that were in linkage disequilibrium with 1097 GWAS SNPs associated with autoimmune disease at r2 >0.8 [59]. These SNPs were overlapped with the imbalanced ChIP-seq variants we identified above. For the samples whose phased genotype data were not available, we used the phasing information from the 1000 Genomes Project [60] to identify the matched, imbalanced RNA-seq variants in the same phase ($|r| > 0.8$) [61]. For the samples whose phased genotype data were available, we leveraged long-range chromatin interactions predicted in LCLs [62] to phase-match the disease-associated SNPs in distal transcriptional enhancers with the RNA-seq ASE genes. Finally, we identified 94 pairs between the GWAS-associated imbalanced ChIP-seq variants and the imbalanced RNA-seq variants being in the relationship whereby the alleles increasing the active histone marks (H3K4me1, H3K4me3, and H3K27ac) were in the same phase with the more abundant RNA-seq alleles, and the variants increasing the repressive histone mark (H3K27me3) were associated with the less expressed RNA-seq alleles.

These 94 pairs overlapped 290 open chromatin loci whose FAIRE-based accessibility was >2-fold different between twin siblings across 36 pairs of the MZ twins that were discordant for immunological traits, mostly involving allergic symptoms [54]. In one of these cases, RNA-seq alleles from the ALCAM (activated leukocyte cell adhesion molecule) gene was associated with ChIP-seq alleles from a H3K27ac peak. The activating H3K27ac allele, G, was in the same phase with the activated RNA allele, T. Chromatin accessibility at the H3K27ac locus was significantly different between five pairs of discordant twin siblings. The H3K27ac variant was in linkage disequilibrium with a tag SNP identified from a GWAS as a risk locus for an immune disease. The allelic imbalance pattern indicates that H3K27ac activity is potentially associated with the disease trait. This locus acts as a distal enhancer to an immune-related gene, and more RNA is expressed from the chromosome that has a higher enhancer activity. In five out of 36 immunologically discordant twin pairs, the enhancer locus shows differential chromatin accessibility. Therefore, the differential expression of the relevant gene between the twin siblings may be regarded as a causal molecular event that drives phenotypic discordance between them. Without such additional evidence from genetic studies, most of epigenetic or gene expression differences should be viewed as secondary consequences of disease onset and progression. This illustrates how causal epigenetic variation between MZ twins can be inferred by leveraging trait-associated genetic variants.

4. Understanding regulatory roles of somatic mutations using MZ twins

Somatic DNA changes, which can occur at twinning or during later developmental stages, can trigger genetic differences in MZ twins, in some cases causing discordance for disease susceptibility [55,56]. When considering changes in DNA regardless their heritability, MZ twins offer an opportunity to observe the effects of somatic mutations in an organism. It has been shown that somatic mutations significantly affect discordance of chromatin accessibility in twins and that the levels of chromatin discordance vary according to density and location of mutations inside the accessible chromatin region [54]. Furthermore, somatic mutations that disrupt TF binding sites particularly increase chromatin discordance, and most importantly, chromatin discordance between twins leads to differential gene expression [54].

Along with these findings, examples that illustrate allelic imbalance due to somatic mutation were also found [54]. However, unlike polymorphisms, somatic mutations can occur in later stages of organ development or cell differentiation, and, in this case, only a small proportion of cells exhibit the mutations. This phenomenon is called somatic mosaicism [57–59]. Somatic mosaicism can only be found by high-depth sequencing. It needs to be further examined whether the observed allelic imbalance due to somatic mutations [54] is a consequence of variation in chromatin accessibility or simply that of somatic mosaicism.

Another study [63] genotyped 66 healthy MZ twins at >500,000 polymorphic sites and tested a selected subset of candidate mutations for somatic mosaicism by targeted high-depth sequencing. Based on allelic ratios at the mutation sites in heterozygous twins, the authors concluded that there is little evidence of mosaicism and that these mutations most likely occurred at the twinning, during early embryonic development, or in somatic stem cells or progenitor cells. This work provides evidence that early somatic mutations do occur and can cause differences in genomes between otherwise identical twins. Somatic mutations arising in progenitor

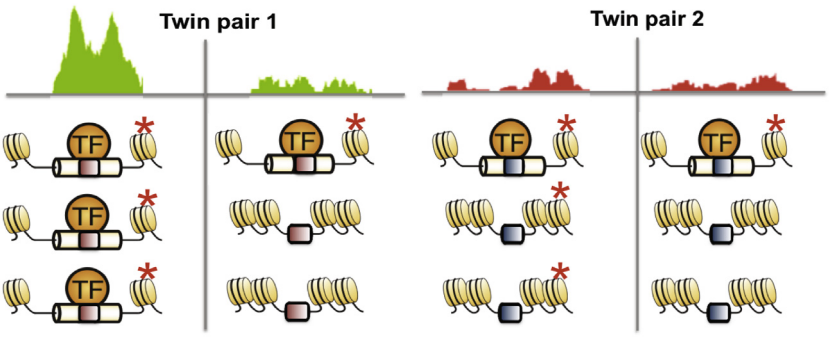

Fig. 3. Illustration of a possible mechanism by which shared genotypes at TF binding sites create chromatin discordance. Twin pair 1 inherited the red genotype that has a high affinity for the relevant TF, whereas twin pair 2 inherited the blue genotype, which hinders TF binding. Environmental differences cause differential histone modifications (red marks) between siblings. This epigenetic variation is manifested by differential TF binding in twin pair 1 (chromatin accessibility signals in green on top) but masked by low TF binding in twin pair 2 (chromatin accessibility signals in red on top).

cells are especially important because the associated chromatin states can be continuously transmitted to daughter cells in a similar manner as chromatin status altered by genetic polymorphism is maintained in offspring [21]. Further studies are needed to assess the degree of DNA changes in MZ twins due to environmental or stochastic differences and whether these DNA changes result in discordant phenotypes through gene expression altered by epigenetic variations at the chromatin level.

Unlike those of polymorphisms, regulatory roles of non-coding somatic DNA changes have not been studied extensively. MZ twins provide an excellent model not only to study the general properties of naturally occurring somatic mutations in living organisms but also to understand the functional effects of a majority of those changes arising in non-coding regions. In the same manner with polymorphisms, chromatin structural changes can be used as a barometer to infer the functionality or causality of these acquired DNA changes.

5. The interplay between genetic and environmental factors induces MZ twin discordance

MZ twins provide a unique opportunity to detect various types of gene-environment interactions, in which different genotypes respond differently to the same environment. The concept of "variability genes" as opposed to "level genes" was introduced based on the observation that intrapair variance for cholesterol levels, not the levels themselves, in MZ pairs who were blood group NN was significantly higher than in MZ pairs who were blood group MM or MN [64]. Level genes affect the mean expression of a trait (e.g., cholesterol level) and are the usual target of association studies. By contrast, variability genes may have no effect on the mean expression level but affect the variance of expression. In a follow-up study that examined the Kidd blood group locus in 142 MZ twin pairs, the co-twin difference in total cholesterol was lower in MZ pairs who were heterozygous for the locus or homozygous for the Jk(b) allele than in those who were homozygous for the Jk(a) allele [65]. In another study [66], the cholesteryl ester transfer protein locus was identified as the functional variability gene with respect to total and LDL cholesterol variability. These findings indicate that variability genes act by influencing phenotypic sensitivity to particular environmental stimuli. In the above example, the reduction of serum cholesterol in response to a low fat diet was greatest in those who were blood group NN and least in MN heterozygotes [67]. This explains the intrapair variance for cholesterol in MZ twins of blood group NN [64]. In agriculture, much of the increase in crop yields can be attributed to selection for genes able to respond to increased fertilizer doses rather than genes with higher yields in the average environment.

A previous study [54] attempted to identify variability genes that increase chromatin discordance between twin siblings. To this end, array-based genotyping was performed across the genomes of 36 pairs of MZ twins, searching for genetic polymorphisms that are shared by twin siblings and increase co-twin differences in chromatin accessibility. This approach should identify cases in which certain chromatin sites are more differentially accessible between twin siblings who share a particular allele than between other siblings with different alleles. Quantitative trait loci (QTL) mapping was performed by associating the intra differences in FAIRE signals with the genotypes shared by each twin pair. At an FDR of 0.01, a total of 10,195 local (cis) associations were identified for 1325 chromatin loci. A possible mechanism by which the variability genotypes may create chromatin discordance is illustrated in Fig. 3. A twin pair (twin pair 1 on the left) shares the red genotype that increases affinity for the relevant TF, while another twin pair (twin pair 2 on the right) carries the blue genotype that decreases affinity for the TF. Intrapair differences in chromatin states can be caused by non-genetic factors that reflect different histories of environmental exposure between twin siblings. In the illustration, active histone modifications (red marks) are found more frequently in one sibling than in the other. These epigenetic differences will be manifested by differential TF binding in twin pair 1 in the form of differential chromatin accessibility (green signals on top) but are masked by genetically low TF binding to the TFBS in twin pair 2 (red signals on top). In addition to histone modifications, the activity of chromatin regulators or the expression levels of TFs can serve as the epigenetic differences that reflect different histories of environmental exposure between twin siblings.

This illustrates that the previously proposed concept of a variability gene may be prevalent in the human genome, potentially affecting various phenotypes beside cholesterol levels. Nevertheless, we are unsure of how much of the examined trait, in this case, the chromatin accessibility, influences physiological phenotype through transcription processes. In the previous section, we introduced methods to predict causal chromatin differences that can affect actual phenotype discordance by using DNA variants discovered by GWAS fine-mapping. If these functional GWAS variants exist at consistently discordant chromatin loci in multiple twin pairs, like those that can be identified through our QTL mapping approach, it will be a superb example of applying a twin model to study disease mechanisms based on interactions between genetics and environment. In particular, it will be a powerful explanation for the missing heritability shown in GWASs.

6. Concluding remarks

There have been multiple lines of research that study the phenotypic contribution of non-genetic factors such as the environment by using MZ twins. However, most of this research has focused on DNA methylation. DNA methylation that occurs on the parental strand is replicated in the daughter strand. Because of the well-known mechanisms by which DNA methylation is maintained during cell division, twin researchers have been focusing on DNA methylation as the major epigenetic mechanism. However, recent studies reveal mechanisms by which chromatin configuration including histone modifications can be transmitted during cell division independently of DNA methylation [68]. Moreover, chromatin structure coupled with somatic changes to underlying DNA can be permanently inherited by daughter cells, in a similar manner that genetic polymorphisms render chromatin accessibility a heritable feature [21]. In addition, in this review, we demonstrated that finding the mechanisms of actions of DNA variants discovered in GWASs by leveraging multi-omics data is applicable to MZ twin models. When applying this principle to genome-wide, transcriptome-wide, and epigenome-wide association studies using a large MZ twin cohort, a more systematic research of gene-environmental interaction can be expected.

Acknowledgments

This work was supported by the KAIST Future Systems Healthcare Project, a grant [HI13C2143] funded through the Korea Health Industry Development Institute, and a grant [2013M3A9C4078139] through the National Research Foundation of Korea.

References

[1] A. Baccarelli, V. Bollati, Epigenetics and environmental chemicals, Curr. Opin. Pediatr. 21 (2009) 243–251. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3035853&tool=pmcentrez&rendertype=abstract (accessed January 21, 2015)..
[2] ENCODE Project Consortium, The ENCODE (ENCyclopedia Of DNA Elements) Project, Science 306 (2004) 636–640, http://dx.doi.org/10.1126/science.1105136.
[3] The ENCODE Project Consortium, A user's guide to the encyclopedia of DNA elements (ENCODE), PLoS Biol. 9 (2011) e1001046, http://dx.doi.org/10.1371/journal.pbio.1001046.
[4] E. Birney, J.A. Stamatoyannopoulos, A. Dutta, R. Guigó, T.R. Gingeras, E.H. Margulies, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, Nature 447 (2007) 799–816, http://dx.doi.org/10.1038/nature05874.
[5] B.E. Bernstein, E. Birney, I. Dunham, E.D. Green, C. Gunter, M. Snyder, An integrated encyclopedia of DNA elements in the human genome, Nature 489 (2012) 57–74, http://dx.doi.org/10.1038/nature11247.
[6] M. Kellis, B. Wold, M.P. Snyder, B.E. Bernstein, A. Kundaje, G.K. Marinov, et al., Defining functional DNA elements in the human genome, Proc. Natl. Acad. Sci. U.S.A. 111 (2014) 6131–6138, http://dx.doi.org/10.1073/pnas.1318948111.
[7] W.F. Doolittle, Is junk DNA bunk? A critique of ENCODE, Proc. Natl. Acad. Sci. U. S.A. 110 (2013) 5294–5300, http://dx.doi.org/10.1073/pnas.1221376110.
[8] O. Bell, V.K. Tiwari, N.H. Thomä, D. Schübeler, Determinants and dynamics of genome accessibility, Nat. Rev. Genet. 12 (2011) 554–564, http://dx.doi.org/10.1038/nrg3017.
[9] A.P. Boyle, S. Davis, H.P. Shulha, P. Meltzer, E.H. Margulies, Z. Weng, et al., High-resolution mapping and characterization of open chromatin across the genome, Cell 132 (2008) 311–322.
[10] L. Song, Z. Zhang, L.L. Grasfeder, A.P. Boyle, P.G. Giresi, B.-K. Lee, et al., Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity, Genome Res. 21 (2011) 1757–1767.
[11] L.E. Berchowitz, S.E. Hanlon, J.D. Lieb, G.P. Copenhaver, A positive but complex association between meiotic double-strand break hotspots and open chromatin in *Saccharomyces cerevisiae*, Genome Res. 19 (2009) 2245–2257.
[12] B. Audit, L. Zaghloul, C. Vaillant, G. Chevereau, Y. d'Aubenton-Carafa, C. Thermes, Open chromatin encoded in DNA sequence is the signature of "master" replication origins in human cells, Nucleic Acids Res. 37 (2009) 6064–6075.
[13] T.S. Furey, ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions, Nat. Rev. Genet. 13 (2012) 840–852, http://dx.doi.org/10.1038/nrg3306.
[14] R.E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M.T. Maurano, E. Haugen, et al., The accessible chromatin landscape of the human genome, Nature 489 (2012) 75–82.
[15] S. Neph, J. Vierstra, A.B. Stergachis, A.P. Reynolds, E. Haugen, B. Vernot, et al., An expansive human regulatory lexicon encoded in transcription factor footprints, Nature 489 (2012) 83–90.
[16] P.G. Giresi, J. Kim, R.M. McDaniell, V.R. Iyer, J.D. Lieb, FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin, Genome Res. 17 (2007) 877–885.
[17] H. Waki, M. Nakamura, T. Yamauchi, K. Wakabayashi, L. Hirose-Yotsuya, K. Take, et al., Global mapping of cell type–specific open chromatin by FAIRE-seq reveals the regulatory role of the NFI family in adipocyte differentiation, PLoS Genet. 7 (2011) e1002311.
[18] K.J. Gaulton, T. Nammo, L. Pasquali, J.M. Simon, P.G. Giresi, M.P. Fogarty, et al., A map of open chromatin in human pancreatic islets, Nat. Genet. 42 (2010) 255–259.
[19] A.J. Smith, P. Howard, S. Shah, P. Eriksson, S. Stender, C. Giambartolomei, et al., Use of allele-specific FAIRE to determine functional regulatory polymorphism using large-scale genotyping arrays, PLoS Genet. 8 (2012) e1002908.
[20] K. Lee, S.C. Kim, I. Jung, K. Kim, J. Seo, H.-S. Lee, et al., Genetic landscape of open chromatin in yeast, PLoS Genet. 9 (2013) e1003229, http://dx.doi.org/10.1371/journal.pgen.1003229.
[21] R. McDaniell, B.-K. Lee, L. Song, Z. Liu, A.P. Boyle, M.R. Erdos, Heritable individual-specific and allele-specific chromatin signatures in humans, Science 328 (2010) 235–239.
[22] J.F. Degner, A.A. Pai, R. Pique-Regi, J.B. Veyrieras, D.J. Gaffney, J.K. Pickrell, et al., DNase I sensitivity QTLs are a major determinant of human expression variation, Nature 482 (2012) 390–394.
[23] M.T. Maurano, R. Humbert, E. Rynes, R.E. Thurman, E. Haugen, H. Wang, et al., Systematic localization of common disease-associated variation in regulatory DNA, Science 337 (2012) 1190–1195, http://dx.doi.org/10.1126/science.1222794.
[24] J. Ernst, P. Kheradpour, T.S. Mikkelsen, N. Shoresh, Mapping and analysis of chromatin state dynamics in nine human cell types, Nature 473 (2011) 43–49.
[25] D. Hnisz, B.J. Abraham, T.I. Lee, A. Lau, V. Saint-André, A.A. Sigova, et al., Super-enhancers in the control of cell identity and disease, Cell 155 (2013) 934–947, http://dx.doi.org/10.1016/j.cell.2013.09.053.
[26] K.K.-H. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W.J. Housley, S. Beik, et al., Genetic and epigenetic fine mapping of causal autoimmune disease variants, Nature 518 (2015) 337–343, http://dx.doi.org/10.1038/nature13835.
[27] T. Pastinen, Genome-wide allele-specific analysis: insights into regulatory variation, Nat. Rev. Genet. 11 (2010) 533–538, http://dx.doi.org/10.1038/nrg2815.
[28] D. Huang, I. Ovcharenko, Identifying causal regulatory SNPs in ChIP-seq enhancers, Nucleic Acids Res. 43 (2014) 225–236, http://dx.doi.org/10.1093/nar/gku1318.
[29] M. Kasowski, S. Kyriazopoulou-Panagiotopoulou, F. Grubert, J.B. Zaugg, A. Kundaje, Y. Liu, Extensive variation in chromatin states across humans, Science 342 (2013) 750–752, http://dx.doi.org/10.1126/science.1242510.
[30] G. McVicker, B. van de Geijn, J.F. Degner, C.E. Cain, N.E. Banovich, A. Raj, et al., Identification of genetic variants that affect histone modifications in human cells, Science 342 (2013) 747–749, http://dx.doi.org/10.1126/science.1242429.
[31] H. Kilpinen, S.M. Waszak, A.R. Gschwind, S.K. Raghav, R.M. Witwicki, A. Orioli, et al., Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription, Science 342 (2013) 744–747, http://dx.doi.org/10.1126/science.1242463.
[32] T. Furey, P. Sethupathy, Genetics driving epigenetics, Science (80-.) 342 (2013) 705–706. http://www.sciencemag.org/content/342/6159/705.short (accessed January 6, 2015).
[33] G.A. Heap, J.H.M. Yang, K. Downes, B.C. Healy, K.A. Hunt, N. Bockett, et al., Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing, Hum. Mol. Genet. 19 (2010) 122–134, http://dx.doi.org/10.1093/hmg/ddp473.
[34] B. DeVeale, D. van der Kooy, T. Babak, Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective, PLoS Genet. 8 (2012) e1002600, http://dx.doi.org/10.1371/journal.pgen.1002600.
[35] C. Gregg, J. Zhang, B. Weissbourd, S. Luo, G.P. Schroth, D. Haig, et al., High-resolution analysis of parent-of-origin allelic expression in the mouse brain, Science 329 (2010) 643–648, http://dx.doi.org/10.1126/science.1190830.
[36] X. Wang, Q. Sun, S.D. McGrath, E.R. Mardis, P.D. Soloway, A.G. Clark, Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain, PLoS One 3 (2008) e3839, http://dx.doi.org/10.1371/journal.pone.0003839.
[37] E. Grundberg, K.S. Small, Å.K. Hedman, A.C. Nica, A. Buil, S. Keildson, et al., Mapping cis- and trans-regulatory effects across multiple tissues in twins, Nat. Genet. 44 (2012) 1084–1089, http://dx.doi.org/10.1038/ng.2394.
[38] Y.S. Ju, J.-I. Kim, S. Kim, D. Hong, H. Park, J.-Y. Shin, et al., Extensive genomic and transcriptional diversity identified through massively parallel DNA and RNA sequencing of eighteen Korean individuals, Nat. Genet. 43 (2011) 745–752, http://dx.doi.org/10.1038/ng.872.
[39] J.C. Almlöf, P. Lundmark, A. Lundmark, B. Ge, S. Maouche, H.H.H. Göring, et al., Powerful identification of cis-regulatory SNPs in human primary monocytes using allele-specific gene expression, PLoS One 7 (2012) e52260, http://dx.doi.org/10.1371/journal.pone.0052260.
[40] D.J. Verlaan, B. Ge, E. Grundberg, R. Hoberman, K.C.L. Lam, V. Koka, et al., Targeted screening of cis-regulatory variation in human haplotypes, Genome Res. 19 (2009) 118–127, http://dx.doi.org/10.1101/gr.084798.108.

[41] L. Conde, P.M. Bracci, R. Richardson, S.B. Montgomery, C.F. Skibola, Integrating GWAS and expression data for functional characterization of disease-associated SNPs: an application to follicular lymphoma, Am. J. Hum. Genet. 92 (2013) 126–130, http://dx.doi.org/10.1016/j.ajhg.2012.11.009.
[42] J.T. Bell, T.D. Spector, DNA methylation studies using twins: what are they telling us ?, Genome Biol 13 (2012) 172.
[43] J.T. Bell, T.D. Spector, A twin approach to unraveling epigenetics, Trends Genet. 27 (2011) 116–125.
[44] M.F. Fraga, E. Ballestar, M.F. Paz, S. Ropero, F. Setien, M.L. Ballestar, et al., Epigenetic differences arise during the lifetime of monozygotic twins, Proc. Natl. Acad. Sci. 102 (2005) 10604–10609.
[45] Z.A. Kaminsky, T. Tang, S.-C. Wang, C. Ptak, G.H.T. Oh, A.H.C. Wong, et al., DNA methylation profiles in monozygotic and dizygotic twins, Nat. Genet. 41 (2009) 240–245.
[46] A.F. McRae, J.E. Powell, A.K. Henders, L. Bowdler, G. Hemani, S. Shah, et al., Contribution of genetic variation to transgenerational inheritance of DNA methylation, Genome Biol. 15 (2014) R73, http://dx.doi.org/10.1186/gb-2014-15-5-r73.
[47] B.M. Javierre, A.F. Fernandez, J. Richter, F. Al-Shahrour, J.I. Martin-Subero, J. Rodriguez-Ubreva, et al., Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus, Genome Res. 20 (2010) 170–179, http://dx.doi.org/10.1101/gr.100289.109.
[48] Y. Liu, M.J. Aryee, L. Padyukov, M.D. Fallin, E. Hesselberg, A. Runarsson, et al., Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis, Nat. Biotechnol. 31 (2013) 142–147, http://dx.doi.org/10.1038/nbt.2487.
[49] I. Keshet, J. Lieman-Hurwitz, H. Cedar, DNA methylation affects the formation of active chromatin, Cell 44 (1986) 535–543, http://dx.doi.org/10.1016/0092-8674(86)90263-1.
[50] H.H. Ng, A. Bird, DNA methylation and chromatin modification, Curr. Opin. Genet. Dev. 9 (1999) 158–163. http://www.ncbi.nlm.nih.gov/pubmed/10322130 (accessed January 9, 2015).
[51] H. Cedar, Y. Bergman, Linking DNA methylation and histone modification: patterns and paradigms, Nat. Rev. Genet. 10 (2009) 295–304, http://dx.doi.org/10.1038/nrg2540.
[52] W. Fischle, Talk is cheap–cross-talk in establishment, maintenance, and readout of chromatin modifications, Genes Dev. 22 (2008) 3375–3382, http://dx.doi.org/10.1101/gad.1759708.
[53] J.-S. Lee, E. Smith, A. Shilatifard, The language of histone crosstalk, Cell 142 (2010) 682–685, http://dx.doi.org/10.1016/j.cell.2010.08.011.
[54] K. Kim, H.-J. Ban, J. Seo, K. Lee, M. Yavartanoo, S.C. Kim, et al., Genetic factors underlying discordance in chromatin accessibility between monozygotic twins, Genome Biol. 15 (2014) R72, http://dx.doi.org/10.1186/gb-2014-15-5-r72.
[55] G.D. Gilfillan, T. Hughes, Y. Sheng, H.S. Hjorthaug, T. Straub, K. Gervin, et al., Limitations and possibilities of low cell number ChIP-seq, BMC Genomics 13 (2012) 645, http://dx.doi.org/10.1186/1471-2164-13-645.
[56] A. Letourneau, F.A. Santoni, X. Bonilla, M.R. Sailani, D. Gonzalez, J. Kind, et al., Domains of genome-wide gene expression dysregulation in Down's syndrome, Nature 508 (2014) 345–350, http://dx.doi.org/10.1038/nature13200.
[57] S.E. Baranzini, J. Mudge, J.C. van Velkinburgh, P. Khankhanian, I. Khrebtukova, N.A. Miller, et al., Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis, Nature 464 (2010) 1351–1356.
[58] V.G. Cheung, A. Bruzel, J.T. Burdick, M. Morley, J.L. Devlin, R.S. Spielman, Monozygotic twins reveal germline contribution to allelic expression differences, Am. J. Hum. Genet. 82 (2008) 1357–1360, http://dx.doi.org/10.1016/j.ajhg.2008.05.003.
[59] A.D. Johnson, R.E. Handsaker, S.L. Pulit, M.M. Nizzari, C.J. O'Donnell, P.I. de Bakker, SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap, Bioinformatics 24 (2008) 2938–2939.
[60] The 1000 Genomes Project Consortium, An integrated map of genetic variation from 1092 human genomes, Nature 491 (2012) 56–65, http://dx.doi.org/10.1038/nature11632.
[61] R. Lewontin, On measures of gametic disequilibrium, Genetics 852 (1988) 849–852. http://www.genetics.org/content/120/3/849.short (accessed January 12, 2015).
[62] B. He, C. Chen, L. Teng, K. Tan, Global view of enhancer-promoter interactome in human cells, Proc. Natl. Acad. Sci. U.S.A. 111 (2014), http://dx.doi.org/10.1073/pnas.1320308111. E2191–9.
[63] R. Li, A. Montpetit, M. Rousseau, S.Y.M. Wu, C.M.T. Greenwood, T.D. Spector, et al., Somatic point mutations occurring early in development: a monozygotic twin study, J. Med. Genet. 51 (2014) 28–34, http://dx.doi.org/10.1136/jmedgenet-2013-101712.
[64] K. Berg, A.L. Børresen, W.E. Nance, Apparent influence of marker genotypes on variation in serum cholesterol in monozygotic twins, Clin. Genet. 19 (1981) 67–70. http://www.ncbi.nlm.nih.gov/pubmed/6936104 (accessed January 10, 2015).
[65] K. Berg, Variability gene effect on cholesterol at the Kidd blood group locus, Clin. Genet. 33 (1988) 102–107. http://www.ncbi.nlm.nih.gov/pubmed/3359662 (accessed January 10, 2015).
[66] K. Berg, I. Kondo, D. Drayna, R. Lawn, Variability gene effect of cholesteryl ester transfer protein (CETP) genes, Clin. Genet. 35 (1989) 437–445. http://www.ncbi.nlm.nih.gov/pubmed/2567644 (accessed January 10, 2015).
[67] A.J. Birley, R. MacLennan, M. Wahlqvist, L. Gerns, T. Pangan, N.G. Martin, MN blood group affects response of serum LDL cholesterol level to a low fat diet, Clin. Genet. 51 (1997) 291–295. http://www.ncbi.nlm.nih.gov/pubmed/9212175 (accessed January 10, 2015).
[68] A.V. Probst, E. Dunleavy, G. Almouzni, Epigenetic inheritance during the cell cycle, Nat. Rev. Mol. Cell Biol. 10 (2009) 192–206, http://dx.doi.org/10.1038/nrm2640.