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].