Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Review

Population epigenetics

John M. Greally

Abstract

The field of epigenetics is maturing, with increased interest in understanding the normal regulation of the genome and the possibility that it becomes reprogrammed aberrantly as part of the cause of disease phenotypes. Applying the current technologies and insights to the study of human populations is potentially a way of understanding mechanisms and consequences of these diseases. When extended to encompass health care disparities, understanding why certain populations are unusually prone to specific conditions, there is certainly some potential for gaining new and valuable insights, but these studies are likely to be unusually prone to the effects of confounding influences and need to be designed, executed and interpreted with extra care.

Addresses

Department of Genetics, Albert Einstein College of Medicine, 1301 Morris Park Avenue, Bronx, NY 10461, USA

E-mail address: john.greally@einstein.yu.edu

Current Opinion in Systems Biology 2017, **1**:84–89

This review comes from a themed issue on **Genomics and epigenomics**

Edited by **Tuuli Lappalainen and Emmanouil Dermitzakis**

For a complete overview see the Issue and the Editorial

Available online 27 January 2017

http://dx.doi.org/10.1016/j.coisb.2017.01.004

Keywords

Epigenetics, population, DNA methylation, transcription factor, diversity.

Introduction

The first uses of the term "population epigenetics" were by Keller [1], who created a conceptual framework for understanding the inheritance of traits mediated by autoregulatory transcription factors, and by Richards [2], whose focus was on the natural variation in DNA methylation that occurs in plants. The intriguing possibility being raised was that this variability could in some way influence and modulate the effects of genetic variability. The implicit assumption at that time was that epigenetic variability was a completely distinct layer of information from that encoded in the genome, an assumption based on the then-current definition of "epigenetics", which back-translates epi- and -genetics as an influence residing above or upon that of the DNA sequence itself [3], usually taken to mean something mediated by a molecular regulator of genomic function.

An implicit hope was probably that we could use the lessons of the field of population genetics to apply to population epigenetics. This has unfortunately proven to be very difficult. As we have recently reviewed [4], the patterns of variability of molecular regulators of genomic function are themselves phenotypes, and subject to multiple influences, unlike the genotype, which is fixed within the individual. Apart from genotype, using any other kind of -omics data in a phenotypic association study is basically correlating a phenotype with a phenotype. This leads to problems when performing the typical cross-sectional study design approach employed when associating molecular genomic regulators with a phenotype — how do you know that the molecular changes observed lead to the phenotype, when the phenotype could instead be leading to the molecular changes? Such reverse causation is now demonstrated to occur for DNA methylation in peripheral blood leukocytes in a study of individuals with increased body mass index [5] and another looking at people with altered blood lipid profiles [6].

Furthermore, in recent years there have been numerous studies that have revealed an interplay between DNA sequence variability and the functional properties of the genome. One source of insight has been through observational studies, such as the identification of loci where the different parental alleles have markedly distinct DNA methylation patterns at a site of DNA sequence polymorphism [7], or larger-scale studies in which DNA sequence variants such as single nucleotide polymorphisms (SNPs) have been significantly correlated with different levels of DNA methylation at local sites, or more distantly or on other chromosomes [8–17]. These have been referred to as methylation quantitative trait loci (meQTLs or mQTLs), and have counterparts described for gene expression (eQTLs) [18] and chromatin states (chrQTLs) [19]. Through twin studies and other approaches, the proportion of inter-individual variability of DNA methylation that can be accounted for by DNA sequence variation has been calculated, and estimated to be between 22% and 80% in humans [7,8,20].

The initial hope that insights into epigenetic variability in a population could complement information about

DNA sequence variability has therefore been complicated by this strong association between the two types of information. However, it should be noted that DNA sequence variability does not account for 100% of variation of DNA methylation, indicating that if we can dissect out the interactions of these molecular events, we should be able to find two interesting types of information — the potential genomic regulatory mechanisms through which DNA sequence variants work, and the independent variation in genomic regulatory mechanisms that may be a source of modulation of sequence-based phenotypes, as originally hoped [2].

Main text

Keys to interpreting DNA methylation variability

The key to a population epigenetics study is to understand the influences affecting inter-individual variability in the genomic regulator being studied, which is usually DNA methylation. This turns out to be a surprisingly complicated area of research, as DNA methylation is influenced by a large number of factors. For example, as DNA methylation differs in each cell type in the body, and DNA methylation assays are performed on pools of cells, any systematic difference in a cell subtype proportion within the pool of cells tested will be reflected by changes in DNA methylation at loci where the pattern is distinctive in that cell type [21]. This is how the testing of pools of cells can generate differences in DNA methylation, without any cells present necessarily having changed their innate patterns of DNA methylation. The influence of cell subtype composition is now a well-recognized problem in studies of DNA methylation and has been the focus of a number of thoughtful analytical approaches [21—24]. Another unexpected problem is a molecular example of reverse causation. While DNA methylation is usually thought to be a regulator of gene expression, the act of transcribing a locus can alter its DNA methylation [25,26], requiring that we concurrently test the samples for their transcriptional profiles when testing DNA methylation.

The results to date of DNA methylation studies associated with phenotypes are not frequently replicated, with the exception of the association between cigarette smoking and DNA methylation in peripheral blood leukocytes, which has revealed the same loci to undergo changes in multiple studies [27—29]. However, even this paradigm of epigenetic association may be undermined by what appear to be substantial effects of DNA sequence variation and cell subtype effects for the informative loci [30—32], raising the possibility that these DNA methylation changes are substantially due to allelic variants for these meQTLs segregating non-randomly into the smoker and non-smoker groups, and blood cell subtypes being altered as a response to cigarette smoking.

When these cell subtype and transcriptional influences are combined with the strong effect of DNA sequence variation on DNA methylation, it becomes clear that any study of a molecular regulator like DNA methylation is by itself uninterpretable [4], and has to be studied with parallel genotyping and transcriptomic studies, and detailed insights into the cell subcomposition present. This makes epigenetic studies complex and demanding of resources, but allows the generation of rich data sets that allow interpretation of the results generated.

Transcription factors

Epigenetics, defined as an influence above or upon that of the DNA sequence itself, also has generally been taken to indicate an influence that can override transcriptional regulatory mechanisms. Observations that DNA methylation could inhibit the binding of DNA-binding proteins like transcription factors (TFs) [33—35] lent support to the implicit model that there exists a generic transcriptional regulatory program that could be overridden by "epigenetic" mechanisms.

The problem is that the evidence for epigenetic reprogramming, usually tested by studying patterns of DNA methylation, involves the same loci changing their patterns in multiple individuals. For the same sequences in the genome to be selected in this way, the mediators have to have the ability to recognize complex DNA sequences, which is not a property of DNA methyltransferases, histone modifying or nucleosomal remodeling enzymes. The potential mediators with the required sequence specificity are transcription factors or possibly some examples of small non-coding RNAs. We have recently noted that TFs not only have a primary role in transcriptional regulation, they drive cell fate choices and maintain cellular identities through autoregulatory mechanisms [36]. Their activities are also influenced by environmental factors, making them very attractive candidates for mediating the cellular reprogramming sought in studies associating molecular reprogramming of cells with phenotypes. As we also note [36], molecular processes like DNA methylation and chromatin modifications are influenced by the local binding of TFs, so that when these kinds of "epigenetic" regulators are noted to be altered in association with a phenotype, they may merely be footprinting where TFs have altered their activities, rather than representing the primary mediators of the regulatory process.

It is therefore essential to consider the possibility that TFs represent the primary mediators of cellular changes associated with phenotypes. Studying DNA methylation would remain of value even in this revised perspective, with its potential to define the sites at which these TF-mediated events are occurring. However, it is probably an over-interpretation to assume that sequence-specific DNA methylation changes occur autonomously.