Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Psychology

Hidden hypotheses in 'hypothesis-free' genome-wide epigenetic associations

Edward D Barker[1], Susanna Roberts[1] and Esther Walton[2]

The recent interest in epigenetics within mental health research, from a developmental perspective, stems from the potential of DNA methylation to index both exposure to adversity and vulnerability for mental health problems. Genome-wide technology has facilitated epigenome-wide association studies (EWAS), permitting 'hypothesis-free' examinations in relation to adversity and/or mental health problems. In EWAS, rather than focusing on *a priori* established candidate genes, the genome is screened for DNA methylation, thereby enabling a more comprehensive representation of variation associated with complex disease. Despite their 'hypothesis-free' label, however, results of EWAS are in fact conditional on several *a priori* hypotheses, dictated by the design of EWAS platforms as well as assumptions regarding the relevance of the biological tissue for mental health phenotypes. In this short report, we review three hidden hypotheses — and provide recommendations — that combined will be useful in designing and interpreting EWAS projects.

Addresses
[1] Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK
[2] Medical Research Council Integrative Epidemiology Unit and Bristol Medical School, Population Health Sciences, University of Bristol, UK

Corresponding author: Barker, Edward D (ted.barker@kcl.ac.uk)

Current Opinion in Psychology 2019, **27**:13–17

This review comes from a themed issue on **Genetics**

Edited by **Brian Boutwell** and **Michael A White**

https://doi.org/10.1016/j.copsyc.2018.07.009

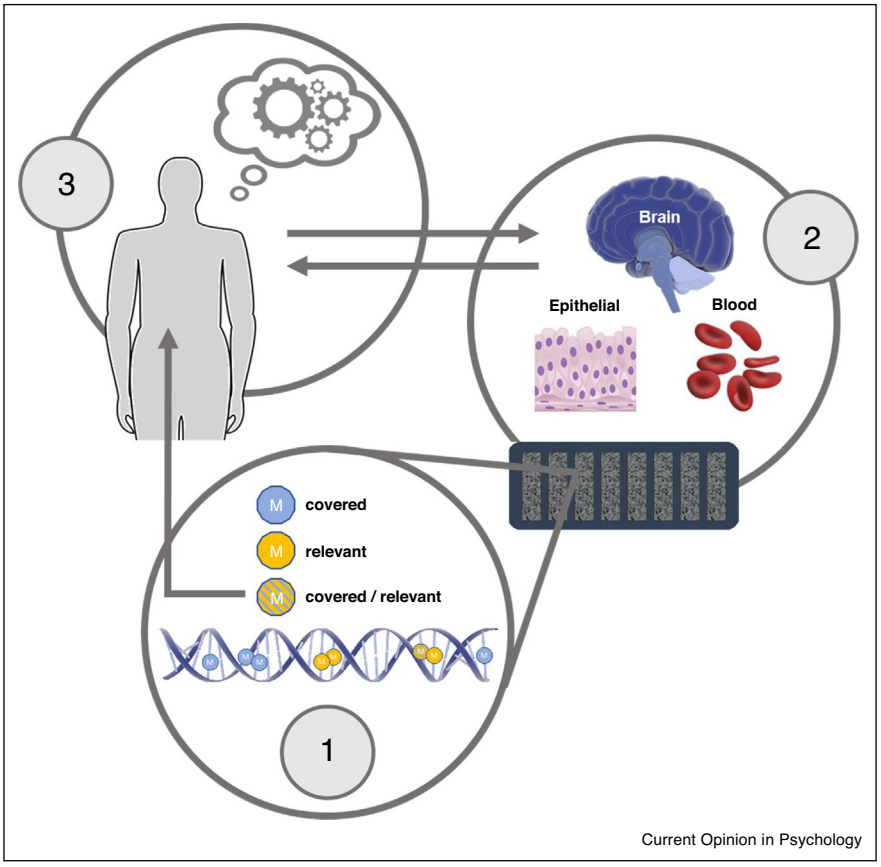
Understanding the biological mechanisms by which early psychosocial adversity associates with long-term mental health problems may have the potential to facilitate the development of effective screening, intervention strategies and health policy decisions [1]. Recent research has focused on the degree to which adversity disrupt gene regulation through epigenetic processes, thereby providing a mechanism by which the environment can have lasting effects on measurable mental health phenotypes [2••]. High profile studies suggest that epigenetic changes associated with early adversities [3,4] and even lifestyle choices [5•] can be observed across the life span, and that these long-term epigenetic modifications are associated with risk for a range of health outcomes [6]. These studies have generally focused on DNA methylation (DNAm) for two reasons: it is currently the best understood epigenetic mechanism and array-based technologies are readily available, which provides coverage of hundreds of thousands of methylation sites across the genome [7]. This combination of basic science and genome-wide technology has facilitated numerous epigenome-wide association studies (EWAS), permitting 'hypothesis-free' examinations in relation to adversity and/or mental health problems.

The logic underlying EWAS is comparable to genome-wide association studies (GWAS [8•]). Rather than focusing on DNAm in proximity to candidate genes, the genome is screened for DNAm, thus enabling a more comprehensive representation of variation associated with complex disease. As with GWAS (e.g. [9,10]), despite their 'hypothesis-free' label, results of EWAS are in fact conditional on several *a priori* hypotheses, dictated by the design of EWAS platforms as well as assumptions regarding the relevance of the biological tissue for the mental health phenotypes under investigation. In this short report, we review three hidden hypotheses (see Figure 1) — and provide recommendations — that combined will be useful in designing and interpreting EWAS projects.

Hidden hypothesis 1: EWAS coverage is sufficient for complex psychiatric problems

Array-based platforms have become widespread in psychology research, largely due to their ease of use, relatively high through-put, and well standardised and validated pipelines for processing, quality control, and analysis techniques. In particular, the Illumina 450k and EPIC arrays feature 480 000–850 000 probes targeting nearly 99% of RefSeq genes, as well as a range of other genomic categories, such as CpG islands, shores and shelves, miRNA promoters and enhancers, where DNAm can be influenced by and/or impact transcription in distal genomic regions [11••]. Compared with the Ilumina 450k, the newer Illumina EPIC 850k array provides much greater coverage of ENCODE and FANTOM5 enhancers [12••], and shows higher genetic influence underlying DNAm probes [13]. Nevertheless, these microarrays are limited in the number of sites they can assess, and thus lack true genome-wide measurements [14].

Figure 1



Hidden hypotheses in epigenome-wide approaches. *Note*: (1) = Hypothesis 1: EWAS coverage is sufficient for complex psychiatric problems; (2) = Hypothesis 2: peripheral tissue is meaningful for mental health problem(s); and (3) = Hypothesis 3: biology can be meaningful to phenotype of interest.

Furthermore, during the design process of the 450k and EPIC arrays, CpG sites were chosen as potentially biologically informative based on consultation with a consortium of DNA methylation experts [15]. Whilst the coverage of genes and CpG islands on these microarrays are comprehensive, it does not represent a complete picture of methylated cytosines across the genome. Selection was, in part, based on data from a number of phenotypes (some medical in nature such as cancer), and thus is not specifically targeted to brain-based, stress-related complex mental health phenotypes. This is an important point: if a sizeable proportion of the CpG sites tested are not relevant to the phenotype of interest, the likelihood of detecting relevant results is reduced.

Hidden hypothesis 2: peripheral tissue is meaningful for mental health problem(s)

The second hidden hypothesis relates to the tissue that is used to quantify DNAm. The majority of mental health research is based on DNAm profiles obtained from peripheral tissues from living persons, such as blood and saliva. When investigating outcomes such as conduct disorder or depression, however, the brain is often the main tissue of interest when it comes to mechanistic interpretations of results [16••]. To this end, research suggests that the correspondence of methylation profiles from blood and saliva to the brain is in fact quite limited, but can be higher with cross-tissue genetic influence [13,17]. This presents a critical disadvantage if the investigator would like to use the peripheral tissue as a surrogate of the central nervous system (CNS; the brain).

One promising avenue is to establish DNAm as a biomarker for mental illness. A biomarker does not have to be mechanistic (i.e. CNS surrogate). Indeed, blood-based biomarkers have been used for diagnostics, predictive risk, disease monitoring and/or treatment response in cancer, cardiovascular and infectious disease [18,19]. However, even within a biomarker framework, the assumption is often that distinct peripheral tissues are interchangeable and equally suited for biomarker detection, when in fact it is highly probable that peripheral tissues themselves correspond differently to environmental adversity and/or disease state [14]. For instance,