Neuroepigenetics 1 (2015) 14–19

Contents lists available at ScienceDirect

Neuroepigenetics

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

New insights and updated guidelines for epigenome-wide association studies☆

Lisa H. Chadwick [a], Akira Sawa [b], Ivana V. Yang [c], Andrea Baccarelli [d], Xandra O. Breakefield [e], Hong-Wen Deng [g], Dana C. Dolinoy [f], M. Daniele Fallin [h], Nina T. Holland [i], E. Andres Houseman [j], Stavros Lomvardas [k], Mahendra Rao [l], John S. Satterlee [m], Frederick L. Tyson [n], Pandurangan Vijayanand [o], John M. Greally [p,*]

[a] Division of Extramural Research and Training, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA
[b] Department of Psychiatry, Johns Hopkins University, Baltimore, MD, USA
[c] Department of Medicine, University of Colorado Denver, Aurora, CO, USA
[d] Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA
[e] Department of Neurology, Massachusetts General Hospital, Boston, MA, USA
[f] Department of Environmental Health Sciences, University of Michigan, Ann Arbor, MI, USA
[g] Department of Biostatistics and Bioinformatics, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA
[h] Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[i] Division of Environmental Health Sciences, School of Public Health, University of California Berkeley, Berkeley, CA, USA
[j] Department of Biostatistics, College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, USA
[k] Department of Anatomy, University of California San Francisco, San Francisco, CA, USA
[l] Laboratory of Stem Cell Biology, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD, USA
[m] Division of Basic Neuroscience and Behavioral Research, National Institute on Drug Abuse, National Institutes of Health, Bethesda, MD, USA
[n] Division of Intramural Research, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA
[o] La Jolla Institute for Allergy and Immunology, La Jolla, CA, USA
[p] Center for Epigenomics, Albert Einstein College of Medicine, Bronx, NY, USA

ARTICLE INFO

Article history:
Received 13 October 2014
Accepted 14 October 2014

Keywords:
Epigenetic
DNA methylation
Disease
Association
Genome-wide

ABSTRACT

Epigenetic dysregulation in disease is increasingly studied as a potential mediator of pathophysiology. The epigenetic events are believed to occur in somatic cells, but the limited changes of DNA methylation in studies to date indicate that only subsets of the cells tested undergo epigenetic dysregulation. The recognition of this subpopulation effect indicates the need for care in design and execution of epigenome-wide association studies (EWASs), paying particular attention to confounding sources of variability. To maximize the sensitivity of the EWASs, ideally, the cell type mediating the disease should be tested, which is not always practical or ethical in human subjects. The value of using accessible cells as surrogates for the target, disease-mediating cell type has not been rigorously tested to date. In this review, participants in a workshop convened by the National Institutes of Health update EWAS design and execution guidelines to reflect new insights in the field.

Epigenetic regulatory mechanisms play a crucial role in normal human development, in part by establishing and maintaining the gene expression programs necessary for cells to perform their unique functional roles. In recent years, there has been growing interest in investigating whether changes in these epigenetic programs contribute to the development of a variety of complex human diseases. Several lines of evidence suggest that this might be the case. These include the potential for the epigenome to mediate environmental influences (reviewed in Cortessis et al., 2012; Hou et al., 2012; Perera & Herbstman, 2011; Reamon-Buettner et al., 2008) or the maintenance of memory of events that occurred in the past, including prenatal exposures influencing adult disease susceptibility (reviewed in Barouki et al., 2012; Gluckman et al., 2011; Warner & Ozanne, 2010; Waterland & Michels, 2007). The field of cancer epigenomics has established a precedent for the silencing or activation of genes being causally involved in neoplasia (Dawson & Kouzarides, 2012; Esteller, 2007), somatic events that are usually limited to the cells or tissue in which the cancer arose, with a few notable exceptions (Cui et al., 2003; Gaston et al., 2001). Unlike the genome, the

☆ Workshop participant Dr. Arturas Petronis from the Centre for Addiction and Mental Health at the University of Toronto, Canada, preferred not to co-author this document.

* Corresponding author. Tel.: +1 17186781234; fax: +1 7184301016.
E-mail address: john.greally@einstein.yu.edu (J.M. Greally).

http://dx.doi.org/10.1016/j.nepig.2014.10.004

epigenome is inherently malleable from a biochemical perspective, and the potential to reverse deleterious epigenetic events has been seen as a major opportunity, especially in the treatment of cancer (Griffiths & Gore, 2013; Popovic & Licht, 2012; Rius & Lyko, 2012).

The field of epigenome-wide association studies (EWASs) is now very active, testing a wide variety of human diseases and other phenotypes. These EWASs have almost all been based on the study of DNA methylation, an epigenetic regulator that is less demanding in terms of clinical sample acquisition than alternatives such as sequencing of DNA from chromatin immunoprecipitation, which maps histone modifications or other chromatin constituents. A consistent outcome of the EWAS studies to date has been the observation of moderate changes in DNA methylation between disease and normal groups, and not a switch between fully unmethylated and methylated states at a given genomic locus. As DNA methylation exists in a binary state at a specific location on an individual allele, moderate changes have to reflect allelic and cell subpopulation changes between the tested groups, an epigenetic mosaicism that may be of pathophysiological significance if subsets of cells can reasonably be proposed to mediate the organ's disease state.

This emerging observation of cell subpopulation effects has forced a re-evaluation of how we should approach EWAS design and execution. With limited effect sizes associated with the disease or phenotype, it is necessary to pay increased attention to other sources of variability potentially affecting the study. It is also essential that we invest our efforts in a cell type that is likely to manifest the differences sought. Epigenomic dysregulation associated with human disease is generally thought of in terms of somatic rather than germline events, raising the issue whether the cell type(s) mediating the disease (target cell type) needs to be sampled, or whether a more accessible surrogate cell type can yield sufficiently useful information.

The influence of cell type on epigenetic variability is highlighted by large-scale epigenomic mapping efforts such as that led by the National Institutes of Health (NIH) Roadmap Epigenomics Program (Bernstein et al., 2010). There are other influences on epigenome-wide studies to consider, such as technical influences, DNA sequence polymorphism, and human subject characteristics (age, sex, exposure history), all potentially confounding the ability to recognize genuine effects associated with a disease or phenotype and leading, in some cases, to misinterpretation of the results obtained.

In this review, we describe the broader issue of rigorous EWAS design and execution, updating prior excellent EWAS overviews (Bell & Spector, 2011; Mill & Heijmans, 2013; Rakyan et al., 2011; Satterlee et al., 2010; Verma, 2012), with an emphasis on the specific issue of the value of studies based on surrogate cell use, and define 3 areas of research priority that could further improve our design and interpretation of EWASs.

Updated guidelines for rigorous EWASs

Taking a cue from the history of genome-wide association studies (GWASs), it is essential that rigorous standards are developed for EWASs, which are likely to be more complex than GWASs, involving many different types of epigenetic regulatory mechanisms, cell types, and likely confounding influences. As a starting point, we provide an overview of updated suggested best practices for EWASs, described below and in Fig. 1, building upon prior recommendations for how best to design and perform such studies (Bell & Spector, 2011; Rakyan et al., 2011; Satterlee et al., 2010; Verma, 2012).

Begin with an explicit biological hypothesis: Linking epigenetic changes to disease or phenotypic causation, or as a marker or mediator of environmental exposure, should be stated in terms of an explicit hypothesis. Although there is clearly value to exploratory or pilot studies to test whether a certain phenotype or condition has any evidence for nonrandomly associated epigenetic changes, when performing a definitive study, it is essential to define the underlying hypothetical mechanism involving epigenetic perturbation. Having such a clear hypothesis allows the experimental design and, in particular, the analysis and interpretation to be focused productively. For example, the timing of sample collection or the choice of cross-sectional versus longitudinal studies and the rationale for choosing a specific surrogate cell type for study will be dictated to a major extent by the underlying hypothesis.

Purified cell types are preferable for epigenetic studies: It is highly desirable that single cell types are used where possible. Although it has been appreciated for some time that different cell types have distinctive epigenetic profiles (Shen et al., 2012; Varley et al., 2013), it has recently been confirmed that the presence of different proportions of cell subtypes in mixed populations of cells can generate distinctive DNA methylation profiles (Houseman et al., 2012), which has been found for some loci to explain up to 40% of DNA methylation differences between individuals (Adalsteinsson et al., 2012). If there is a systematic bias in cell subpopulation composition between the groups of individuals being tested (e.g., between disease states, exposures, or phenotypes), DNA methylation assays performed will identify differences between groups, but these will not necessarily represent altered epigenetic patterns within each cell type associated with the disease. It is therefore preferable that pure cell samples be tested when feasible, often a difficult issue when cell numbers from purified cell samples yield less material than can usually be tested in genome-wide assays. It should be recognized that "pure" does not mean homogeneous in terms of function, so purification does not eliminate the possibility of cell subtype composition exerting an influence. However, the selection of purified cells ensures that as similar cell types as possible are compared between groups and makes it easier for other groups to reproduce the experiment, both valuable considerations.

If it is not possible to purify cells, it should be attempted to account for subpopulation effects in statistical models, either through analysis of the samples collected (for example, histological studies to quantify cell proportions or measurement of differential white blood cell count in leukocyte-based studies) or through the development of new analytical techniques that use DNA methylation signatures to measure cell subpopulation proportions (Houseman et al., 2012). Examples of this kind of approach were published recently (Guintivano et al., 2013; Liu et al., 2013), demonstrating a major reduction in association signals after adjustment for the estimated cell proportions in the blood samples tested, emphasizing both the potential for variability in cell subpopulations to exert strong effects on the DNA methylation signals as well as the potential for sophisticated analytical approaches to account for these effects.

Target versus surrogate cell types: An underlying assumption in many EWASs is that epigenetic changes associated with a particular disease are likely acquired in somatic cells during development or during aging (as opposed to through the germline or extremely early in development). It follows that these epigenetic changes may not be observed in all cell types in the body. The choice of cell type is thus of great importance in human disease studies. The ideal situation would be to acquire those cells directly affected by or mediating the disease, cells we refer to as the *target* cell type, purified to the greatest extent possible. For many diseases, obtaining such samples can be very challenging. As an example, target cell types in disorders of the central nervous system generally can only be studied in postmortem specimens.