DOCUMENT SUMMARY This review from *PLOS Genetics* critiques the simplistic model that environmental influences on disease are mediated solely through direct "epigenetic" reprogramming. Using nonalcoholic fatty liver disease (**NAFLD**) as a paradigm, the authors argue that results from epigenome-wide association studies (**EWAS**) are often confounded by unappreciated factors like changes in cell subtype proportions and underlying genetic variation (**meQTLs**, **eQTLs**). The paper advocates for embracing these "confounders" as valuable sources of insight and proposes a more integrated approach that combines **epigenomics** with functional genomics and single-cell techniques to better understand disease pathogenesis.

FILENAME
wattacheril_2023_research_review_epigenomics_environmental_influences_disease.md

METADATA Category: RESEARCH Type: review Relevance: Core Update Frequency: Static Tags: #epigenomics #epigenetics #nafld #ewas #dna_methylation #environmental_exposures #gene_environment_interaction #functional_genomics #qtl #dohad Related Docs: Documents related to genetics, chronic illness (especially liver disease), developmental health, and advanced research methodologies. Supersedes: N/A

FORMATTED CONTENT

Abstract

It is a generally accepted model that environmental influences can exert their effects, at least in part, by changing the molecular regulators of transcription that are described as **epigenetic**. As there is biochemical evidence that some epigenetic regulators of transcription can maintain their states long term and through cell division, an epigenetic model encompasses the idea of maintenance of the effect of an exposure long after it is no longer present. The evidence supporting this model is mostly from the observation of alterations of molecular regulators of transcription following exposures. With the understanding that the interpretation of these associations is more complex than originally recognised, this model may be oversimplistic; therefore, adopting novel perspectives and experimental approaches when examining how environmental exposures are linked to phenotypes may prove worthwhile. In this review, we have chosen to use the example of **nonalcoholic fatty liver disease (NAFLD)**, a common, complex human disease with strong environmental and genetic influences. We describe how **epigenomic** approaches combined with emerging functional genetic and single-cell genomic techniques are poised to generate new insights into the pathogenesis of environmentally influenced human disease phenotypes exemplified by **NAFLD**.

Introduction

Many human diseases have a clear environmental contribution, and for decades it has been assumed that the environment influences gene expression through "

epigenetic" mechanisms like **DNA methylation** and chromatin states. This model posits that

epigenetic alterations are the primary mediators of environmental influences, potentially propagating these effects long after the exposure ceases.

However, our interpretation of genome-wide studies of these mediators, known as

epigenome-wide association studies (EWAS), has become more critical. The once-simple interpretation that a change in

DNA methylation reflects cellular reprogramming is now recognized as too simplistic. This review uses

nonalcoholic fatty liver disease (NAFLD), including **nonalcoholic steatohepatitis (NASH)**, as a paradigm for a common disease with known environmental and genetic risk factors to re-examine the evidence.

We will review how systematic changes in cell subtype proportion or genetic variants can lead to what we term "molecular genomic" changes, offering alternative insights into disease pathophysiology. Furthermore, single-cell genomic assays are revealing unexpected heterogeneity within cell types and how DNA sequence variants can have effects restricted to specific cell subtypes. This review aims to prompt innovative ways of thinking about how to gain therapeutic insights into environmentally driven diseases like

NAFLD using novel molecular and cellular approaches.

The Epigenome and the Environment

Foundational Studies

Early studies supported the idea that **DNA methylation** responds to environmental influences. For example, feeding rats a methyl-deficient diet was associated with decreased

DNA methylation in their livers. A highly influential model was the "viable yellow" mouse, where maternal diet during pregnancy was shown to influence the offspring's adult phenotype by altering

DNA methylation at a specific gene locus. This work helped form the foundation for the field of the

Developmental Origins of Health and Disease (DOHaD). Similarly, studies on the Dutch Famine of 1944-1945 suggested that intrauterine nutritional deprivation was associated with later-life obesity, providing a model for

epigenetic mediation of environmental exposures long after the exposure has ended.

How Molecular Regulators May Respond to the Environment

The environment may act through regulators of chromatin and modifiers of DNA to influence gene expression.

DNA methylation is the paradigm for a regulator that can maintain a biochemical memory through cell division. However, some histone modifications and variants can also directly mediate environmental responses.

Environmental exposure	Molecular regulatory	Effect	Genomic location of effects

	changes		
Glucose	GlcNAcylation of EZH2	Enhances protein stability and enzymatic activity	Heterochromatin (H3K27me3)
Lactic acid	Lactylation of histones	Possibly gene activation	Active gene promoters
Ethanol	Acetylation of histones	Induction of expression of genes involved in signal transduction and learning and memory	New loci of H3K9ac and H3K27ac formed (promoters/enhancers)
Short chain fatty acids	Inhibition of histone deacetylases (HDACs)	Increased histone crotonylation	Histone crotonylation is enriched at gene promoters
Dietary folic acid, vitamins B6 and B12	Increased production of S-adenosyl methionine (SAM)	Increased methylation of DNA and histones	Genome-wide
Vitamin C	TET cofactor	Increased 5-hydroxymethylcytosine, decreased 5-methylcytosine	Genome-wide

A key problem is

cellular memory: with the exception of **DNA methylation**, there are no clear biochemical mechanisms for the self-propagation of these molecular events through cell division. Another problem is that these global regulators lack

sequence specificity and cannot choose specific loci for their activity, implying that sequence-specific **transcription factors (TFs)** likely play a primary role.

Epigenetic Association Studies and NAFLD

An

epigenome-wide association study (EWAS) is the typical approach for identifying **epigenetic** changes associated with a disease phenotype. Most

EWAS involve testing human blood leukocytes using **DNA methylation** microarrays to find statistically significant differences between groups.

NAFLD is the most common form of liver disease worldwide, affecting an estimated 6 to 30 million people in the United States. It is a spectrum of disease ranging from simple fat accumulation (steatosis) to an inflammatory phenotype,

NASH, which can progress to fibrosis and cirrhosis. The pathogenesis involves both environmental exposures (e.g., Western diet, gut microbiome) and genetic predisposition. Importantly, the environmental exposure to the organism (e.g., high-fructose corn syrup) may not be the same as the exposure experienced by the cells in the affected tissue (e.g., the inflammatory cytokine TGFB).

EWAS Design Issues

Several major issues complicate the interpretation of **EWAS** results:

1. **Surrogate tissue sampling:** Many studies use accessible tissues like peripheral blood leukocytes to study diseases in inaccessible organs like the liver. For these results to be informative, the genes and pathways would have to be active and dysregulated in the same way in both very different tissue types, a rationale that is often missing from study reports.
2. **Cohort sizes:** The studies on **NAFLD** have relatively small sample sizes (median of 61.5 for liver samples), meaning they likely detected only a limited subset of the actual **DNA methylation** changes.
3. **Reverse causation:** In cross-sectional studies, it is difficult to determine if the identified **DNA methylation** changes *cause* the phenotype or are *caused by* it. Longitudinal studies, such as sampling before and after bariatric surgery, can help address this by showing whether changes are reversible.
4. **Cell subtype proportional composition:** A tissue's composition can change with disease (e.g., influx of immune cells in **NASH**), leading to apparent **DNA methylation** changes in bulk tissue analysis without any single cell being reprogrammed. Deconvolution approaches can estimate and correct for these changes, revealing, for example, an increase in immune cells with the progression of fibrosis in **NAFLD**.
5. **Genetic effects on DNA methylation:** DNA sequence variation can be associated with **DNA methylation** differences, a phenomenon known as **methylation quantitative trait loci (meQTLs)**. These genetic effects can be very strong, accounting for 14% to 80% of **DNA methylation** variation. If not accounted for, genetic differences between cases and controls, especially if ancestries differ, can be misinterpreted as disease-associated **epigenetic** changes.

The potential outcome that results from failing to account for these problems is that changes in

DNA methylation could be overinterpreted, assumed to represent cellular reprogramming responding to an environmental provocation, and causative of the phenotype, whereas the **DNA methylation** change could instead be due to changes in cell subtype composition within the tissue, genetic differences between individuals, or the consequence of the hepatic phenotype.

Embracing the Sources of Variation

The influences that complicate

EWAS interpretation (cell subtype changes, genetic variants) are not merely artifacts; for them to create a consistent signal, they must be nonrandomly distributed between the groups. This means they reveal systematic cellular or genetic associations with the disease, offering valuable insights into pathogenesis.

Genetic Variation Modifying Environmental Response

Different people respond to the same environmental exposure in different ways, often due to genetic variation. This is the basis for

pharmacogenetics. This concept of

gene-environment (GxE) interaction can be extended beyond drugs to other exposures.

A striking example in **NAFLD** is the interaction between the *PNPLA3* genotype and Body Mass Index (BMI). The risk of fatty liver disease conferred by the

PNPLA3 variant is significantly amplified by increased BMI, vividly demonstrating a GxE interaction.

An even more powerful discovery is that of "

response eQTLs"—genetic variants whose effects on gene expression are only revealed after an environmental challenge, such as infection. These response

eQTLs have been identified for multiple exposures and are enriched at loci implicated by GWAS for relevant diseases, highlighting how genetic variation can mediate individual differences in environmental responses.

Insights from the Epigenome into Cell Signalling

Why do specific genomic loci undergo **epigenetic** changes in response to the environment? The targeting implies a primary role for

transcription factors (TFs). Many environmental exposures (e.g., hypoxia, hyperglycemia, ethanol) are known to influence cell signalling pathways, which in turn can alter TF activity and nuclear localization.

> This suggests a
>
> **TF-centric perspective**: the primary response to an environmental influence is the activation of a cell signalling pathway that regulates TFs. The TFs then select specific loci in the genome for altered function, leading to the
>
> **epigenomic** and transcriptional changes that we measure, which can be seen as a *secondary response*.

This model helps explain the sequence specificity of environmental responses and reveals upstream cell signalling pathways as potential targets for therapeutic intervention.

Conclusions and Future Directions

The study of the

epigenome's response to environmental exposures is more complex than initially thought, but also more promising. Rather than viewing influences like DNA sequence variation and cell subtype changes as errors, we should embrace them as sources of insight into disease mechanisms.

Future work should:

- Leverage advances in iPSC differentiation to create cell types and organoids for studying molecular responses to in vitro exposures in controlled systems.
- Use
  epigenomic approaches to infer upstream cell signalling regulators and TFs as therapeutic targets.
- Employ deep whole genome sequencing to identify the ultra-rare variants that may mediate a significant proportion of interindividual differences in environmental responses.
- Focus on associating the polymorphism of *molecular genomic phenotypes* (like chromatin accessibility or gene expression) with disease first, and then link these back to the multiple rare DNA variants that may cause them. This "collapses" many rare variants into a common functional outcome, increasing statistical power.
- Broaden pluripotent stem cell resources to represent more diverse racial and ethnic groups.

Sources