Newcastle University ePrints

Gautrey HE, van Otterdijk SD, Cordell HJ, Mathers JC, Stratthdee G, Newcastle 85 Study Core Team. [DNA methylation abnormalities at gene promoters are extensive and variable in the elderly and phenocopy cancer cells.](#)

The FASEB Journal 2014, 28(7), 3261-3272.

Copyright:

The definitive version of this article, published by the Federation of American Societies for Experimental Biology, 2014, is available at:

<http://dx.doi.org/10.1096/fj.13-246173>

Always use the definitive version when citing.

Further information on publisher website: <http://www.fasebj.org/>

Date deposited: 26th November 2014

Version of article: Submitted

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](#)

ePrints – Newcastle University ePrints

<http://eprint.ncl.ac.uk>

DNA methylation abnormalities at gene promoters are extensive and variable in the elderly and phenocopy cancer cells

Hannah E Gautrey¹, Sanne D. van Otterdijk¹, Heather J. Cordell², Newcastle 85+ study core team⁴, John C. Mathers³ and Gordon Strathdee¹

¹Northern Institute for Cancer Research, Newcastle University, Newcastle Upon Tyne, NE1 4LP;

²Institute for Genetic Medicine, Newcastle University, Newcastle Upon Tyne, NE1 3BZ; ³ Human Nutrition Research Centre, Institute for Ageing and Health, Campus for Ageing and Vitality, Newcastle University, Newcastle Upon Tyne, NE4 5PL

⁴Newcastle 85+ Study Core Team: John Bond, Joanna Collerton, Karen Davies (previously Barrass), Martin Eccles, Carol Jagger, Oliver James, Tom Kirkwood, Carmen Martin-Ruiz, Louise Robinson, Thomas von Zglinicki

Correspondence: Dr Gordon Strathdee
Northern Institute for Cancer Research
Newcastle University
5th Floor, Sir James Spence Institute
Royal Victoria Infirmary, Queen Victoria Road
Newcastle Upon Tyne, NE1 4LP
Gordon.Strathdee@ncl.ac.uk
Tel No.: 0191 282 1327

Running title: DNA methylation patterns in aging phenocopy cancer

Non Standard abbreviations

WBC – white blood cells

MNC – mononuclear cells

ALL - acute lymphoblastic leukemia

SNP – single nucleotide polymorphism

Abstract

Abnormal patterns of DNA methylation are one of the hallmarks of cancer cells. The process of aging has also been associated with similar, albeit less dramatic, changes in methylation patterns, leading to the hypothesis that age-related changes in DNA methylation may partially underlie the increased frequencies of cancer in the elderly. Here we studied 377 participants aged 85 years from the Newcastle 85+ Study, to investigate the extent of, and inter-individual variation in, age-related changes in DNA methylation at specific CpG islands. Using highly quantitative pyrosequencing analysis, we found extensive and highly variable methylation of promoter associated CpG islands with levels ranging from 4-35%, even at known tumor suppressor genes such as *TWIST2*.

Furthermore, the inter-individual differences in methylation seen across this elderly population phenocopies multiple features of the altered methylation seen in cancer cells. Both aging and cancer-related methylation can occur at similar sets of genes, both result in the formation of densely methylated, and likely transcriptionally repressed, alleles and both exhibit co-ordinate methylation across multiple loci. In addition, high methylation levels were associated with subsequent diagnosis of leukemia or lymphoma during a 3 year follow-up period ($p=0.00008$). These data suggest that the accumulation of age-related changes in promoter methylation may contribute to the increased cancer risk seen during aging.

Key words: epigenetics, aging, CpG island, cell type specific

Introduction

Cancer is one of the leading causes of death world-wide. The risk of almost all types of cancer increases with age and population aging is expected to be one of the primary drivers of the predicted dramatic increases in cancer incidence in the next 20 years (1). Thus, despite the improved success of treatment of most types of cancer, it is likely that the burden of cancer in the world's population will continue to increase.

DNA methylation, which occurs almost exclusively at CpG dinucleotides, is an important epigenetic mechanism for controlling gene expression (2). In particular, hypermethylation of gene promoters which have high densities of CpG sites, known as CpG islands, is associated with chromatin compaction and consequent loss of gene expression (3). This is a critical mechanism for control of expression of imprinted genes, for X-chromosome inactivation and for many germ cell specific genes (2). However, although the majority of human genes possess a promoter associated CpG island, these generally remain methylation free, even in tissues where the associated gene is not expressed (3). The development of cancer is associated with dramatic shifts in the pattern of DNA methylation (4). These include global reduction in genome wide levels of DNA methylation, highly variable methylation at CpG island shores and increased methylation of promoter-associated CpG islands. This increased CpG island methylation is often associated with gene silencing and the altered DNA methylation has been shown to be required for maintaining transcriptional inactivation of the linked gene (5). Many genes known to be important in cancer development, including the tumor suppressor genes *MLH1* and *BRCA1* (6) are inactivated by this mechanism.

However, these methylation changes are not unique to cancer and similar, albeit less dramatic, changes are also seen in apparently healthy individuals during aging (7). Cross-sectional studies have reported that global levels of DNA methylation reduce with age (8), although cohort studies, in which methylation was measured in the same individual several years apart, have shown that the accumulation of age-related methylation changes can be highly individual (9). Furthermore,

promoter associated CpG islands can exhibit increasing levels of DNA methylation with increasing age (7).

Based on these observations we hypothesized that the gradual acquisition of abnormalities in the pattern of DNA methylation during aging increases cancer risk and thus partially underlies the strong link between older age and greater risk of most types of cancer (10). In addition, it has been speculated that the altered methylation will likely result in age-related changes in gene expression which may have subsequent impacts on the development of other age-related illnesses (11). Several age-related illnesses are associated with changes in DNA methylation at specific genes, suggesting a potentially significant role for altered DNA methylation in the development of these disorders (11).

If the acquisition of age-related methylation changes varies widely between individuals, this could play a key role in differential susceptibility to the development of cancer and other diseases associated with aging. However, the extent of, and inter-individual variability in, CpG island methylation in apparently healthy people across the life-course is not well understood. To begin to assess whether differences in the acquisition of altered methylation patterns during aging impacts on health, we have examined CpG island methylation in peripheral blood DNA from a large cohort of 85 year olds from the Newcastle 85+ Study. This strategy has the advantage that, as all participants are the same age, all variation detected represents inter-individual variation with no confounding due to differences in chronological age. Furthermore, by studying a population of the oldest old, any inter-individual differences in age-associated methylation changes are likely to be maximized.

Materials and Methods

Study population

DNA was extracted from peripheral blood leukocytes from 377 participants from the Newcastle 85+ study. The analysis was initially performed as a pilot study on 50 samples (consisting of 25 survivors (at 3 years follow-up) and 25 non-survivors). After successful analysis of the pilot study samples the analysis was extended to all participants with a previous cancer history (n=113), all participants with an incident cancer diagnosis (at 3 years follow-up, n=72, 20 of whom were also in the previous cancer history group) and a control group without a history of cancer or an incident diagnosis (n=212). The methodology for the Newcastle 85+ Study has been reported (12, 13). In brief, members of the 1921 birth cohort living in Newcastle or North Tyneside (North-East England) were recruited at approximately age 85 through general practice patient lists. People living in institutions and those with cognitive impairment were included. The research complied with the requirements of the Declaration of Helsinki. Ethical approval was obtained from the Newcastle and North Tyneside 1 Research Ethics Committee (reference number 06/Q0905/2). Written informed consent was obtained from participants and where people lacked capacity to consent, for example because of dementia, a formal written opinion was sought from a relative or carer. Information on previous cancer history and incident cancer diagnoses (at 3 years) was obtained from the general practice medical records. In the UK, patients are registered with a single general practice which acts as a gatekeeper to secondary care and receives details of all hospital admissions and outpatient attendances. The review of general practice records included hospital correspondence to ensure that all recorded disease diagnoses were extracted irrespective of where and when the diagnosis was made.

To allow analysis of methylation at different ages another 150 DNA samples were obtained. These consisted of 50 DNA samples extracted from cord blood from newborns and 50 extracted from peripheral blood of their mothers (median age 28 years old) from the North Cumbria

Community Genetics Project (14) and 50 peripheral blood DNA samples from the 50 year old participants from the Newcastle Thousand Families Study (15).

DNA modification and PCR

200ng of genomic DNA was modified with sodium bisulfite using the Methylamp™ One-Step DNA Modification Kit (Epigentek, Brooklyn, NY, USA) as per the manufacturer's instructions. All samples were re-suspended in 15µl of TE and 1µl of this was used for subsequent PCR reactions. The samples were amplified in 25µl volumes containing 1X manufacturer's buffer, 1 unit of FastStart Taq polymerase (Roche, Welwyn Garden City, UK), 1-4mM MgCl₂, 10mM dNTPs, and 75ng of each primer. PCR was performed with one cycle of 95°C for 6min, 40 cycles of 95°C for 30sec, 57-63°C for 30sec and 72°C for 30sec, followed by one cycle of 72°C for 5min. For each set of primers (listed in Table 1) one of the forward or reverse primers included a 5' biotin label to allow for subsequent analysis by pyrosequencing.

Quantitative DNA methylation analysis using pyrosequencing

Following PCR amplification sequencing was performed using a PSQ 96MA pyrosequencer (Qiagen, Hilden, Germany), as per manufacturer's protocol. For all loci, assays were performed in duplicate and values averaged between the duplicates. Only samples that were passed by the pyrosequencer were included and to further ensure a high degree of accuracy only runs in which single peak heights were in excess of 200 were included. Samples in which the repeats varied in excess of 2% methylation were repeated one additional time and the average taken across the three repeats.

For each locus between 3 and 6 consecutive CpG sites were measured and the methylation value for each loci was taken as the mean of all CpG sites measured at that loci. Methylation values at consecutive CpG sites were strongly correlated, with correlation coefficients between consecutive

CpG sites ranging from 0.93-0.98 (*TWIST2*), 0.87-0.93 (*TUSC3*), 0.63-0.87 (*HAND2*), 0.80-0.86 (*HOXD4*), 0.42-0.64 (*EPHA10*) and 0.96-0.98 (*HOXA5*).

Primer design was performed using the manufacturer's provided PyroMark software. For initial assessment of all assays pyrosequencing was carried out on samples of known methylation status (5,10, 15, 20, 50, 70 and 100% methylated), which were produced by diluting peripheral blood derived DNA (from a young adult volunteer) into in vitro methylated DNA (Millipore, Watford, UK). . Only assays which produced correlations with R^2 of ≥ 0.90 were used for further analysis.

Derivation of overall methylation level

To produce a more accurate estimate of the comparative levels of age-related CpG island methylation in the samples, the values for each of the 5 age-related loci examined in the full sample set (*EPHA10*, *HOXD4*, *HAND2*, *TUSC3*, *TWIST2*) were averaged. The values at the 5 loci were weighted prior to averaging to ensure an equal contribution from each locus to the final average.

The percentage of lymphocytes in each peripheral blood sample was highly variable (8-56%) and was found to correlate strongly with overall methylation levels. To correct for this effect a correlation analysis was performed between overall methylation and lymphocyte percentage. The slope of the regression line produced was calculated to determine the average increase in methylation per 1% increase in lymphocyte percentage (slope =0.1811). This was then used to correct all values back to the average percentage of lymphocytes in the population (29.0%).

Bisulfite sequencing

Sodium bisulphite modification and PCR was carried out as above. PCR products were then cloned using the TA Cloning® Kit (Invitrogen Ltd, Paisley, UK), as per manufacturer's protocol. 2µl of the ligation product was transformed into One Shot TOP10 Chemically Competent E. coli cells

(Invitrogen Ltd, Paisley, UK) and single colonies obtained by plating on LB agar plates containing kanamycin (50µg/ml). Vector DNA was isolated and purified by lysis of bacterial cells and adsorption of DNA using QIAprep Spin Miniprep kit (Qiagen, Crawley, UK) according to the manufacturer's protocol and eluted in 50µl of elution buffer. The presence of the PCR amplified region of interest was confirmed by EcoRI (Fermentas, York, UK) restriction enzyme digestion, which cuts the vector either side of the incorporated PCR product. Sequencing of the inserts was then carried out using the Sp6 or T7 sequencing primers (Source Bioscience, Nottingham, UK).

Isolation of specific hematopoietic cell types

CD19⁺ B-lymphocytes and CD14⁺ monocytes were isolated using the DynaBeads system (Invitrogen, Paisley, UK), according to manufacturer's protocol. These cell populations were isolated from mononuclear cell populations, which were obtained after density gradient centrifugation of whole blood samples using Ficoll-Hypaque (BiochromAG, Berlin, Germany). 15 ml of whole blood was collected in an EDTA tube and re-suspended in 15 ml PBS. The blood-PBS mixture was carefully layered on top of 15 ml of Ficoll-Hypaque and centrifuged for 20 minutes at 800 x g. This resulted in several layers of cell types, including a layer of peripheral blood mononuclear cells. This mononuclear cell layer was carefully removed and put into a fresh tube. Mononuclear cells were then split into three and used directly for DNA isolation or for isolation of CD19⁺ or CD14⁺ cells.

The presence of the DynaBeads still attached to the isolated cells prevented assessment of purity by FACS analysis of cell surface markers and so purity of the isolated cell populations was validated using RT-PCR for the specific cell surface markers used to isolate the populations.

The levels of lymphocyte subpopulations in peripheral blood samples taken from Newcastle 85+ participants were determined using flow-cytometry (Becton Dickinson FACScan Flow Cytometer) for B Cells (CD19⁺) and T cells (CD3⁺, CD56⁻), as described before (16).

Statistical analysis

Correlations between methylation levels and fractions of specific cell types and correlations in methylation levels between the individual genes assessed were investigated by calculating the Pearson correlation coefficient. The Pearson correlation coefficient was also used to investigate relatedness in methylation at neighboring CpG sites at individual loci, as determined by pyrosequencing analysis. The Student t-test was used to assess differences in methylation at specific loci or average methylation (of *TWIST2*, *TUSC3*, *HOXD4*, *HAND2* and *EphA10*) between cohorts of different ages, with different histories of cancer diagnosis and between sorted cell types. Coefficients of variation in the initial 50 samples assessed were calculated by dividing the standard deviation by the mean methylation at each individual locus. The Mann-Whitney U test and Fisher's exact test were also used where indicated.

Results

Genes methylated in hematological malignancies show highly variable DNA methylation in peripheral blood from 85 year olds.

DNA methylation levels in peripheral blood samples were assessed using pyrosequencing analysis at 15 candidate genes (see Table 2), in an initial subset of 50 participants from the Newcastle 85+ Study. Pyrosequencing provides robust quantification of methylation levels so that even low levels of inter-individual variation can be quantified. The assays interrogated CpG sites in CpG islands surrounding the transcriptional start sites of the respective genes (except for *IGF2* where it targeted the imprinting control region), since these regions would be most likely to influence expression of the corresponding gene. The candidate genes were selected to include: genes known to be methylated in hematological malignancies (*EPHA10*, *HOXD4*, *HAND2*, *MNX1*, *TUSC3*, *TWIST2*, *p15* (17-21), and the authors' unpublished data); genes known to exhibit age or malignancy- associated methylation in other tissues (*ESR1*, *RARB2*, *RASSF1A*, *MLH1* (22-25)); and genes implicated in other age-related diseases (*APOE* (26)). The imprinted gene *IGF2* and *HOXA4* and *HOXA5*, which exhibit high levels of methylation throughout life (27, 28), were included as positive controls.

Methylation levels (averaged across all CpG sites assayed) were relatively high for several candidate genes including *EPHA10*, *HOXD4*, *HAND2*, *TUSC3* and *TWIST2*, which had maximal methylation levels of 18 - 35%. However, inter-individual variation in methylation levels was also substantial (see Fig. 1A-C for examples). Furthermore, such higher levels of methylation were not seen for all genes examined and methylation levels for some genes (such as *MLH1*, *RASSF1A* and *RARB2*) exhibited lower or near undetectable levels (Fig. 1 D-F). As shown in Table 2, higher levels of CpG island methylation were more common in genes known to be methylated in hematological malignancies. In contrast, genes known to exhibit age or malignancy associated methylation in other tissues or those implicated in non-malignant disease generally had lower levels

of methylation ($p=0.009$, Mann Whitney U test). Despite these inter-gene differences in absolute levels of methylation, inter-individual variation in methylation was similar across all genes (as quantified by the respective coefficients of variation (Table 2, $p=0.07$ for comparison of leukemia-related versus non-leukemia related genes, Mann Whitney U test)). As expected, high levels of methylation were found in the three control genes assessed (*HOXA5*, *HOXA4* and *IGF2* (Table 2)). The patterns of genes hypermethylated in different types of cancer are known to be cell type specific (29) and the results presented here are consistent with the hypothesis that the patterns of genes that accumulate methylation during aging are similarly cell type specific and may target similar sets of genes to those that are highly methylated in cancers derived from the same cell type.

Methylation levels in this initial subset of 50 participants were correlated significantly with the percentage of lymphocytes in the WBC population ($r^2=0.148$, $p=0.006$). Consequently, for all studies presented below, methylation levels were corrected for lymphocyte percentage as described in the Experimental Procedures section.

Genes which exhibit variable methylation in 85 year olds also exhibit age-related acquisition of methylation across the life-course.

Methylation levels at the five most highly methylated loci, the control loci and a largely unmethylated locus (*MLH1*) were also assessed in DNA from newborns, young adults and fifty-year olds. While methylation of the five highly methylated loci was detectable at all ages, this increased significantly with increasing age (i.e. newborns < young adults < 50 year olds < 85 year olds), confirming that the acquisition of methylation at these loci was age-related (Fig. 2, Supplemental Table S1). In contrast, the levels of methylation at the control loci, although consistently high, did not show age-related increases across the life-course (Supplemental Fig. S2, Supplemental Table S1) and methylation of *MLH1* remained low/undetectable methylation in all age groups (Fig. 2G).

Age-related methylation exhibits multiple features of malignancy-associated methylation.

As described above, there was significant overlap between the genes exhibiting higher methylation in DNA from peripheral blood in the subset of 50 participants from the Newcastle 85+ Study and those genes known to exhibit cancer-related methylation. To provide a more rigorous examination of this apparent concordance between aging and tumorigenesis, we examined DNA from an expanded set of samples derived from an additional 327 participants in the Newcastle 85+ Study. For this analysis we quantified methylation of the 5 genes which were most highly methylated during aging (*EPHA10*, *HOXD4*, *HAND2*, *TUSC3*, and *TWIST2*) together with the highly methylated, but not age-related, gene *HOXA5*. This expanded panel of samples confirmed the extensive and highly variable methylation at all five loci (Supplemental Fig. S1). One of the notable features of methylation in cancer is that methylation across multiple CpG islands is strongly correlated (30). To assess if a similar pattern is seen for age-related methylation, methylation levels (corrected for lymphocyte percentage) at all five loci were investigated for evidence of co-methylation. As shown in Table 3, we found highly significant correlations between methylation of each of the five genes with methylation levels of any of the other genes. In contrast, no significant correlation was found between methylation levels for any of these five loci and that for the highly methylated, but not age-related, gene *HOXA5*.

A further feature of cancer cells is the presence of densely methylated alleles, in which most or all CpG sites in a promoter associated CpG island are methylated. Such alleles have invariably been associated with lack of gene expression (31). To begin to investigate if the identified age-related methylation can lead to the appearance of densely methylated alleles, similar to those observed in cancer cells, we used bisulfite sequencing to investigate methylation at the allelic level in one of the participants with higher methylation at two loci (*TWIST2* and *HOXD4*). As shown in Fig. 3A, methylation was unequally distributed, with some alleles exhibiting dense methylation at all or most CpG sites assessed, while other alleles remained largely methylation free. This

demonstrates that age-related methylation can lead to the development of cancer-like densely methylated alleles that would likely be transcriptionally inactive. Consequently, clonal expansion of the cells that contain these cancer-like methylated alleles would result in cancer-like methylation patterns, at least at some loci, without the requirement for further alterations in DNA methylation during disease development.

High methylation levels may be detectable prior to the diagnosis of lymphoid malignancy

The similarities between the processes of age-associated and cancer-associated DNA methylation raises the possibility that altered methylation in apparently healthy cells may precede the development of clinically detectable disease and that age-associated methylation may predispose individuals to the development of cancer (32). To provide an overall estimate of the extent of age-related CpG island methylation in each sample, the methylation values for all 5 age-related genes were combined (as described in Experimental Procedures) to produce an overall methylation level. Interestingly, when the distribution of these combined methylation levels was examined, four of the participants (who exhibited the highest methylation levels) were found to lie outwith the normal distribution (Fig. 4). Of these, two participants were subsequently diagnosed with leukemia or lymphoma compared with none out of 373 participants without unusually high methylation levels ($p=0.00008$, Fisher exact test). As the diagnosis of malignant disease in these two patients was up to 2.5 years after the baseline blood sample was taken, this illustrates that increased methylation can be detectable for a significant period of time before the clinical diagnosis of disease.

To determine if the patterns of age-associated methylation detected in this population were more generally associated with cancer development, methylation levels were compared between participants who had had a diagnosis of any type of cancer prior to the study (excluding non-melanoma skin tumors, $n=113$) and participants with no history of cancer diagnosis ($n=212$).

Individuals with a previous diagnosis of cancer had significantly higher methylation levels, however the difference between individuals with and without a history of cancer was relatively small (10.40% vs. 10.06%, $p=0.04$). During a three year follow-up period, 72 individuals were diagnosed with cancer and methylation levels in this group of individuals were compared with those in the 212 cancer-free participants. This confirmed the link between increased methylation and cancer diagnosis, as individuals with an incident cancer diagnosis also exhibited significantly higher methylation levels at study baseline (10.59% vs. 10.06% respectively, $p=0.008$). However, as for participants with a previous cancer diagnosis, the absolute increase was small. Overall, the results suggest that dramatic increases in methylation levels may be detectable prior to cancer diagnosis only if examined in the corresponding tissue or cell type (e.g. peripheral blood for prediction of hematological malignancies). Small differences in the methylation levels of *HOXA5* were also seen (mean methylation levels 62.2% for no cancer controls, and 60.4% and 64.19%, for previous cancer and incident cancer groups respectively). However, these differences were not statistically significant and the direction of change was not consistent in the previous cancer and incident cancer groups.

DNA methylation is cell type specific and largely restricted to cells of lymphoid origin.

As described above DNA methylation levels correlated with the proportion of lymphocytes in the WBC population ($r^2=0.32$, $n=377$, $p\text{-value}<1\times10^{-40}$ for the whole cohort (Supplemental Fig. S3)). DNA from individual cell types from participants within the 85+ Study was not available. However we obtained whole blood samples from young adults. Multiple different cell populations were then purified from these samples and methylation was assessed in DNA extracted from the different cell populations. This analysis showed that for the five loci exhibiting age-related methylation, DNA from total lymphocytes exhibited higher methylation than total peripheral blood leukocytes and that methylation levels in purified myeloid cells (monocytes) were significantly

lower (examples in Fig. 3B). Even within the lymphoid compartment, methylation levels differed between cell types with B lymphocytes exhibiting significantly higher levels of CpG island methylation (Fig. 3B, Supplemental Table S2). While slight differences were seen in the non-age-related genes *HOXA4* and *HOXA5*, these were far less pronounced (Supplemental Table S2). These results are again consistent with a link between age-related and malignancy-related methylation, as all five loci are highly methylated specifically in lymphoid and not myeloid malignancies ((20, 33), unpublished data).

The above analysis is consistent with the correlation which we observed between total methylation levels and proportion of lymphocytes in the WBC population in the Newcastle 85 + study samples. However, as B lymphocytes exhibit relatively high levels of methylation, it further suggests that this correlation may be primarily driven by the fraction of B lymphocytes. Thus we examined the correlations between methylation levels and the proportion of B lymphocytes and T lymphocytes in the WBC population in the Newcastle 85 + study samples. We observed significant correlations for both cell fractions ($r^2=0.11$, $p\text{-value}<0.0001$, for B lymphocytes and $r^2=0.22$, $p\text{-value}<0.0001$, for T lymphocytes (Supplemental Fig. S3)), but these associations were not as strong as for the proportion of lymphocytes as a whole and there was no evidence that the association was specifically with B (as opposed to T) lymphocytes. Furthermore, the relative proportion of B lymphocytes within the total lymphocyte population did not correlate with methylation levels ($r^2=0.005$, $p\text{-value}=0.12$), nor did the relative proportion of T lymphocytes ($r^2=0.002$, $p\text{-value}=0.31$). The lack of correlation with the relative proportion of B cells is likely due to the comparatively low number of B lymphocytes as a proportion of total lymphocytes (average 8.4%).

Discussion

Both tumor development and aging are associated with similar alterations in the patterns of genomic DNA methylation. Both processes result in loss of genome wide methylation alongside gains in promoter associated CpG island methylation. This has raised the possibility that accumulation of age-related methylation changes may increase the likelihood of malignant transformation and thus partly explain the strong link between older age and incidence of most cancer types (34). Here we sought to quantify the extent of, and inter-individual variability in, CpG island methylation occurring during aging. This analysis demonstrated that even in cancer free individuals, methylation levels at promoter associated CpG islands can reach high levels, even at presumptive tumor suppressor genes such as *TWIST2* and *TUSC3*. For some genes, we observed extensive variation in methylation not only between different age groups but also between individuals of the same age, raising the possibility that differential levels of accumulation of age-related methylation could play an important role in determining susceptibility to the development of age-associated disorders, particularly cancer.

This analysis found a clear similarity between the patterns of CpG island methylation resulting from aging and those previously attributed to the development of cancer. This included the altered methylation in a common set of genes, co-ordinate methylation of multiple loci at multiple sites throughout the genome and the presence of densely methylated, and likely transcriptionally inactive, alleles. This raises two interesting possibilities. One potential explanation for this would be that the mechanisms responsible for cancer-related and aging-related alterations in DNA methylation may be largely the same. Alternatively, it may be that a significant portion of the abnormal methylation found in cancer cells is not acquired during the disease process, but could instead be produced by clonal expansion of cells which already contain the abnormal methylation patterns. If true, this would suggest that cells with altered methylation states may be at much greater

risk for malignant transformation, as suggested by the epigenetic progenitor origin of cancer hypothesis (35).

Several studies have previously demonstrated an overlap between age-related and cancer-related methylation. These include studies showing that both processes preferentially involve genes regulated by the polycomb group complexes (36), genes which exhibit bivalent chromatin domains (37) and genes associated with inter-individual epialleles (38). Our results agreed with this link between age-related and cancer-related methylation, as methylation levels were significantly higher in genes with a known link to leukemia compared with those chosen due to age/cancer-related methylation in other tissues. However, by using highly quantitative methylation analysis our results also emphasize that the development of age-related methylation is not inevitable. This is true for certain genes, such as *MLH1*, for which there was no increased methylation across the life-course from newborns to 85 years olds. In addition, it was also true for certain individuals even in genes displaying clear age-related methylation. The lower range of methylation seen in the 85 year old population was equivalent to the average level of methylation in the newborn samples. Indicating that some individuals can accumulate very little age-related methylation (at least at the loci assessed here) even by 85 years old. This lack of inevitability of age-related methylation suggests that a combination of genetic background and environmental/dietary factors may be able to limit or prevent any negative impact the accumulation of age-related methylation changes has on health.

Co-ordinate methylation across multiple promoter associated CpG islands has been found in many types of cancer. For example, it has been suggested that there is a CpG island methylator phenotype in colorectal cancer in which a subset of tumors accumulate much more widespread methylation of CpG islands (39). Although many other studies have described a CpG island methylator phenotypes in other cancers, in the main these have been limited to identifying co-methylation of genes, i.e. tumors in which high methylation of one gene is associated with increased likelihood of higher methylation at the other genes assessed in the study (40). Indeed in a genome

wide study of childhood acute lymphoblastic leukemia (ALL), we have found that >1000 loci (including all 5 age-related genes investigated in the present study) exhibited coordinated methylation levels (i.e. their levels of methylation were significantly correlated, unpublished data). Thus the co-ordinate methylation of genes seen in the present study (as detailed in Table 3) appears to be analogous to the co-ordinate methylation previously reported in multiple types of cancer. Furthermore, as pointed out by Teschendorff and colleagues in a recent review, a key question regarding age-related methylation changes is whether they occur predominantly at individual CpG sites or are spatially related as in cancer (41). Here we show that not only do neighboring CpG sites show strong correlations in methylation levels, but that these correlations extend to multiple loci located on separate chromosomes. Thus the potential impact of age-related methylation on biological function has to be viewed not just in the context of changes at a single locus, but also co-ordinate changes in methylation across many loci.

The variability in methylation even among healthy individuals which we report here may be driven, at least in part, by underlying genetic variation. Twin and population studies have demonstrated that epigenetic variation can be associated with genetic variation (42, 43). However, these studies have predominantly found SNPs to be associated *in cis* with methylation at closely linked loci (43). In this study we found that variability at the age-related loci was correlated and the loci exhibit co-ordinate patterns of methylation. Thus any genetic variation underlying this variability would have to influence methylation levels at multiple loci, unlike the site specific correlations observed to date. The possibility of SNPs related to methylation at a genome wide level has also been investigated using candidate based approaches targeting genes regulating one carbon metabolism or the DNA methyltransferase genes, but no clear impact on DNA methylation levels has been identified (44). An alternative explanation is that the variation observed is driven by extrinsic factors and a number of environmental exposures and dietary factors have been reported to alter epigenetic states (45).

The detection of higher levels of methylation in individuals prior to clinical diagnosis of leukemia/lymphoma raises the possibility that this could be used for early detection or even identification of high risk individuals. This would be consistent with the observation that the two cases of leukemia/lymphoma that were subsequently diagnosed in the study population occurred in the participants with the highest baseline levels of methylation. It is also notable that while methylation of the genes examined increased with age, aberrant promoter methylation of the same genes was detectable in PBL samples at all ages (even at birth), suggesting that cells with pre-existing methylation abnormalities could underlie the development of cancer at all stages throughout the life-course.

In addition, CpG island methylation levels were higher in participants with a cancer diagnosis (previous diagnosis or incident cancer). While the increases observed (versus participants with no cancer history) were statistically significant, the absolute differences were very small and thus assessing methylation levels at the genes in this study would not provide useful information regarding overall cancer risk. However, analysis of an expanded set of loci or genome-wide methylation studies may identify loci whose methylation status is more predictive of risk of non-hematological cancers. Indeed, Brennan et al have recently demonstrated that intragenic methylation at the ATM locus in peripheral blood samples is associated with risk of breast cancer development (46). No statistically significant correlation was found with methylation of the *HOXA5* gene and cancer diagnosis. This may have been due to a lack of power in the study, as small differences were observed. However, these changes were not consistent (participants with a previous diagnosis had lower methylation levels, whilst those with an incident diagnosis had higher levels) and the average methylation levels at *HOXA5* across all participants with a cancer diagnosis was almost identical to controls.

A potential explanation for the similarity in genes showing age-related and cancer-related methylation could be that non-expressed or functionally unimportant genes may be the predominant

targets for both processes. Indeed a number of recent studies have found relatively low correlations between aging associated changes in methylation and gene expression (47, 48) and studies in cancer have also found that hypermethylation of promoter associated CpG islands is predominantly associated with genes that are not expressed in the corresponding normal tissue (49). However, this is clearly not always the case. The *TWIST2* gene exhibits tumor suppressor like properties in B cell leukemia and is expressed in normal differentiated B lymphocytes (20, 50) and yet methylation of this gene was not only present in peripheral blood samples, but is predominantly seen in B lymphocytes; the same cell type in which it appears to play a role in preventing malignant transformation. Subsequent bisulfite sequencing showed that this methylation was characterized by the presence of densely methylated alleles. We have observed similar densely methylated alleles in ALL and shown that this methylation maintains transcriptional repression of the gene (20). As age-related methylation, like cancer-related methylation, can affect a large number of loci across the genome (36, 37), only a fraction of these epigenetic changes would need to target functionally relevant genes to result in significant biological effects.

Methylation levels varied between the different cell types in sorted peripheral blood samples from younger individuals and were largely lymphoid specific. Furthermore, even within in lymphoid cells, methylation levels were higher in B lymphocytes. This raises two interesting aspects. Firstly, although B cells are one of the less common cell types in the blood, malignancies derived from B cells make up over half of hematological cancers. The propensity of B lymphocytes for malignant transformation may be explained, in part, by the intrinsically high levels of methylation observed in this cell type, further strengthening the link between age-related and cancer-related methylation. Secondly, the conceptually simplest explanation for the development of age-related methylation is that such changes accumulate in stem cells over time, leading to gradual and permanent increases in CpG island methylation in the tissue derived from these stem cells. However, the cell type specificity seen here in cells derived from the same pool of stem cells is not

fully compatible with this hypothesis and suggests that alternative mechanisms are at least partly responsible for age-related methylation. These could involve preferential lymphoid differentiation of stem cells with a high CpG island methylation content, differential expansion of short “seeds” of methylation in the different cell types (with the “seeds” accumulating in stem cells) or gradual deterioration of the fidelity of replicating DNA methylation patterns with increasing age, specifically in B lymphocytes. Methylation levels could be further influenced by the lifespan of the different cell types, as lymphoid cells are generally more long-lived than myeloid cells.

Overall these results strengthen the link between cancer-related and age-related methylation patterns and show that age-related methylation in specific sets of genes can phenocopy the main features of altered methylation patterns seen in cancer cells, implying that the two processes may be interlinked.

Acknowledgements

Thanks are especially due to the 85 year olds of Newcastle and North Tyneside, and their families and carers, for the generous donation of their time and personal information to make the study possible. In addition we thank the: research nurses (Brenda Balderson, Sally Barker, Julie Burrows, June Edwards, Julie Ferguson, Gill Hedley, Joan Hughes, Judith Hunt, Julie Kimber, and Victoria Raynor); biomarker technicians (Sam Jameson, Claire Kolenda, Craig Parker, and Anna Tang); data manager (Pauline Potts); project secretary (Lucy Farfort); and the North of England Commissioning Support Unit (formerly NHS North of Tyne, working on behalf of Newcastle PCT, North Tyneside PCT and Northumberland Care Trust) and local general practices.

The baseline phase of the Newcastle 85+ Study was funded by a combined grant from the UK Medical Research Council and the Biotechnology and Biological Sciences Research Council (G0500997), and a grant from the Newcastle Healthcare Charity. The research was also supported by the National Institute for Health Research Newcastle Biomedical Research Centre, based at Newcastle upon Tyne Hospitals NHS Foundation Trust and Newcastle University. The Rank Prize Funds supported purchase of the pyrosequencer. HG was funded by a PhD studentship from the Biotechnology and Biological Sciences Research Council and SvO was funded by a PhD studentship from the National Institute for Health Research Newcastle Biomedical Research Centre.

References

1. Bray, F., Jemal, A., Grey, N., Ferlay, J., and Forman, D. (2012) Global cancer transitions according to the Human Development Index (2008-2030): a population-based study. *Lancet Oncol* **13**, 790-801
2. Costello, J. F., and Plass, C. (2001) Methylation matters. *J Med Genet* **38**, 285-303
3. Bird, A. P., and Wolffe, A. P. (1999) Methylation-induced repression--belts, braces, and chromatin. *Cell* **99**, 451-454
4. Kulis, M., and Esteller, M. (2010) DNA methylation and cancer. *Adv Genet* **70**, 27-56
5. Fahrner, J. A., Eguchi, S., Herman, J. G., and Baylin, S. B. (2002) Dependence of histone modifications and gene expression on DNA hypermethylation in cancer. *Cancer Res* **62**, 7213-7218
6. Jacinto, F. V., and Esteller, M. (2007) Mutator pathways unleashed by epigenetic silencing in human cancer. *Mutagenesis* **22**, 247-253
7. Fraga, M. F., and Esteller, M. (2007) Epigenetics and aging: the targets and the marks. *Trends Genet* **23**, 413-418
8. Bollati, V., Schwartz, J., Wright, R., Litonjua, A., Tarantini, L., Suh, H., Sparrow, D., Vokonas, P., and Baccarelli, A. (2009) Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech Ageing Dev* **130**, 234-239
9. Bjornsson, H. T., Sigurdsson, M. I., Fallin, M. D., Irizarry, R. A., Aspelund, T., Cui, H., Yu, W., Rongione, M. A., Ekstrom, T. J., Harris, T. B., Launer, L. J., Eiriksdottir, G., Leppert, M. F., Sapienza, C., Gudnason, V., and Feinberg, A. P. (2008) Intra-individual change over time in DNA methylation with familial clustering. *JAMA* **299**, 2877-2883
10. DePinho, R. A. (2000) The age of cancer. *Nature* **408**, 248-254
11. Portela, A., and Esteller, M. (2010) Epigenetic modifications and human disease. *Nat Biotechnol* **28**, 1057-1068
12. Collerton, J., Barrass, K., Bond, J., Eccles, M., Jagger, C., James, O., Martin-Ruiz, C., Robinson, L., von Zglinicki, T., and Kirkwood, T. (2007) The Newcastle 85+ study: biological, clinical and psychosocial factors associated with healthy ageing: study protocol. *BMC Geriatr* **7**, 14
13. Collerton, J., Davies, K., Jagger, C., Kingston, A., Bond, J., Eccles, M. P., Robinson, L. A., Martin-Ruiz, C., von Zglinicki, T., James, O. F., and Kirkwood, T. B. (2009) Health and disease in 85 year olds: baseline findings from the Newcastle 85+ cohort study. *BMJ* **339**, b4904
14. Chase, D. S., Tawn, E. J., Parker, L., Jonas, P., Parker, C. O., and Burn, J. (1998) The North Cumbria Community Genetics Project. *J Med Genet* **35**, 413-416
15. Pearce, M. S., Unwin, N. C., Parker, L., and Craft, A. W. (2009) Cohort profile: the Newcastle Thousand Families 1947 birth cohort. *Int J Epidemiol* **38**, 932-937
16. Martin-Ruiz, C., Jagger, C., Kingston, A., Collerton, J., Catt, M., Davies, K., Dunn, M., Hilkins, C., Keavney, B., Pearce, S. H., den Elzen, W. P., Talbot, D., Wiley, L., Bond, J., Mathers, J. C., Eccles, M. P., Robinson, L., James, O., Kirkwood, T. B., and von Zglinicki, T. (2011) Assessment of a large panel of candidate biomarkers of ageing in the Newcastle 85+ study. *Mech Ageing Dev* **132**, 496-502
17. Ferguson, S., Gautrey, H. E., and Strathdee, G. (2011) The dual role of HLXB9 in leukemia. *Pediatr Blood Cancer* **56**, 349-352
18. Herman, J. G., Jen, J., Merlo, A., and Baylin, S. B. (1996) Hypermethylation-associated inactivation indicates a tumor suppressor role for p15INK4B. *Cancer Res* **56**, 722-727
19. Kuang, S. Q., Bai, H., Fang, Z. H., Lopez, G., Yang, H., Tong, W., Wang, Z. Z., and Garcia-Manero, G. (2010) Aberrant DNA methylation and epigenetic inactivation of Eph receptor tyrosine kinases and ephrin ligands in acute lymphoblastic leukemia. *Blood* **115**, 2412-2419
20. Thathia, S. H., Ferguson, S., Gautrey, H. E., van Otterdijk, S. D., Hili, M., Rand, V., Moorman, A. V., Meyer, S., Brown, R., and Strathdee, G. (2012) Epigenetic inactivation of TWIST2 in acute lymphoblastic leukemia modulates proliferation, cell survival and chemosensitivity. *Haematologica* **97**, 371-378
21. Tong, W. G., Wierda, W. G., Lin, E., Kuang, S. Q., Bekele, B. N., Estrov, Z., Wei, Y., Yang, H., Keating, M. J., and Garcia-Manero, G. (2010) Genome-wide DNA methylation profiling of chronic

- lymphocytic leukemia allows identification of epigenetically repressed molecular pathways with clinical impact. *Epigenetics* **5**, 499-508
22. Di Gioia, S., Bianchi, P., Destro, A., Grizzi, F., Malesci, A., Laghi, L., Levrero, M., Morabito, A., and Roncalli, M. (2006) Quantitative evaluation of RASSF1A methylation in the non-lesional, regenerative and neoplastic liver. *BMC Cancer* **6**, 89
 23. Kanaya, T., Kyo, S., Maida, Y., Yatabe, N., Tanaka, M., Nakamura, M., and Inoue, M. (2003) Frequent hypermethylation of MLH1 promoter in normal endometrium of patients with endometrial cancers. *Oncogene* **22**, 2352-2360
 24. Kwabi-Addo, B., Chung, W., Shen, L., Ittmann, M., Wheeler, T., Jelinek, J., and Issa, J. P. (2007) Age-related DNA methylation changes in normal human prostate tissues. *Clin Cancer Res* **13**, 3796-3802
 25. Tamura, G. (2004) Promoter methylation status of tumor suppressor and tumor-related genes in neoplastic and non-neoplastic gastric epithelia. *Histol Histopathol* **19**, 221-228
 26. Yu, C. E., Cudaback, E., Foraker, J., Thomson, Z., Leong, L., Lutz, F., Gill, J. A., Saxton, A., Kraemer, B., Navas, P., Keene, C. D., Montine, T., and Bekris, L. M. (2013) Epigenetic signature and enhancer activity of the human APOE gene. *Hum Mol Genet*
 27. Strathdee, G., Sim, A., Soutar, R., Holyoake, T. L., and Brown, R. (2007) HOXA5 is targeted by cell-type-specific CpG island methylation in normal cells and during the development of acute myeloid leukaemia. *Carcinogenesis* **28**, 299-309
 28. Sullivan, M. J., Taniguchi, T., Jhee, A., Kerr, N., and Reeve, A. E. (1999) Relaxation of IGF2 imprinting in Wilms tumours associated with specific changes in IGF2 methylation. *Oncogene* **18**, 7527-7534
 29. Costello, J. F., Fruhwald, M. C., Smiraglia, D. J., Rush, L. J., Robertson, G. P., Gao, X., Wright, F. A., Feramisco, J. D., Peltomaki, P., Lang, J. C., Schuller, D. E., Yu, L., Bloomfield, C. D., Caligiuri, M. A., Yates, A., Nishikawa, R., Su Huang, H., Petrelli, N. J., Zhang, X., O'Dorisio, M. S., Held, W. A., Cavenee, W. K., and Plass, C. (2000) Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* **24**, 132-138
 30. Toyota, M., and Issa, J. P. (1999) CpG island methylator phenotypes in aging and cancer. *Semin Cancer Biol* **9**, 349-357
 31. Jones, P. A., and Laird, P. W. (1999) Cancer epigenetics comes of age. *Nat Genet* **21**, 163-167
 32. Tapp, H. S., Commane, D. M., Bradburn, D. M., Arasaradnam, R., Mathers, J. C., Johnson, I. T., and Belshaw, N. J. (2013) Nutritional factors and gender influence age-related DNA methylation in the human rectal mucosa. *Aging Cell* **12**, 148-155
 33. Scholz, C., Nimmrich, I., Burger, M., Becker, E., Dorken, B., Ludwig, W. D., and Maier, S. (2005) Distinction of acute lymphoblastic leukemia from acute myeloid leukemia through microarray-based DNA methylation analysis. *Ann Hematol* **84**, 236-244
 34. Issa, J. P. (1999) Aging, DNA methylation and cancer. *Crit Rev Oncol Hematol* **32**, 31-43
 35. Feinberg, A. P., Ohlsson, R., and Henikoff, S. (2006) The epigenetic progenitor origin of human cancer. *Nat Rev Genet* **7**, 21-33
 36. Teschendorff, A. E., Menon, U., Gentry-Maharaj, A., Ramus, S. J., Weisenberger, D. J., Shen, H., Campan, M., Noushmehr, H., Bell, C. G., Maxwell, A. P., Savage, D. A., Mueller-Holzner, E., Marth, C., Kocjan, G., Gayther, S. A., Jones, A., Beck, S., Wagner, W., Laird, P. W., Jacobs, I. J., and Widschwendter, M. (2010) Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* **20**, 440-446
 37. Rakyan, V. K., Down, T. A., Maslau, S., Andrew, T., Yang, T. P., Beyan, H., Whittaker, P., McCann, O. T., Finer, S., Valdes, A. M., Leslie, R. D., Deloukas, P., and Spector, T. D. (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* **20**, 434-439
 38. Gemma, C., Ramagopalan, S. V., Down, T. A., Beyan, H., Hawa, M. I., Holland, M. L., Hurd, P. J., Giovannoni, G., David Leslie, R., Ebers, G. C., and Rakyan, V. K. (2013) Inactive or moderately active human promoters are enriched for inter-individual epialleles. *Genome Biol* **14**, R43
 39. Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J. G., Baylin, S. B., and Issa, J. P. (1999) CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* **96**, 8681-8686

40. Hughes, L. A., Melotte, V., de Schrijver, J., de Maat, M., Smit, V. T., Bovee, J. V., French, P. J., van den Brandt, P. A., Schouten, L. J., de Meyer, T., van Criekinge, W., Ahuja, N., Herman, J. G., Weijenberg, M. P., and van Engeland, M. (2013) The CpG Island Methylator Phenotype: What's in a Name? *Cancer Res*
41. Teschendorff, A. E., West, J., and Beck, S. (2013) Age-associated epigenetic drift: implications, and a case of epigenetic thrift? *Hum Mol Genet*
42. Zhang, D., Cheng, L., Badner, J. A., Chen, C., Chen, Q., Luo, W., Craig, D. W., Redman, M., Gershon, E. S., and Liu, C. (2010) Genetic control of individual differences in gene-specific methylation in human brain. *Am J Hum Genet* **86**, 411-419
43. Boks, M. P., Derks, E. M., Weisenberger, D. J., Strengman, E., Janson, E., Sommer, I. E., Kahn, R. S., and Ophoff, R. A. (2009) The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One* **4**, e6767
44. Huidobro, C., Fernandez, A. F., and Fraga, M. F. (2013) The role of genetics in the establishment and maintenance of the epigenome. *Cell Mol Life Sci* **70**, 1543-1573
45. Mathers, J. C., Strathdee, G., and Relton, C. L. (2010) Induction of epigenetic alterations by dietary and other environmental factors. *Adv Genet* **71**, 3-39
46. Brennan, K., Garcia-Closas, M., Orr, N., Fletcher, O., Jones, M., Ashworth, A., Swerdlow, A., Thorne, H., Riboli, E., Vineis, P., Dorronsoro, M., Clavel-Chapelon, F., Panico, S., Onland-Moret, N. C., Trichopoulos, D., Kaaks, R., Khaw, K. T., Brown, R., and Flanagan, J. M. (2012) Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk. *Cancer Res* **72**, 2304-2313
47. Beerman, I., Bock, C., Garrison, B. S., Smith, Z. D., Gu, H., Meissner, A., and Rossi, D. J. (2013) Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging. *Cell Stem Cell* **12**, 413-425
48. Bock, C., Beerman, I., Lien, W. H., Smith, Z. D., Gu, H., Boyle, P., Gnirke, A., Fuchs, E., Rossi, D. J., and Meissner, A. (2012) DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Mol Cell* **47**, 633-647
49. Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R. A., Niveleau, A., Cedar, H., and Simon, I. (2006) Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat Genet* **38**, 149-153
50. Raval, A., Lucas, D. M., Matkovic, J. J., Bennett, K. L., Liyanarachchi, S., Young, D. C., Rassenti, L., Kipps, T. J., Grever, M. R., Byrd, J. C., and Plass, C. (2005) TWIST2 demonstrates differential methylation in immunoglobulin variable heavy chain mutated and unmutated chronic lymphocytic leukemia. *J Clin Oncol* **23**, 3877-3885

Tables

Table 1 Primer sequences used for pyrosequencing assays

Gene	Forward Primer	Reverse Primer	Sequencing Primer	CpG Sites	Distance from TSS
APOE	GGAGTTTATAATTGGATAAGT	AAAATCCCAACTCTTCT (Bio)	GGGATTTTGGATTTTATT	3	-24 → +123
EPHA10	ATTATTTATGGGAATTAATTATTAG	CCCACCAATTAATATTC (Bio)	TATTTATGGGAATTAATTATTAAGTA	3	+28 → +143
ESR1	GGAGATTAGTATTTAAAGTTGGAGGT	CCTAAAAAACAACACCC (Bio)	TGGGATTGTATTTGTTTT	5	+92 → +294
HAND2	CCAAATTTAATTATCTTATA (Bio)	GGATTTTGTAGTAAGATTT	ATTTTATTACCCAAACTC	4	-29 → +185
HLXB9	TAGTTYGGGTTGTTAGGAT	TACRCTACAACCTCAAA (Bio)	GAAGTYGYGTTGAGTTTAAAG	4	-161 → +61
HOXA4	TACACTTCACAAATTAATAACCATAACTC (Bio)	GTTGTTGTAGYGGTAGGTGTTG	AACCCAAATCCCTCCCTT	4	+31 → +219
HOXA5	CCACAAATCAAAACACATATC (Bio)	GGTTGGTGTATTTGGGT	CTACRACTACAATAACATAAATCT	5	+30 → +327
HOXD4	GAAATTAATGGTTATGAGT	CCCRAACTAAAAATCTA (Bio)	ATTTGTAAGGYGGTTATTTAGGYGAG	5	+242 → +398
IGF2	GGGTTAAGGTAGTTTTTTGGGAATG	AAATATAAAACCTCTCCACCTCC (Bio)	AGGGGGTTTATTTTTTAGGAA	3	DMR 0
MLH1	AGTTTTTTTTTAGGAGTGAAGG	ATAAAACCTTATACCTATCTATC (Bio)	GTA GTATTYGTGTTTAGTTT	6	-569 → -375
TUSC3	GAATAGGATGTTTGTAGTT (Bio)	TAACTAAACCTCCCAATA	CCAAATAACAACCACTTAC	6	+25 → +116
p15	GGTTTGGGGTTTYTAGTGG	CRACCCCTTAACCAACTAAAAAC (Bio)	GAGGAA GTAGTTTAGTTTAAAG	5	-207 → +41
RARB2	TATAGAGGAATTTAAAGTGGGTG	AAAACCTAAAACTCCCAACAACCTC (Bio)	GTTTTTAAGGTGAGAAATAG	4	-416 → -118
RASSF1A	GGGGGA GTTTGAGTTTATT GA	CTACCCCTTAACTACCCCTTC (Bio)	GGGTYGTATTYGGTT GGAG	5	+136 → +434
TWIST2	AACAACATTTTAAACCAACCAAC	GGGYGAGTTGGAGTTTTTTTTATGG (Bio)	CAAAACCTTTCAACAAC	4	-26 → +208

Table 2 - Variation in CpG island methylation levels

Gene	Methylation range (%)	Mean (%)	SD (%)	CoV ¹	Associated with leukemia/lymphoma
HOXD4	7-35	14.8	6	0.399	Yes
EPHA10	7-31	12	3.9	0.318	Yes
TUSC3	4-22	8.9	3.4	0.38	Yes
TWIST2	4-20	8.4	3.3	0.395	Yes
HAND2	2-18	7.3	2.8	0.382	Yes
HLXB9	1-15	3.3	2.1	0.371	Yes
APOE	6-13	8.4	1.9	0.22	No
p15	2-12	4.9	1.9	0.381	Yes
ESR1	4-11	5.5	1.6	0.292	No
RARB2	1-5	2.5	1	0.397	No
RASSF1A	1-4	2.5	0.7	0.257	No
MLH1	1-2	1.3	0.3	0.256	No
Control genes					
HOXA5	28-92	61.5	15.1		
HOXA4	13-64	41.5	11.8		
IGF2	30-50	40.2	4.4		

¹CoV; Coefficient of variance

Table 3 - Co-methylation of genes in 85+ samples

	<i>TWIST2</i>	<i>TUSC3</i>	<i>HOXD4</i>	<i>HAND2</i>	<i>EPHA10</i>
<i>TUSC3</i>	¹ <i>r</i> = 0.60 <i>p</i> < 1x10 ⁻⁷				
<i>HOXD4</i>	<i>r</i> = 0.55 <i>p</i> < 1x10 ⁻⁷	<i>r</i> = 0.55 <i>p</i> < 1x10 ⁻⁷			
<i>HAND2</i>	<i>r</i> = 0.29 <i>p</i> < 1x10 ⁻⁷	<i>r</i> = 0.31 <i>p</i> < 1x10 ⁻⁷	<i>r</i> = 0.38 <i>p</i> < 1x10 ⁻⁷		
<i>EPHA10</i>	<i>r</i> = 0.32 <i>p</i> < 1x10 ⁻⁷	<i>r</i> = 0.33 <i>p</i> < 1x10 ⁻⁷	<i>r</i> = 0.20 <i>p</i> < 0.0001	<i>r</i> = 0.32 <i>p</i> < 1x10 ⁻⁷	
<i>HOXA5</i>	<i>r</i> = 0.04 <i>p</i> = 0.62	<i>r</i> = -0.03 <i>p</i> = 0.72	<i>r</i> = -0.002 <i>p</i> = 0.98	<i>r</i> = -0.02 <i>p</i> = 0.77	<i>r</i> = 0.003 <i>p</i> = 0.97

¹Pearson correlation coefficient

Figure Legends

Figure 1. Examples of variation in methylation in peripheral blood samples.

Examples of genes with high levels (A-C) or low levels (D-F) of age-related methylation in Newcastle 85+ study samples. Methylation levels were assessed by pyrosequencing in each of the indicated genes in an initial sample set from 50 85+ study participants. Y-axis scale is the same for all genes (0-25%), except for *HOXD4* (0-40%), as this gene exhibited the highest levels of methylation. Error bars indicate standard error of the mean.

Figure 2. Methylation levels of 5 most variable genes are age-related. Methylation levels at the indicated genes were assessed by pyrosequencing in individuals from 3 younger age groups (50 participants in each, median ages 0, 27 and 50 years), in addition to the Newcastle 85+ participants. The combination of the 5 most variable genes (A), as well as each of the 5 genes individually (B-F) exhibit increasing levels of methylation with increasing age. In contrast the *MLH1* gene (G) exhibits no significant differences between the age groups.

Figure 3. DNA methylation is non-randomly distributed in peripheral blood. A)

Bisulfite sequencing at the *TWIST2* and *HOXD4* loci. To assess methylation at the allelic level, bisulfite sequencing was performed at two loci (*TWIST2* and *HOXD4*, as indicated) in a sample with high levels of methylation at both loci. The CpG sites within each allele are represented by the horizontal rows of circles. A filled in circle indicates methylated and an open circle the absence of methylation. For *TWIST2*, 23 consecutive CpG sites, between +38 and +181 relative to the transcriptional start site, were sequenced. For *HOXD4*, 13 consecutive CpG sites, between +262 and +377 relative to the transcriptional

start site, were sequenced. **B)** DNA methylation is highly variable between cell types. Specific cell subsets were isolated from 10 peripheral blood samples. These included total peripheral blood leukocytes (PBL), mononuclear cells (MNC), B lymphocytes (B-cells), monocytes and total lymphocytes. All five genes exhibited very similar distributions across the cell types as illustrated by *TWIST2* and *TUSC3*. Methylation levels for all genes assessed are shown in Supplemental Table S2.

Figure 4. Q-Q plot to assess the normality of distribution of methylation data.

Methylation for all samples, averaged across the five genes and normalized for percentage of lymphocytes, was found to exhibit a largely normal distribution. Four samples with the highest methylation were found to lie above the line of normal distribution. As indicated two of those individuals were subsequently diagnosed with leukemia/lymphoma.

Figure 1

Figure 2

Figure 3

A

B

Figure 4

Supplemental Fig. 1

Supplemental Figure S1. Methylation levels of 5 age-related and one control gene in expanded sample set. Pyrosequencing analysis of 5 age-related genes and *HOXA5* in expanded sample set, confirming high levels of methylation and extensive variation across the Newcastle 85+ sample set.

Supplemental Table 1 - Methylation changes significantly between individuals of different ages

Gene name	<i>TWIST2</i>	<i>TUSC3</i>	<i>HOXD4</i>	<i>EPHA10</i>	<i>HAND2</i>	5 genes combined ¹	<i>MLH1</i>
Newborn vs. young adult (27Y) ²	0.886 ³	1.19X10 ⁻¹⁰	5.21X10 ⁻¹⁷	0.003	5.14X10 ⁻¹⁸	1.02X10 ⁻¹⁹	0.704
young adult (27Y) vs. older adult (50Y)	5.07X10 ⁻⁰⁷	1.16X10 ⁻¹⁴	7.08X10 ⁻⁰⁸	1.66X10 ⁻⁰⁶	5.75X10 ⁻¹⁰	1.05X10 ⁻¹⁶	0.214
Older adult (50Y) vs. elderly (85Y)	2.93X10 ⁻⁰⁷	0.008	1.04X10 ⁻⁰⁴	3.97X10 ⁻⁰⁸	0.641	1.42X10 ⁻⁰⁶	0.002
Control Gene name	<i>HOXA5</i>	<i>HOXA4</i>	<i>IGF2</i>				
Newborn vs. young adult (27Y)	0.021	1.68X10 ⁻⁰⁷	3.75X10 ⁻⁰⁵				
young adult (27Y) vs. older adult (50Y)	0.861	0.246	0.678				
Older adult (50Y) vs. elderly (85Y)	0.059	0.344	0.060				

¹Average methylation for all samples across the 5 genes (*TWIST2*, *TUSC3*, *HOXD4*, *EPHA10*, *HAND2*).

²Groups compared. Median ages 0 years (newborn), 27 years (young adult), 50 years (older adult) and 85 years (elderly).

³p-value, as determined by one way ANOVA, for comparison of methylation between indicated groups. Values highlighted in red indicate significant increases with age, highlighted in blue indicate significant decreases.

Supplemental Fig. 2

Supplemental Fig. 3

Supplemental Figure S2. Methylation levels in control genes in different age groups. Pyrosequencing analysis of the three control genes (*HOXA5*, *HOXA4*, *IGF2*), as indicated, in groups of individuals with median ages 0, 27, 50 and 85 years old (n=50 for each group). For all three genes methylation levels remain high at all age groups tested and methylation levels do not show clear evidence of increasing across the life-course (with the exception of an increase in *HOXA4* methylation early in life).

Supplemental Figure S3. Correlation between methylation levels and lymphocyte cell populations. Scatter plot indicating the correlations between averaged methylation levels and the percentage of total lymphocytes (upper panel), T-lymphocytes (middle panel) and B-lymphocytes (lower panel) (as a fraction of total WBC count). R^2 values (coefficient of determination) for each of the correlations and the corresponding p-values are indicated on the respective plots.

Supplemental Table 2 - Variation in CpG island methylation by cell type

	Mean ¹ (%)	<i>TWIST2</i>	<i>TUSC3</i>	<i>HOXD4</i>	<i>EPHA10</i>	<i>HAND2</i>	<i>HOXA5</i>	<i>HOXA4</i>
PBL	8.05	5.24	6.62	10.70	9.23	8.46	53.54	38.5
MNC	9.07	5.61	8.48	13.01	9.80	8.45	52.62	39.62
B cells	15.40 ²	12.94	15.99	21.83	15.79	10.47	57.35	42.88
Monocytes	5.80 ³	3.13	4.94	7.42	7.48	6.02	52.44	37.05
lymphocytes	9.36 ⁴	6.57	8.21	12.62	10.91	8.50	56.51	40.95

¹Mean average of the 5 genes which exhibit age-related methylation

²Significantly increased vs. PBL (p<0.001) and vs. lymphocytes (p<0.001)

³Significantly decreased vs. PBL (p <0.001)

⁴Significantly increased vs. PBL (p <0.05)