Using epigenome-wide association scans of DNA methylation in age-related complex human traits

With rapid technological advancements emerging epigenetic studies of complex traits have shifted from candidate gene analyses towards epigenome-wide association studies (EWAS). EWAS aim to systematically identify epigenetic variants across the genome that associate with complex phenotypes. Recent EWAS using case–control and disease-discordant identical twin designs have identified phenotype-associated differentially methylated regions for several traits. However, EWAS still face many challenges related to methodology, design and interpretation, owing to the dynamic nature of epigenetic variants over time. This article reviews analytical considerations in conducting EWAS and recent applications of this approach to human aging and age-related complex traits.

KEYWORDS: age-related complex trait ▪ differentially methylated regions ▪ DNA methylation ▪ epigenetics ▪ epigenome-wide association studies

Pei-Chien Tsai[1], Tim D Spector[1] & Jordana T Bell*[1]

[1]Department of Twin Research & Genetic Epidemiology, King's College London, London, UK
**Author for correspondence: jordana.bell@kcl.ac.uk*

Human complex traits can be described as the outcomes of interactions among multiple susceptibility factors, including genetic, environmental, stochastic and lifestyle effects. Epigenetic marks are heritable cellular modifications that can also respond to environmental changes. Epigenetic variants have downstream effects on gene expression and, consequently, play an important role in normal development and disease. Therefore, epigenetic mechanisms may present a key link between genetic and environmental risk factors that underlie complex trait susceptibility.

Many epigenetic studies of complex diseases have been performed, but until recently, the majority of these have focused on candidate genes. However, candidate gene selection is based on known or (usually) hypothetical disease mechanisms, and so these studies are limited to *a priori* knowledge and specific candidate gene-based hypotheses. With recent advances in genome-wide approaches to assay epigenetic variation, emphasis is fast shifting towards high-resolution genome-wide comprehensive scans to characterize epigenetic variation, similar to genome-wide association studies (GWAS) in 2007 [1]. The association between epigenetic variants measured across the genome and disease forms the basis of the epigenome-wide association study (EWAS). An EWAS, analogous to a GWAS, aims to systematically discover the association between epigenetic variants and disease. To date, efforts have focused on characterizing genome-wide DNA methylation variation, but there is potential for EWAS in the near future to examine additional epigenetic marks, such as histone modifications and chromatin structure changes [2]. Recent EWAS have identified disease-associated differentially methylated regions (DMRs) for several traits [3–15], but many aspects of this approach require careful consideration owing to the unique features of DNA methylation [16,17]. These include DNA methylation tissue-specificity, variability over time and effects of genetic, environmental and stochastic factors on DNA methylation itself. This review focuses on the analytical considerations and challenges in performing EWAS on human complex traits, with examples of recent EWAS for age-related complex traits and chronological age.

EWAS analytical considerations

This section provides an overview of EWAS methodology: specifically, choice of assay platform, study design and statistical aspects regarding EWAS of DNA methylation variants in complex traits. The review focuses on DNA methylation-based EWAS, because the majority of EWAS examine DNA methylation, potentially due to the availability of genome-wide technologies to assay DNA methylation marks, which currently surpasses the choice available for other epigenetic variants.

DNA methylation is the most widely studied epigenetic mark. Methylation primarily occurs at CpG dinucleotides, which tend to cluster in the genome in regions known as CpG islands (CGIs). DNA methylation in CGI shores, which are regions of lower CpG density on the borders of CGIs, has been shown to be functionally important in gene expression, tissue

Future Medicine part of fsg

differentiation and disease [18]. DNA methylation correlates well with other epigenetic marks such as histone modification [19] and chromatin structure. For example, DNA methylation levels negatively correlate with histone marks that target active genes (such as H3K4me3 and H3K9ac [20]) and enhancers [21]. In general, DNA methylation of promoter regions is negatively associated with gene expression [22–24], whereas gene-body methylation has been reported to positively correlate with gene-expression levels in specific tissues [25,26]. DNA methylation can also occur at cytosines outside of CpG contexts [27], although this phenomenon is thought to be rare. There is evidence that DNA methylation at different loci across the genome can be affected by genetic [28], environmental [29,30] and stochastic variation [31,32]. Evidence for DNA methylation heritability comes from observations that monozygotic (MZ; identical) twins have more similar DNA methylation levels than dizygotic (DZ; nonidentical) twins [33], and that methylation patterns segregate within families [31]. Recently, many underlying genetic variants (methylation quantitative trait loci; me-QTLs) have been discovered to significantly associate with DNA methylation levels at specific CpG sites, with the majority of associations occurring in *cis* [20,34–38]. Another important feature of DNA methylation is that heterogeneity in methylation levels has been observed among different cells and tissues [39–42]. DNA methylation levels have also been reported to associate with environmental variants such as smoking [10], and are therefore thought to change in response to environmental variation over time [31,32,43].

Choice of assay platform for DNA methylation-based EWAS

Current commonly used DNA methylation platforms can be broadly divided into three main categories [16,44,45]: microarray-based approaches; enrichment-based platforms followed by sequencing; and bisulfite sequencing-based methods. Array-based platforms, such as the Illumina Infinium® HumanMethylation27 (Illumina 27k) [46], and HumanMethylation450 (Illumina 450k) [47] bead arrays, and comprehensive high-throughput arrays for relative methylation (CHARM) [48], are hybridization-based microarrays based on either bead-anneal genotyping assays of bisulfite-converted DNA or restriction enzymes, respectively. The previously available Illumina GoldenGate® Methylation Cancer Panel I (Illumina GoldenGate), which covered approximately 1500 CpG sites, was also used by some EWAS. These approaches assay only parts of the methylome, for example, the approximately 27,000 CpG sites in promoter-specific regions (Illumina 27k), or approximately 485,000 CpG sites predominantly located near genes (Illumina 450k) out of approximately 10^7 possible CpG sites across the genome. Nevertheless, array-based methods can provide a cost-effective method to characterize DNA methylation variants, and if widely used, will provide standardized comparisons and allow meta-analysis across EWAS on a common platform, such as the Illumina 450k [16]. However, the most appropriate approach to preprocessing and normalizing array-based DNA methylation data remains to be determined.

Enrichment-based platforms capture the methylated sections of the genome and assay these by array or sequencing methods, which include methylated DNA immunoprecipitation sequencing (MeDIP-seq), methylated DNA capture by affinity purification sequencing (MeCAP-seq), and methylated DNA binding domain sequencing (MBD-seq). These approaches may exhibit a CpG-density dependent bias in binding affinity [49] and do not provide single-CpG-level resolution of methylation, because they are limited by the DNA fragment size from the sonication or size-selection steps in the protocols. Bisulfite-sequencing-based methods include whole-genome bisulfite sequencing (WGBS) and reduced representation bisulfite sequencing. WGBS is typically seen as the gold standard for DNA methylation assays; however, this approach is also imperfect because certain regions of the genome may be difficult to bisulfite sequence and the method does not differentiate between 5-hydroxymethylation and 5-methylation [50]. Sample requirements for DNA quality and quantity will also impact the choice of DNA methylation platform, as previously discussed [45]. For example, array-based platforms require less DNA (e.g., 500 ng for Illumina 27k and 450k arrays [46,47]) compared with enrichment-based or enzyme-based platforms [45]. The methylation-profiling technologies outlined in this section have successfully been applied by multiple studies [20,21,49,51–53]; however, each strategy can be subject to pitfalls and bias. Therefore, it is important to validate the DNA methylation levels at key DMRs of interest using multiple DNA methylation assays.

In an EWAS, DNA methylation levels measured by any of the technologies described in this section may still be affected by confounders, for example, batch effects have previously

been detected in array-based DNA methylation estimates [20]. Different blocking factors may apply to multiple platforms (such as chip effects in array studies and flow-cell effects in sequencing studies), and appropriate experimental design should be used to minimize these effects. Randomizing samples in the experimental design and performing analyses to adjust for measured and unmeasured confounders should be routine EWAS methods. Approaches previously applied to assess heterogeneity in high-throughput technologies (e.g., surrogate-variable analysis in gene-expression studies [54], or principal component analysis and other methods [55]) should also be applied to avoid bias. Initial data checks should include assessment of correlation patterns in genome-wide DNA methylation estimates across the entire sample of individuals and within cases and controls separately, as well as across genome-wide loci and within autosomes and sex chromosomes separately. Data analysis using unprocessed signals, as well as using data transformation and normalization approaches, such as variance stabilization transformations [20,56], should be applied [45]. Experimental and computational approaches to control for confounders and technical artifacts can increase power to detect biological effects, and reduce spurious signals and incorrect conclusions.

EWAS designs

As outlined by Rakyan *et al.,* multiple study designs can be used in the EWAS setting [16]. Cross-sectional EWAS designs can be split into two general groups: population- and family-based (FIGURE 1). In population-based designs, such as the case–control design, one can retrospectively identify cases and controls for a particular disease and compare their DNA methylation patterns to detect disease-related DMRs. At a proportion of sites across the genome, DNA methylation levels are heritable and it is possible that EWAS in cases and controls will detect DMRs at heritable methylation sites. Therefore, case–control EWAS may detect genetic associations with the phenotype that are mediated via DNA methylation.

In family-based designs, the twin design is most well characterized for epigenetic studies [57]. Since 1924, the twin study has traditionally been used to estimate phenotype heritability, based on comparing phenotypic correlation between MZ and DZ twins. However, in the context of EWAS, the disease-discordant MZ twin design is most commonly sought after, because identical twins share nearly 100% of their genetic variation, as well as many environmental and lifestyle factors, but have different epigenomes. Disease-discordant MZ twins are widely used to assess the contribution of environmental and lifestyle risk factors to many complex traits and phenotypes [44], and recent EWAS examples have focused on this design. The aim of these analyses is to clarify potential environmentally driven or stochastic epigenetic changes present in the case but not the control twin. Another option for twin-based EWAS is to compare reared-together and reared-apart identical disease-discordant twins, because epigenetic comparisons in these two groups may reveal additional clues about potential environmental or stochastic triggers of the epigenetic mark. Other family-based study designs can include comparisons across MZ, DZ, sibling and half-sibling pairs, and across parent–offspring pairs, trios and multigenerational families. These extended designs may give insight into the extent of genetic regulation of the epigenetic mark at a particular locus within and across generations. The main limitation of family-based studies is the difficulty of obtaining sufficient samples, such as disease-discordant MZ twins – particularly for rare diseases as only one in 250 individuals has an identical twin [58]. Additional experimental designs can be considered in EWAS of cancer, where an alternative is to include healthy and diseased tissue from the same individual, and if extending to discordant twin designs, to include healthy tissue from both affected and unaffected twins as two separate control samples. In addition, if the aim of an EWAS is to identify extreme hypo- and hyper-methylation effects, rather than differential methylation over a continuous range of methylation levels, DNA methylation analyses of pooled cases and controls may also be pursued.

One disadvantage of cross-sectional EWAS is that the causality of methylation modification in disease cannot be established, because methylation can be both causal and also a consequence of disease. Longitudinal samples, although difficult to obtain, can help address causality, because epigenetic information is available prior to and after the development of disease.

Tissue-shared & -specific effects

Epigenetic variants can be tissue-specific or shared across tissues [20,34–36]. Interestingly, more differences have been found in different tissues from the same individual than in the same type of cell from the same tissue from two unrelated individuals [24]. In the context of EWAS, it is important to identify the correct tissue that is

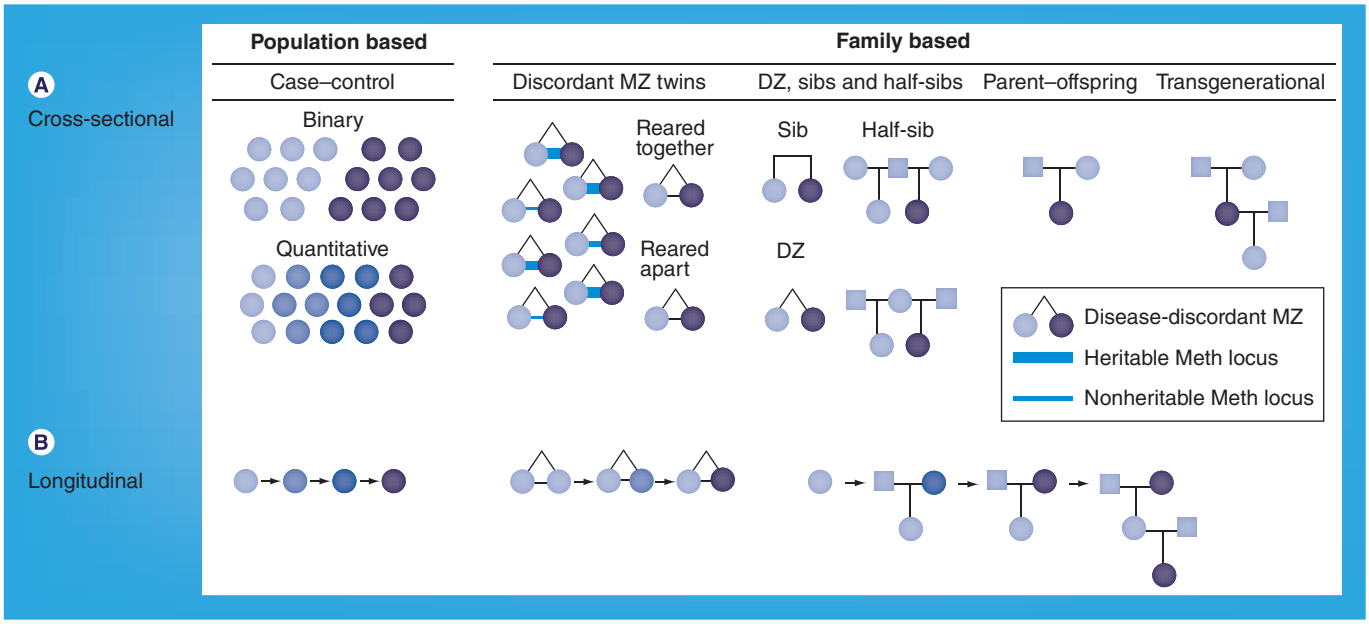

Figure 1. Epigenome-wide association study designs. (A) Cross-sectional and **(B)** longitudinal study designs for epigenome-wide association studies using population- and family-based samples. Circles and squares represent female and male individuals, respectively, where filled data points are individuals with disease and empty data points are controls. For quantitative phenotypes, the grayscale represents the quantile of the phenotypic distribution. Lines between MZ twin pairs show different levels of DNA methylation heritability at the CpG site of interest.
DZ: Dizygotic; Meth: Methylation; MZ: Monozygotic; sib: Sibling.

most relevant for the phenotype of interest. In the majority of studies (particularly in retrospective studies) only whole-blood DNA is often available and the suitability of whole blood as a surrogate for methylation levels across multiple tissues is extensively debated. We recently compared CpG sites from the Illumina 27k array that were identified to be associated with me-QTLs in whole blood [28], in lymphoblastoid cell lines [20] and four brain tissues [34] in different samples. We observed that out of the 1537 CpG sites with me-QTLs obtained in whole blood, 444 (28%) were also observed to have me-QTLs in brain samples and 61 (34%) were observed to have me-QTLs lymphoblastoid cell lines [28]. The results suggest that methylation levels at these CpG sites are more heritable, stable over time and conserved across tissues. At CpG sites with me-QTLs, DNA methylation in whole blood may be a good surrogate for DNA methylation in multiple tissues. Therefore, DMR studies of whole-blood samples should study CpG sites with me-QTLs, but not to the exclusion of the rest of methylome.

The power of EWAS

Power to detect DMRs in an EWAS design depends on many factors. These include, but are not limited to, study design, sample size, DMR effect size, DNA methylation assay coverage and sensitivity, and the stability of the DNA methylation variant over time. Relatively few studies have assessed EWAS power across study designs. Power of the discordant twin design has been examined for a particular genome-wide methylation assay covering 12,192 clones targeting CGI regulatory elements [4]. Power calculations were based on DNA methylation levels at a single locus, *DLX1*, using the DNA methylation fold change within co-twins. The results suggested that only a small sample of disease-discordant twins is required to achieve reasonable power to detect moderate effect sizes. For example, 21 pairs of twins were sufficient to reach greater than 80% power to detect a 1.15-fold change in the methylation signal. However, these power estimates are not generalizable to different assays of DNA methylation, because the sensitivity and specificity of the methylation assay will also affect the power to detect DMRs. The majority of current methylation assays aim to assess methylation levels at single CpG sites, while the previous twin DMR power estimates were based on estimates of methylation levels over larger genomic regions [4]. Although there is evidence for comethylation across the genome [20,28], significant variability in methylation levels at nearby CpG sites has also been observed [20]. It is likely that larger samples of

discordant twins will be needed to achieve sufficient power to detect DMRs of modest effect sizes at single-CpG-site resolution.

The power of the case–control EWAS has recently been examined in two studies. Wang proposed an Uniform-Normal-mixture model to categorize DNA methylation levels into unmethylated, hemimethylated and methylated groups [59]. The study compared the performance of this approach in EWAS with standard parametric and nonparametric tests used in EWAS, and suggested that greater power can be achieved if not only mean differences between groups, but also the variance of methylation levels are taken into account. In terms of standard t-test-based EWAS power estimates, 86.6% power was achieved for a mean difference of 9% methylation using 250 cases and 250 controls. Similarly, a more recent study also suggested taking into account not only the difference, but also the variance in methylation levels across groups [16]. For example, with methylation odds ratios (ORs) of 1.25 and a sample size of 800 cases and 800 controls, 88% power can be obtained at a significance level of 10^{-6} if the locus is primarily methylated, but power is reduced to <10% for more variable methylation levels. The results suggest that greater power could be achieved if the individuals in the sample have similar levels of methylation, and, in general, the DMR effect size measured by the methylation OR (methOR) was found to be a better predictor of power compared with the methylation difference. Power of 80% in EWAS could be achieved with either a high methOR of 2.11 in at least 200 cases and 200 controls, or for a relatively smaller and perhaps more realistic methOR of 1.49, increased sample sizes of at least 800 cases and 800 controls are required [16].

Multiple testing correction for EWAS

Correcting for multiple testing in EWAS is usually based on the total number of CpG sites examined. For example, the Illumina 450k array includes approximately 485,000 CpG sites across the genome, and would result in a Bonferroni-adjusted significance threshold of p-value of approximately 10^{-7}. However, patterns of comethylation across the genome indicate nonindependence in methylation levels at unique CpG sites located close together. Therefore, a Bonferroni correction is probably overconservative. A false-discovery-rate correction for multiple comparisons can also be considered. Currently, permutation-based approaches would be most suitable in initial studies until consensus is reached.

Replication & validation

As in GWAS it is important to follow-up initial discoveries by replicating DMR effects in an independent sample. This is particularly important in the context of determining whether DMRs are causal or consequences of the disease state. Replication guidelines for EWAS have previously been discussed [16]. Furthermore, an additional consideration specific to EWAS is the importance of validating DNA methylation levels at the regions of interest using an alternative method to assay DNA methylation. Each high-throughput assay can be subject to bias and noise, which can result in incorrect DNA methylation levels. For example, recent reanalyses of RNA-sequencing data suggests that high-throughput methods may overestimate the frequencies of both genomic imprinting and RNA editing and require careful methodology [60–64]. To date, most DNA methylation studies have used bisulfite sequencing for validation of candidate regions (TABLES 1 & 2). However, custom validation assays will become necessary for the large-scale studies that are currently underway. A standard custom DNA methylation validation assay may be bisulfite sequencing based and include the identified disease-related methylation regions in the study, and potentially also other known DMRs, heritable CpG sites or CpG sites with meQTLs, and highly variable methylation regions. On the other hand, designing disease-specific validation assays may also be appropriate, similar to custom Illumina Infinium genotyping assays, such as the Immunochip or Metabochip. For example, a cancer-specific panel may target CGI shores, cancer DMRs and cancer-related hypomethylation blocks [65,66], while common disease panels may include promoter-specific regions and previously identified disease-associated DMRs.

Examples of EWAS for age-related phenotypes

Many studies have examined the role of epigenetic variation at individual genes in the context of complex disease. In the past decade, epigenetic variation at particular genes was associated with several complex diseases in case–control and disease-discordant twin studies [3,67–71]. Recently published EWAS have identified DMRs for multiple common complex traits using genome-wide arrays or sequencing-based DNA methylation assays [3–9,11–15]. TABLE 1 lists recent EWAS

Table 1. Recent epigenome-wide association studies for age-related human complex traits.

Disease	Tissue/sample	Assay	a-DMR analysis (significance threshold)†	Major findings	Validation/ replication	Ref.
Disease-discordant MZ EWAS						
SLE, RA and DM	WBCs from five discordant MZ pairs per disease, 30 unrelated controls (discovery) and 17 SLE discordant sib pairs (replication)	Illumina GoldenGate®	Student t-test (10% mean difference, $p < 0.05$ or FDR <0.25)	49 SLE-DMRs	Pyrosequencing (seven genes)‡§	[3]
Multiple sclerosis	$CD4^+$ cells from three discordant MZ pairs	RRBS	No formal statistical framework	Discordance not linked to epigenetic differences	NA	[77]
SZ and BD	WBCs from 11 SZ-discordant and 11 BD-discordant pairs (discovery and validation), and 45 post-mortem brain case–control samples (replication)	Illumina 27k	Paired t-test (FDR <0.05, Bonferroni correction, top-ranked)	Significant hypomethylation in *ST6GALNAC1*	Bisulfite-PCR and EpiTYPER® (one gene)‡§	[5]
Psoriasis	$CD4^+$ cells from 17 and $CD8^+$ from 13 discordant MZ pairs	Illumina 27k	Paired t-test (FDR <0.05)	No genome-wide significant DMRs	NA	[6]
Childhood-onset T1D	$CD14^+$ cells from 15 (discovery and validation) and four (replication) discordant MZ pairs	Illumina 27k	Wilcoxon signed-rank test ($p < 0.01$)	132 T1D-MVPs (DMRs)	Pyrosequencing and Illumina 27k‡	[7]
Case–control EWAS						
Diabetic nephropathy	WBCs from 96 cases and 96 controls	Illumina 27k	Multivariate Cox regression (FDR <0.05)	19 DMRs	NA	[8]
Congenital heart defect	WBCs from 180 affected and 187 unaffected pregnant women	Illumina 27k	Permutation-based multiple linear regression ($p < 0.005$ and FDR <0.05)	425 DMRs at $p < 0.005$, but not FDR <0.05	NA	[9]
T2D	WBCs from 710 cases and 459 controls	DNA methylation at 13,728 genes [25]	Two-proportion z-tests; paired t-tests; binomial logistic regression ($p < 0.05$ or $p < 0.005$)	Significant hypomethylation in *FTO*	NA	[11]
LOAD	Brain tissue from 12 pairs (discovery) and 25 case–control pairs (replication)	Illumina 27k	Linear models ($p < 0.05$)	948 LOAD DMRs	Pyrosequencing (one association with age, but not LOAD)‡§	[12]

†*FDR and Q-value both represent the FDR of significance cutoff.*
‡*Validation using same samples but different technologies.*
§*Replication using different samples and different technologies.*
a-DMR: Age-associated differentially methylated region; BD: Bipolar disorder; DM: Diabetes mellitus; DMR: Differentially methylated region; EWAS: Epigenome-wide association study; FDR: False-discovery rate; Illumina 27k: Illumina Infinium® Human Methylation27 BeadChip; Illumina GoldenGate: Illumina GoldenGate® Methylation Panel I; LOAD: Late-onset Alzheimer's disease; MVP: Methylation variable position; MZ: Monozygotic; NA: Not available; RA: Rheumatoid arthritis; RRBS: Reduced representation bisulfite sequencing; sib: Sibling; SLE: Systemic lupus erythematosus; SZ: Schizophrenia; T1D: Type 1 diabetes; T2D: Type 2 diabetes; WBC: Whole-blood cell.

examples of case–control and disease-discordant MZ twin designs, which are predominantly of limited samples sizes and so far lack the replication expected of GWAS. This article reviews examples of recently published EWAS for age-related traits, including EWAS in disease-discordant MZ twins, cancer case–control EWAS and EWAS for chronological age.

EWAS for age-related complex disease in MZ twins

Disease discordance rates vary greatly in MZ twins [3,70,72–75], suggesting that environmental or epigenetic components may play a significant role in trait etiology. For example, the discordance rate for systemic lupus erythematosus is as high as 70% in MZ twins [76]. Correspondingly, a recent EWAS identified epigenetic differences in autoimmune diseases that were present in systemic lupus erythematosus-discordant MZ twins, but not in rheumatoid arthritis, and dermatomyositis-discordant twins [3]. Additional EWAS of autoimmune diseases have been performed in psoriasis-discordant MZ twins [6] and in MZ twins discordant for multiple sclerosis [77]; however, genome-wide, significant, disease-related DMRs were not identified. Overall, the studies described so far examined small sample sizes (less that 17 MZ pairs per trait) and therefore have limited power to detect DMR effects in complex disease.

More recent EWAS in twins have used larger samples and validation. A genome-wide methylation association study in 44 whole-blood major psychosis-discordant MZ twin samples found that the top-ranked DMR genes were related to psychiatric disorders and neurodevelopment, and were validated using a different technology [5]. Rakyan *et al.* performed an EWAS for childhood-onset Type 1 diabetes in $CD14^+$ monocytes [7]. The DMRs obtained from the discovery stage of the analyses were replicated in two further samples, including longitudinal samples. In total, they identified 132 Type 1 diabetes-associated methylation variable positions that probably occur at an early stage in the etiological process.

EWAS of allele-specific methylation (ASM) have also been performed. ASM typically results in the allele-specific transcriptional silencing of one copy of the gene. For example, imprinted genes exhibit ASM in a parent-of-origin manner; however, ASM does not occur exclusively at imprinted regions, but can be widespread and tissue specific [78,79]. A recent study of ASM found that specific haplotypes associate with DMRs in homologous chromosomes, suggesting that genotypes have greater influence on ASM compared with gametic imprinting [37], and multiple previous studies have also identified genetic variants that associate with ASM in *cis* [78,79]. Recent EWAS have reported ASM changes between MZ twins with autism spectrum disorders and their control co-twins [80], and allele-specific expression results suggest that examining ASM in discordant MZ samples may prove informative to DMR discovery [77]. The prospect of performing ASM-based EWAS in disease is an exciting opportunity in future studies, because these analyses may provide more insight into the disease model underlying epigenetic effects in complex traits.

Recent EWAS in cancer

Epigenetic modifications play a clear role in cancer pathogenesis. Multiple studies have found that tumor suppressor genes tend to have hyper-methylated promoters leading to transcriptional silencing or, more recently, have identified blocks of hypomethylation across the genome [65,66]. A number of recent EWAS have identified hundreds to thousands of differentially methylated CpG sites associated with prostate [15,81], breast [82], ovarian [14] and colon cancers [65]. The findings show that multiple epigenetic changes differentiate cancer and normal tissues; however, unless longitudinal data are available, it is difficult to establish whether the epigenetic marks found in cancer tissue are causal or consequent to disease. The effects of disease on genetic mutations and rearrangements in cancer cells may significantly alter the epigenetic landscape of the cancer cell [83]. Attempts to assess whether cancer-specific DMRs (c-DMRs) are shared across multiple cancer tissues have also been explored. A recent study tested whether c-DMRs found in colon cancer were also differentially methylated in other multiple cancer tissues [65]. The results suggested that methylation at c-DMRs occurs stochastically and is most probably affected by environmental exposures in the later stages of life.

Recent cancer EWAS address validation, and in some cases, replication and timing of the epigenetic changes relative to disease progression, but are nevertheless limited because they are based on relatively sparse DNA methylation arrays that only cover <0.1% of CpG sites in the genome. Many current EWAS are underway for both complex diseases and

Table 2. Recent epigenome-wide association studies for chronological age.

Age range† (years)	Tissue/sample	Assay	a-DMR analysis (significance threshold)‡	Major findings	Validation/ replication	Ref.
14–89	139 multiple tissue samples (discovery) and 112 blood samples (validation)	Illumina GoldenGate®	Generalized linear model ($Q < 0.05$ and $p < 0.05$)	120 hypermethylated a-DMRs and 175 hypomethylated a-DMRs ($Q < 0.05$), 424 a-DMRs ($p < 0.05$)	Pyrosequencing (two genes)§	[29]
26–68 (male); 19–72 (female)	Skin samples from ten male and 20 female subjects	Illumina 27k	Two sample Wilcoxon test ($\beta \geq 0.2$ and p(BH) <0.01)	43 hypermethylated a-DMRs	Bisulfite sequencing (three genes)¶	[30]
52–78 (healthy); 50–84 (cancer patients)	WBCs from 148 healthy and 113 cancer female subjects	Illumina 27k	Linear regression models (FDR <0.05)	226 hypermethylated a-DMRs; 363 hypomethylated a-DMRs	NA	[90]
49–75 (cohort I); 20–69 (cohort II); 16–69 (buccal cohort)	WBCs from 93 subjects (cohort I); $CD4^+$ and $CD14^+$ cells from 25 subjects (cohort II); buccal swabs of ten subjects (buccal cohort)	Illumina 27k	Permutation-based rank correlation ($p < 0.01$)	213 hypermethylated a-DMRs and 147 hypomethylated a-DMRs (cohort I)	Illumina 27k (131 DMRs)#	[91]
1–102	Brain tissues from 150 (stage I) and 237 Caucasians (stage II)	Illumina 27k	Multivariate linear regression ($p < 1.8 \times 10^{-6}$)	Ten a-DMRs in four brain regions	Illumina 27k# (ten a-DMRs)	[92]
21–55 (discovery); 18–70 (replication)	Saliva from 34 healthy male MZ twin pairs (discovery), and 22 twins and 60 unrelated subjects (replication)	Illumina 27k	Absolute correlation coefficients >0.57 ($Q < 0.05$)	69 hypermethylated a-DMRs; 19 hypomethylated a-DMRS	Pyrosequencing (three genes)§	[93]
16–72 (training); 0–78 (validation)	Five data sets from 130 control subjects and 13 cell types (training); 766 subjects with multiple cell types (validation)	Illumina 27k	Pavlidis template matching	431 hypermethylated a-DMRs and 25 hypomethylated a-DMRs ($R > 0.4$, $p < 10^{-5}$), 19 hypermethylated a-DMRs ($R > 0.6$, $p < 10^{-13}$)	Illumina 27k (five genes)§	[94]
Newborns (maternal age: 18–39)	WBCs from 168 newborns (discovery; 92 of these subjects used for validation)	Illumina 27k	Rank correlation ($p < 1.8 \times 10^{-6}$)	144 a-DMRs for maternal age at birth	Pyrosequencing (three genes; two with different effect)¶	[95]
6–23 and 60–73 (fibroblasts); 21–50 and 53–85 (MSCs)	Fibroblasts from nine young and six elderly donors; bone marrow MSCs from four young and four elderly donors (discovery and validation)	Illumina 27k	Methylation difference between young and elderly donors at 15% threshold, rank-based adjustment ($Q < 0.05$)	75 CpG sites in fibroblasts; 1060 CpG sites in MSCs	Pyrosequencing (one gene)¶	[97]

†*Age <0 years: subjects were in gestational weeks.*
‡*FDR and Q-value both represented the FDR of significance cutoff.*
§*Replication using different samples and different technologies.*
¶*Validation using same samples but different technologies.*
#*Replication using different samples but same technology.*
a-DMR: Age-associated DMR; DMR: Differentially methylated region; FDR: False-discovery rate; Illumina 27k: Illumina Infinium® Human Methylation27 BeadChip; Illumina 450k: Illumina Infinium® HumanMethylation450; Illumina GoldenGate: Illumina GoldenGate® Methylation Panel I; MSC: Mesenchymal stromal cell; MZ: Monozygotic; NA: Not available; p(BH): p-value with Benjamini–Hochberg adjustment; WBC: Whole-blood cell.

Table 2. Recent epigenome-wide association studies for chronological age (cont.).

Age range[†] (years)	Tissue/sample	Assay	a-DMR analysis s(significance threshold)[‡]	Major findings	Validation/ replication	Ref.
32–80 (discovery); 20–61 (replication)	WBCs from 172 females (discovery); WBCs from 22 MZ twin pairs (replication)	Illumina 27k	Linear mixed effect models (permutation-based FDR <0.05)	490 significant a-DMRs (98% hypermethylated a-DMRs)	Illumina 27k (184 DMRs)[#]	[28]
-0.5–83.64	Post-mortem human brain tissues from 108 controls (30 fetal samples, 15 samples from children and 63 adult samples)	Illumina 27k	Linear regression (FDR <0.05)	586, 1900 and 3452 hypermethylated a-DMRs; and 279, 3606 and 7126 hypomethylated a-DMRs across the three age groups	NA	[36]
3–17	WBCs from 398 (discovery, 75 of these subjects used for validation) and 78 pediatric subjects (replication)	Illumina 27k and Illumina 450k	Linear regression model (FDR <0.01; $p < 0.001$)	1601 hypermethylated a-DMRs; 477 hypomethylated a-DMRs	Illumina 450k (63% a-DMRs)[§]; Pyrosequencing (two a-DMRs)[¶]	[96]

[†]Age <0 years: subjects were in gestational weeks.
[‡]FDR and Q-value both represented the FDR of significance cutoff.
[§]Replication using different samples and different technologies.
[¶]Validation using same samples but different technologies.
[#]Replication using different samples but same technology.
a-DMR: Age-associated DMR; DMR: Differentially methylated region; FDR: False-discovery rate; Illumina 27k: Illumina Infinium® Human Methylation27 BeadChip; Illumina 450k: Illumina Infinium® HumanMethylation450; Illumina GoldenGate: Illumina GoldenGate® Methylation Panel I; MSC: Mesenchymal stromal cell; MZ: Monozygotic; NA: Not available; p(BH): p-value with Benjamini–Hochberg adjustment; WBC: Whole-blood cell.

cancer, and ongoing efforts are moving towards using next-generation sequencing approaches for better coverage across the whole genome, longitudinal validation addressing causality of epigenetic variants in disease and replication efforts.

EWAS of chronological age

Longitudinal DNA methylation studies

Several studies have looked at DNA methylation within individuals over time. The results indicate that in adults longitudinal changes in methylation can occur, but appear to have generally modest effects over the scale of 3–10 years [84,85]. Conversely, in children, longitudinal changes methylation have much greater effects over a 5-year period [43]. These results demonstrate that changes in single-locus DNA methylation levels likely occur at different rates during life, and may contribute to aging and potential longevity.

Recent EWAS for age-associated DMRs

Epigenetic patterns vary during aging and development, but the rate of change and contribution to biological aging are poorly understood. Global changes in DNA methylation have been reported in individuals with increasing age [31,86], and may be due in part to a progressive loss in DNA methylation in repetitive sequences throughout the genome [87]. However, at specific regions of the genome, individual CpG sites may also significantly change with age. Several studies identified age-associated DMRs (a-DMRs) in whole-blood cells (WBCs), blood cell subtypes and other tissues using the relatively sparse Illumina GoldenGate assay (~1500 CpG sites). Boks *et al.* conducted an EWAS to discover a-DMRs in whole-blood samples from 46 twin pairs and 96 healthy controls on the Illumina GoldenGate platform [88]. They found 58 CpG sites strongly correlated with age, of which six demonstrated significant effects in both samples. a-DMR effects were also examined in 139 subjects using the same methylation assay in blood and brain, lung, pleura, head and neck tissues, and 295 a-DMRs were identified at a false discovery rate of 5% [29]. A more recent study examined 1628 human samples, which included controls, cancer and noncancerous disease samples, with the Illumina GoldenGate array, and found 43 hypermethylated and 25 hypomethylated a-DMRs in 180 leukocyte samples from healthy donors [89]. More recent a-DMR reports have examined DNA methylation using the more extensive Illumina 27k array [28,30,36,90–97].

Next, nine studies [28,36,90–96] that have identified a-DMRs using the Illumina 27k array in multiple tissue samples from cohorts that included healthy subjects have been selected (Table 2). Three studies identified a-DMRs in WBCs and blood cell subtypes. Teschendorff *et al.* examined DNA methylation in WBCs from 113 ovarian cancer subjects and 148 controls, and identified altogether 589 a-DMRs in the overall sample of 261 subjects, of which 61.6% were hypomethylated with age [90]. Rakyan *et al.* examined a-DMRs in WBCs from 93 subjects and replicated 131 hypermethylated a-DMRs across WBCs, CD14+ and $CD4^+$ cells, suggesting that these changes occur among the precursor hematopoietic stem cells, prior to the divergence of the myeloid and lymphoid lineages [91]. Bell *et al.*, recently identified 490 a-DMRs in whole-blood samples from 172 middle-aged female twins, the majority of which were hypermethylated with age [28]. Together, these studies provide strong consistent evidence for a-DMRs in WBCs and cell subtypes, and show a-DMR overlap across studies (see below and Figure 2).

Several EWAS have examined a-DMRs in samples other than WBCs, including multiple tissues. In brain tissue, ten a-DMRs were identified from 150 neurologically normal Caucasian subjects in four brain regions [92]. Another recent study with 130 subjects identified 19 hypermethylated a-DMRs across fibroblasts, keratinocytes, epithelial cells, $CD4^+$ T cells and $CD14^+$ monocytes, and validated a proportion of these in additional cell types [94]. In saliva, 88 a-DMRs were identified in 34 healthy male MZ twin pairs and a-DMR effects were replicated at three genes in another 60 unrelated subjects [93]. These studies show that a-DMR effects exist across tissues and provide the opportunity to test for tissue-specific and shared effects, which may give insights into different components of aging mechanisms.

The EWAS outlined above focus predominantly on adult human samples, but recent aging EWAS have also examined a-DMR effects across wider age ranges, including in children and newborns. A recent study in 398 boys (aged 3–17 years) using peripheral blood reported 1601 hypermethylated and 477 hypomethylated a-DMRs [96]. The authors replicated 62.7% of the initial findings using the Illumina 450k array in another 78 pediatric samples. Numata

Studies†	589 a-DMRs [90]	490 a-DMRs [28]‡	131 a-DMRs [91]§	88 a-DMRs [93]	19 a-DMRs [97]§¶	10 a-DMRs [92]§
589 a-DMRs [90]	CGIs: 379 Non-CGIs: 210	81 (78, 3)	30 (30, 0)	42 (30, 12)	7 (7, 0)	4 (4, 0)
490 a-DMRs [28]	16.5% (92.3%, 7.7%)	CGIs: 484 Non-CGIs: 6	75 (75, 0)	36 (34, 2)	11 (11, 0)	3 (3, 0)
131 a-DMRs [91]	22.9% (100%, 0%)	57.3% (100%, 0%)	CGIs: 126 Non-CGIs: 5	10 (10, 0)	4 (4, 0)	3 (3, 0)
88 a-DMRs [93]	47.7% (71.4%, 29.6%)	40.9% (94.4%, 5.6%)	11.4% (100%, 0%)	CGIs: 73 Non-CGIs: 15	9 (9, 0)	1 (1, 0)
19 a-DMRs [97]	36.8% (100%, 0%)	57.9% (100%, 0%)	21.1% (100%, 0%)	47.4% (100%, 0%)	CGIs: 19 Non-CGIs: 0	0
10 a-DMRs [92]	40% (100%, 0%)	30.0% (100%, 0%)	30.0% (100%, 0%)	10.0% (100%, 0%)	0%	CGIs: 9 Non-CGIs: 1

Figure 2. Human age-associated differentially methylated regions that overlap among six studies.
†Each column and row represent a-DMRs published by six a-DMR studies. The diagonal represents the overall number of a-DMRs within each study and their relative location with respect to CGIs. The upper-right diagonal shows the numbers of overlapping a-DMRs between each pair of studies. The lower-left diagonal shows the percentage of the overlapping a-DMRs, calculated as the number of overlapping differentially methylated regions over the minimum of the total number of differentially methylated regions presented by each study. The numbers and percentages shown in brackets represent the effect directions of a-DMRs: on the left are hypermethylated a-DMRs, and on the right are hypomethylated a-DMRs.
‡This study contains subjects from [91].
§These studies provide hypermethylated a-DMRs only.
¶This study contains subjects from [90,91].
a-DMR: Age-associated differentially methylated region; CGI: CpG island.

et al. used prefrontal cortex tissues in 108 non-psychiatric male and female controls to identify a-DMRs in three age groups: 30 subjects during fetal development, 15 at childhood and 63 at age greater than 10 [36]. They found twice as many age-related hypermethylated changes in childhood and postchildhood compared with the fetal group, and observed that the rate of age-related changes in methylation is greatest during the fetal period. Finally, a study by Adkins *et al.* examined blood samples from 168 newborns and found methylation differences at 144 CpG sites associated with maternal age [95]. Altogether, the aging EWAS indicate that many a-DMRs are triggered at an early age; the rate of age-related methylation changes is highest in the fetal period; and DNA methylation may also be associated with maternal age at birth.

Overlap among human a-DMR studies across tissues

Based on the availability of the a-DMR results provided, the overlap across a-DMRs was assessed using results from six studies that used Illumina 27k arrays in adult human samples. The six studies examined included a-DMRs identified in WBCs and cell subtypes [28,90,91,94], saliva [93] and multiple brain tissues [92], where all the subjects were at least 20 years of age. In four out of the six studies, samples partially overlapped (see Figure 2). In total, 1093 unique a-DMRs were identified in the six studies, 0.2% overlapped across five studies, 1.1% overlapped in four studies, and 4.5% and 15.5% overlapped in at least three and two studies, respectively (Figure 2). None of the a-DMRs successfully replicated in all six studies, but two a-DMRs (near the *NPTX2* and *PDE4C* genes) overlapped across five studies, and 12 a-DMRs (near the *GLRA1*, *TMEM179*, *GCM2*, *TRIM58*, *PTGER3*, *ATP8A2*, *MYOD1*, *BRUNOL6*, *GRIA2*, *KCNK12* and *B3GALT6* genes) overlapped across four studies. These regions represent good candidates for epigenetic biomarkers of ageing, and should be included in future studies modeling chronological age from DNA methylation

patterns. Most of the a-DMRs (84.5%) were identified only in one study, which is probably in part due to different analyses, sample size and significance criteria, and also in part due to tissue-specific effects or type I error. It was observed that more than 93% of the a-DMRs that overlapped across four a-DMR sets were located in CGIs, suggesting that tissue-shared a-DMRs tend to be located in CGIs. Ultimately, the evidence that a-DMRs identified using the sparse Illumina 27k array overlap across studies shows that in some regions DNA methylation may play a general role in the aging process.

Although it is clear that the epigenetic landscape changes during our lifetime, the patterns of epigenetic variation in normal tissues due to aging remain poorly understood [98]. Aside from age-related epigenetic changes at individual genes, the relevance of the noncoding genome to aging remains unclear and the existence of an epigenetic memory with transgenerational inheritance may also impact this field. Several reviews have addressed specific mechanisms of epigenetic effects in somatic tissue aging [98–100]. These include discussions of global methylation changes with age [98,100], hypermethylation of specific tumor suppressor genes [98] and the potential of epigenetic mechanisms to mediate environmental factors that delay aging, such as dietary restrictions [99].

Interpretation of EWAS findings

Once a phenotype-associated DMR is identified in a discovery sample, specific follow-up analyses are crucial to characterize the DMR effect and stability. These include validation of the methylation signal using a different methylation technology, replication of the DMR effect size and direction in an independent sample, and longitudinal analyses to establish whether the DMR variant arises prior to disease onset (and is therefore potentially causal to disease) or as a consequence of disease state. Once these are performed, follow-up studies can interpret the role of the methylation variant in disease and aging.

One option of follow-up is to integrate genetic and epigenetic information in a shared analytical framework, to assess whether the DMR mediates genotype–phenotype associations. If so, DMR effects are likely to be detected in a sample of unrelated individuals (or individuals of different degree of relatedness), but may not be observed in samples of disease-discordant MZ twins. Examples of such integrative analyses include the genotype–epigenotype interaction in the *FTO* gene in Type 2 diabetes [101], and combining me-QTL, GWAS and DMR findings across multiple phenotypes related to ageing [28]. Conversely, if the DMR is suspected to mediate environmental effects on the phenotype, it should be detected in samples of disease-discordant identical twins, but may be more difficult to detect in samples of genetically heterogeneous individuals unless large sample sizes are used.

To explore the effect of epigenetic variants on complex traits, different epigenetic models of disease risk have been proposed [102]. The majority of EWAS to date only examine the methylation threshold model – that is, comparing phenotype variation to overall levels of DNA methylation at a locus of interest. However, it is possible that additional epigenetic models of disease risk exist. For example, cases of ASM, where methylation at one allele is more likely to result in disease (e.g., a dominant epigenetic model with reduced penetrance) may not be detected in standard analyses, which assume a methylation threshold model and use assays that measure overall levels of methylation at a locus (e.g., the Illumina 450k array). Technologies that can detect ASM (e.g., Me-DIP-seq, WGBS and RRBS) will be useful to establish whether more complex scenarios exist.

Several exploratory options exist to clarify the functional role of the DMR variant [17]. Characterizing the location of the DMR relative to the gene and, if possible, directly obtaining gene-expression estimates may both inform the functional role of the DMR on gene expression. For example, a study of psoriasis-discordant MZ twins found no differentially methylated or expressed genes; however, the genes with greatest correlation in methylation and gene-expression differences between unaffected and affected twins were biologically relevant to psoriasis [6]. Therefore, gene-expression levels should be evaluated where possible to gain insight into the functional effect of the methylation variant. If data are not available for the tissue relevant to disease, comparing methylation and expression patterns at the locus across multiple tissues may inform the tissue-specificity of the variant. Additional analyses scanning for transcription factor binding sites in the vicinity of the variant or genomic repeats may also prove informative. Lastly, comparing patterns of multiple levels of epigenetic regulation (e.g., from the ENCODE project [201]) in a region may also be informative. These options have previously been discussed elsewhere [17].

Conclusion & future perspective

In conclusion, epigenetics is an exciting fast-growing field with recent studies identifying

DMRs for a vast array of complex diseases. The EWAS approach provides a promising strategy for uncovering disease-associated epigenetic variation, but there are methodological challenges in studying dynamic epigenetic marks. These accentuate the need to perform longitudinal validation and replication in large EWAS samples. In the near future, upcoming assays of multiple levels of epigenetic variation will allow for more extensive and detailed EWAS scans, leading towards a more extensive understanding of the epigenomic basis of human disease. Integrative analysis across multiple sources of genomic, epigenomic, functional and phenotypic data will help disentangle the underlying biological mechanisms in human disease.

Financial & competing interests disclosure

The authors are supported by the European Research Council (ERC 250157). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Background

- Epigenomic studies of complex traits have progressed towards epigenome-wide association studies (EWAS).
- Ongoing EWAS primarily focus on identifying differentially methylated regions (DMRs) in complex disease.
- The unique dynamic features of DNA methylation should be taken into consideration in the EWAS design.

EWAS analytical considerations

- Several choices for DNA methylation assays exist, but it is important to validate DNA methylation signals at regions of interest using multiple technologies.
- Most commonly used cross-sectional EWAS study designs can be broadly split into case–control and disease-discordant identical twin designs.
- Longitudinal EWAS are crucial to characterizing the role of DMRs in disease susceptibility.
- Tissue specificity in DNA methylation patterns should be explored.
- Maximizing EWAS power, correcting for multiple testing and undertaking replication are important features in EWAS.

Recent EWAS for age-related phenotypes have identified DMRs in disease-discordant monozygotic twins

- Allele-specific methylation may be important in disease.
- Major DNA methylation effects have been reported in cancer, and some of these studies include validation, replication and longitudinal analysis of DMRs.

EWAS of chronological age

- Many age-associated DMRs have been reported across samples from multiple tissues, cell types and age ranges.
- Age-associated DMRs overlap across studies.

Interpretation of EWAS findings

- Discovery of disease DMRs should be followed-up by methylation technology validation, replication and longitudinal studies.
- Allele-specific methylation levels should be explored, because they allow for the estimation of different risk models of disease susceptibility.
- The functional role of DMRs can be explored by obtaining an overview of the epigenetic and transcriptional variation in a region of interest.

References

Papers of special note have been highlighted as:
- ▪ of interest
- ▪▪ of considerable interest

1 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145), 661–678 (2007).

2 Schones DE, Zhao K. Genome-wide approaches to studying chromatin modifications. *Nat. Rev. Genet.* 9(3), 179–191 (2008).

3 Javierre BM, Fernandez AF, Richter J *et al.* Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res.* 20(2), 170–179 (2010).

▪▪ **First epigenome-wide association study (EWAS) conducted in autoimmune disease-discordant monozygotic twins that identified significant differentially methylated regions for systemic lupus erythromatosis.**

4 Kaminsky Z, Petronis A, Wang SC *et al.* Epigenetics of personality traits: an illustrative study of identical twins discordant for risk-taking behavior. *Twin Res. Hum. Genet.* 11(1), 1–11 (2008).

5 Dempster EL, Pidsley R, Schalkwyk LC *et al.* Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Hum. Mol. Genet.* 20(24), 4786–4796 (2011).

6 Gervin K, Vigeland MD, Mattingsdal M *et al.* DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes. *PLoS Genet.* 8(1), e1002454 (2012).

7 Rakyan VK, Beyan H, Down TA *et al.* Identification of Type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet.* 7(9), e1002300 (2011).

8 Bell CG, Teschendorff AE, Rakyan VK, Maxwell AP, Beck S, Savage DA. Genome-wide DNA methylation analysis for diabetic nephropathy in Type 1 diabetes mellitus. *BMC Med. Genomics* 3, 33 (2010).

9 Chowdhury S, Cleves MA, Macleod SL, James SJ, Zhao W, Hobbs CA. Maternal DNA hypomethylation and congenital heart defects. *Birth Defects Res. A Clin. Mol. Teratol.* 91(2), 69–76 (2011).

10 Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am. J. Hum. Genet.* 88(4), 450–457 (2011).

11 Toperoff G, Aran D, Kark JD *et al.* Genome-wide survey reveals predisposing diabetes Type 2-related DNA methylation variations in human peripheral blood. *Hum. Mol. Genet.* 21(2), 371–383 (2012).

12 Bakulski KM, Dolinoy DC, Sartor MA *et al.* Genome-wide DNA methylation differences between late-onset Alzheimer's disease and cognitively normal controls in human frontal cortex. *J. Alzheimer's Dis.* 29(3), 571–588 (2012).

13 Cheung HH, Lee TL, Davis AJ, Taft DH, Rennert OM, Chan WY. Genome-wide DNA methylation profiling reveals novel epigenetically regulated genes and non-coding RNAs in human testicular cancer. *Br. J. Cancer* 102(2), 419–427 (2010).

14 Michaelson-Cohen R, Keshet I, Straussman R, Hecht M, Cedar H, Beller U. Genome-wide *de novo* methylation in epithelial ovarian cancer. *Int. J. Gynecol. Cancer* 21(2), 269–279 (2011).

15 Kim JH, Dhanasekaran SM, Prensner JR *et al.* Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer. *Genome Res.* 21(7), 1028–1041 (2011).

16 Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* 12(8), 529–541 (2011).

▪▪ **First comprehensive review of EWAS methodology, which also provides power estimates for the case–control design.**

17 Heijmans BT, Mill J. Commentary: the seven plagues of epigenetic epidemiology. *Int. J. Epidemiol.* 41(1), 74–78 (2012).

▪▪ **Details important caveats and challenges of performing EWAS.**

18 Doi A, Park IH, Wen B *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* 41(12), 1350–1353 (2009).

19 Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* 10(5), 295–304 (2009).

20 Bell JT, Pai AA, Pickrell JK *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* 12(1), R10 (2011).

21 Lister R, Pelizzola M, Dowen RH *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271), 315–322 (2009).

22 Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447(7143), 425–432 (2007).

23 Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature* 447(7143), 433–440 (2007).

24 Eckhardt F, Lewin J, Cortese R *et al.* DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* 38(12), 1378–1385 (2006).

25 Aran D, Toperoff G, Rosenberg M, Hellman A. Replication timing-related and gene body-specific methylation of active human genes. *Hum. Mol. Genet.* 20(4), 670–680 (2011).

26 Jjingo D, Conley AB, Yi SV, Lunyak VV, Jordan IK. On the presence and role of human gene-body DNA methylation. *Oncotarget* 3(4), 462–474 (2012).

27 Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl Acad. Sci. USA* 97(10), 5237–5242 (2000).

28 Bell JT, Tsai PC, Yang TP *et al.* Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.* 8(4), e1002629 (2012).

29 Christensen BC, Houseman EA, Marsit CJ *et al.* Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet.* 5(8), e1000602 (2009).

30 Gronniger E, Weber B, Heil O *et al.* Aging and chronic sun exposure cause distinct epigenetic changes in human skin. *PLoS Genet.* 6(5), e1000971 (2010).

31 Bjornsson HT, Sigurdsson MI, Fallin MD *et al.* Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 299(24), 2877–2883 (2008).

32 Fraga MF, Ballestar E, Paz MF *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA* 102(30), 10604–10609 (2005).

33 Kaminsky ZA, Tang T, Wang SC *et al.* DNA methylation profiles in monozygotic and dizygotic twins. *Nat. Genet.* 41(2), 240–245 (2009).

▪▪ **First study to provide genome-wide estimates of DNA methylation twin-heritabilities in monozygotic and dizygotic twins.**

34 Gibbs JR, Van Der Brug MP, Hernandez DG *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* 6(5), e1000952 (2010).

35 Gamazon ER, Badner JA, Cheng L *et al.* Enrichment of *cis*-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Mol. Psychiat.* doi:10.1038/mp.2011.174 (2012) (Epub ahead of print).

36 Numata S, Ye T, Hyde TM *et al.* DNA methylation signatures in development and aging of the human prefrontal cortex. *Am. J. Hum. Genet.* 90(2), 260–272 (2012).

37 Gertz J, Varley KE, Reddy TE *et al.* Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLoS Genet.* 7(8), e1002228 (2011).

38 Schilling E, El Chartouni C, Rehli M. Allele-specific DNA methylation in mouse strains is mainly determined by *cis*-acting sequences. *Genome Res.* 19(11), 2028–2035 (2009).

39 Rakyan VK, Down TA, Thorne NP *et al.* An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res.* 18(9), 1518–1529 (2008).

40 Thompson RF, Atzmon G, Gheorghe C *et al.* Tissue-specific dysregulation of DNA methylation in aging. *Aging Cell* 9(4), 506–518 (2010).

41 Heijmans BT, Tobi EW, Lumey LH, Slagboom PE. The epigenome: archive of the prenatal environment. *Epigenetics* 4(8), 526–531 (2009).

42 Ollikainen M, Smith KR, Joo EJ *et al.* DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. *Hum. Mol. Genet.* 19(21), 4176–4188 (2010).

43 Wong CC, Caspi A, Williams B *et al.* A longitudinal study of epigenetic variation in twins. *Epigenetics* 5(6), 516–526 (2010).

44 Bell JT, Spector TD. A twin approach to unraveling epigenetics. *Trends Genet.* 27(3), 116–125 (2011).

45 Laird PW. Principles and challenges of genomewide DNA methylation analysis. *Nat. Rev. Genet.* 11(3), 191–203 (2010).

46 Bibikova M, Le J, Barnes B *et al.* Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics* 1(1), 177–200 (2009).

47 Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the infinium methylation 450K technology. *Epigenomics* 3(6), 771–784 (2011).

48 Irizarry RA, Ladd-Acosta C, Carvalho B *et al.* Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res.* 18(5), 780–790 (2008).

49 Down TA, Rakyan VK, Turner DJ *et al.* A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat. Biotechnol.* 26(7), 779–785 (2008).

50 Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE* 5(1), e8888 (2010).

51 Sandoval J, Heyn H, Moran S *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 6(6), 692–702 (2011).

52 Gervin K, Hammero M, Akselsen HE *et al.* Extensive variation and low heritability of DNA methylation identified in a twin study. *Genome Res.* 21(11), 1813–1821 (2011).

53 Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 33(18), 5868–5877 (2005).

54 Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* 3(9), 1724–1735 (2007).

55 Leek JT, Scharpf RB, Bravo HC *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* 11(10), 733–739 (2010).

▪▪ **Overview of techniques to handle experimental design and technical artefacts in high-throughput data sets.**

56 Du P, Zhang X, Huang CC *et al.* Comparison of β-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 11, 587 (2010).

57 Bell JT, Saffery R. The value of twins in epigenetic epidemiology. *Int. J. Epidemiol.* 41(1), 140–150 (2012).

58 Bulmer MG. *The Bology of Twinning in Man*. Clarendon Press, Oxford, UK (1970).

59 Wang S. Method to detect differentially methylated loci with case–control designs using Illumina arrays. *Genet. Epidemiol.* 35(7), 686–694 (2011).

60 Deveale B, Van Der Kooy D, Babak T. Critical evaluation of imprinted gene expression by RNA-seq: a new perspective. *PLoS Genet.* 8(3), e1002600 (2012).

61 Kelsey G, Bartolomei MS. Imprinted genes ... and the number is? *PLoS Genet.* 8(3), e1002601 (2012).

62 Kleinman CL, Majewski J. Comment on "widespread RNA and DNA sequence differences in the human transcriptome". *Science* 335(6074), 1302; author reply 1302 (2012).

63 Pickrell JK, Gilad Y, Pritchard JK. Comment on "widespread RNA and DNA sequence differences in the human transcriptome". *Science* 335(6074), 1302; author reply 1302 (2012).

64 Lin W, Piskol R, Tan MH, Li JB. Comment on "widespread RNA and DNA sequence differences in the human transcriptome". *Science* 335(6074), 1302; author reply 1302 (2012).

65 Hansen KD, Timp W, Bravo HC *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nat. Genet.* 43(8), 768–775 (2011).

66 Berman BP, Weisenberger DJ, Aman JF *et al.* Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* 44(1), 40–46 (2012).

67 Mill J, Tang T, Kaminsky Z *et al.* Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *Am. J. Hum. Genet.* 82(3), 696–711 (2008).

68 Weksberg R, Shuman C, Caluseriu O *et al.* Discordant *KCNQ1OT1* imprinting in sets of monozygotic twins discordant for Beckwith–Wiedemann syndrome. *Hum. Mol. Genet.* 11(11), 1317–1325 (2002).

69 Handel AE, Ebers GC, Ramagopalan SV. Epigenetics: molecular mechanisms and implications for disease. *Trends Mol. Med.* 16(1), 7–16 (2010).

70 Handunnetthi L, Handel AE, Ramagopalan SV. Contribution of genetic, epigenetic and transcriptomic differences to twin discordance in multiple sclerosis. *Expert Rev. Neurother.* 10(9), 1379–1381 (2010).

71 Marsit CJ, Koestler DC, Christensen BC, Karagas MR, Houseman EA, Kelsey KT. DNA methylation array analysis identifies profiles of blood-derived DNA methylation associated with bladder cancer. *J. Clin. Oncol.* 29(9), 1133–1139 (2011).

72 Poulsen P, Esteller M, Vaag A, Fraga MF. The epigenetic basis of twin discordance in age-related diseases. *Pediatr. Res.* 61(5 Pt 2), R38–R42 (2007).

73 Ballestar E. Epigenetics lessons from twins: prospects for autoimmune disease. *Clin. Rev. Allergy Immunol.* 39(1), 30–41 (2010).

74 Spector TD, Cicuttini F, Baker J, Loughlin J, Hart D. Genetic influences on osteoarthritis in women: a twin study. *BMJ* 312(7036), 940–943 (1996).

75 Macgregor AJ, Snieder H, Rigby AS *et al.* Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum.* 43(1), 30–37 (2000).

76 Arnett FC, Reveille JD. Genetics of systemic lupus erythematosus. *Rheum. Dis Clin. N. Am.* 18(4), 865–892 (1992).

77 Baranzini SE, Mudge J, van Velkinburgh JC *et al.* Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* 464(7293), 1351–1356 (2010).

78 Kerkel K, Spadola A, Yuan E *et al.* Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat. Genet.* 40(7), 904–908 (2008).

79 Schalkwyk LC, Meaburn EL, Smith R *et al.* Allelic skewing of DNA methylation is widespread across the genome. *Am. J. Hum. Genet.* 86(2), 196–212 (2010).

80 Wong CCY, Schalkwyk LC, Meaburn EL *et al.* Genome-wide DNA methylation profiling of monozygotic twins discordant for autism spectrum disorder. Presented at: *12th International Congress of Human Genetics/61st Annual Meeting of the American Society of Human Genetics*. Montreal, Canada, 11–15 October 2011 (Abstract 25).

81 Kobayashi Y, Absher DM, Gulzar ZG *et al.* DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer. *Genome Res.* 21(7), 1017–1027 (2011).

82 Fackler MJ, Umbricht CB, Williams D *et al.* Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. *Cancer Res.* 71(19), 6195–6207 (2011).

83 Feinberg AP, Tycko B. The history of cancer epigenetics. *Nat. Rev. Cancer* 4(2), 143–153 (2004).

84 Madrigano J, Baccarelli A, Mittleman MA *et al.* Aging and epigenetics: longitudinal changes in gene-specific DNA methylation. *Epigenetics* 7(1), (2012).

85 Talens RP, Boomsma DI, Tobi EW *et al.* Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology. *FASEB J.* 24(9), 3135–3144 (2010).

86 Heyn H, Li N, Ferreira HJ *et al.* Distinct DNA methylomes of newborns and centenarians. *Proc. Natl Acad. Sci. USA* (2012).

87 Bollati V, Schwartz J, Wright R *et al.* Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mechanisms Ageing Dev.* 130(4), 234–239 (2009).

88 Boks MP, Derks EM, Weisenberger DJ *et al.* The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS ONE* 4(8), e6767 (2009).

89 Fernandez AF, Assenov Y, Martin-Subero JI *et al.* A DNA methylation fingerprint of 1628 human samples. *Genome Res.* 22(2), 407–419 (2012).

90 Teschendorff AE, Menon U, Gentry-Maharaj A *et al.* Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res.* 20(4), 440–446 (2010).

91 Rakyan VK, Down TA, Maslau S *et al.* Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res.* 20(4), 434–439 (2010).

92 Hernandez DG, Nalls MA, Gibbs JR *et al.* Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum. Mol. Genet.* 20(6), 1164–1172 (2011).

93 Bocklandt S, Lin W, Sehl ME *et al.* Epigenetic predictor of age. *PLoS ONE* 6(6), e14821 (2011).

94 Koch CM, Wagner W. Epigenetic-aging-signature to determine age in different tissues. *Aging* 3(10), 1018–1027 (2011).

95 Adkins RM, Thomas F, Tylavsky FA, Krushkal J. Parental ages and levels of DNA methylation in the newborn are correlated. *BMC Med. Genet.* 12, 47 (2011).

96 Alisch RS, Barwick BG, Chopra P *et al.* Age-associated DNA methylation in pediatric populations. *Genome Res.* 22(4), 623–632 (2012).

97 Koch CM, Suschek CV, Lin Q *et al.* Specific age-associated DNA methylation changes in human dermal fibroblasts. *PLoS ONE* 6(2), e16679 (2011).

98 Berdasco M, Esteller M. Hot topics in epigenetic mechanisms of aging: 2011. *Aging Cell* 11(2), 181–186 (2012).

99 Liu L, Rando TA. Manifestations and mechanisms of stem cell aging. *J. Cell Biol.* 193(2), 257–266 (2011).

100 Rodrigues AA Jr, Suaid HJ, Fazan VP *et al.* Histologic study of urethral extracellular matrix and collagen from aging and long-term alloxan-induced diabetic male rats. *Urology* 77(2), 510.e6–e11 (2011).

101 Bell CG, Finer S, Lindgren CM *et al.* Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the *FTO* Type 2 diabetes and obesity susceptibility locus. *PLoS ONE* 5(11), e14040 (2010).

102 Slatkin M. Epigenetic inheritance and the missing heritability problem. *Genetics* 182(3), 845–850 (2009).

▪ **Description of a generalized model of epigenetic disease susceptibility risk in a mathematical framework.**

Website

201 Encyclopedia of DNA Elements. http://genome.ucsc.edu/ENCODE