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

Andrew E. Teschendorff,[1] Usha Menon,[2] Aleksandra Gentry-Maharaj,[2] Susan J. Ramus,[2] Daniel J. Weisenberger,[3] Hui Shen,[3] Mihaela Campan,[3] Houtan Noushmehr,[3] Christopher G. Bell,[1] A. Peter Maxwell,[4] David A. Savage,[4] Elisabeth Mueller-Holzner,[5] Christian Marth,[5] Gabrijela Kocjan,[6] Simon A. Gayther,[2] Allison Jones,[2] Stephan Beck,[1] Wolfgang Wagner,[7] Peter W. Laird,[3] Ian J. Jacobs,[2] and Martin Widschwendter[2,8]

[1]Medical Genomics Group, UCL Cancer Institute, University College London, London WC1E 6BT, United Kingdom; [2]Department of Gynecological Oncology, UCL Elizabeth Garrett Anderson Institute for Women's Health, University College London, London W1T 7DN, United Kingdom; [3]USC Epigenome Center, University of Southern California, Keck School of Medicine, Los Angeles, California 90089-9601, USA; [4]Nephrology Research Group, Centre for Public Health, Queen's University Belfast, Belfast BT9 7AB, Northern Ireland; [5]Department of Obstetrics and Gynaecology, Innsbruck Medical University, Innsbruck 6020, Austria; [6]Department of Histopathology, University College London, London WC1E 6JJ, United Kingdom; [7]Helmholtz Institute for Biomedical Engineering–Cell Biology, Aachen University Medical School, 52074 Aachen, Germany

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], $P < 10^{-10}$), 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 in human embryonic stem cells (hESC) are far more likely to undergo cancer-specific promoter DNA hypermethylation 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; Bjornsson 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, $0 < \beta < 1$) 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, $P = 2 \times 10^{-17}$). We estimated that PCGT loci were approximately fivefold (odds ratio [OR]) more likely (median unbiased mid-p test, $P < 10^{-12}$, 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 OR 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 than 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

Set	Sample size	Cell type	Sample characteristics	Age range (yr)	Methylation assay[a]	Use	Reference
Set-1	148	WB	Healthy women	52–78	Infin. 27k	Training	Song et al. 2009
Set-2	113	WB	Ovarian cancer before treatment	50–84	Infin. 27k	Training	Song et al. 2009
Set-3	108	WB	Healthy women	52–78	Infin. 27k	Validation	Song et al. 2009
Set-4	122	WB	Ovarian cancer after treatment	49–91	Infin. 27k	Validation	Song et al. 2009
Set-5	188	WB	Type-1 diabetics (93 male + 95 female)	24–74	Infin. 27k	Validation	—
Set-6	177	Ovarian cancer	Women with ovarian cancer	24–88	Infin. 27k	Validation	Widschwendter et al. 2004
Set-7	8	Mesenchymal stem cells	Bone marrow from healthy donors	21–85	Infin. 27k	Validation	Bork et al. 2010
Set-8	110	Solid tissues	Normal tissue from healthy donors + cancer patients	Top age-CpGs provided	GG 1.5k v2	Validation	Christensen et al. 2009
Set-9	46	Lung	Normal adjacent + lung cancer tissue from 23 patients	NA	GG 1.5k v1	Validation	Bibikova et al. 2006
Set-10	48	Cervix	LBC samples	26–43	Infin. 27k	Validation	—

[a]Illumina Infinium 27k or GoldenGate-GG 1.5k v1.
(WB) Whole blood, (NA) not available.
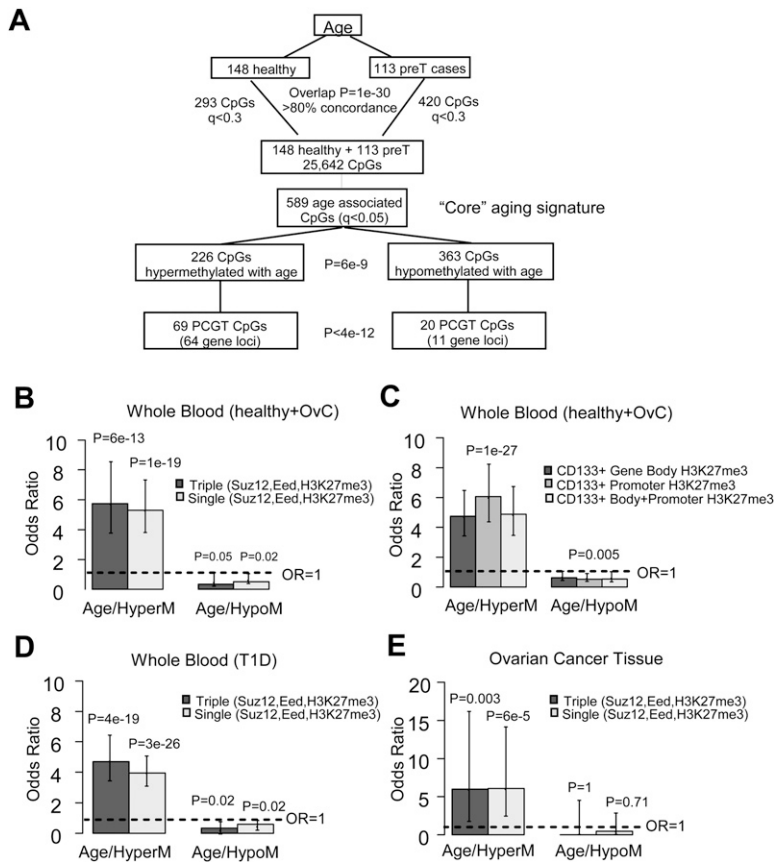
Figure 1. DNAm signatures for aging and enrichment of PCGTs. (*A*) Flowchart depicting the derivation of the "core" DNA methylation signature for aging. First, the supervised analysis was performed separately for the blood samples from 148 healthy and 113 ovarian cancer cases. This yielded 293 CpGs and 420 CpGs passing a FDR (*q*) cut-off of 0.3. There was a strong overlap between these two signatures (Fisher's exact test, $P = 10^{-30}$) 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, $P = 6 \times 10^{-9}$). 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, $P < 4 \times 10^{-12}$). 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*,*C*) Enrichment odds ratios with 95% confidence intervals for PCGTs (*B*) and for H3K27me3 marks (*C*), among the 226 age-hypermethylated and 363 age-hypomethylated CpGs. H3K27me3 marks were defined by trimethylation of H3K27 within gene body, promoter, and gene body + promoter regions in CD133+ hematopoietic stem cells (HSC) (Cui et al. 2009). (*D*,*E*) Independent validation: enrichment odds ratios with 95% confidence intervals for PCGTs among CpGs undergoing significant hyper- and hypomethylation with age in 188 blood 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.

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 DNAm signature, exhibiting the same directional DNAm 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 preneoplastic lesion for ovarian cancer, we used the uterine cervix as a model to test this hypothesis. We performed DNAm 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

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 DNAm 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

the HPV-positive samples exhibiting dysplasia compared with HPV-positive and -negative normal samples (Fig. 3A). In contrast, DNAm of PCGT CpGs 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. 3C). In only 0.5% of 10,000 random choices of other 69 PCGTs CpGs 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).
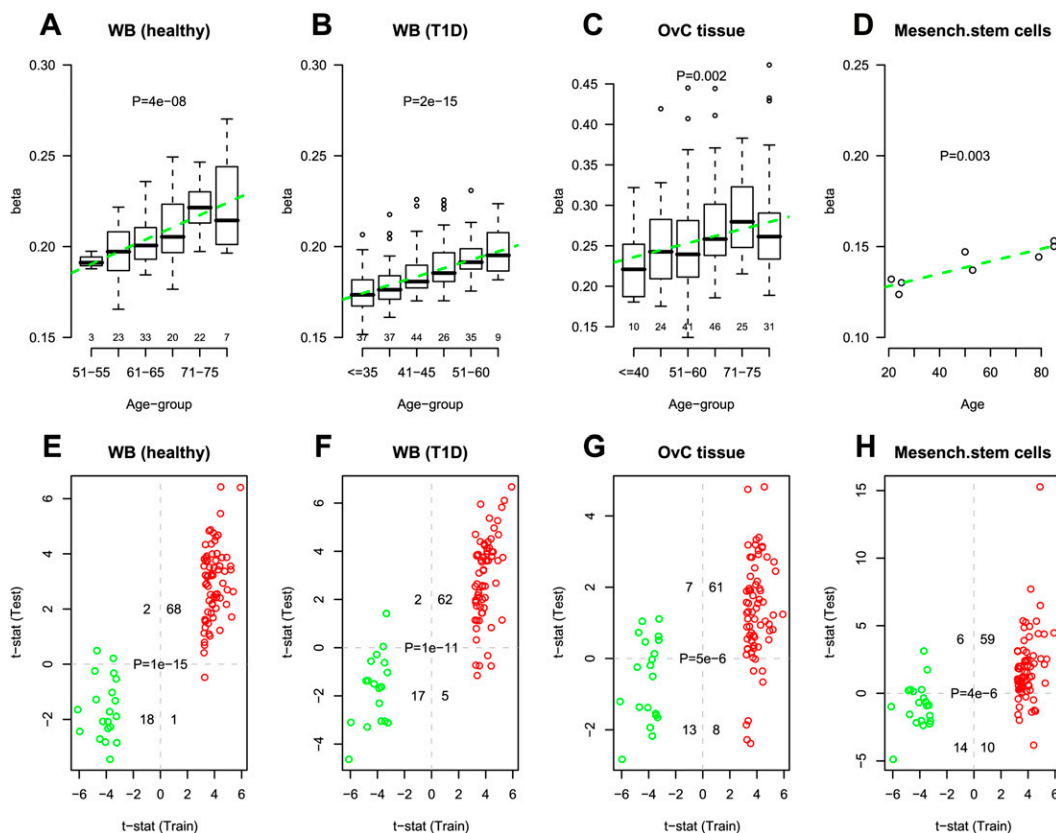
Figure 2. External validation of specific age-associated PCGT DNAm signature. (*A–D*) Average beta-methylation values over the 69 age-hyper-methylated PCGTs (*y*-axis) as a function of age (*x*-axis) in validation data sets. Number of samples in each age group are given *above* the *x*-axis. *t*-test *P*-values for linear trend derived from a robust linear regression are given; (green dashed line) best linear fit. (*E–H*) Validation of age-associated (69 hypermethylated and 20 hypomethylated) PCGT CpGs in test sets. (*X*-axis) *t*-statistic of the linear regression test of age vs. methylation in the training set (blood samples from 148 healthy + 113 pretreatment ovarian cancer cases). Colors reflect directionality: (red) hypermethylated, (green) hypomethylated. (*Y*-axis) *t*-statistic of the linear regression test of age vs. methylation in the test set. We provide the number of CpGs displaying significant hyper/hypo-methylation in the training set and hyper/hypomethylation in the test set, as well as the corresponding Fisher's exact test *P*-value. (*A,E*) Test set of blood samples from an independent set of 108 healthy individuals spanning an age range of 50–80 yr. In *A*, age was categorized into six age groups (50–55, 56–60, 61–65, 66–70, 71–75, >75). (*B,F*) Test set of blood samples from 188 T1D patients spanning an age range of 24–74 yr. In *B*, age was categorized into six age groups (≤35, 36–40, 41–45, 46–50, 51–60, >60). (*C,G*) A test set of ovarian cancer samples from 177 ovarian cancer patients spanning an age range 24–88 yr. In *C*, age was categorized into six age groups (≤40, 41–50, 51–60, 61–70, 71–75, >75). (*D,H*) A test set of eight bone marrow mes-enchymal stromal cell samples from healthy donors of the following ages: 21, 24, 25, 50, 53, 79, 85, 85 (Bork et al. 2010).

Of the 64 age-PCGT genes, many have been reported to un-dergo hypermethylation in cancer (Ongenaert et al. 2008). No-tably, *TP73* and *SFRP1* have been reported to undergo hyper-methylation 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 cer-vical 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 signifi-cantly 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 re-spective 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 DNAm with age, characterized by hypermethylation of PCGTs (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 64 PCGTs exhibiting a clear trend toward hypermethylation with age across multiple cell types (blood, ovarian cancer, cervix, mesenchymal stem cells) indicates that a component of the identified signature is largely nonspecific (Fig. 2; Supplemental Figs. 6–8). It is also very unlikely that the identified age-PCGT signature is caused by 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-asso-ciated hypomethylation signature in blood with changes in blood cell-type composition, but not so for the age-hypermethylated
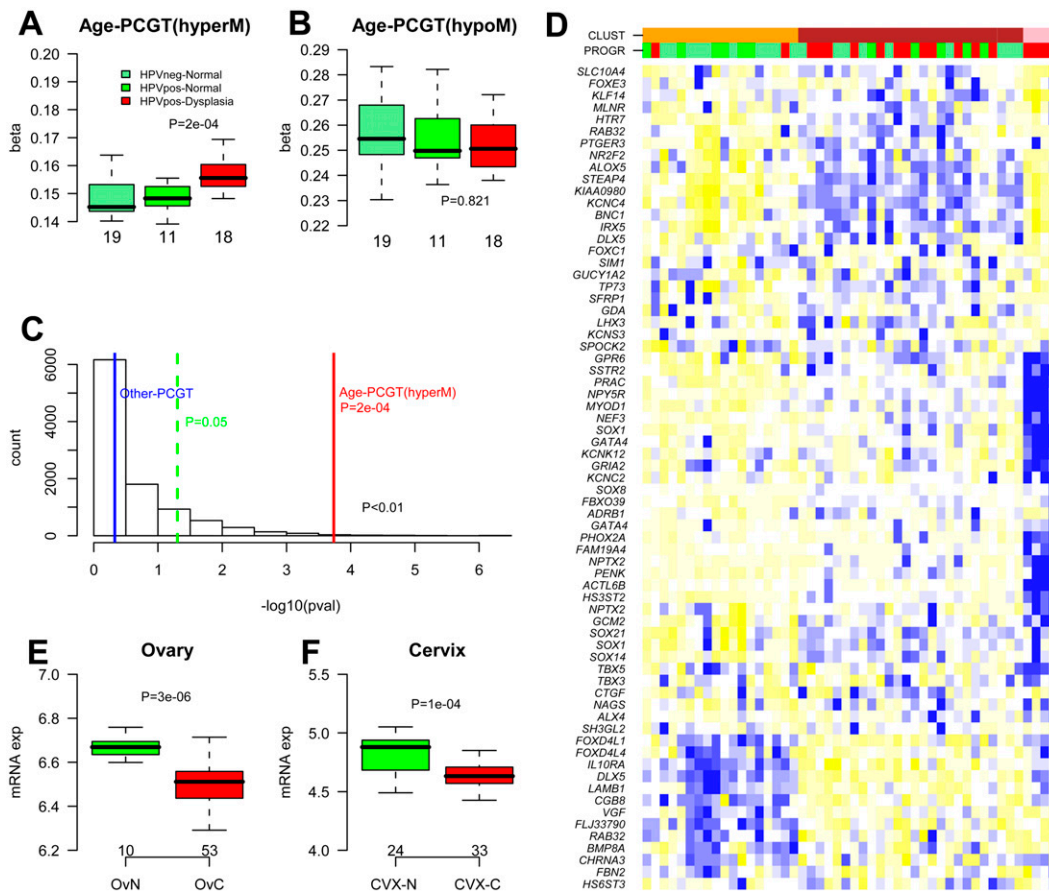
Figure 3. Biological and clinical significance of age-PCGT DNAm signature. (*A,B*) Average methylation values of the 69 age-hypermethylated and 20 age-hypomethylated PCGT CpGs as a function of disease status in 48 cervical cytology samples. (HPVneg) Normal cervical sample not infected with HPV, (HPVpos) normal cervical sample infected with HPV, (HPVpos-Dysplasia) samples infected with HPV and displaying dysplasia. Wilcoxon test *P*-value between normal and dysplastic condition is given. Number of samples in each group given *below* boxplots. (*C*) Histogram distribution of $-\log_{10}(P$-values) from 1000 randomly selected 69 non-age-associated PCGT CpGs. *P*-values were derived from the Wilcoxon test. (Red line) $-\log_{10}(P$-value) for the 69 age-hypermethylated PCGT CpGs, (blue line) $-\log_{10}(P$-value) for PCGT CpGs not mapping to age-PCGTs. In less than 0.5% of runs ($P < 0.01$) were *P*-values as extreme as the observed one, indicating that the age-PCGTs discriminate the dysplastic condition better than a random set of PCGTs. (*D*) Heatmap of the 48 cervical samples over the 69 age-hypermethylated PCGT CpGs. Samples were clustered using a Gaussian mixture model and three optimal clusters were inferred using the Bayesian Information Criterion (see Supplemental material). (Orange, brown, pink) Distinct clusters. The disease status of samples is labeled as a color bar (PROGR): (light green) HPVneg, (green) HPVpos+normal, (red) HPVpos+dysplasia. CpGs were clustered according to hierarchical clustering with a Pearson correlation metric. Prior to sample and CpG clustering, methylation profiles of invividual CpGs were renormalized to mean zero and unit standard deviation. Heatmap reflects, for each CpG, relative methylation levels across samples as determined by the renormalized methylation profile. (Blue) Relative high methylation, (yellow) relative low methylation. (*E,F*) Average gene expression intensity (Affymetrix) values for the 64 age-hypermethylated PCGTs in normal ovarian (OvN) and ovarian cancer tissue (OvC) and in normal cervix (CVX-N) and cervical cancer (CVX-T). Number of samples of each type and Wilcoxon test *P*-values are given.

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 sup-

porting 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, *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). *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. *GATA4* is expressed in human ovarian surface epithelial cells and is important for the formation and maintenance of the differentiated state of these cells (Capo-chichi et al. 2003; Caslini et al. 2006). Loss of *GATA4* expression precedes neoplastic transformation of ovarian surface epithelia (Cai et al. 2009), and *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 polyploidy 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). (3) Finally, alongside transcription factors, there are numerous 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 [posT] 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 specifc CpG site was calculated from the intensity of the methylated (M) and unmethylated (U) alleles, as the ratio of fluorescent signals $\beta = \text{Max}(M,0)/[\text{Max}(M,0) + \text{Max}(U,0) + 100]$. On this scale, $0 < \beta < 1$, 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 unwanted 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.

References

Ahuja N, Li Q, Mohan AL, Baylin SB, Issa JP. 1998. Aging and DNA methylation in colorectal mucosa and cancer. *Cancer Res* **58:** 5489–5494.
Baylin SB, Ohm JE. 2006. Epigenetic gene silencing in cancer—a mechanism for early oncogenic pathway addiction? *Nat Rev Cancer* **6:** 107–116.
Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, Weinberg RA. 2008. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* **40:** 499–507.
Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, Doucet D, Thomas NJ, Wang Y, Vollmer E, et al. 2006. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res* **16:** 383–393.
Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, Yu W, Rongione MA, Ekstroem TJ, Harris TB, et al. 2008. Intra-individual change over time in DNA methylation with familial clustering. *JAMA* **299:** 2877–2883.
Bloushtain-Qimron N, Yao J, Snyder EL, Shipitsin M, Campbell LL, Mani SA, Hu M, Chen H, Ustyansky V, Antosiewicz JE, et al. 2008. Cell type-specific DNA methylation patterns in the human breast. *Proc Natl Acad Sci* **105:** 14076–14081.
Bork S, Pfister S, Witt H, Horn P, Korn B, Ho AD, Wagner W. 2010. DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. *Aging Cell* **9:** 54–63.

Cai KQ, Caslini C, Capo-chichi CD, Slater C, Smith ER, Wu H, Klein-Szanto AJ, Godwin AK, Xu XX. 2009. Loss of GATA4 and GATA6 expression specifies ovarian cancer histological subtypes and precedes neoplastic transformation of ovarian surface epithelia. *PLoS One* **4:** e6454. doi: 10.1371/journal.pone.0006454.

Capo-chichi CD, Roland IH, Vanderveer L, Bao R, Yamagata T, Hirai H, Cohen C, Hamilton TC, Godwin AK, Xu XX. 2003. Anomalous expression of epithelial differentiation-determining GATA factors in ovarian tumorigenesis. *Cancer Res* **63:** 4967–4977.

Caslini C, Capo-chichi CD, Roland IH, Nicolas E, Yeung AT, Xu XX. 2006. Histone modifications silence the GATA transcription factor genes in ovarian cancer. *Oncogene* **25:** 5446–5461.

Catalano A, Rodilossi S, Caprari P, Coppola V, Procopio A. 2005. 5-Lipoxygenase regulates senescence-like growth arrest by promoting ROS-dependent p53 activation. *EMBO J* **24:** 170–179.

Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, et al. 2009. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* **5:** e1000602. doi: 10.1371/journal.pgen.1000602.

Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K. 2009. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* **4:** 80–93.

Fraga MF, Esteller M. 2007. Epigenetics and aging: The targets and the marks. *Trends Genet* **23:** 413–418.

Fraga MF, Agrelo R, Esteller M. 2007. Cross-talk between aging and cancer: The epigenetic language. *Ann N Y Acad Sci* **1100:** 60–74.

Hoeijmakers JH. 2009. DNA damage, aging, and cancer. *N Engl J Med* **361:** 1475–1485.

Irwin M, Marin MC, Phillips AC, Seelan RS, Smith DI, Liu W, Flores ER, Tsai KY, Jacks T, Vousden KH, et al. 2000. Role for the p53 homologue p73 in E2F-1-induced apoptosis. *Nature* **407:** 645–648.

Issa JP, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, Baylin SB. 1994. Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat Genet* **7:** 536–540.

Issa JP, Vertino PM, Boehm CD, Newsham IF, Baylin SB. 1996. Switch from monoallelic to biallelic human IGF2 promoter methylation during aging and carcinogenesis. *Proc Natl Acad Sci* **93:** 11757–11762.

Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, et al. 2006. Control of developmental regulators by polycomb in human embryonic stem cells. *Cell* **125:** 301–313.

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

Leek JT, Storey JD. 2008. A general framework for multiple testing dependence. *Proc Natl Acad Sci* **105:** 18718–18723.

Mok SC, Bonome T, Vathipadiekal V, Bell A, Johnson ME, Wong KK, Park DC, Hao K, Yip DK, Donninger H, et al. 2009. A gene signature predictive for outcome in advanced ovarian cancer identifies a novel survival factor: Microfibril-associated glycoprotein 2. *Cancer Cell* **16:** 521–532.

Moll UM, Slade N. 2004. p63 and p73: Roles in development and tumor formation. *Mol Cancer Res* **2:** 371–386.

Myatt SS, Lam EW. 2007. The emerging roles of forkhead box (Fox) proteins in cancer. *Nat Rev Cancer* **7:** 847–859.

Nakagawa H, Nuovo GJ, Zervos EE, Martin EW Jr, Salovaara R, Aaltonen LA, de la Chapelle A. 2001. Age-related hypermethylation of the 5′ region of MLH1 in normal colonic mucosa is associated with microsatellite-unstable colorectal cancer development. *Cancer Res* **61:** 6991–6995.

Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, Yu W, et al. 2007. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet* **39:** 237–242.

Ongenaert M, Van Neste L, De Meyer T, Menschaert G, Bekaert S, van Criekinge W. 2008. PubMeth: A cancer methylation database combining text-mining and expert annotation. *Nucleic Acids Res* **36:** D842–D846.

Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, et al. 2007. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* **39:** 232–236.

Scotto L, Narayan G, Nandula SV, Arias-Pulido H, Subramaniyam S, Schneider A, Kaufmann AM, Wright JD, Pothuri B, Mansukhani M, et al. 2008. Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: Potential role in progression. *Genes Chromosomes Cancer* **47:** 755–765.

So K, Tamura G, Honda T, Homma N, Waki T, Togawa N, Nishizuka S, Motoyama T. 2006. Multiple tumor suppressor genes are increasingly methylated with age in non-neoplastic gastric epithelia. *Cancer Sci* **97:** 1155–1158.

Song H, Ramus SJ, Tyrer J, Bolton KL, Gentry-Maharaj A, Wozniak E, Anton-Culver H, Chang-Claude J, Cramer DW, DiCioccio R, et al. 2009. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat Genet* **41:** 996–1000.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci* **100:** 9440–9445.

Talos F, Nemajerova A, Flores ER, Petrenko O, Moll UM. 2007. p73 suppresses polyploidy and aneuploidy in the absence of functional p53. *Mol Cell* **27:** 647–659.

Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, Jones A, Lechner M, Beck S, Jacobs I, et al. 2009. An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS One* **4:** e8274. doi: 10.1371/journal.pone.0008274.

Truty MJ, Lomberk G, Fernandez-Zapico ME, Urrutia R. 2009. Silencing of the transforming growth factor-β (TGFβ) receptor II by Kruppel-like factor 14 underscores the importance of a negative feedback mechanism in TGFβ signaling. *J Biol Chem* **284:** 6291–6300.

Wakana K, Akiyama Y, Aso T, Yuasa Y. 2006. Involvement of GATA-4/-5 transcription factors in ovarian carcinogenesis. *Cancer Lett* **241:** 281–288.

Weisenberger DJ, den Berg DV, Pan F, Berman BP, Laird PW. 2008. *Comprehensive DNA methylation analysis on the Illumina Infinium assay platform. Technical report.* Illumina, Inc., San Diego. http://www.illumina.com/Documents/products/appnotes/appnote_infinium_methylation.pdf.

Widschwendter M, Jiang G, Woods C, Müller HM, Fiegl H, Goebel G, Marth C, Mueller-Holzner E, Zeimet AG, Laird PW, et al. 2004. DNA hypomethylation and ovarian cancer biology. *Cancer Res* **64:** 4472–4480.

Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I, et al. 2007. Epigenetic stem cell signature in cancer. *Nat Genet* **39:** 157–158.

Wurmbach E, Chen YB, Khitrov G, Zhang W, Roayaie S, Schwartz M, Fiel I, Thung S, Mazzaferro V, Bruix J, et al. 2007. Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. *Hepatology* **45:** 938–947.