Ng et al. Genome Biology 2012, 13:246
http://genomebiology.com/2012/13/6/246

Genome Biology

REVIEW

The role of longitudinal cohort studies in epigenetic epidemiology: challenges and opportunities

Jane WY Ng[1,2], Laura M Barrett[1], Andrew Wong[3], Diana Kuh[1], George Davey Smith[4] and Caroline L Relton[1]*

Abstract

Longitudinal cohort studies are ideal for investigating how epigenetic patterns change over time and relate to changing exposure patterns and the development of disease. We highlight the challenges and opportunities in this approach.

Keywords Epigenomics, DNA methylation, life course, longitudinal studies

Introduction

Interest in the role of epigenetic processes in common complex diseases continues to increase [1,2]. Epigenetics is a potentially major mechanism by which environmental factors can affect physiological function and disease risk. Research into epigenetics promises to reveal many of the causes that remain undiscovered after extensive investigation of common genetic variation [3].

Epidemiological approaches can be used to identify whether epigenetic processes are involved in mediating the association between risk factors (environmental, genetic, lifestyle, socioeconomic and so on) and common complex disease [4,5]. For example, longitudinal cohort studies have been a cornerstone of observational epidemiology for many years. Long-term follow-up of adult cohorts has identified important risk factors for cardiovascular disease, chronic bronchitis, and cancers, and follow-up of cohorts from birth or childhood has been equally successful at identifying the importance of early exposures (especially the childhood social environment) and developmental characteristics for adult health (for example, [6-10]). Longitudinal studies, particularly those that start in early life, can contribute to our understanding of how the epigenome changes over time, as a result of varying environmental exposures, and how disease phenotypes evolve. Longitudinal studies are costly to instigate and maintain, and cross-sectional studies (a less expensive alternative study design) have more often been used to assess the relationship between exposures and the epigenome and/or the epigenome and disease. However, cross-sectional studies cannot capture the dynamic nature of epigenetic mechanisms [11], making it difficult to identify the influences of the environment and/or disease state (or sub-clinical features of disease) on the epigenome and thus establish the direction of causality. As a result of this, study designs that make use of multiple time points are being increasingly recognized as the most suitable to analyze the epigenetics of common complex diseases. Because longitudinal studies track the same cohort at multiple time points throughout their lifetime, enabling the temporal relationship between exposure and disease to be established, they are ideally placed for exploitation in epigenetic investigations.

Advances in genomic technologies have opened up the possibility of large-scale population-based assessment of epigenetic patterns to help understand their influence on disease. How should such studies be conducted to maximize their impact and what can epigenetics researchers learn from previous approaches to population-based studies? Here we focus on how epidemiological approaches, including the design of cohort studies, can help investigate the role of epigenetic variation in common complex disease. Furthermore, the dynamic nature of epigenetic patterns means that they can be altered by disease-related factors (a process called 'reverse causation') as well as a host of confounding factors (such as age, sex, socioeconomic position, diet, or smoking). Many relevant approaches have been developed in the context of both genetic and life course epidemiology that could be fruitfully applied to epigenetics; examples are methods for dealing with biases, confounding, and reverse causation and also longitudinal statistical modeling techniques [12,13]. We first assess what epigenetic markers have been measured within existing life course studies before

*Correspondence: caroline.relton@ncl.ac.uk
[1]Institute of Genetic Medicine, Newcastle University, NE1 3BZ, UK
Full list of author information is available at the end of the article

discussing how the epidemiologist's toolkit can be applied to epigenomics.

Epigenetic studies within longitudinal cohorts

Since 2010, 34 life course studies have included measurements of DNA methylation, and just four of these have included analysis of epigenetic features at more than one time point (Table 1). In line with the vast majority of other epigenetic studies, the focus is on DNA methylation as this is the most straightforward form of epigenetic modification to measure, and the only currently feasible option in archived DNA samples. Prospective sample collection will permit the analysis of chromatin modifications and microRNA. Three of the studies analyzing more than one time point (Table 1) report findings relating specifically to age-related changes in childhood [14] or adulthood [15,16], and all three focus on gene-specific DNA methylation of a small panel of (different) loci and report differences that were modest in size (generally <5%). A further study considers changes in DNA methylation over a relatively short time period (28 to 180 days) in relation to air pollution exposure [17]. Although there was some indication of lower global DNA methylation in repetitive elements across the genome in this study [17] at 90 days of exposure, there was no evidence of a dose response, casting doubt on the biological importance of this association. In summary, very little has been done in this area.

Table 2 summarizes additional examples in which case-control studies of DNA methylation have been nested within existing large-scale longitudinal cohorts; this approach has been applied so far exclusively in the context of cancer. Analyses in this instance have been limited to gene panels (generally established tumor suppressor or oncogenes) and have been undertaken either (i) to assess the utility of epigenetic signatures as early biomarkers of cancer risk [18-20] or (ii) to consider the determinants of a perturbed methylation state (methylator phenotype), which has been implicated in numerous cancers [21-25]. With improved knowledge of methylation variable regions associated with diseases other than cancer (for example, cardiovascular disease, dementia, and rheumatoid arthritis), the same approach could be adopted in the context of longitudinal cohort studies.

The paucity of DNA methylation measurements undertaken in cohorts that have collected serial samples from the same individuals is clear, indicating that the potential richness of longitudinal data and sampling in these studies has yet to be fully exploited. Few studies have routinely collected serial samples from the same individuals at multiple points in the life course (for example, the Avon Longitudinal study of Parents and Children (ALSPAC) [26,27], and the Normative Aging Study [17,28-32]), but others are planning serial sampling in light of the interest in epigenetics (such as the Medical Research Council National Survey of Health and Development [33] and the Southall And Brent REvisited (SABRE) cohort [34]). Given the temporal variation in epigenetic patterns, serial sampling of any longitudinal cohort would be advised where possible.

Of the studies published so far, the variety of tissues analyzed is limited mainly to easily accessible peripheral blood, cord blood or buccal cells, the studies are modest in size compared with those used for genetic research, and the range of different methods that have been used to quantify DNA methylation have led to an overall lack of comparability between studies. It is clear from these observations that more can be done with respect to the collection and analysis of biological samples from longitudinal cohorts so that they are optimal for epigenetic studies.

Attributes of longitudinal cohort studies

Ideally, longitudinal epigenetic studies should include extensive, prospectively collected data and biological samples at multiple time points across the life course. Many existing longitudinal cohort studies are population-based, although some focus on a specific sub-group of the general population. For example, the SABRE cohort focuses on groups that are first or second generation migrants to the UK of non-European ethnicity to examine particular health issues, in this case the marked discordance in disease risk observed in migrant groups compared with Europeans living in the UK [34]. Longitudinal epigenetic studies can add value to existing resources, such as data from genome-wide association studies - for example, ALSPAC [26,27] and the Relationship between Insulin Sensitivity and Cardiovascular disease (RISC) cohort [35]. Exposures commonly captured in longitudinal studies include lifestyle factors, such as smoking, alcohol intake, diet, and physical activity patterns, and also socioeconomic measures across the life course. Common phenotypes on which longitudinal studies tend to focus include physical and anthropometric measures, cognitive, cardiovascular, metabolic, respiratory, and musculoskeletal function, and a range of blood-based intermediate biomarkers. Of particular value are birth cohorts with trans-generational and across-life samples from birth onwards, allowing an appraisal of epigenetic changes associated with *in utero* and early life exposures, a period when the epigenome is believed to be particularly plastic.

The epidemiological toolkit

Applying principles of life course epidemiology to epigenetic research

Research in life course epidemiology investigates developmental, aging, and risk factor trajectories and how

Table 1. Epigenetic studies in longitudinal cohorts: a summary of recent literature (2010 to 2012)

Cohort	DNA analysis time points	Tissue	Form of DNA methylation analysis	Loci	Exposure (if applicable)	Outcome (if applicable)	n	References
Avon Longitudinal Study of Parents and Children	1	Cord blood	Proxy genotype	*TACSTD2*	Postnatal growth	Childhood adiposity	6,990	[53]
Avon Longitudinal Study of Parents and Children	1	Cord blood	Illumina GoldenGate Cancer Panel	1,576 CpG sites		Childhood body composition	178	[56]
1958 British Birth Cohort Study	1	Peripheral blood	MeDIP-chip	Genome-wide methylation	Socioeconomic position		40	[92]
Columbia Children's Center for Environmental Health Northern Manhattan Mothers and Newborns Study	1	Cord blood	Methylamp	Global methylation	Polycyclic aromatic hydrocarbons, benzo[*a*] pyrene-DNA adducts		164	[93]
Detroit Neighborhood Health Study	1	Peripheral blood	Illumina HM27 BeadChip	*SLC6A4*	Traumatic events	Post-traumatic stress disorder	100	[94]
Detroit Neighborhood Health Study	1	Peripheral blood	Illumina HM27 BeadChip	Genome-wide methylation	Traumatic events	Post-traumatic stress disorder	100	[95]
Detroit Neighborhood Health Study	1	Peripheral blood	Illumina HM27 BeadChip	Genome-wide methylation		Depression	100	[96]
Detroit Neighborhood Health Study	1	Peripheral blood	Illumina HM27 BeadChip	Genome-wide methylation	Post-traumatic stress disorder		100	[97]
Dutch Famine Cohort	1	Peripheral blood	Pyrosequencing MethyLight	*Sat2*, LINE-1, *LUMA*	Pre-natal famine		947	[98]
Environmental Risk (E-Risk) Longitudinal Twin Study	2	Buccal cells	Sequenom MassArray	*DRD4*, *SERT*, *MAOA*	Change over time		182	[14]
Epigenetic Birth Cohort	1	Cord blood and placenta	Pyrosequencing	LINE-1	Gestational age and birth weight		319 mother-child dyads	[99]
Longitudinal Study of Adolescent Health	1	Buccal cells	Sequenom MassArray	*5HTT*		Depression	150	[100]
Longitudinal Study of Child Development	1	Buccal cells	Microarray	Genome-wide methylation	Childhood adversity		109	[101]
Lovelace Smokers' Cohort	1	Sputum	Methylation specific PCR	Lung cancer genes	Wood smoke exposure	Chronic obstructive pulmonary disease	1,827	[102]
Netherlands Twin Registry	2	Peripheral blood and buccal cells	Sequenom MassArray	*IL10*, *NR3C1*, *TNF*, *IGF2R*, *GRB10*, *LEP*, *CRH*, *ABCA1*, *IGF2*, *INSIGGF*, *KCNQ1OT1*, *MEG3*, *APOC1*, *GNASAS*, *GNAS A/B*	Cell counts, change over time		64	[15]
Newcastle Preterm Birth Growth Study	1	Peripheral blood	Pyrosequencing	*TACSTD2*	Postnatal growth	Childhood adiposity	121	[53]
New York Women's Birth Cohort	1	Peripheral blood	Pyrosequencing	*Sat2*, Alu, LINE-1	Prenatal tobacco smoke		90	[103]

Continued overleaf

Table 1. Continued

Cohort	DNA analysis time points	Tissue	Form of DNA methylation analysis	Loci	Exposure (if applicable)	Outcome (if applicable)	n	References
Normative Aging Study	1	Peripheral blood	Pyrosequencing	*CRAT*, *F3*, *GCR*, *ICAM*, *IFNγ*, *IL6*, *iNOS*, *OGG1*, *TLR2*		Lung function	756	[28]
Normative Aging Study	1	Peripheral blood	Pyrosequencing	Alu, LINE-1, *F3*, *TLR2*, *ICAM-1*	Air pollution	Fibrinogen, ICAM-1, VCAM-1, and CRP	704	[29]
Normative Aging Study	1-5	Peripheral blood	Pyrosequencing	*CRAT*, *F3*, *GCR*, *ICAM*, *IFNγ*, *IL6*, *iNOS*, *OGG1*, *TLR2*	Age		784	[16]
Normative Aging Study	1	Peripheral blood	Pyrosequencing	Alu, LINE-1		Cancer	722	[30]
Normative Aging Study	1-3	Peripheral blood	Pyrosequencing	Alu, LINE-1	Air pollution		706	[17]
Normative Aging Study	1	Peripheral blood	Pyrosequencing	LINE-1		Ischemic heart disease and stroke	712	[31]
Normative Aging Study	1	Peripheral blood	Pyrosequencing	LINE-1		Inflammatory markers VCAM-1, ICAM-1 and CRP	593	[32]
North Cumbria Community Genetics Project	1	Cord blood and maternal peripheral blood	Pyrosequencing	*IGF2*, *IGFBP3*, *ZNT5*, *LUMA*	Maternal characteristics, folate metabolism and genotype		430	[104]
Project Viva	1	Cord blood and maternal peripheral blood	Pyrosequencing	LINE-1	Methyl donor nutrients		516 infants and 830 mothers	[105]
Bangladesh Birth Cohort	1	Cord blood and maternal peripheral blood	Pyrosequencing	Alu,LINE-1, *p53*, *p16*	Arsenic exposure		120 mother-child pairs	[106]
Bangladesh Birth Cohort	1	Cord blood and maternal peripheral blood	Pyrosequencing	Alu, LINE-1	Arsenic exposure		114 mother-child pairs	[107]
Psychological, Social and Behavioral Determinants of Ill Health	1	Peripheral blood	Methylamp	Global methylation	Socioeconomic position		239	[108]
Rhode Island Child Health Study	1	Placenta	Pyrosequencing	*HSD11B2*	Maternal characteristics	Neurobehavioral outcomes	185	[109]
Singapore Chinese Health Study	1	Peripheral blood	MethyLight	Alu, *Sat2*	B vitamins	Cardiovascular disease	286	[110]
Southampton Women's Study	1	Umbilical cord	Sequenom MassArray	*eNOS*		Bone mineral content	66	[111]
Southampton Women's Study	1	Umbilical cord	Sequenom MassArray	*RXRA*, *eNOS*, *SOD1*, *PIK3CD*, *IL-8*		Childhood adiposity	66 + 239	[57]
Five studies	1	Peripheral blood	Pyrosequencing	Alu, LINE-1	Age, gender, smoking, alcohol, body mass index		1,465	[67]

Search strategy: a literature search retrieved >350 publications published between 1 January 2010 and 13 April 2012. Only established longitudinal cohort studies were included in the summary table (in as far as this could be ascertained from the information available). Of more than 350 publications retrieved for review only 34 included analyses of DNA methylation in established longitudinal cohort studies and of these only four included analyses at more than one time point. The remaining 30 studies measured DNA methylation at a single time point and considered this with respect to longitudinal data relating to exposures or outcomes.

Table 2. Nested case-control epigenetic studies in longitudinal cohorts: a summary of recent literature (2010 to 2012)

Cohort	DNA analysis time points	Tissue	Form of DNA methylation analysis	Loci	Exposure (if applicable)	Outcome (if applicable)	n	References
EPIC & Breakthrough Generations Study & KConFab	1	Peripheral blood	Pyrosequencing	*ATM*, LINE-1		Breast cancer	1,381	[112]
EPIC-Lung	1	Peripheral blood	Pyrosequencing	*CDKN2A RASSF1A*, *GSTP1*, *MTHFR*, *MGMT*	B vitamins, smoking	Lung cancer	93	[18]
EPIC-Norfolk	1	Tumor tissue	Pyrosequencing	*MLH1*		Colorectal cancer	185	[21]
EPIC-EURGAST	1	Tumor tissue	Pyrosequencing	*CHRNA3*, *DOK1*, *MGMT*, *RASSF1A*, *p14ARF*, *CDH1*, *MLH1*, *ALDH2*, *GNMT*, *MTHFR*		Gastric cancer	162	[22]
Iowa Women's Health Study	1	Tumor tissue	MethyLight	CpG island methylator phenotype	Smoking	Colorectal cancer	555	[23]
Netherlands Cohort Study	1	Tumor tissue	Methylation specific PCR	*CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3*, *SOCS1*	Body size and physical activity	Colorectal cancer	734	[113]
New York University Women's Health Study	1	Peripheral blood	Methylation specific PCR	*RASSF1A*, *GSTP1*, *APC*, *RARβ2*		Breast cancer	200	[19]
Northern Sweden Health and Disease Study	1	Tumor tissue	MethyLight	CpG island methylator phenotype	B vitamins	Colorectal cancer	570	[24]
Nurse's Health Study	1	Tumor tissue	MethyLight	*CACNA1G*, *CDKN2A*, *CRABP1*, *IGF2*, *MLH1*, *NEUROG1*, *RUNX3*, *SOCS1CHFR*, *HIC1*, *IGFBP3*, *MGMT*, *MINT-1*, *MINT-31*, *p14*, *WRN*	B vitamins and alcohol	Colorectal cancer	761	[25]
Shanghai Women's Health Study	1	Peripheral blood	Pyrosequencing	Alu, LINE-1		Gastric cancer	576	[20]

Search strategy: a literature search retrieved >350 publications published between 1 January 2010 and 13 April 2012. Nested case-control studies within established longitudinal cohort studies are included. There are many further examples in which DNA methylation has been analyzed in a case-control study design, cross-sectional study or randomized controlled trial that includes some element of data collection over time (often retrospective), but these have not been included in the tables presented.

dynamic relationships unfold over time, and takes into account potential confounding, mediating, or interactive effects of lifetime biological, psychological, and social risk factors [36]. This conceptual framework is relevant for epigeneticists investigating long-term associations that may be biased, confounded or due to reverse causation. Life course epidemiologists have investigated various different methods for modeling risk factor trajectories (particularly growth trajectories) in relation to later health outcomes and have developed a novel structured approach [37] to distinguish critical, sensitive, and accumulation life course models [38]. They use a range of approaches for modeling repeat continuous and binary outcome measures, such generalized estimating equations or mixed models that consider correlated data such as repeat measures from the same individuals over time, and for modeling time to an event, such as survival and event history analysis. This toolkit is relevant to epigeneticists, whether studying lifetime environmental exposures that promote particular epigenetic signatures over time or how these signatures themselves may affect not just the level (intercept) of function (such as blood pressure) at a point in time but also its rate of change (slope) over time. Such statistical approaches have not been widely applied to epigenetic data, although examples can be found in Madrigano *et al*. [16,17], who illustrate the use of mixed models to analyze changes in methylation over time while accounting for the correlation among measurements within the same individual. Further discussion of this subject is provided below in the section on data analysis considerations.

Several research collaborations involving cohort studies, such as HALCyon (Healthy Aging across the Life Course) [39], FALCon (Function Across the Life Course) [40] and GEoCoDE (Genomic and Epigenomic Complex Disease Epidemiology) [41] have been formed. These have increased the sample size and power to investigate lifetime risk factors on longitudinal phenotypes and to test whether findings are replicated across cohorts in a systematic way, and they will be useful to epigenetics research. The collaborations have developed experience in data harmonization to derive comparable phenotypes across the cohorts, and in cross-cohort methods (for example, [42]). Those running epigenetic studies may want to make use of these collaborations for similar reasons, and a coordinated approach is likely to advance the science and be appealing to funders. Coordinating the cohorts has led to more effective ways of gaining knowledge of the various datasets and metadata as well as facilitating data sharing and encouraging good practice in data management.

From genetic to epigenetic epidemiology

Incorporating epigenetic measures into epidemiological studies is often done in the context of genetic epidemiology resources. However, studying epigenetic factors - which are, partly at least, phenotypic - is more similar to conventional epidemiology than it is to genetic epidemiology. Several aspects of germline genetic variation lead to special-case conditions that allow relaxation of usual epidemiological principles: reverse causation (disease influencing the variable being measured rather than *vice versa*) is clearly not an issue in genetic epidemiology, and confounding - which often vitiates conventional epidemiology - generally relates only to ancestry in genetic epidemiology [43], and this can be accounted for by using principal components from genome-wide data as control variables. Germline genetic variation can be assessed on samples taken at any stage of life, does not change over time, and can be assayed with high precision and low measurement error. Effect sizes for the influence of common genetic variants on common complex diseases tend to be small, which means that very large sample sizes are required. Given these circumstances, the genetic epidemiology study design of choice became large case-control studies, with the controls not being carefully selected to represent the source population - and sometimes (as in the case of the landmark Wellcome Trust Case Control Consortium (WTCCC) [44]) control groups shared for comparison with several disease groups. For example, in the WTCCC the common control groups consisted of blood donors (who are very unrepresentative in terms of factors that would be important confounders in conventional epidemiological studies, such as health-related behaviors and social class) and participants in the 1958 birth cohort - all of the same age, which in some cases barely overlapped with the age of the cases.

However, such study designs are not appropriate for epigenetic epidemiology, as confounding, bias, and reverse causation are all serious problems when studying phenotypic exposures. It is important that the successes of genetic epidemiology are not translated into failures for epigenetic epidemiology [1,5,45]. Prospective studies are the ideal type of study, including documented exposure (epigenetic) measures collected before the outcomes and temporal changes, detailed assessment of confounding factors, and consideration of measurement error. Currently, the effect sizes of associations in epigenetic studies are poorly delineated, but it is likely that, unlike the situation in the early days of molecular genetic epidemiology, the problem will not be one of relatively few robust associations, but rather many real observational associations will exist and the issue will be the separation of causal associations from those generated by confounding and bias. Various methods that have been developed to strengthen causality in conventional epidemiology - including collaborative analysis of multiple cohorts in which confounding structures differ [46], comparisons of plausible and implausible associations

[47,48], and the use of instrumental variables [47] - can be applied to epigenetic epidemiology studies.

An instrumental variables method that uses germline genetic variants as the instruments - Mendelian randomization - is increasingly used to strengthen causality with respect to environmentally modifiable exposures for which genetic variants can serve as proxy measures [49-51]. Mendelian randomization can be extended to the investigation of epigenetic profiles as the potentially modifiable exposure. This method - 'two step epigenetic Mendelian randomization' - is currently under development, and details can be found elsewhere [5,52].

A further complexity of epigenetic studies is the tissue-specific nature of epigenetic patterns. Given that they are integrally involved in the process of cell and tissue differentiation, it is no surprise that epigenetic patterns differ between tissue sources. Genetic comparisons within and between studies can be made using a variety of sources of DNA to generate genotype data; however, this is not the case in an epigenetic context. Population-based studies often have to rely on easily accessible DNA sources (such as blood, saliva, buccal cells; Table 1). These serve as a surrogate for the target tissue involved in the disease of interest, but there is inevitable heterogeneity in both specific cell type represented and sample processing, which may bias epigenetic measurement (see the section below on data analysis considerations). Despite these limitations, epigenetic epidemiological studies are emerging and include strategies such as Mendelian randomization approaches [53] or inter-tissue comparisons [15] to interrogate the functional relevance and casual nature of observations.

Inter-generational epigenetic studies

Family-based sampling of both siblings and multiple generations can have particular value in epigenetic studies. The fact that epigenetic states are often established in early (in particular antenatal) development makes birth cohorts with recruitment and sample collection from pregnant women and sample collection on offspring from birth onwards of particular value [26,27]. There is considerable interest in the role of epigenetic mechanism in the developmental origins of adult disease, to which longitudinal cohort studies are making a valuable contribution [4,53-59].

Data analysis considerations

Most research undertaking longitudinal analysis of molecular biomarker data assumes that there are predictable biological changes over time associated with a given exposure or disease process. However, in the context of epigenetic studies, change over time can be due to technical [60] or genetic factors [61], tissue type [62,63], changes with normal aging, and stochastic changes [64]. These sources of data 'noise' threaten the detection of the biological signal of interest. Thus, as is often the case, the first and most critical step to performing longitudinal DNA methylation analysis is careful study design and data collection with meticulous recording of technical factors and factors that vary between people. Given that data collection may occur months, years or even decades apart, the awareness and/or control of such sources of variability are paramount to making valid conclusions regarding within-individual changes over time as it may be impossible to account for these factors after the fact. Pre-processing of data is often necessary to generate comparable data from samples between and within individuals over time. International initiatives to address and reach consensus on such issues are in progress [65]. Equally important is that many of these methods seek to optimize the signal-to-noise ratio. These two considerations are critical to generating valid and reproducible results. Prudent use of pre-processing that matches the study design and data, and experimentation with several different methods are strongly encouraged. In addition, the threat of time-varying artifacts masquerading as biological signal is constantly present in longitudinal studies. This possibility should be formally tested as an automatic addition to the primary study hypothesis.

An example of a 'noise' source that is just beginning to be understood is the role of genetic factors in determining the degree of variability in DNA methylation over time. This is suggested by familial clustering of DNA methylation variability over time [61]. From the perspective of individual loci, there is also evidence of CpG site-dependent differential stability [15]. This indicates that loci should be carefully selected that demonstrate greater inter- than intra-individual variation over time. The mechanisms underlying this are unknown but could reasonably be related to overlying genetic architecture (for example, interaction with other epigenetic marks and possibly even the DNA itself) or the cellular milieu, as suggested by tissue-specific difference in stability in the same loci [63]. With the success of next-generation sequencing and its falling costs, we can look forward to a clearer view of the effect of genetic factors on DNA methylation and time-dependent variability.

As alluded to earlier, the vast majority of longitudinal cohort studies that are in a position to consider including epigenetic assessment have used biological specimens collected from peripheral blood. Reliance on leukocyte DNA extracted from peripheral blood introduces a potential source of measurement error [66]. Given the labile nature of leukocyte subtype populations over time, this variation may make an important contribution to intra-individual changes in DNA methylation. For instance, shifts in leukocyte populations can occur as a result of normal development and aging, inflammation

from infectious, rheumatological, or oncological diseases, or normal response to medications (such as non-steroidal anti-inflammatory drugs). The most definitive solution is to isolate cell types (for example, through magnetic-activated or fluorescence-activated cell sorting), so as to perform comparisons within relatively homogenous leukocyte populations. However, this is possible only with freshly collected samples; one of the advantages of prospective longitudinal studies is the potential to collect appropriate samples relevant for epigenetic studies.

When analysis of relatively homogeneous cell types was unavailable, Zhu and colleagues [67] used total and differential leukocyte count (from a sample drawn concurrent with the methylation sample) to control for this variation in regression models. These researchers found that the proportion of leukocyte cell types correlated with levels of LINE-1 methylation. Importantly though, statistical adjustment for this did not alter the association between LINE-1 and Alu methylation levels and individual characteristics (age, gender, smoking habits, alcohol intake, and body mass index). Candidate gene studies of methylation have reached similar conclusions [15,16]. This could mean that leukocyte populations contribute a negligible amount of variance relative to the specified model factors. Alternatively, it may be that controlling for leukocyte population in this manner inadequately captures the effect of this noise. The possibility that using the direct measure of an unwanted variable in a regression equation may sub-optimally reduce noise was explored by Teschendorff and colleagues [60]. Using Illumina HumanMethylation27 BeadChip data, they proposed a variation of surrogate variable analysis in which confounders are modeled as statistically independent components. Using these components instead of the original measures in regression analysis, they found a stronger association between methylation of Polycomb-family gene loci and their phenotype of interest, age. From this, they concluded that the effect of confounders on the DNA methylation data was better represented by independent components than the original covariates.

Lastly, in cases where no information on cell counts is available, a potential solution may arise from the DNA methylation data itself. Such a possibility is presented by Houseman and colleagues through their software methylSpectrum [68]. The authors propose an algorithm to infer the contribution of different leukocyte subpopulations to whole blood DNA methylation patterns. This software is not designed to examine changes over time and requires a suitable reference sample from which to make inferences, which would reasonably require multiple age-appropriate references in a longitudinal study setting.

In summary, we need formal comparisons of these methods in heterogeneous and homogeneous samples from the same specimen. International efforts to create reference epigenomes from homogeneous cell samples will be highly beneficial [65]. However, variation due to cellular and tissue heterogeneity is just one example of the wide breadth of issues regarding noise that require detailed and systematic study.

Modeling epigenetic change over time

There are several issues that need to be considered when analyzing epigenetic change over time, such as the unit of DNA methylation change under examination (Box 1) and the analytic technique. The unit of analysis must consider several issues. For example, how is DNA methylation measured? What is the question under investigation? Is the research focused on testing site-specific changes in DNA methylation related to exposures and/or outcomes or is it seeking to explore a network of gene regulation? What type of *a priori* information is available? How does this information contribute to understanding of error or covariance of methylation measurements? Are individuals compared using categorical or continuous variables?

Guided by the selected unit of DNA methylation change, we now turn to examples of modeling intra-individual variation over time that is due to disease and/or environmental factors. The selection of an appropriate modeling technique has important implications for study power and calculations of statistical significance. We limit this discussion to longitudinal studies with three or more time points, as two time points can at most infer a difference rather than the nature of change. Much of this work is borrowed from other fields, particularly gene expression studies, and uses data-driven or knowledge-driven techniques, or combinations of both.

Several techniques use comparisons between two groups (such as controls versus cases) to determine differential time courses [69,70]. Some of these methods can be extended to comparisons between more than two groups (for example, [71]). An alternative to this individual-based approach is to find time course patterns that distinguish one group of individuals from another (for example, [72,73]). Methods that capitalize on other biological knowledge (such as genomes, transcriptomes, or nucleosomes) may allow us to better infer the nature of methylation in the context of how functional regulation of the genome relates to exposures or disease processes. This is especially powerful to detect signals that are expected to be subtle but consistent among jointly regulated loci [74]. An example is longitudinal gene set analysis [75] using annotations from databases such as Gene Ontology. The parallel analysis of different sources of high-throughput data has so far only been explored in cross-sectional methylation studies but could in theory be applied to longitudinal analysis. However, such longitudinal analysis will require advanced multi-dimensional

Box 1: Potential units of change to examine epigenetic mechanisms

- A single gene or gene region of interest
- Single gene loci that have different temporal patterns between biological groups
- A family of genes of known biological or clinical importance (such as those previously known to show exposure-related differential methylation)
- A group of functionally related genes (for example, as identified by Gene Ontology or Kyoto Encyclopedia of Genes and Genomes (KEGG) terms)
- A network of co-regulated genes (for example, using intersection with concurrent gene expression data or from previous literature)
- Genes related by their linear proximity on the DNA strand (such as regional grouping, as done to examine differential methylation between and within individuals [70])
- Genes related to the overlying chromatin architecture (such as knowledge of nucleosome position or histone modifications)
- Genes that show similar patterns of change (for example, gene curve [71])

techniques (Box 2). These techniques require pre-processed data that are relatively free of noise. Another approach may use data reduction techniques to extract meaningful features from data noise while simultaneously considering the time-varying nature of DNA methylation. For example, group-independent component analysis with temporal concatenation of microarray data would assume that there are common sites of epigenetic activity but that the course of change may be different for each individual. Most experience in this type of technique comes from the analysis of neuroimaging data, where the goal is to uncover areas of the brain that are activated similarly among individuals in an experimental group over time [76]. The translation of such ideas to molecular data, which often have far lower temporal resolution but higher 'spatial' resolution (gene loci as opposed to areas of the brain), would be a challenging but also potentially promising avenue.

The promise of epigenetic studies of longitudinal cohorts

Future longitudinal epigenetic studies will undoubtedly integrate greater levels of genomic, biologic and/or phenomic information. For example, our expanding knowledge of factors influencing chromatin architecture may soon allow the analysis of methylation marks within context of the broader chromatin state. Examples of such data are nucleosome mapping [77], histone modifications [78], and chromosome conformation capture [79]. The influence of the underlying and overlying chromatin architecture (interaction with protein, RNA, and DNA primary and secondary 'structure' [80]) on differential locus stability over time remains to be elucidated. Analysis of DNA methylation is clearly only scratching the surface of the epigenetic information that regulates gene expression, but longitudinal cohort studies provide a tractable opportunity to contribute to our knowledge base in this area and, as our understanding of the wider epigenome improves, additional epigenetic features may also be added to such studies.

Increasingly, studies are pushing to provide a broader mechanistic picture of cellular function and regulation by juxtaposing data from two or more kinds of high-throughput data [81,82]. So far, these data are often extracted from different materials or individuals (such as DNA methylation from whole blood and RNA from cell culture). This limits interpretation of functional relevance. However, advances in biotechnology that reduce the amount of specimen required and increase automation, in conjunction with falling costs, are likely to overcome this problem. Biobanked samples, such as plasma, DNA, and RNA from longitudinal cohorts, could make a valuable contribution to developments in this area. Furthermore, the development of nested recall studies for intensive phenotyping within established cohorts will greatly enhance research opportunities in this area.

As multi-dimensional datasets evolve and the ability to mine the information within them improves, it will be imperative that this information is made as accessible as possible to the wider scientific community. Although it is currently possible to access some information relating epigenetic data to common genetic variation and gene expression, providing an integrative approach, this is not available at multiple time points. Longitudinal studies can offer considerable added value in these settings and profiling using a comprehensive range of high-throughput methods can be overlaid on a wealth of exposure and phenotypic data, allowing researchers to explore specific hypotheses *in silico* and thus helping to prioritize resources for more detailed investigations.

In summary, longitudinal cohorts can offer a great deal in the context of epigenetic epidemiology, including identification of the major determinants of epigenetic variation in populations and a better understanding of the relationship between genetic and epigenetic variation. They provide an unprecedented opportunity to increase our understanding of the dynamic nature of epigenetic patterns and how changes occur in response to a wide range of environmental, lifestyle, and behavioral factors. Population-based studies will improve our knowledge of the extent and topography of inter-individual variation in epigenetic patterns and permit assessment of effect sizes of shifts in epigenetic patterns on health-related

Box 2: Longitudinal modeling strategies for high-dimensional data

Many techniques determine differential time courses based on comparison of two groups of variables (for example, [69,70,84-86]). When there are more than two groups, Yuan and colleagues [71] have demonstrated the utility of their method using hidden Markov models. Multi-group comparisons are also possible; Yuan and colleagues have demonstrated the utility of hidden Markov models to classify genes based upon their temporal expression patterns, which, rather than ignoring, takes advantage of the information contained in time course data. If no groups are present, an alternative is to group genes that show similar temporal patterns (for example, [72]). Another approach is to group genes using *a priori* knowledge of biological similarities and reduce the amount of multiple comparisons. Using Gene Ontology annotation to group 'functionally' related genes, Zhang *et al.* [75] developed a non-parametric longitudinal gene set analysis of gene expression data to detect time-exposure interaction effects. This method is suitable for unbalanced data with missing time points. It is also appropriate for heteroscedastic variance (where variance is uneven across a given data distribution) and non-normal data distributions.

Another consideration is the anticipated type of time course. If a cyclical pattern is expected - for instance, in the study of circadian rhythms or cell cycles - Li *et al.* [73] propose functional clustering using an autoregressive moving-average process. If the goal is to identify groups of co-expressed genes showing gradual changes over time that may be linked to disease progression, Qiu *et al.* [87] have developed a method to study gene expression in cancer tissue at various stages of malignant transformation, which may be applicable to epigenetic data.

Units that consider genes as groups or networks may require a transition from viewing DNA methylation data as a two-dimensional entity (such as disease group and time) to a three-dimensional one (such as disease group, gene locus and time), or even data 'blocks' with greater dimensions. The family of matrix and tensor decompositions (such as independent component analysis, canonical correlation analysis, non-negative tensor factorization, and canonical-decomposition/parallel factor analysis) used in areas such as psychometrics and chemometrics have been proposed as powerful representations of biological multi-dimensional data [88,89]. Translation of such methods to DNA methylation is sure to follow.

Although having multiple time points is advantageous for several reasons, a complication is that similar patterns of change in any group of people can start at different times (such as onset of puberty). This may obscure detection of meaningful but overlapping patterns. This can be unraveled using methods that account for lag between individuals, such as by using parallel factor analysis-related models [90] or spline-based models [91].

outcomes. A wealth of statistical approaches can be borrowed and adapted from related fields and be applied to longitudinal epigenetic analysis - an area of biostatistics that is likely to grow exponentially as high-throughput datasets become increasingly multi-dimensional. Insights into the temporal relationship between changes in epigenetic patterns and functional and health-related outcomes that can be gleaned from longitudinal studies will assist in defining causality. This, and other epidemiological methods to strengthen causal inference, will contribute to the identification of predictive epigenetic biomarkers and modifiable targets for intervention.

The ultimate goal of observational data generated in epidemiological investigations is to feed forward into clinical practice or public health. There is already evidence of translation of longitudinal biological data to clinical applications [83]. The incorporation of epigenetic biomarkers to enhance clinical tools for prediction and prognosis is beginning to emerge [5] (Table 2), and longitudinal cohorts will undoubtedly help in this domain.

Abbreviations

ALSPAC, Avon Longitudinal Study of Parents and Children; SABRE, Southall And Brent REvisited; WTCCC, Wellcome Trust Case Control Consortium.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

All authors contributed to the preparation of the manuscript.

Acknowledgements

JN receives funding support from the Clinical Investigator Program at University of British Columbia and GEoCoDE (EUFP7); LB is an MRC-funded PhD student; AW and DK are supported by the MRC; GDS is funded from multiple sources, including the MRC, Wellcome Trust and NIH; CR is supported by funding from multiple sources, including BBSRC, MRC, EUFP7, NIHR and the Wellcome Trust. This review was not supported from any single source.

Author details

[1]Institute of Genetic Medicine, Newcastle University, NE1 3BZ, UK. [2]Clinician Investigator Program, University of British Columbia, Vancouver, BC V6Z 1Y6, Canada. [3]MRC Unit for Lifelong Health & Aging, London, WC1B 5JU, UK. [4]MRC Centre for Causal Analyses in Translational Epidemiology (CAiTE), School of Social and Community Medicine, University of Bristol, Bristol BS8 2BN, UK.

Published: 29 June 2012

References

1. Relton CL, Davey Smith G: **Is epidemiology ready for epigenetics?** *Int J Epidemiol* 2012, **41:**5-9.
2. Davey Smith G: **Epigenetics for the masses: more than Audrey Hepburn and yellow mice?** *Int J Epidemiol* 2012, **41:**303-308.
3. Maher B: **Personal genomes: The case of the missing heritability.** *Nature* 2008, **456:**18-21.
4. Michels KB: **The promises and challenges of epigenetic epidemiology.** *Exp Gerontol* 2010, **45:**297-301.
5. Relton CL, Davey Smith G: **Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment.** *PLoS Med* 2010, **7:**e1000356.
6. Pearson H: **Children of the 90s: coming of age.** *Nature* 2012, **484:**155-158.
7. Moayyeri A, Hammond CJ, Valdes AM, Spector TD: **Cohort Profile: TwinsUK and Healthy Ageing Twin Study.** *Int J Epidemiol* 2012. doi:10.1093/ije/dyr207.
8. Deary IJ, Gow AJ, Pattie A, Starr JM: **Cohort profile: The Lothian Birth Cohorts of 1921 and 1936**. *Int J Epidemiol* 2011. doi: 10.1093/ije/dyr197.
9. Wadsworth M, Kuh D, Richards M, Hardy R: **Cohort profile: the 1946 National Birth Cohort (MRC National Survey of Health and Development)**. *Int J Epidemiol* 2006, **35:**49-54.

10. Power C, Elliott J: **Cohort profile: 1958 British birth cohort (National Child Development Study).** *Int J Epidemiol* 2006, **35:**34-41.
11. Foley DL, Craig JM, Morley R, Olsson CA, Dwyer T, Smith K, Saffery R: **Prospects for epigenetic epidemiology.** *Am J Epidemiol* 2009, **169:**389-400.
12. De Stavola BL, Nitsch D, dos Santos Silva I, McCormack V, Hardy R, Mann V, Cole TJ, Morton S, Leon DA: **Statistical issues in life course epidemiology.** *Am J Epidemiol* 2006, **163:**84-96.
13. Kuh D, Ben-Shlomo Y, Lynch J, Hallqvist J, Power C: **Life course epidemiology.** *J Epidemiol Community Health* 2003, **57:**778-783.
14. Wong CC, Caspi A, Williams B, Craig IW, Houts R, Ambler A, Moffitt TE, Mill J: **A longitudinal study of epigenetic variation in twins.** *Epigenetics* 2010, **5:**516-526.
15. Talens RP, Boomsma DI, Tobi EW, Kremer D, Jukema JW, Willemsen G, Putter H, Slagboom PE, Heijmans BT: **Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology.** *FASEB J* 2010, **24:**3135-3144.
16. Madrigano J, Baccarelli A, Mittleman MA, Sparrow D, Vokonas PS, Tarantini L, Schwartz J: **Aging and epigenetics: longitudinal changes in gene-specific DNA methylation.** *Epigenetics* 2012, **7:**63-70.
17. Madrigano J, Baccarelli A, Mittleman MA, Wright RO, Sparrow D, Vokonas PS, Tarantini L, Schwartz J: **Prolonged exposure to particulate pollution, genes associated with glutathione pathways, and DNA methylation in a cohort of older men.** *Environ Health Perspect* 2011, **119:**977-982.
18. Vineis P, Chuang SC, Vaissiere T, Cuenin C, Ricceri F, Johansson M, Ueland P, Brennan P, Herceg Z: **DNA methylation changes associated with cancer risk factors and blood levels of vitamin metabolites in a prospective study.** *Epigenetics* 2011, **6:**195-201.
19. Brooks JD, Cairns P, Shore RE, Klein CB, Wirgin I, Afanasyeva Y, Zeleniuch-Jacquotte A: **DNA methylation in pre-diagnostic serum samples of breast cancer cases: results of a nested case-control study.** *Cancer Epidemiol* 2010, **34:**717-723.
20. Gao Y, Baccarelli A, Shu XO, Ji BT, Yu K, Tarantini L, Yang G, Li HL, Hou L, Rothman N, Zheng W, Gao YT, Chow WH: **Blood leukocyte Alu and LINE-1 methylation and gastric cancer risk in the Shanghai Women's Health Study.** *Br J Cancer* 2012, **106:**585-591.
21. Gay LJ, Arends MJ, Mitrou PN, Bowman R, Ibrahim AE, Happerfield L, Luben R, McTaggart A, Ball RY, Rodwell SA: **MLH1 promoter methylation, diet, and lifestyle factors in mismatch repair deficient colorectal cancer patients from EPIC-Norfolk.** *Nutr Cancer* 2011, **63:**1000-1010.
22. Balassiano K, Lima S, Jenab M, Overvad K, Tjonneland A, Boutron-Ruault MC, Clavel-Chapelon F, Canzian F, Kaaks R, Boeing H, Meidtner K, Trichopoulou A, Laglou P, Vineis P, Panico S, Palli D, Grioni S, Tumino R, Lund E, Bueno-de-Mesquita HB, Numans ME, Peeters PH, Ramon Quirós J, Sánchez MJ, Navarro C, Ardanaz E, Dorronsoro M, Hallmans G, Stenling R, Ehrnström R, *et al.*: **Aberrant DNA methylation of cancer-associated genes in gastric cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC-EURGAST).** *Cancer Lett* 2011, **311:**85-95.
23. Limsui D, Vierkant RA, Tillmans LS, Wang AH, Weisenberger DJ, Laird PW, Lynch CF, Anderson KE, French AJ, Haile RW, Harnack LJ, Potter JD, Slager SL, Smyrk TC, Thibodeau SN, Cerhan JR, Limburg PJ: **Cigarette smoking and colorectal cancer risk by molecularly defined subtypes.** *J Natl Cancer Inst* 2010, **102:**1012-1022.
24. Van Guelpen B, Dahlin AM, Hultdin J, Eklof V, Johansson I, Henriksson ML, Cullman I, Hallmans G, Palmqvist R: **One-carbon metabolism and CpG island methylator phenotype status in incident colorectal cancer: a nested case-referent study.** *Cancer Causes Control* 2010, **21:**557-566.
25. Schernhammer ES, Giovannucci E, Baba Y, Fuchs CS, Ogino S: **B vitamins, methionine and alcohol intake and risk of colon cancer in relation to BRAF mutation and CpG island methylator phenotype (CIMP).** *PLoS ONE* 2011, **6:**e21102.
26. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G: **Cohort Profile: The 'Children of the 90s' – the index offspring of the Avon Longitudinal Study of Parents and Children.** *Int J Epidemiol* 2012. doi:10.1093/ije/dys064.
27. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, Henderson J, Macleod J, Molloy L, Ness A, Ring S, Nelson SM, Lawlor DA: **Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort.** *Int J Epidemiol* 2012. doi:10.1093/ije/dys066.
28. Lepeule J, Baccarelli A, Motta V, Cantone L, Litonjua AA, Sparrow D, Vokonas PS, Schwartz J: **Gene promoter methylation is associated with lung function in the elderly: The Normative Aging Study.** *Epigenetics* 2012, **7:**261-269.
29. Bind MA, Baccarelli A, Zanobetti A, Tarantini L, Suh H, Vokonas P, Schwartz J: **Air pollution and markers of coagulation, inflammation, and endothelial function: associations and epigene-environment interactions in an elderly cohort.** *Epidemiology* 2012, **23:**332-340.
30. Zhu ZZ, Sparrow D, Hou L, Tarantini L, Bollati V, Litonjua AA, Zanobetti A, Vokonas P, Wright RO, Baccarelli A, Schwartz J: **Repetitive element hypomethylation in blood leukocyte DNA and cancer incidence, prevalence, and mortality in elderly individuals: the Normative Aging Study.** *Cancer Causes Control* 2011, **22:**437-447.
31. Baccarelli A, Wright R, Bollati V, Litonjua A, Zanobetti A, Tarantini L, Sparrow D, Vokonas P, Schwartz J: **Ischemic heart disease and stroke in relation to blood DNA methylation.** *Epidemiology* 2010, **21:**819-828.
32. Baccarelli A, Tarantini L, Wright RO, Bollati V, Litonjua AA, Zanobetti A, Sparrow D, Vokonas P, Schwartz J: **Repetitive element DNA methylation and circulating endothelial and inflammation markers in the VA Normative Aging Study.** *Epigenetics* 2010, **5:**222-228.
33. Kuh D, Pierce M, Adams J, Deanfield J, Ekelund U, Friberg P, Ghosh AK, Harwood N, Hughes A, Macfarlane PW, Mishra G, Pellerin D, Wong A, Stephen AM, Richards M, Hardy R; NSHD scientific and data collection team: **Cohort Profile: updating the cohort profile for the MRC National Survey of Health and Development: a new clinic-based data collection for ageing research.** *Int J Epidemiol* 2011, **40:**e1-e9.
34. Tillin T, Forouhi NG, McKeigue PM, Chaturvedi N: **Southall And Brent REvisited: Cohort profile of SABRE, a UK population-based comparison of cardiovascular disease and diabetes in people of European, Indian Asian and African Caribbean origins.** *Int J Epidemiol* 2012, **41:**33-42.
35. Hills SA, Balkau B, Coppack SW, Dekker JM, Mari A, Natali A, Walker M, Ferrannini E: **The EGIR-RISC STUDY (The European group for the study of insulin resistance: relationship between insulin sensitivity and cardiovascular disease risk): 1. Methodology and objectives.** *Diabetologia* 2004, **47:**566-570.
36. Kuh D, Ben-Shlomo Y: *A Life Course Approach to Chronic Disease Epidemiology: Tracing the Origins of Ill-health from Early to Adult Life.* 2nd edition. Oxford: Oxford University Press; 2004.
37. Mishra G, Nitsch D, Black S, De Stavola B, Kuh D, Hardy R: **A structured approach to modelling the effects of binary exposure variables over the life course.** *Int J Epidemiol* 2009, **38:**528-537.
38. Kuh D Ben-Shlomo Y, Lynch J, Hallqvist J, Power C: **Glossary for life course epidemiology.** *J Epidemiol Community Health* 2003, **57:**778-783
39. **Healthy Ageing across the Life Course** [http://www.halcyon.ac.uk]
40. **FALCon project** [http://www.nshd.mrc.ac.uk/collaborations/falcon.aspx]
41. **Bristol University: MRC Centre for Causal Analyses in Translational Epidemiology: GEoCoDE** [http://www.bristol.ac.uk/caite/geocode/]
42. Wills AK, Lawlor DA, Matthews FE, Sayer AA, Bakra E, Ben-Shlomo Y, Benzeval M, Brunner E, Cooper R, Kivimaki M, Kuh D, Muniz-Terrera G, Hardy R: **Life course trajectories of systolic blood pressure using longitudinal data from eight UK cohorts.** *PLoS Med* 2011, **8:**e1000440.
43. Davey Smith G, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S: **Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology.** *PLoS Med* 2007, **4:**e352.
44. Wellcome Trust Case Control Consortium: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature* 2007, **447:**661-678.
45. Heijmans BT, Mill J: **Commentary: The seven plagues of epigenetic epidemiology.** *Int J Epidemiol* 2012, **41:**74-78.
46. Brion MJ, Zeegers M, Jaddoe V, Verhulst F, Tiemeier H, Lawlor DA, Davey Smith G: **Intrauterine effects of maternal prepregnancy overweight on child cognition and behavior in 2 cohorts.** *Pediatrics* 2011, **127:**e202-211.
47. Davey Smith G: **Assessing intrauterine influences on offspring health outcomes: can epidemiological studies yield robust findings?** *Basic Clin Pharmacol Toxicol* 2008, **102:**245-256.
48. Davey Smith G: **Negative control exposures in epidemiologic studies.** *Epidemiology* 2012, **23:**350-351; author reply 351-352.
49. Davey Smith G, Ebrahim S: **'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?** *Int J Epidemiol* 2003, **32:**1-22.
50. Davey Smith G: **Use of genetic markers and gene-diet interactions for interrogating population-level causal influences of diet on health.** *Genes Nutr* 2011, **6:**27-43.
51. Timpson NJ, Wade KH, Davey Smith G: **Mendelian randomization:**

application to cardiovascular disease. *Curr Hypertens Rep* 2012, **14:**29-37.
52. Relton CL, Davey Smith G: **Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease.** *Int J Epidemiol* 2012, **41:**161-176.
53. Groom A, Potter C, Swan DC, Fatemifar G, Evans DM, Ring SM, Turcot V, Pearce MS, Embleton ND, Smith GD, Mathers JC, Relton CL: **Postnatal growth and DNA methylation are associated with differential gene expression of the TACSTD2 gene and childhood fat mass.** *Diabetes* 2012, **61:**391-400.
54. Waterland RA, Michels KB: **Epigenetic epidemiology of the developmental origins hypothesis.** *Annu Rev Nutr* 2007, **27:**363-388.
55. Waterland RA: **Epigenetic epidemiology of obesity: application of epigenomic technology.** *Nutr Rev* 2008, **66 Suppl 1:**S21-S23.
56. Relton CL, Groom A, St Pourcain B, Sayers AE, Swan DC, Embleton ND, Pearce MS, Ring SM, Northstone K, Tobias JH, Trakalo J, Ness AR, Shaheen SO, Davey Smith G: **DNA methylation patterns in cord blood DNA and body size in childhood.** *PLoS ONE* 2012, **7:**e31821.
57. Godfrey KM, Sheppard A, Gluckman PD, Lillycrop KA, Burdge GC, McLean C, Rodford J, Slater-Jefferies JL, Garratt E, Crozier SR, Emerald BS, Gale CR, Inskip HM, Cooper C, Hanson MA: **Epigenetic gene promoter methylation at birth is associated with child's later adiposity.** *Diabetes* 2011, **60:**1528-1534.
58. Gabory A, Attig L, Junien C: **Developmental programming and epigenetics.** *Am J Clin Nutr* 2011, **94(6 Suppl):**1943S-1952S.
59. Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Slagboom PE, Lumey LH: **Persistent epigenetic differences associated with prenatal exposure to famine in humans.** *Proc Natl Acad Sci U S A* 2008, **105:**17046-17049.
60. Teschendorff AE, Zhuang J, Widschwendter M: **Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies.** *Bioinformatics* 2011, **27:**1496-1505.
61. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, Yu W, Rongione MA, Ekström TJ, Harris TB, Launer LJ, Eiriksdottir G, Leppert MF, Sapienza C, Gudnason V, Feinberg AP: **Intra-individual change over time in DNA methylation with familial clustering.** *JAMA* 2008, **299:**2877-2883.
62. Tawa R, Ueno S, Yamamoto K, Yamamoto Y, Sagisaka K, Katakura R, Kayama T, Yoshimoto T, Sakurai H, Ono T: **Methylated cytosine level in human liver DNA does not decline in aging process.** *Mech Ageing Devel* 1992, **62:**255-261.
63. Ono T, Tawa R, Shinya K, Hirose S, Okada S: **Methylation of the c-myc gene changes during aging process of mice.** *Biochem Biophys Res Comm* 1986, **139:**1299-1304.
64. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suñer D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, Carlsson E, Poulsen P, Vaag A, Stephan Z, Spector TD, Wu YZ, Plass C, Esteller M: **Epigenetic differences arise during the lifetime of monozygotic twins.** *Proc Natl Acad Sci U S A* 2005, **102:**10604-10609.
65. Eckhardt F, Beck S, Gut IG, Berlin K: **Future potential of the Human Epigenome Project.** *Expert Rev Mol Diagn* 2004, **4:**609-618.
66. Martin GM: **Epigenetic drift in aging identical twins.** *Proc Natl Acad Sci U S A* 2005, **102:**10413-10414.
67. Zhu ZZ, Hou L, Bollati V, Tarantini L, Marinelli B, Cantone L, Yang AS, Vokonas P, Lissowska J, Fustinoni S, Pesatori AC, Bonzini M, Apostoli P, Costa G, Bertazzi PA, Chow WH, Schwartz J, Baccarelli A: **Predictors of global methylation levels in blood DNA of healthy subjects: a combined analysis.** *Int J Epidemiol* 2010. doi:10.1093/ije/dyq154.
68. **Software by E. Andres Houseman: methylSpectrum** [http://people.oregonstate.edu/~housemae/software/]
69. Ma P, Zhong W, Liu J: **Identifying differentially expressed genes in time course microarray data.** *Stat Biosci* 2009, **1:**144-159.
70. Storey JD, Xiao W, Leek JT, Tompkins RG, Davis RW: **Significance analysis of time course microarray experiments.** *Proc Natl Acad Sci U S A* 2005, **102:**12837-12842.
71. Yuan M, Kendziorski C: **Hidden Markov models for microarray time course data in multiple biological conditions.** *J Am Stat Assoc* 2006, **101:**1323-1332.
72. Yuan Y, Li CT, Wilson R: **Partial mixture model for tight clustering of gene expression time-course.** *BMC Bioinformatics* 2008, **9:**287.
73. Li N, McMurry T, Berg A, Wang Z, Berceli SA, Wu R: **Functional clustering of periodic transcriptional profiles through ARMA(p,q).** *PLoS ONE* 2010, **5:**e9894.
74. Choi H, Pavelka N: **When one and one gives more than two: challenges and opportunities of integrative omics.** *Front Genet* 2011, **2:**105.
75. Zhang K, Wang H, Bathke AC, Harrar SW, Piepho HP, Deng Y: **Gene set analysis for longitudinal gene expression data.** *BMC Bioinformatics* 2011, **12:**273.
76. Cole DM, Smith SM, Beckmann CF: **Advances and pitfalls in the analysis and interpretation of resting-state fMRI data.** *Front Syst Neurosci* 2010, **4:**8.
77. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K: **Dynamic regulation of nucleosome positioning in the human genome.** *Cell* 2008, **132:**887-898.
78. Cedar H, Bergman Y: **Linking DNA methylation and histone modification: patterns and paradigms.** *Nat Rev Genet* 2009, **10:**295-304.
79. Tiwari VK, McGarvey KM, Licchesi JDF, Ohm JE, Herman JG, Schübeler D, Baylin SB: **PcG Proteins, DNA methylation, and gene repression by chromatin looping.** *PLoS Biol* 2008, **6:**e306.
80. Edwards JR, O'Donnell AH, Rollins RA, Peckham HE, Lee C, Milekic MH, Chanrion B, Fu Y, Su T, Hibshoosh H, Gingrich JA, Haghighi F, Nutter R, Bestor TH: **Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns.** *Genome Res* 2010, **20:**972-980.
81. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, Gilad Y, Pritchard JK: **DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines.** *Genome Biol* 2011, **12:**R10.
82. Schadt EE, Bjorkegren JL: **NEW: network-enabled wisdom in biology, medicine, and health care.** *Sci Transl Med* 2012, **4:**115rv111.
83. Zhang Y, Tibshirani RJ, Davis RW: **Predicting patient survival from longitudinal gene expression.** *Stat Appl Genet Mol Biol* 2010, **9:**Article41.
84. Minas C, Waddell SJ, Montana G: **Distance-based differential analysis of gene curves.** *Bioinformatics* 2011, **27:**3135-3141.
85. Wang Y, Xu M, Wang Z, Tao M, Zhu J, Wang L, Li R, Berceli SA, Wu R: **How to cluster gene expression dynamics in response to environmental signals.** *Brief Bioinform* 2012, **13:**162-174.
86. Tai YC, Speed TP: **On gene ranking using replicated microarray time course data.** *Biometrics* 2009, **65:**40-51.
87. Qiu P, Gentles AJ, Plevritis SK: **Discovering biological progression underlying microarray samples.** *PLoS Comput Biol* 2011, **7:**e1001123.
88. Rubingh CM, Bijlsma S, Jellema RH, Overkamp KM, van der Werf MtJ, Smilde AK: **Analyzing longitudinal microbial metabolomics data.** *J Proteme Res* 2009, **8:**4319-4327.
89. Phan AH, Cichocki A: **Tensor decompositions for feature extraction and classification of high dimensional data.** *Nonlinear Theory Applications* 2010, **1:**37-68.
90. Puig AT, Wiesel A, Zaas AK, Woods CW, Ginsburg GS, Fleury G, Hero AO: **Order-Preserving factor analysis - application to longitudinal gene expression.** *IEEE Trans Signal Process* 2011, **59:**4447-4458.
91. Smith AA, Vollrath A, Bradfield CA, Craven M: **Similarity queries for temporal toxicogenomic expression profiles.** *PLoS Comput Biol* 2008, **4:**e1000116.
92. Borghol N, Suderman M, McArdle W, Racine A, Hallett M, Pembrey M, Hertzman C, Power C, Szyf M: **Associations with early-life socio-economic position in adult DNA methylation.** *Int J Epidemiol* 2012, **41:**62-74.
93. Herbstman JB, Tang D, Zhu D, Qu L, Sjödin A, Li Z, Camann D, Perera FP: **Prenatal exposure to polycyclic aromatic hydrocarbons, benzo[a]pyrene-DNA adducts, and genomic DNA methylation in cord blood.** *Environ Health Perspect* 2012, **120:**733-738.
94. Koenen KC, Uddin M, Chang SC, Aiello AE, Wildman DE, Goldmann E, Galea S: **SLC6A4 methylation modifies the effect of the number of traumatic events on risk for posttraumatic stress disorder.** *Depress Anxiety* 2011, **28:**639-647.
95. Uddin M, Galea S, Chang SC, Aiello AE, Wildman DE, de los Santos R, Koenen KC: **Gene expression and methylation signatures of MAN2C1 are associated with PTSD.** *Dis Markers* 2011, **30:**111-121.
96. Uddin M, Koenen KC, Aiello AE, Wildman DE, de los Santos R, Galea S: **Epigenetic and inflammatory marker profiles associated with depression in a community-based epidemiologic sample.** *Psychol Med* 2011, **41:**997-1007.
97. Uddin M, Aiello AE, Wildman DE, Koenen KC, Pawelec G, de Los Santos R, Goldmann E, Galea S: **Epigenetic and immune function profiles associated with posttraumatic stress disorder.** *Proc Natl Acad Sci U S A* 2010, **107:**9470-9475.
98. Lumey L, Terry MB, Delgado-Cruzata L, Liao Y, Wang Q, Susser E, McKeague I, Santella RM: **Adult global DNA methylation in relation to pre-natal nutrition.** *Int J Epidemiol* 2012, **41:**116-123.
99. Michels KB, Harris HR, Barault L: **Birthweight, maternal weight trajectories and global DNA methylation of LINE-1 repetitive elements.** *PloS ONE* 2011, **6:**e25254.

100. Olsson CA, Foley DL, Parkinson-Bates M, Byrnes G, McKenzie M, Patton GC, Morley R, Anney RJ, Craig JM, Saffery R: **Prospects for epigenetic research within cohort studies of psychological disorder: a pilot investigation of a peripheral cell marker of epigenetic risk for depression.** *Biol Psychol* 2010, **83:**159-165.
101. Essex MJ, Thomas Boyce W, Hertzman C, Lam LL, Armstrong JM, Neumann SM, Kobor MS: **Epigenetic vestiges of early developmental adversity: childhood stress exposure and DNA methylation in adolescence.** *Child Dev* 2011. doi: 10.1111/j.1467-8624.2011.01641.x.
102. Sood A, Petersen H, Blanchette CM, Meek P, Picchi MA, Belinsky SA, Tesfaigzi Y: **Wood smoke exposure and gene promoter methylation are associated with increased risk for COPD in smokers.** *Am J Respir Crit Care Med* 2010, **182:**1098-1104.
103. Flom JD, Ferris JS, Liao Y, Tehranifar P, Richards CB, Cho YH, Gonzalez K, Santella RM, Terry MB: **Prenatal smoke exposure and genomic DNA methylation in a multiethnic birth cohort.** *Cancer Epidemiol Biomarkers Prev* 2011, **20:**2518-2523.
104. McKay JA, Groom A, Potter C, Coneyworth LJ, Ford D, Mathers JC, Relton CL: **Genetic and non-genetic influences during pregnancy on infant global and site specific DNA methylation: role for folate gene variants and vitamin B12.** *PLoS ONE* 2012, **7:**e33290.
105. Boeke CE, Baccarelli A, Kleinman KP, Burris HH, Litonjua AA, Rifas-Shiman SL, Tarantini L, Gillman M: **Gestational intake of methyl donors and global LINE-1 DNA methylation in maternal and cord blood: prospective results from a folate-replete population.** *Epigenetics* 2012, **7:**253-260.
106. Kile ML, Baccarelli A, Tarantini L, Hoffman E, Wright RO, Christiani DC: **Correlation of global and gene-specific DNA methylation in maternal-infant pairs.** *PloS ONE* 2010, **5:**e13730.
107. Kile ML, Baccarelli A, Hoffman E, Tarantini L, Quamruzzaman Q, Rahman M, Mahiuddin G, Mostofa G, Hsueh YM, Wright RO, Christiani DC: **Prenatal arsenic exposure and DNA methylation in maternal and umbilical cord blood leukocytes.** *Environ Health Perspect* 2012. http://dx.doi.org/10.1289/ehp.1104173.
108. McGuinness D, McGlynn LM, Johnson PC, Macintyre A, Batty GD, Burns H, Cavanagh J, Deans KA, Ford I, McConnachie A, McGinty A, McLean JS, Millar K, Packard CJ, Sattar NA, Tannahill C, Velupillai YN, Shiels PG: **Socio economic status is associated with epigenetic differences in the pSoBid cohort.** *Int J Epidemiol* 2012, **41:**151-160.
109. Marsit CJ, Maccani MA, Padbury JF, Lester BM: **Placental 11-Beta hydroxysteroid dehydrogenase methylation is associated with newborn growth and a measure of neurobehavioral outcome.** *PLoS ONE* 2012, **7:**e33794.
110. Kim M, Long TI, Arakawa K, Wang R, Yu MC, Laird PW: **DNA methylation as a biomarker for cardiovascular disease risk.** *PLoS ONE* 2010, **5:**e9692.
111. Harvey NC, Lillycrop KA, Garratt E, Sheppard A, McLean C, Burdge G, Slater-Jefferies J, Rodford J, Crozier S, Inskip H, Emerald BS, Gale CR, Hanson M, Gluckman P, Godfrey K, Cooper C: **Evaluation of methylation status of the eNOS promoter at birth in relation to childhood bone mineral content.** *Calcif Tissue Int* 2012, **90:**120-127.
112. Brennan K, Garcia-Closas M, Orr N, Fletcher O, Jones M, Ashworth A, Swerdlow A, Thorne H; KConFab Investigators, Riboli E, Vineis P, Dorronsoro M, Clavel-Chapelon F, Panico S, Onland-Moret NC, Trichopoulos D, Kaaks R, Khaw KT, Brown R, Flanagan JM: **Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk.** *Cancer Res* 2012, **72:**2304-2313.
113. Hughes LA, Simons CC, van den Brandt PA, Goldbohm RA, de Goeij AF, de Bruine AP, van Engeland M, Weijenberg MP: **Body size, physical activity and risk of colorectal cancer with or without the CpG island methylator phenotype (CIMP).** *PLoS ONE* 2011, **6:**e18571.

doi:10.1186/gb-2012-13-6-246
Cite this article as: Ng JWY, *et al.*: **The role of longitudinal cohort studies in epigenetic epidemiology: challenges and opportunities.** *Genome Biology* 2012, **13:**246.