Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer.


Published in:
Genome Research

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer

Andrew E. Teschendorff, Usha Menon, Aleksandra Gentry-Maharaj, et al.

*Genome Res.* 2010 20: 440-446 originally published online March 10, 2010
Access the most recent version at doi:10.1101/gr.103606.109
Polycomb group proteins (PCGs) are involved in repression of genes that are required for stem cell differentiation. Recently, it was shown that promoters of PCG target genes (PCGTs) are 12-fold more likely to be methylated in cancer than non-PCGTs. Age is the most important demographic risk factor for cancer, and we hypothesized that its carcinogenic potential may be referred by irreversibly stabilizing stem cell features. To test this, we analyzed the methylation status of over 27,000 CpGs mapping to promoters of ~14,000 genes in whole blood samples from 261 postmenopausal women. We demonstrate that stem cell PCGTs are far more likely to become methylated with age than non-targets (odds ratio = 5.3 [3.8–7.4], \( P < 10^{-8} \)), independently of sex, tissue type, disease state, and methylation platform. We identified a specific subset of 69 PCGT CpGs that undergo hypermethylation with age and validated this methylation signature in seven independent data sets encompassing over 900 samples, including normal and cancer solid tissues and a population of bone marrow mesenchymal stem/stromal cells (\( P < 10^{-5} \)). We find that the age-PCGT methylation signature is present in preneoplastic conditions and may drive gene expression changes associated with carcinogenesis. These findings shed substantial novel insights into the epigenetic effects of aging and support the view that age may predispose to malignant transformation by irreversibly stabilizing stem cell features.

[Supplemental material is available online at http://www.genome.org. The microarray data from this study have been submitted to the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession nos. GSE19711, GSE20067, and GSE20080.]

Targets of polycomb group proteins (PCGTs) are repressed in human embryonic and adult stem cells (Lee et al. 2006). The repression mechanism involves chromatin modifications and is reversible, allowing stem cells and multipotent progenitors to differentiate into committed cell lineages through expression of specific PCGTs. Recently, we and others have demonstrated that stem cell PCGTs are far more likely to become methylated than non-targets, suggesting a stem-cell origin model of cancer. In this model, PCGTs in stem cells would gradually undergo de novo methylation, irreversibly locking cells in an undifferentiated state of self-renewal and thereby predisposing them to subsequent malignant transformation (Ohm et al. 2007; Schlesinger et al. 2007; Widschwendter et al. 2007). However, the mechanisms and factors contributing to this de novo methylation are not yet known.

Age is by far the strongest demographic risk factor for cancer. Besides time-dependent DNA damage (Hoeijmakers 2009), there is now also substantial evidence that aging affects DNA methylation (DNAm) of specific loci, including cancer-related genes (Issa et al. 1994, 1996; Ahuja et al. 1998; Nakagawa et al. 2001; So et al. 2006; Fraga and Esteller 2007; Fraga et al. 2007; Björnsson et al. 2008; Christensen et al. 2009). Based on these observations, we hypothesized that age may induce DNAm of PCGTs, and thereby predispose to cancer. Although blood and epithelial cells originate from different germ layers, we speculated that genes that are mandatory for the differentiation of epithelial cells are more likely to become methylated with increasing age in non-epithelial tissue such as blood. Hence, in order to identify age-dependent CpGs that may be important in the biology of epithelial cancers, we first retrieved an age-dependent signature from peripheral blood cells, then validated the age signature in independent blood samples and normal epithelial tissues, and
finally tested the biological relevance of this signature in epithelial neoplasias.

Results

Age-dependent hypermethylation of PCGTs is independent of cell type

We first performed DNAm profiling (Illumina Infinium 27k) (Weisenberger et al. 2008) of peripheral blood samples drawn from 261 postmenopausal women spanning a 30-yr age range (Supplemental Fig. 1; Supplemental Tables 1, 2). A stringent quality control and interarray normalization procedure resulted in a normalized data matrix of methylation scores (β-values, 0 ≤ β < 1) across 261 blood samples (148 from healthy women [Set-1], 113 from ovarian cancer cases [Set-2]) and 25,642 CpG sites (Table 1; see Methods, Supplemental material). Unsupervised analysis using singular value decomposition (SVD) revealed significant components of variation associated with age (Supplemental Fig. 2). Next, using linear regressions, we derived a DNAm signature for aging. To see if this signature would be dependent on disease status, this analysis was performed separately for cases and controls. We observed that the age-associated DNAm signature was very similar regardless of disease status (Fig. 1A). We thus combined the samples (n = 261) to derive a core DNA methylation signature for aging (589 CpGs passed a false discovery rate [FDR] threshold of 0.05; Fig. 1A, Supplemental Table 3). While the majority of CpGs were hypomethylated with age, we observed that CpGs mapping to promoters of PCGTs (defined by single occupancy of SUZ12, EED, or H3K27me3 in human embryonic stem cells [hESC] [Lee et al. 2006]) were preferentially hypermethylated (Fig. 1A). Specifically, we identified 69 CpGs mapping to 64 unique PCGT loci (Supplemental Table 4), which was significantly more than the 20 unique gene loci expected by chance (Fisher’s exact test, P = 2 × 10−17). We estimated that PCGT loci were approximately fivefold more likely (median unbiased mid-p test, P < 10−10, Fig. 1B) to be hypermethylated with age than non-PCGTs, defined by genes that lack occupancy of SUZ12, EED, or H3K27me3 marks in hESC (Lee et al. 2006). Similarly, we observed a fivefold enrichment of H3K27me3 marks (Fig. 1C) in hematopoietic stem cells (HSC) (Cui et al. 2009). In contrast, only 11 PCGTs were hypomethylated with age, which was somewhat less expected by chance (Fig. 1A,B). We verified that PCGT enrichment among hypermethylated CpGs was not due to an overrepresentation of PCGT CpGs within CpG islands, by showing that the enrichment remained when restricting the comparison to those CpGs located within CpG islands (OR = 4.2 [3.0–5.7], P < 10−10). The 69 PCGT CpGs displayed an average methylation profile that increased monotonically over an age range spanning >25 yr (50–80 yr) (Supplemental Fig. 3).

To investigate the generality of this epigenetic phenomenon, we next applied the same linear regression approach to derive DNAm signatures for aging in two independent data sets (Table 1, Set-5 and Set-6): whole blood (WB) samples from 188 patients (95 women and 93 men) with type 1 diabetes (T1D), and tumor tissue samples from 177 women with ovarian cancer (OvC). Using the same FDR cutoff of 0.05, we observed many age-associated CpGs in WB and OvC tissue (Supplemental Tables 5, 6), with a highly significant OR enrichment of PCGTs among CpGs undergoing hypermethylation with age, but not so among CpGs undergoing hypomethylation (Fig. 1D,E). For the WB T1D samples, we verified that PCGT enrichment was independent of sex (Supplemental Fig. 4).

In addition, using data generated on a different platform, with a different set of CpGs (Goldengate assay; Christensen et al. 2009), we confirmed that PCGTs undergo preferential hypermethylation with age in normal tissues other than blood, including normal pleura and lung samples (Supplemental Fig. 5).

Given the common enrichment of PCGTs across multiple tissue types, we next asked if this result could be due to a specific “core” subset of PCGTs, or if instead the age-PCGT signature is largely tissue-specific. To address this question, we took the specific subset of 69 PCGT CpGs, as identified in the training set of 261 WB samples, and asked if they showed a consistent pattern of increased methylation with age in the validation data sets (Table 1). We found that the average methylation profile of the 69 CpGs correlated significantly with age in blood samples from 108 healthy individuals (Fig. 2A, Set-3), 122 ovarian cancer cases (Supplemental Fig. 3, Set-4), 188 patients with T1D (Fig. 2B, Set-5), and in ovarian cancer tissue from 177 women (Fig. 2C, Set-6). Moreover, we observed that the 69 PCGT CpGs exhibited a significant skew toward hypermethylation with age in all validation sets examined (Fig. 2E–G; Supplemental Fig. 3; Supplemental Table 4). The skew toward hypermethylation remained significant relative to random choices of 69 CpGs. Furthermore, we observed that the 69 PCGT

---

**Table 1. Main methylation data sets used in this study**

<table>
<thead>
<tr>
<th>Set</th>
<th>Sample size</th>
<th>Cell type</th>
<th>Sample characteristics</th>
<th>Age range (yr)</th>
<th>Methylation assay</th>
<th>Use</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set-1</td>
<td>148</td>
<td>WB</td>
<td>Healthy women</td>
<td>52–78</td>
<td>Infin. 27k</td>
<td>Training</td>
<td>Song et al. 2009</td>
</tr>
<tr>
<td>Set-2</td>
<td>113</td>
<td>WB</td>
<td>Ovarian cancer before treatment</td>
<td>50–84</td>
<td>Infin. 27k</td>
<td>Training</td>
<td>Song et al. 2009</td>
</tr>
<tr>
<td>Set-3</td>
<td>108</td>
<td>WB</td>
<td>Healthy women</td>
<td>52–78</td>
<td>Infin. 27k</td>
<td>Validation</td>
<td>Song et al. 2009</td>
</tr>
<tr>
<td>Set-4</td>
<td>122</td>
<td>WB</td>
<td>Ovarian cancer after treatment</td>
<td>49–91</td>
<td>Infin. 27k</td>
<td>Validation</td>
<td>Song et al. 2009</td>
</tr>
<tr>
<td>Set-5</td>
<td>188</td>
<td>WB</td>
<td>Type-1 diabetics (93 male + 95 female)</td>
<td>24–74</td>
<td>Infin. 27k</td>
<td>Validation</td>
<td>Widschwendter et al. 2004</td>
</tr>
<tr>
<td>Set-6</td>
<td>177</td>
<td>Ovarian cancer</td>
<td>Women with ovarian cancer</td>
<td>24–88</td>
<td>Infin. 27k</td>
<td>Validation</td>
<td>—</td>
</tr>
<tr>
<td>Set-7</td>
<td>8</td>
<td>Mesenchymal stem cells</td>
<td>Bone marrow from healthy donors</td>
<td>21–85</td>
<td>Infin. 27k</td>
<td>Validation</td>
<td>Bork et al. 2010</td>
</tr>
<tr>
<td>Set-8</td>
<td>110</td>
<td>Solid tissues</td>
<td>Normal tissue from healthy donors + cancer patients</td>
<td>26–43</td>
<td>Infin. 27k</td>
<td>Validation</td>
<td>Christensen et al. 2009</td>
</tr>
<tr>
<td>Set-9</td>
<td>46</td>
<td>Lung</td>
<td>Normal adjacent + lung cancer tissue from 23 patients</td>
<td>NA</td>
<td>GG 1.5k v1</td>
<td>Validation</td>
<td>Bibikova et al. 2006</td>
</tr>
<tr>
<td>Set-10</td>
<td>48</td>
<td>Cervix</td>
<td>LBC samples</td>
<td>26–43</td>
<td>Infin. 27k</td>
<td>Validation</td>
<td>—</td>
</tr>
</tbody>
</table>

*Illumina Infinium 27k or GoldenGate-GG 1.5k v1. (WB) Whole blood, (NA) not available.*
of the 69 age-hypermethylated CpGs demonstrating corresponding increases in methylation, while 14 of the 20 age-hypomethylated CpGs demonstrated coordinate decreases (Fisher’s exact test, \( P = 4 \times 10^{-6} \); Fig. 2H; Supplemental Table 4).

All these results demonstrate that although the magnitude of methylation changes differed between studies and tissues (Supplemental Table 4), the 69 PCGT CpGs (henceforth “age-PCGT” CpGs) defined a robust age-related DNA methylation signature, exhibiting the same directional DNA methylation changes independently of disease state, sex, tissue, and cell type.

The age-PCGT signature discriminates normal from preinvasive and invasive cancer

We observed that in ovarian cancer tissue, methylation levels of age-PCGT CpGs were higher than those of PCGT CpGs not associated with age (Supplemental Fig. 7). This suggested to us that the implicated genes could be contributing to carcinogenesis. We therefore hypothesized that this age-PCGT signature could be present in preinvasive lesions. As there is still debate over the cell of origin, and there is no well-defined preeclamptic lesion for ovarian cancer, we used the uterine cervix as a model to test this hypothesis. We performed DNA methylation profiling of 48 age-matched cervical smear samples from premenopausal women (Table 1, Set-10) with normal smears (HPV-positive and -negative) and smears exhibiting dysplasia (all HPV-positive; Supplemental material). We verified that the age of samples with dysplasia did not differ from the normal smears (Wilcoxon test, \( P = 0.86 \)). Despite the relatively small sample size and narrow age range of this premenopausal sample set, we found that PCGTs and our 69 age-PCGT CpG subset were preferentially hypermethylated with age (Supplemental Fig. 8). In addition, we observed that the 69 age-PCGT CpGs were more highly methylated in the HPV-positive samples exhibiting dysplasia compared with HPV-positive and -negative normal samples (Fig. 3A). In contrast, DNA methylation of PCGTs that underwent hypomethylation in whole blood did not correlate with progression (Fig. 3B), and non-age-associated PCGT CpGs also did not exhibit methylation differences between dysplasia and normal conditions (\( P = 0.47 \), Fig. 3D). We verified that the age of samples with dysplasia did not differ from the normal smears (Wilcoxon test, \( P = 0.86 \)). Despite the relatively small sample size and narrow age range of this premenopausal sample set, we found that PCGTs and our 69 age-PCGT CpG subset were preferentially hypermethylated with age (Supplemental Fig. 8). In addition, we observed that the 69 age-PCGT CpGs were more highly methylated in the HPV-positive samples exhibiting dysplasia compared with HPV-positive and -negative normal samples (Fig. 3A). In contrast, DNA methylation of PCGTs that underwent hypomethylation in whole blood did not correlate with progression (Fig. 3B), and non-age-associated PCGT CpGs also did not exhibit methylation differences between dysplasia and normal conditions (\( P = 0.47 \), Fig. 3D). In only 0.5% of 10,000 random choices of other 69 PCGTs did we observe an association as strong as the one provided by the age-PCGTs (\( P < 0.01 \), Fig. 3C). Clustering the 48 samples over the 69 CpG methylation profiles also demonstrated that inferred clusters correlated significantly with dysplasia (Fisher’s exact test, \( P < 0.001 \), Fig. 3D).

Figure 1. DNA methylation signatures for aging and enrichment of PCGTs. (A) Flowchart depicting the derivation of the “core” DNA methylation signature for aging. First, the supervised analysis was performed separately for the blood samples from 148 healthy and 113 ovarian cancer cases. This yielded 293 CpGs and 420 CpGs passing a FDR (q) cut-off of 0.3. There was a strong overlap between these two signatures (Fisher’s exact test, \( P = 10^{-18} \)) with \( 80\% \) concordance. Healthy and pretreatment samples were thus combined and supervised analysis was performed on this larger set to identify with more confidence a DNA methylation signature for aging. This gave 589 age-associated CpGs (\( q < 0.05 \)), termed the “core” aging signature. Distribution of these 589 CpGs in terms of hyper- and hypomethylation patterns demonstrated a skew toward hypomethylation (binomial test, \( P = 6 \times 10^{-8} \)). Among the 226 hypermethylated CpGs, 69 mapped to polycomb group targets (PCGTs) (64 unique gene loci), while among the 363 hypomethylated CpGs this number was only 20 (11 unique gene loci). Thus, relative to the core aging signature, PCGTs were preferentially hypermethylated (69 vs. 20 compared with 226 vs. 363, Fisher’s exact test, \( P = 4 \times 10^{-10} \)). Here, PCGTs were defined by promoter occupancy of any one of SUZ12, EED, or H3K27me3 in human embryonic stem cells (Lee et al. 2006). (B–E) Independent validation: enrichment odds ratios with 95% confidence intervals for PCGTs (\( P = 0.01 \), Fig. 3C). Clustering the 48 samples from patients with type-1 diabetes (D) and 177 ovarian cancer samples (E). (Dashed line) Line of unit odds ratio. Two-tailed \( P \)-values of enrichment (i.e., deviation from this line) are given.

CpGs exhibited higher levels of methylation than non-age-associated PCGT CpGs and that the difference in methylation between these two groups increased with age in all validation sets examined (Supplemental Fig. 6).

We next asked if the 69 PCGT CpG DNA methylation signatures could be reflected in multipotent progenitor and stem cell pools. To this end, we investigated the DNA methylation profiles of cultured mesenchymal stromal/stem cells (MSC) derived from the bone marrow of eight healthy individuals spanning a wide age range (21–85 yr; Set-7) (Bork et al. 2010). Despite the small sample size, the average methylation profile exhibited a significant linear increase with age (t-test for linear trend, \( P = 0.003 \), Fig. 2D), with 59
Of the 64 age-PCGT genes, many have been reported to undergo hypermethylation in cancer (Ongenaert et al. 2008). Notably, TP73 and SFRP1 have been reported to undergo hypermethylation in not less than 10 different cancers (Supplemental Table 7). In line with this, we also observed that methylation levels of age-PCGTs discriminated other cancers from their normal counterparts (Supplemental Fig. 9; Bibikova et al. 2006). In addition to frequent hypermethylation, PCGTs also exhibit frequent underexpression in cancer (Ben-Porath et al. 2008). We therefore compared gene expression profiles in ovarian and cervical cancer samples with their respective normal tissues (Scotto et al. 2008; Mok et al. 2009). In both cases, we observed that age-PCGTs exhibited average expression profiles that were significantly lower in cancer compared with normal tissue (Fig. 3E,F). We also observed that age-PCGTs were generally better discriminators of ovarian cancer than 1000 random choices of other 64 PCGTs ($P = 0.06$). Clustering over age-PCGTs further confirmed their power to discriminate ovarian and cervical cancer from their respective normal tissues (Supplemental Fig. 10). Interestingly, age-PCGT mRNA expression also showed a gradual decrease with cancer progression in a data set including preneoplastic lesions (Supplemental Fig. 11; Wurmbach et al. 2007).

**Discussion**

In this paper we have described a consistent directional change of DNA methylation of 64 age-PCGTs with age, characterized by hypermethylation in various cancers (Fig. 1; Supplemental Figs. 4, 8). While effect and sample sizes were not large enough for us to ascertain which genes undergo age-associated hypermethylation in a tissue-specific manner, the fact that we were able to identify a subset of age-PCGTs discriminated other cancers from their normal counterparts (Supplemental Table 7; Bibikova et al. 2006). In addition to frequent hypermethylation, PCGTs also exhibit frequent underexpression in cancer (Ben-Porath et al. 2008). We therefore compared gene expression profiles in ovarian and cervical cancer samples with their respective normal tissues (Scotto et al. 2008; Mok et al. 2009). In both cases, we observed that age-PCGTs exhibited average expression profiles that were significantly lower in cancer compared with normal tissue (Fig. 3E,F). We also observed that age-PCGTs were generally better discriminators of ovarian cancer than 1000 random choices of other 64 PCGTs ($P = 0.06$). Clustering over age-PCGTs further confirmed their power to discriminate ovarian and cervical cancer from their respective normal tissues (Supplemental Fig. 10). Interestingly, age-PCGT mRNA expression also showed a gradual decrease with age-related variation in cell-type composition. Indeed, as demonstrated in our recent work (Teschendorff et al. 2009), we were able to correlate the age-associated hypermethylation signature in blood with changes in blood cell-type composition, but not so for the age-hypermethylated
signature. Consistent with this, we were also not able to validate the age hypomethylation signature in tissues other than blood (Supplemental Fig. 12), or to implicate it in carcinogenesis (Fig. 3B; Supplemental Fig. 13). In contrast, the age-hypermethylated PCGT signature was able to discriminate preneoplastic from normal cells and was found to be aggravated in invasive cancer leading to reduced expression of affected genes (Fig. 3; Supplemental Figs. 9–11). Besides the 64 age-PCGTs identified here, it is likely that other non-PCGT genes that also undergo hypermethylation with age in blood may also be broadly implicated in aging and carcinogenesis (Supplemental Figs. 12, 13).

To obtain direct functional proof that simultaneous silencing of age-PCGT genes predisposes a cell to become malignant is not possible with currently available technology. In the absence of a functional test, there are, however, other lines of evidence supporting the role of age-PCGTs in carcinogenesis: (1) 36% (24/64) of the 64 age-PCGT genes have already been published to be aberrantly methylated and deregulated in cancer; (2) 34% (22/64) of these genes are transcription factors known to be involved in normal differentiation. For instance, \textit{FOXC1} has been shown to play an essential role in development (Myatt and Lam 2007) and is also implicated in cancer (Bloushtain-Qimron et al. 2008). \textit{GATA4} belongs to the family of zinc finger–containing GATA transcription factors, which play critical roles in cell lineage specification during early embryonic development and organ formation. \textit{GATA4} is expressed in human ovarian surface epithelial cells and is important for the formation and maintenance of the differentiated state of these cells (Cai et al. 2009), and \textit{GATA4} is also
heavily methylated in ovarian cancer (Wakana et al. 2006). Another age-PCGT transcription factor is DLX5, which is increasingly methylated and silenced in MSC by both replicative senescence in vitro and aging in vivo (Bork et al. 2010). As a final example, TP73 (also known as p73) shares many functional properties with the TP53 (also known as p53) tumor suppressor and is also involved in mediating DNA damage-induced apoptosis as well as suppressing polyplody and aneuploidy when p53 is inactivated, which suggests that age-dependent methylation and suppression of TP73 may potentially lead to genetic alterations and increased predisposition to cancer (Irwin et al. 2000; Moll and Slade 2004; Talos et al. 2007).

Finally, alongside transcription factors, there are other genes in the age-PCGT panel that have been demonstrated to be involved in carcinogenesis, including ALOX5 (Catalano et al. 2005), SFRP1 (Wnt pathway) (Baylin and Ohm 2006), and KLF14 (TGF-beta signaling) (Truty et al. 2009).

In summary, we have found that age may contribute to carcinogenesis by irreversibly silencing genes that are suppressed in stem cells. To our knowledge, this constitutes the first report of a molecular (epigenetic) signature common to the processes of aging and carcinogenesis. Our findings may have broad implications for cancer prevention, risk prediction, detection, prognosis, and therapy.

Methods

Clinical samples

All DNAm data sets used in the study are summarized in Table 1. The primary sample set consisted of 491 whole blood samples drawn from the United Kingdom Ovarian Cancer Population Study (UKOPS) (Table 1; Supplemental Table 1, Data Sets 1–4; Song et al. 2009). Blood samples were taken at ages spanning a wide age range (50–85 yr) (Table 1). A total of 256 samples were from healthy postmenopausal women (Set-1 and Set-3). The remaining samples (n = 235) consisted of postmenopausal women diagnosed with primary epithelial ovarian cancer. About half of these (pre-treatment [preT] cases; n = 113; Set-2) gave their blood at the time of their diagnosis prior to treatment, and the other half (post-treatment [postT] cases; n = 122; Set-4) gave their blood at some stage during their follow-up visits after primary treatment (mean 2.4 ± 2.7 yr between diagnosis and blood sample taken). The distribution of all these samples across batches is given in Supplemental Table 2. Set-5 consisted of 188 whole blood samples from patients with type 1 diabetes mellitus (CG Bell, AE Teschendorff, V Rakyan, AP Maxwell, S Beck, and DA Savage, in prep.). Set-6 consisted of 177 ovarian cancer tissue specimens from pre-and postmenopausal women. Clinical characteristics of this cohort are provided in Supplemental Table 8. Details of the age distribution of samples per study is shown in Supplemental Figure 1. Full experimental methods and descriptions of other sample sets used in this study and any associated references are available in the Supplemental material. Ethical approval has been obtained for all sample sets.

DNA methylation profiling and quality control

Methylation analysis was performed using the validated Illumina Infinium Human Methylation27 BeadChip (Weisenberger et al. 2008). The methylation status of a specific CpG site was calculated from the intensity of the methylated (M) and unmethylated (U) alleles, as the ratio of fluorescent signals β = Max(M)0/[Max(M) + Max(U)] + 100]. On this scale, 0 < β < 1, with β-values close to 1 (0) indicating methylation (no methylation). Quality control procedures are described in the Supplemental material. After quality control, singular value decompositions (SVD) were used to assess unbalanced variation caused by experimental factors (variable bisulfite conversion efficiency, plate and chip effects) and to test the efficiency of interarray normalization procedures (full details are in the Supplemental material).

All primary data used in this study are available at the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE19711, GSE20067, and GSE20080.

Statistical analysis

Unsupervised analysis was performed using singular value decomposition (SVD), adapted to the methylation data, to determine the number of significant components of variation and their association with phenotypes (here, age). Supervised analyses were performed for each CpG site separately, using a robust linear regression model with age as the response and DNAm as the predictor, including covariates to model the batch, DNA input, and bisulfite conversion efficiency effects. FDRs were evaluated analytically (q-values) (Storey and Tibshirani 2003) as well as using random permutation of sample labels to take potential correlations between CpG sites into account. When data from potentially confounding experimental factors were not available, we used the surrogate variable analysis (SVA) framework (Leek and Storey 2007, 2008) to perform the supervised analysis and FDR estimation. Further details of methodology and software used are available in the Supplemental material.

Acknowledgments

We thank all the individuals who took part in this study and all the researchers, clinicians, and administrative staff who have enabled the many studies contributing to this work. In particular, we thank Andy Ryan, Jeremy Ford, Eva Wozniak, and Nyaladzi Balogun. This work was supported by the Eye Appeal and undertaken at UCLH/UCL, which received a proportion of its funding from the Department of Health NIHR Biomedical Research Centres funding scheme. A.E.T. was supported by a Heller Research Fellowship. S.B. was supported by the Wellcome Trust. D.S and P.M acknowledge the support of the Renal Unit Fund, Belfast Health and Social Care Trust. This work was supported in part by the Ovarian Cancer Research Fund (OCRF) and NIH/NCI grant R01-CA096958 (P.W.L.). We also thank Keji Zhao and Chongzhi Zang for sending us processed ChIP-seq data.

References


