

RESEARCH ARTICLE

Open Access



Aging-associated DNA methylation changes in middle-aged individuals: the Young Finns study

L. Kananen^{1,2*}, S. Marttila^{1,2}, T. Nevalainen^{1,2}, J. Jylhävä^{1,2}, N. Mononen³, M. Kähönen⁴, O. T. Raitakari⁵, T. Lehtimäki^{3,6} and M. Hurme^{1,2,3}

Abstract

Background: Chronological aging-associated changes in the human DNA methylome have been studied by multiple epigenome-wide association studies (EWASs). Certain CpG sites have been identified as aging-associated in multiple studies, and the majority of the sites identified in various studies show common features regarding location and direction of the methylation change. However, as a whole, the sets of aging-associated CpGs identified in different studies, even with similar tissues and age ranges, show only limited overlap. In this study, we further explore and characterize CpG sites that show close relationship between their DNA methylation level and chronological age during adulthood and which bear the relationship regardless of blood cell type heterogeneity.

Results: In this study, with a multivariable regression model adjusted for cell type heterogeneity, we identified 1202 aging-associated CpG sites (a-CpGs, FDR < 5 %), in whole blood in a population with an especially narrow age range (40 - 49 years). Repeatedly reported a-CpGs located in genes *ELOVL2*, *FHL2*, *PENK* and *KLF14* were also identified. Regions with aging-associated hypermethylation were enriched regarding several gene ontology (GO) terms (especially in the cluster of developmental processes), whereas hypomethylated sites showed no enrichment. The genes with higher numbers of a-CpG hits were more often hypermethylated with advancing age. The comparison analysis revealed that of the 1202 a-CpGs identified in the present study, 987 were identified as differentially methylated also between nonagenarians and young adults in a previous study (The Vitality 90+ study), and importantly, the directions of changes were identical in the previous and in the present study.

Conclusions: Here we report that aging-associated DNA methylation features can be identified in a middle-aged population with an age range of only 9 years. A great majority of these sites have been previously reported as aging-associated in a population aged 19 to 90 years. Aging is associated with different types of changes in DNA methylation, clock-like as well as random. We speculate that the a-CpGs identified here in a population with a narrow age-range represent clock-like changes, as they showed concordant methylation behavior in population spanning whole adulthood as well.

Keywords: Aging-associated, DNA methylation, EWAS, CpG sites, Adulthood, Hypermethylation, Blood cell type heterogeneity

* Correspondence: laura.kananen@uta.fi

¹Department of Microbiology and Immunology, School of Medicine, University of Tampere, Tampere, Finland

²Gerontology Research Center, Tampere, Finland

Full list of author information is available at the end of the article



Background

The epigenome includes DNA methylation (DNAm), post-translational histone modifications and chromatin remodeling. Tens of millions of nucleotides referred to as CpG sites, which are prone to DNAm, exist in the haploid human genome. Furthermore, the genome-wide DNAm profile is maintained through cell divisions. DNA methyltransferases apply methyl groups on CpG sites to form 5-methylcytosine, whereas demethylation may occur either passively due to dysfunction of the transferring enzyme or actively through 5-hydroxymethylcytosine formation. Genomic regions spanning approximately 0.5 kilobases with a high density of CpG sites are called CpG islands, and these are commonly localized near transcription start sites. CpG sites in such islands are often less methylated; thus, the genes are available for initiation of transcription. Moreover, DNAm plays crucial roles in gene expression by not only blocking the promoter region but also altering the activities of regulatory elements, such as enhancers and insulators. Alternatively, gene body methylation may influence alternative splicing [1, 2]. Thus, the cell identity is in part determined and maintained by a cell type-specific genome-wide methylation pattern, which may therefore be used in the laboratory as a marker to characterize the cell types [3–5].

The genome-wide DNAm profile of the cell changes; DNAm patterns are altered in diseases, such as Alzheimer disease, cancer and type 2 diabetes, and are also influenced by the accumulating effects of environmental factors such as toxin exposure and diet [1, 6, 7]. Single CpG sites undergo hypo- and hypermethylation either randomly by stochastic factors or via more systematic mechanisms [1]. For example, exposure to environmental factors such as smoking induces hypomethylation of a well-characterized single CpG site in the gene *F2RL3*; this represents an example of a non-random change in DNAm because the magnitude of the change is dose and exposure-time dependent [8, 9].

Furthermore, the epigenome is modified by the biological aging process. As also Heyn et al. [10] reported and Zampieri et al. [1] reviewed, in general, aging induces a decrease in average DNA methylation level genome-wide (global hypomethylation). This was demonstrated by whole-genome bisulfite sequencing of newborns and centenarians with as high as ~90 % genomic coverage. The comparison of methylation states between the two extremes of the human lifespan also revealed how the systematic methylation patterns of the CpG sites are eventually lost and how inter-individual differences increase with advanced age. In addition, hypermethylation in regions near promoters can cause down-regulation of essential genes that influence vitally important pathways; Heyn et al. [10] reported that aging-accelerated hypermethylation events occurred in

13 % of the CpG sites among the millions of sites in the genome. Therefore, methylation alterations may be considered as one important factor in the development of aging-associated diseases [1, 10].

Many studies have addressed the aging-associated DNAm changes in blood cells using Illumina array technology-based methods, which cover 27000 or 485000 CpG sites in the genome [1]. The methylation levels of specific CpG sites are known to be associated with chronological aging in a wide variety of tissues [11–13]. However, as a whole, the sets of aging-associated CpGs identified in different studies, even with comparable tissues and age ranges, show limited overlap. Only few EWASs on age have taken the cell type heterogeneity into account [14–17]. We and others [4] hypothesize that lack of cell type adjustment may have potentially distorted the results obtained, and this may have contributed to the lack of concordance observed between the studies.

In this study, we aimed to discover and characterize regions where the DNAm levels are associated with chronological age (a-CpGs) in a middle-aged population (aged 40–49 years) through analysis where the cell type heterogeneity was adjusted for. Middle-aged individuals were selected from the Young Finns Study (YFS) [18] follow-up in 2011; the selection in the present study is a balanced sample (i.e. the number of subjects in each age group was equal and the groups had similar sex-distribution), and it therefore provides an excellent opportunity to inspect the effects of aging on DNA methylome. Furthermore, this sample comprises individuals in an extremely narrow age range of only nine years. The subjects' DNA methylomes were characterized using Illumina Infinium HumanMethylation450 BeadChips and the cell type heterogeneity and sex were adjusted for in the analysis.

Additionally, our findings were interpreted together with compatible data obtained using the same 450BeadChip technology, including our previous results obtained from an EWAS on age (The Vitality 90 + Study, V90+), in which the subjects' ages ranged from 19 to 90 years [15], as well as other results compiled by Steegenga et al. [19]. The results from the YFS were interpreted by considering that rates of aging-associated DNAm changes fluctuate, especially during the growth period before adulthood and at the end of the lifespan [11, 20]. Accordingly, the a-CpGs found in the YFS that overlap with those established from adult samples with wider age ranges, such as V90+ study, may be speculated to be DNAm regions with constant rate of change throughout adulthood. Thus, we aimed to explore the a-CpGs where level of methylation changes in a clocklike fashion throughout adulthood from those that show a more random aging-associated pattern.

Results

Aging-associated alterations in DNA methylation

In this study, the genome-wide DNAm levels in whole blood samples of middle-aged individuals were measured using 450BeadChip technology. The sample heterogeneity (i.e., the proportions of CD8T and CD4T cells, monocytes, granulocytes, and NK and B cells) were estimated by comparing DNAm profiles to the reference dataset [4] (Additional file 1: Figure S1). The cell type proportions were verified as important determinants of variation in DNAm using Spearman's correlation analysis, in which the cell type proportions were correlated with the main principal components (PCs). The PCs were defined with principal component analysis (PCA) from the DNAm data without cell subtype adjustment (Additional file 2: Table S1a). The analysis revealed that PC1 to PC6 together explained a large proportion (24 %) of the variance in the DNA methylome data. Among those PCs, several PCs had considerable large ($-0.5 > r > 0.5$) correlation coefficients; thus, adjustments for the cell type proportion in the analysis were mandatory. The hypothesis whether DNAm level of a CpG site is associated with chronological age was tested at each CpG

site using generalized linear regression analysis ('beta regression'), where sex and cell type proportions were adjusted for.

We found 1202 a-CpGs (i.e. CpG sites where age was a statistically significant variable in the multivariable regression model, $FDR < 5\%$) in middle-aged individuals (aged 40–49 years), of which 622 (52 %) were hypomethylated and 580 (48 %) were hypermethylated with advancing age. These hypo- and hypermethylated sites were annotated on 440 and 437 genes, respectively. Lists of the most significant aging-associations in YFS are shown in Tables 1 and 2 and in Additional file 3: Table S4. Frequently reported CpG sites (summarized by Steegenga et al. [19]) located in the *ELOVL2* (cg16867657, cg24724428 and cg21572722), three sites in the *FHL2* (cg06639320, cg22454769 and cg24079702), two sites in the *PENK* (cg16219603, cg16419235), and two sites in the *KLF14* (cg08097417, cg09499629 and cg07955995) were also identified as hypermethylated in the present study.

Interestingly, similar to correlation analysis results shown in Additional file 2: Table S1a, the cell type proportions were important determinants of variation in

Table 1 The top 20 hypermethylated a-CpGs in middle-aged individuals. The hypermethylated and hypomethylated a-CpGs are shown separately in Tables 1 and 2, respectively. The top-ranking hypermethylated a-CpGs were selected with the following criteria: 1) direction of the association based on the value of beta regression (denoted as 'betareg') estimate of age; 2) more than one hit identified per gene (q -value < 0.05 which corresponds to false discovery rate $< 5\%$) and 3) the top-ranking p -values. The full list of a-CpGs is shown in Additional file 3: Table S4. The q -value denotes the Benjamini-Hochberg-corrected p -value

ProbelID	Gene name	CHR	Coordinate	Betareg estimate of age	q -value
cg16867657	<i>ELOVL2</i>	6	11152863	0.022	0.00E+00
cg24724428	<i>ELOVL2</i>	6	11152874	0.021	4.80E-07
cg21572722	<i>ELOVL2</i>	6	11152880	0.013	3.46E-06
cg06639320	<i>FHL2</i>	2	105382171	0.018	3.46E-06
cg00059225	<i>GLRA1</i>	5	151284550	0.013	5.13E-06
cg08097417	<i>KLF14</i>	7	130069673	0.020	1.87E-05
cg22454769	<i>FHL2</i>	2	105382199	0.021	5.03E-05
cg07553761	<i>TRIM59</i>	3	161650671	0.016	6.12E-05
cg01588592	<i>ETV3L</i>	1	155335949	0.011	1.14E-04
cg11176990	<i>LOC375196</i>	2	39041037	0.014	1.54E-03
cg09499629	<i>KLF14</i>	7	130069676	0.018	1.54E-03
cg22158769	<i>LOC375196</i>	2	39041043	0.020	2.43E-03
cg18898125	<i>NEFM</i>	8	24826286	0.012	2.49E-03
cg21911021	<i>ZIK1</i>	19	62786823	0.020	3.07E-03
cg27217742	<i>RGS12</i>	4	3335078	0.013	3.07E-03
cg17737681	<i>DLX1</i>	2	172660382	0.015	3.29E-03
cg24079702	<i>FHL2</i>	2	105382203	0.015	5.99E-03
cg16219603	<i>PENK</i>	8	57523140	0.013	7.00E-03
cg23930856	<i>TFAP2B</i>	6	50919683	0.013	7.22E-03
cg11152943	<i>TRAPPC9</i>	8	141318170	0.013	7.57E-03

Table 2 The top 20 hypomethylated a-CpGs in middle-aged individuals. The hypermethylated and hypomethylated a-CpGs are shown separately in Tables 1 and 2, respectively. The top-ranking hypomethylated a-CpGs were selected with the following criteria: 1) direction of the association based on the value of beta regression (denoted as 'betareg') estimate of age; 2) more than one hit identified per gene (q-value < 0.05 which corresponds to false discovery rate < 5 %) and 3) the top-ranking p-values. The full list of a-CpGs is shown in Additional file 3: Table S4. The q-value denotes the Benjamini-Hochberg-corrected p-value

ProbelD	Gene name	CHR	Coordinate	Betareg estimate of age	q-value
cg00791074	<i>MTHFD1L</i>	6	151227862	-0.018	7.51E-04
cg18618815	<i>COL1A1</i>	17	45630323	-0.018	5.99E-03
cg14169886	<i>PRDM16</i>	1	3101709	-0.014	5.99E-03
cg01820374	<i>LAG3</i>	12	6752344	-0.014	9.24E-03
cg19421125	<i>LAG3</i>	12	6753117	-0.022	1.02E-02
cg14829066	<i>NTRK3</i>	15	86360145	-0.013	1.49E-02
cg03290281	<i>C6orf195</i>	6	2577606	-0.021	1.49E-02
cg05561193	<i>DCLK2</i>	4	151218492	-0.017	1.96E-02
cg20249566	<i>NWD1</i>	19	16691739	-0.024	1.97E-02
cg23928726	<i>PEX10</i>	1	2334858	-0.014	1.97E-02
cg20007894	<i>SCAND3</i>	6	28648421	-0.019	2.08E-02
cg16355231	<i>PEX10</i>	1	2334839	-0.019	2.14E-02
cg15058210	<i>HDAC4</i>	2	239861814	-0.018	2.16E-02
cg06030846	<i>TMEM108</i>	3	134581182	-0.011	2.16E-02
cg25994988	<i>UBASH3B</i>	11	122157592	-0.011	2.16E-02
cg18345924	<i>NCAM2</i>	21	21294102	-0.016	2.18E-02
cg00638021	<i>COL1A1</i>	17	45622061	-0.013	2.26E-02
cg19344626	<i>NWD1</i>	19	16691749	-0.024	2.36E-02
cg01288258	<i>ITFG2</i>	12	2792128	-0.011	2.41E-02
cg05221385	<i>TAF10</i>	11	6590080	-0.010	2.43E-02

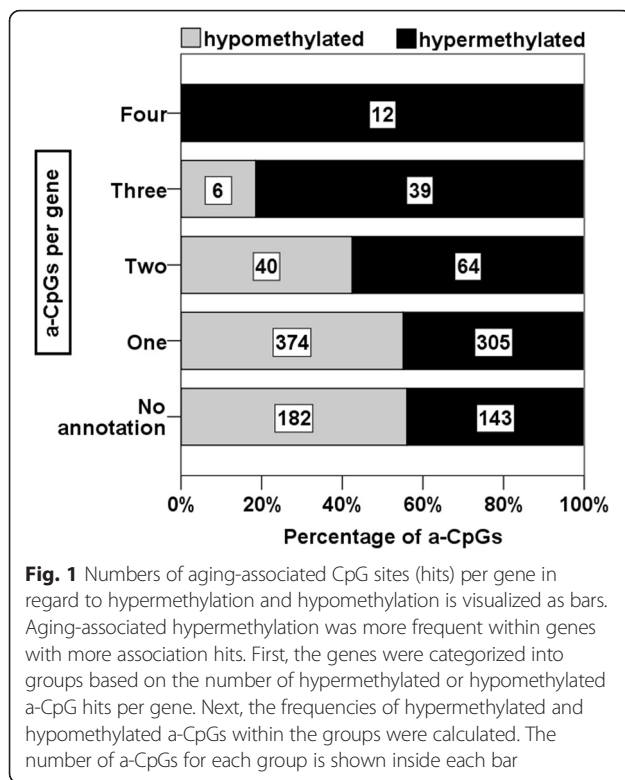
DNAmet levels of the 1202 a-CpGs as well (Additional file 2: Table S1b). In this second correlation analysis, the PCs were defined with PCA from DNA methylation data of the 1202 a-CpGs (aging-associated CpG sites, FDR < 5 %); methylation data in PCA were not adjusted for the cell subtype heterogeneity. Correlation analysis revealed that PC1-PC6 determined more than 50 % of variance in methylation levels of these a-CpGs and these PCs correlated clearly with age and the cell counts. It is also worth of mentioning that of the 1202 a-CpGs in our initial aging-association analysis, there were 526 multivariable regression models (corresponding 526 CpG sites) where all cell count variables (monocytes, granulocytes, NK, CD8T and CD4T cells) were detected as statistically significant (FDR < 5 %) predictors of DNA methylation levels.

The importance of the cell count considerations was explored with an additional set of regression models, where the DNA methylation level in each CpG site genome-wide was explained with age and sex only while the cell counts were not adjusted for. In this analysis, only 56 sites were classified as aging-associated (FDR < 5 %) and these sites were all included to the original

pool of 1202 a-CpGs. The 56 a-CpGs are pointed out in the Additional file 3: Table S4.

Aging-associated hypermethylation and hypomethylation differ in their features. The exploration of aging-associations in the YFS revealed that hypermethylation was more frequent within genes with more association hits as shown in Additional file 4: Table S5 and Fig. 1). Specifically, there were 70 genes in total either with more than one hypomethylated or more than one hypermethylated a-CpGs per gene. Of those, 22 genes comprised more than one hypomethylated a-CpGs per gene and 48 genes comprised more than one hypermethylated a-CpGs per gene as shown in Additional file 4: Table S5.

Next, the genomic locations of the a-CpGs were investigated, revealing that 388 of the 1202 a-CpGs were located on CpG islands rather than island shores, shelves or non-island regions, and a majority (N = 331) of those were hypermethylated (Additional file 1: Figure S2). The remaining sites were distributed to shores, shelves and non-island regions with opposite manner as shown in Additional file 1: Figure S2; the aging-associated hypomethylation was more abundant on those regions. The a-CpG locations on genes were also investigated; no



enrichment of a-CpGs was detected in the regions of 3' untranslated regions (UTRs), 5'UTRs or close distances to transcription start sites or gene bodies (Additional file 1: Figure S3a and b). The distributions of the a-CpGs on chromosomes were also investigated; hypermethylated a-CpGs were over-represented on chromosome 18, whereas hypomethylated sites were not enriched on any chromosome (hypergeometric test, nominal p -value of 0.05) (Additional file 1: Figure S3c). In addition, we ensured using visual examination that there were no spatial local cluster(s) of a-CpGs on Chr-18.

Sex specificity of the aging-associated CpG sites

To evaluate the sex specificity of the aging-associations, an interaction model with variables corresponding to sex, age and the interaction of sex and age (age*sex) was constructed. No sex-specific a-CpGs were identified, as analysis revealed that no interaction term had a false discovery rate (FDR) below 5 % (q -value < 0.05) in the interaction models. Furthermore, we analyzed women ($N = 111$) and men ($N = 73$) separately as well: sex-specific a-CpGs were explored among all CpG sites with an multivariable regression model ('beta regression') where age and cell type proportion variables were used to predict DNA methylation level in each CpG site. These analyses revealed that there were 105 and 173 a-CpGs (FDR < 5 %) among men and women, respectively; these CpG sites were all included to our original pool of

1202 a-CpGs which were detected using whole sample ($N = 184$). Importantly, as shown in Additional file 1: Figure S5, when the directions of change among the 1202 a-CpGs were cross-compared between men and women (without p -value cut-off), all sites, except one, showed concordant behavior regarding hypermethylation or hypomethylation during aging (i.e. whether the estimate of age variable in the regression model was negative or positive value). This behavior was also identical to the directions of change among the 1202 a-CpGs in the initial analysis ($N = 184$). As a conclusion, these results were in line with our interaction analysis: there were no significantly sex-specific a-CpGs among middle-aged individuals.

Functional roles of a-CpGs in the YFS

The gene ontology (GO) functions and processes of the genes with a-CpGs were investigated using the Gene Ontology enRiChment anaLysis and visualiZation (GORilla) tool [21]. The analysis was conducted separately for genes with hypermethylated a-CpGs and for hypomethylated a-CpGs ($N = 440$ and $N = 437$, respectively). The analysis revealed an unambiguous differences between hypo- and hypermethylated a-CpGs, as 73 GO process terms and to 8 GO function terms were enriched to genes with hypermethylated a-CpGs (Tables 3 and 4, respectively; Additional file 2: Table S2.), whereas there was no enrichment of terms among the genes with hypomethylated a-CpGs (Bonferroni-adjusted p -value threshold of 0.05). The most statistically significant processes were anatomical structure development (GO:0048856, $p = 1.02 \times 10^{-11}$) and morphogenesis (GO:0009653, $p = 5.02 \times 10^{-10}$), both of which cluster under the term 'developmental process'.

In addition, Pscan [22] was used to predict whether there were common regulators for groups of genes. The hypermethylation-associated genes were predicted to be regulated by 11 common transcription factors (Additional file 2: Table S3), several of which were zinc coordinating. For hypomethylation-associated genes, no common transcription factors were found. A large proportion of the 11 regulators of genes with hypermethylated a-CpGs in the YFS were zinc coordinating, and four (E2F1, EGR1, SP1, TFAP2A) were identical to those identified in the V90+ study [15].

Comparisons to other studies

In the explorative cross-comparison analysis, the a-CpGs identified in middle-aged individuals of the YFS were compared to aging-associated DNA methylome alterations between nonagenarians and 19–30-year-old individuals evidenced in our previous study (the V90+ study) [15]. The a-CpGs identified in the V90+ study were strongly associated with aging while the cell type heterogeneity was adjusted for in the analysis. A total of

Table 3 Several GO process terms were enriched within genes with hypermethylated a-CpGs in the analysis with GOrilla [21, 43]. This table represents the main clusters of processes (53 redundant GO terms were filtered out of 73 terms using REVIGO [44]). The full list of processes is shown in Additional file 2: Table S2

GO term	Description of the process	<i>p</i> -value (-log ₁₀)
GO:0048856	Anatomical structure development	10.9914
GO:0050794	Regulation of cellular process	8.9788
GO:0007389	Pattern specification process	8.2343
GO:0032502	Developmental process	8.2041
GO:0009893	Positive regulation of metabolic process	8.0511
GO:0044708	Single-organism behavior	7.5544
GO:0035108	Limb morphogenesis	7.5544
GO:0003002	Regionalization	7.3585
GO:0051239	Regulation of multicellular organismal process	7.301
GO:0006357	Regulation of transcription from RNA polymerase II promoter	7.2248
GO:0065007	Biological regulation	7.1675
GO:0007610	Behavior	7.08
GO:0048598	Embryonic morphogenesis	7.0778
GO:0048518	Positive regulation of biological process	6.8761
GO:0048519	Negative regulation of biological process	6.7122
GO:0008285	Negative regulation of cell proliferation	6.4921
GO:0048523	Negative regulation of cellular process	5.8827
GO:0010842	Retina layer formation	5.8041
GO:0051961	Negative regulation of nervous system development	5.7423
GO:0032774	RNA biosynthetic process	5.4225

the 1202 a-CpGs established in the YFS cohort, 999 a-CpGs were also aging-associated in the V90+ sample (FDR < 5 %, Additional file 3: Table S4). Of these 999 a-CpGs, 464 (46 %) were hypermethylated, and 535 (54 %) were hypomethylated with advancing age. Furthermore, in 987 of the overlapping 999 a-CpGs the direction of the aging-associated change was the same: in the present and

in the V90+ study, 455 a-CpGs were hypermethylated, and 532 were hypomethylated with advancing age (Fig. 2).

Finally, a-CpGs that were characterized from whole blood samples as aging-associated using 450 BeadChip technology and previously reported by Hannum et al. (number of hits, 89) [13], Garagnani et al. (number of hits, 9) [12] and Florath et al. (number of hits, 162) [23] and presented as summary table in Steegenga et al [19] were further compared with our data. The corresponding age of the samples ranged between 19–101, 9–83 and 50–75 years, respectively. The comparison revealed 21 common CpG sites out of the 999 a-CpGs in two or more studies in addition to the YFS and the V90+ study (Fig. 3).

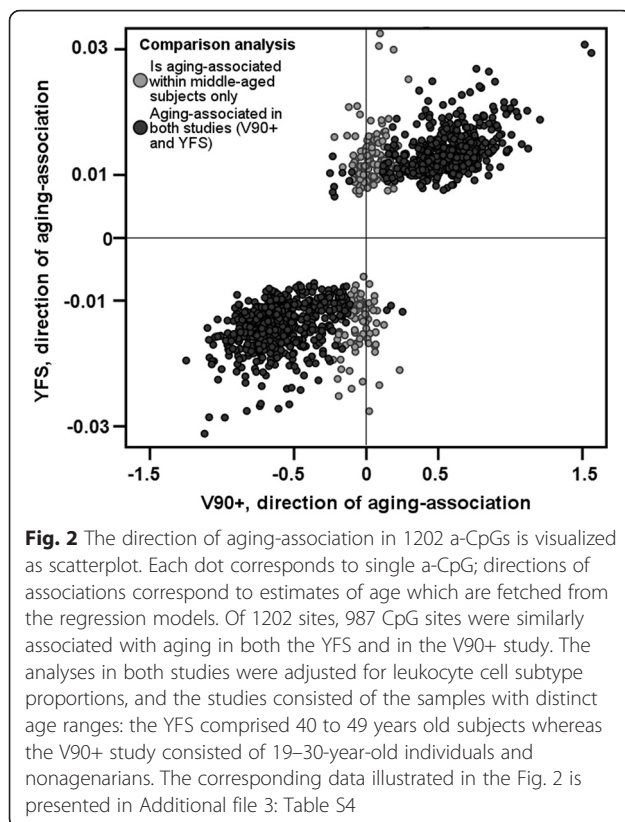
Discussion

In this study, we identified 1202 a-CpGs where the DNAm level was associated with aging in middle-aged individuals (i.e. with an age range of 40 to 49 years), in whom the growth and development of youth has ended yet old age and its associated diseases had not begun. Of the 1202 a-CpGs, 622 (52 %) were hypomethylated, and 580 (48 %) were hypermethylated with advancing age, with annotations on 440 and 437 different genes, respectively. In general, the functional features of these aging-associated sites are mostly similar to those identified from cohorts with larger age differences. Our study highlights also that a large number of sites undergo aging-associated DNAm level changes throughout adulthood and we speculate that a great proportion of those probably change with a clock-like manner.

A large fraction of the DNAm sites are altered during the lifespan, as shown by previous studies performed using 450BeadChip technology [15, 24] and whole-genome bisulfite sequencing [10]. Furthermore, the rates of these changes may fluctuate at different stages of the lifespan. Studies have shown that a-CpGs behave differently during the growth period before adulthood and at the end of the lifespan [11, 20]. Nonetheless, there are genes (*ELOVL2*, *SFMBT1*, *KLF14*, *PENK*, and *FHL2*) with CpG sites that are consistently detected as being aging-associated despite of differences in sample tissue

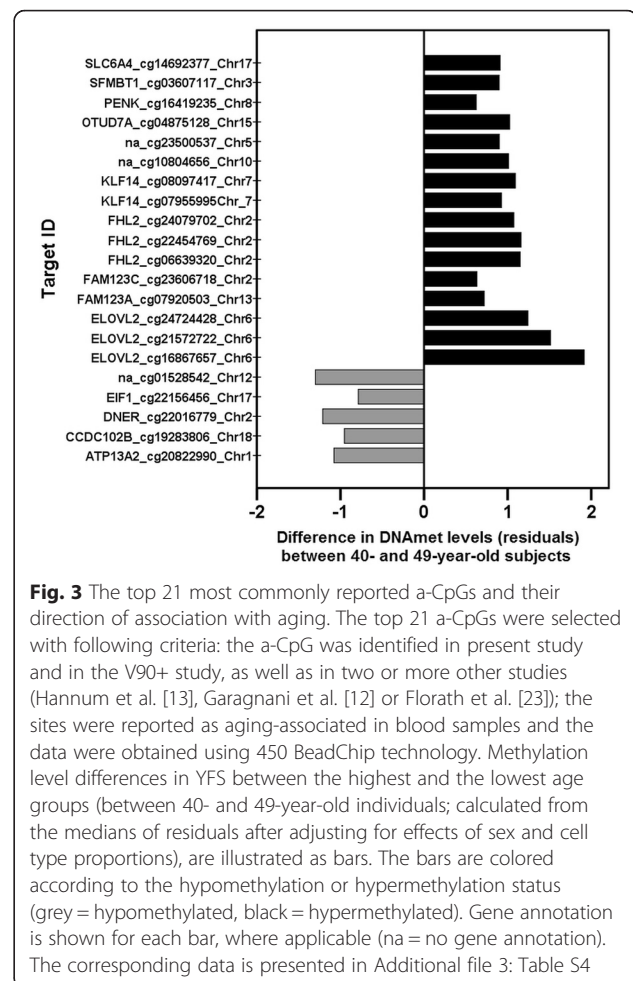
Table 4 GO function terms were enriched within genes with hypermethylated a-CpGs in the analysis with GOrilla. Table contains the full list of enriched GO function terms (Bonferroni-adjusted $p < 0.05$) obtained from analysis with GOrilla [21, 43]

GO term	Description of the function	<i>p</i> -value (-log ₁₀)
GO:0043565	Sequence-specific DNA binding	10.001
GO:0000981	Sequence-specific DNA binding RNA polymerase II transcription factor activity	7.322
GO:0001071	Nucleic acid binding transcription factor activity	6.721
GO:0003700	Sequence-specific DNA binding transcription factor activity	6.721
GO:0003677	DNA binding	6.625
GO:0005326	Neurotransmitter transporter activity	5.148
GO:0005488	Binding	4.967



types or age distributions [11–13, 15]; notably, these genes were also identified in the present study as being aging-associated (Tables 1 and 2; Additional file 3: Table S4). However, a recent meta-analysis on three DNAm levels data sets obtained using 450BeadChip illustrated discrepancies in the lists of regions where DNAm levels were altered during the entire human lifespan, ranging from 0 to 100 years of age [2]. Because blood sample heterogeneity has been shown to have a great impact on EWASs [4, 15], our speculation is that the discrepancies might be due to the presence of different cell types.

In the primary analysis, we aimed to identify a-CpGs in middle-aged individuals representing general population with age range of only one decade. Then, we cross-compared the results to those obtained with similar analysis pipeline from a population aged 19 to 90 years (Vitality 90+ study) [15]. Among the 1202 a-CpGs characterized from the YFS with an age range of nine years, 987 sites had an identical association direction as detected in the Vitality 90+ study, as shown in Fig. 2 and in Additional file 3: Table S4. We hypothesize that sites displaying aging-associated methylation changes in both populations possibly represent sites where the change in DNA methylation follows a clock-like pattern. We further speculate that the non-overlapping CpG sites identified in the population with a wider age range (19 to 90 years of age) may possibly represent



sites where the aging-associated change is accelerated in either early or late adulthood; the a-CpGs identified only when comparing group of nonagenarians to young adults may represent changes that reflect e.g. aging-associated pathologies or accumulation of aging-associated impairments.

As aging influences the immune system of men and women differently and as the risk rates of several diseases between sexes are unequal [25, 26], 1) an interaction analysis was performed to address the sex specificity of a-CpGs, and 2) the aging-associations were also evaluated in separate analyses among men and women. These analyses revealed no sex-specific single a-CpGs; thus, the identified a-CpGs are universally altered in both men and women. These results are in accordance with our previous results from the V90+ study, in which the DNAm states of nonagenarians were compared with 19–30-year-old individuals [15], and with results published by others [24, 27]. However, studies have shown that as a whole, the DNA methylomes of males age more rapidly than those of females [13, 28].

Aging-accelerated hypomethylation may be thought as an erosion-like event, whereas hypermethylation may be thought as an actively guided process. In practice, the difference between these features is manifested, for example, through the enrichment of GO terms for groups of genes and for signaling pathways [1, 15]. The distinct roles of the methylation status were demonstrated in the present study with the numbers of a-CpG hits in a gene, as we observed notable enrichment of hypermethylation events located in genes with more than one a-CpG (Fig. 1). The functional roles of genes with a-CpGs were established by GO term enrichment analysis, which revealed obvious difference between hypo- and hypermethylated a-CpGs, even though the analysis was conducted with an equal number of genes in the GO term analyses. A high number of GO terms were enriched to genes with hypermethylated a-CpGs (Tables 3 and 4; Additional file 2: Table S2), whereas there was no GO term enrichment within genes with hypomethylated a-CpGs. The most statistically significant processes enriched to genes with hypermethylated a-CpGs were 'anatomical structure development' and 'morphogenesis', both of which cluster under the term 'developmental process'. The enrichment of hypermethylated a-CpGs to these processes has been reported previously [14, 23, 24, 29]. Reynolds [30] and Yuan [16] reported also that the CpG sites hypermethylated during aging are enriched to common processes and exhibit shared features, whereas hypomethylated a-CpGs are a less homogenous group. Furthermore, age-associated hypermethylation interactome hotspots have been reported [31].

In addition to the details mentioned above, we observed other similar hypermethylation characteristics in the YFS, as those reported in previous studies [1, 15]. For example, the majority (85 % out of 388) of a-CpGs localized in CpG-islands (instead of shores, shelves or other regions) were hypermethylated, and an excess of hypermethylated a-CpGs were also found on chromosome 18. However, there was no enrichment of a-CpGs on chromosome 19. In the V90+ study, the hypermethylated a-CpGs located in the genes encoding zinc-associated proteins were more abundant on chromosome 19 [15], where zinc-finger genes are clustered. The zinc-finger genes (such as *ZNF154*) located in chromosome 19 are proposed to be repressors of endogenous retroviruses (ERVs) [32], and the repressor activity may be disturbed by hypermethylation. Interestingly, CpG sites located in the gene *ZNF154* and almost all other genes encoding zinc-fingers on chromosome 19 were absent from our pool of 1202 a-CpGs. Thus, as the hypermethylation of CpG sites located in genes encoding zinc-fingers was observed in the oldest age group, we hypothesize that rates of methylation level changes at the CpG sites located in ERV repressor genes

(e.g. *ZNF154*) may fluctuate throughout the lifespan and that the rates may be enhanced in association with other senescence-related factors. Therefore, it is possible that DNAm-based dysfunction of the repression system might explain the increased expression of ERVs in old age [33]. Future studies are required to address these questions.

To further inspect the roles of the genes with aging-accelerated DNAm changes, analysis of the common regulators (transcription factors) of groups of genes with hypermethylated and hypomethylated a-CpGs was conducted with Pscan [22]. The results were again surprisingly concordant with those in the V90+ study. There were 11 regulators with unique identifiers for hypermethylated a-CpGs (Additional file 2: Table S3), whereas hypomethylated a-CpGs had no common regulators. A great proportion of the 11 regulators of genes with hypermethylated a-CpGs in the YFS were zinc coordinating, and four (E2F1, EGR1, SP1, and TFAP2A) were identical to those identified in the V90+ study results [15]. Overall, the results from analysis of the functional roles of the genes with a-CpGs were surprisingly well in line with the observations from the V90+ study and supported the proposition that aging-associated hypermethylation is a more tightly regulated process, whereas aging-associated hypomethylation is induced more by environmental effects and stochastic factors.

Finally, we demonstrated the lack of concordance in previously reported pools of a-CpGs by comparing three published lists of overlapping a-CpGs produced using 450BeadChips from whole blood samples from subjects with age ranges of 50–75, 19–101 and 9–83 [12, 13, 23]. Although 987 of the a-CpGs in the YFS showed similar association directions as in the V90+ study (Fig. 2 and Additional file 3: Table S4), we observed only 61 overlapping a-CpGs in the YFS and the V90+ study, which were also reported as aging-associated in one or more other robustly compatible studies (same sample type and array technology). Of these, only 21 a-CpGs were observed in two or more of the studies in the comparison (Fig. 3). To the best of our knowledge [4, 15], the main factor that contributes to the DNAm profiles in blood cells is cell type heterogeneity; thus, we speculate that the lack of cell type adjustments may account for the majority of disparity in the cross-comparisons. The results of aging-association analysis and combined PCA-correlation analysis in this study supports our speculation. Cell type heterogeneity should be taken into account when analyzing samples composed of mixed cell types, but a limited number of such studies have been conducted [4, 14–17].

Notably, our study had an obvious limitation, it would substantially benefit from being a follow-up; therefore, future studies are needed. Nevertheless, the analysis is

powered by well-designed sample characteristics because each age group was matched by sex and sample size and because adjustments were made for cell type heterogeneity. Thus, the analysis was sensitive enough to detect DNAm changes within an age range spanning nine years.

Conclusions

Here we report that aging-associated DNA methylation changes can be identified in a middle-aged population with a narrow age range of 9 years. Aging-associated DNAm changes are not uniform, but occur due to different reasons, at different rates and directions in different parts of the genome and are not alike in all cell types. Thus, due to this diverse nature of aging-associated DNA methylation changes, all confounding factors should be accounted for in the analysis, in order to obtain comparable results. Our results support the notion that cell type heterogeneity should be adjusted for when analyzing tissues consisting of mixed cell types. Moreover, our results imply that considerable proportion of DNAm changes show clock-like behavior throughout adulthood.

Methods

Study population

The Young Finns study (YFS) comprises a series of six cohorts, representing general population, born in 1962, 1965, 1968, 1971, 1974 and 1977 from five cities with university hospitals in Finland (Helsinki, Kuopio, Oulu, Tampere and Turku) [18]. A subsample of 184 individuals was randomly assigned from a follow-up in 2011. The sample collection in 2011 is described in more detail elsewhere [34]. The categories of age in the methylation analysis were 40, 43, 46 and 49 years old, with group sizes of 50, 44, 55 and 35, in which 58 %, 68.2, 56.4 and 60 % were women, respectively. All of the participants were of western European descent. The study followed the guidelines of the Declaration of Helsinki and was approved by the Ethical Review Committee of Turku University Hospital. All participants provided informed consent.

DNA methylome quantification

Sample preparations

Leukocyte DNA of the YFS cohort was obtained from EDTA-blood samples using a Wizard® Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA) according to the manufacturer's instructions. Genome-wide DNA methylation levels were obtained using Illumina Infinium HumanMethylation450 BeadChips [35–37] in the Core Facility at the Institute of Molecular Medicine Finland (FIMM), University of Helsinki according to the protocol by Illumina.

The methylation data set was preprocessed identically with a previously described analysis pipeline which was used in the DNA methylation analysis of the V90+ study samples [15, 38, 39]. Briefly, methylation signal data was preprocessed as a methylumi object using R software ($R \geq 2.15.3$) with array-specific algorithms implemented in the R package `wateRmelon` [40] and `BMIQ` [38]. The resulting β values ranged linearly from 0 (non-methylated, 0 %) to 1 (completely methylated, 100 %). The quality of DNA samples and methylation data was carefully ensured by standard examinations with principal component analysis (PCA) and visualizations with density plots, boxplots and dotplots. Three of the YFS samples were excluded due to atypically low probe intensities compared with control probe intensities.

The YFS sample was lacking leukocyte cell type characterizations; thus, the proportions were determined by the estimation algorithm implemented in the `estimateCellCounts` function of the `minfi` Bioconductor package [4] using R software ($R \geq 2.15.3$). The algorithm utilizes the selection of 600 control probes that represents specific signatures of CD8T and CD4T cells, monocytes, granulocytes, and NK and B cells (Additional file 1: Figure S1). The reference data used in the estimation is available in the `FlowSorted.Blood.450K` Bioconductor package [4].

Quality control of the DNA methylome data

As the cell type proportions contribute to most of the variation in genome-wide DNAm [4, 15], the significance of the estimated cell counts in the DNAm data was investigated by PCA, and the main PCs of DNAm were correlated with the cell counts (Additional file 2: Table S1a). Spearman's correlation analysis indicated a clear connection between methylation profiles and estimated cell proportions. Thus, the estimated cell counts as well as the genome-wide methylation data was shown to behave as expected.

As part of the quality control step, a well-known CpG site with phenotype association was selected. Smoking is strongly associated with the hypomethylation of cg03636183, located in the gene *F2RL3* [8, 9]; our data from the YFS replicated this finding, as we observed a difference between daily smokers and others (Wilcoxon rank sum-test, $P = 2.4 \times 10^{-6}$; Additional file 1: Figure S4). Analysis with multivariable regression model (function `lm()` in R) revealed that the cell type heterogeneity, age or sex of the samples did not alter the finding of cg0363618.

Detection of aging-associated methylation regions

Aging-associated CpG sites, the a-CpGs, were explored using a generalized linear regression model, referred to as the 'variable dispersion beta regression' in an iterative

manner for each methylation locus (CpG site). The age (categories of 40, 43, 46 and 49) was employed as a variable to predict the site-specific methylation outcome in the form of a β value (ranging from 0 to 1); this was done in each equation using the mean model and a linker function of *logit*. The cellular heterogeneity was adjusted in the initial multivariable regression analyses: in addition to age and sex variables, variables corresponding to each estimated blood cell subtype proportion (CD8T and CD4T cells, monocytes, granulocytes, NK and B cells; all ranging linearly from 0 to 1) were included to the regression models as predictors of DNA methylation level. Additionally, sex-specific a-CpGs were explored among all CpG sites using two approaches: 1) with an interaction model where age, sex, sex*age and cell type proportion variables were used to predict DNA methylation level, and 2) with a regression model where age and cell type proportion variables were used to predict DNA methylation level separately for men and women. Furthermore, to explore the relevance of the cell count considerations in the regression analyses, an additional set of age-association analyses was performed. In these regression models, the DNA methylation level of each CpG site was explained with age and gender variables only and the cell proportions were not adjusted for. The analyses were performed using R software ($R > 2.15.3$), and the regression analyses were mainly conducted with algorithms implemented in the *betareg* package [41]. The nominal Benjamini-Hochberg adjusted p -value (q -value) was set to 0.05. The a-CpGs were annotated based on the assembly provided by the R package, *FDb.InfiniumMethylation.hg19* [42]. For the purpose of visualization in Fig. 3, standardized weighted residual values of the methylation levels were extracted for each CpG site from regression models in which only sex and cell type proportion variables were set as predictors.

Analysis of the functional roles of a-CpGs

The enriched gene ontology (GO) terms of the genes with a-CpGs were discovered using *GOrilla* [21, 43], and the significant terms were further clustered by *REViGO* [44]. The *GOrilla* analysis was performed for the process, function and component categories with two un-ranked lists, of which the first list comprised genes with hypomethylated or hypermethylated a-CpGs (Additional file 3: Table S4), and the second comprised the genes in the background ($N = 20,902$; analysis date, 9.3.2015). Furthermore, the prediction of common transcription factors of the groups of genes with either hypermethylated or hypomethylated a-CpGs (as two separate analyses) was conducted using *Pscan* with the default settings (*JASPAR* database; analysis date, 10.3.2015) [22]. The nominal

p -value was set to at the Bonferroni-corrected value of 0.05 in each analysis.

Availability of supporting data

The methylation data presented in this manuscript have been submitted to the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) under the accession number GSE69270.

Additional files

Additional file 1: Figures S1-S5. 1) A figure of estimated proportions of CD8T, CD4T, NK, B cell, monocyte and granulocyte cells of peripheral blood samples in YFS. Proportions are visualized as boxplots, categorized by age group and organized to separate panels by sex. 2) A figure of aging-associated CpG site locations in regard to CpG islands (CGIs). Number of aging-associated CpG sites are visualized with stacked bars. 3) A figure (a-c) presenting locations of a-CpGs. 4) A figure showing results for association of DNA methylation level in cg03636183 with smoking. 5) A figure presenting sex specificity of the aging-associated CpG sites (a-CpGs). (DOCX 357 kb)

Additional file 2: Tables S1-S3. 1) Two summary tables (a and b) of the results from Spearman correlation analyses between age, the cell counts and the first principal components (PCs). PCs were defined from either the whole methylation data or 1202 a-CpGs using PCA. 2) A table of the GO terms of the bio processes that are enriched to genes with aging-associated CpG-sites. 3) A table of common transcription factors for genes with hypermethylated a-CpGs characterized using *Pscan*. (DOCX 31 kb)

Additional file 3: Table S4. A full table of 1202 a-CpGs with detailed information. (XLSX 183 kb)

Additional file 4: Table S5. A summary table where the 70 genes with more than one hypomethylated or more than one hypermethylated a-CpGs per gene are presented. (XLSX 15 kb)

Abbreviations

a-CpGs: aging-associated CpG sites (FDR < 5 %); DNAm: DNA methylation; ERV: endogenous retrovirus; EWAS: epigenome-wide association studies; PC: principal component; UTR: untranslated region; V90+: The Vitality 90+ Study; YFS: Young Finns study.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LK processed the data and performed analyses and was responsible for writing the manuscript. NM, JJ and TN were responsible of the experiments. MH, MK and TL provided reagents and materials and LK, SM, JJ, TL and MH contributed to the design of the study. OTR, MK and TL were responsible for recruiting the subjects in the study. All authors contributed to the writing of manuscript and read and approved the final manuscript.

Acknowledgements

The Young Finns Study has been financially supported by the Academy of Finland: grants 286284 (T.L.), 132704 (M.H.), 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Kuopio, Tampere and Turku University Hospital Medical Funds (grant X51001 for T.L.); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation of Cardiovascular Research (T.L.); Finnish Cultural Foundation; Tampere Tuberculosis Foundation (T.L., M.H.); Emil Aaltonen Foundation (T.L.); and Yrjö Jahnsson Foundation (T.L., M.H.) This work was also supported by grants from the Competitive Research Fund of Pirkanmaa Hospital District (9M017, 9N013 to M.H.). We thank Nina Peltonen and Sinikka Repo-Koskinen for their excellent technical assistance.

Author details

¹Department of Microbiology and Immunology, School of Medicine, University of Tampere, Tampere, Finland. ²Gerontology Research Center, Tampere, Finland. ³Fimlab Laboratories, Tampere, Finland. ⁴Departments of Clinical Physiology, Tampere University Hospital and University of Tampere School of Medicine, Tampere, Finland. ⁵Research Centre of Applied and Preventive Cardiovascular Medicine and the Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Turku, Finland. ⁶Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere, Finland.

Received: 5 June 2015 Accepted: 1 February 2016

Published online: 09 February 2016

References

- Zampieri M, Ciccarone F, Calabrese R, Franceschi C, Burkle A, Caiafa P. Reconfiguration of DNA methylation in aging. *Mech Ageing Dev*. 2015.
- Bacalini MG, Boattini A, Gentilini D, Giampieri E, Pirazzini C, Giuliani C, et al. A meta-analysis on age-associated changes in blood DNA methylation: results from an original analysis pipeline for Infinium 450k data. *Aging (Albany NY)*. 2015;7(2):97–109.
- Accomando WP, Wiencke JK, Houseman EA, Nelson HH, Kelsey KT. Quantitative reconstruction of leukocyte subsets using DNA methylation. *Genome Biol*. 2014;15(3):R50.
- Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 2014;15(2):R31.
- Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86.
- Johnson AA, Akman K, Calimport SR, Wuttke D, Stolzing A, de Magalhaes JP. The role of DNA methylation in aging, rejuvenation, and age-related disease. *Rejuvenation Res*. 2012;15(5):483–94.
- Bacalini MG, Friso S, Olivieri F, Pirazzini C, Giuliani C, Capri M, et al. Present and future of anti-ageing epigenetic diets. *Mech Ageing Dev*. 2014;136–137:101–15.
- Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H. F2RL3 methylation as a biomarker of current and lifetime smoking exposures. *Environ Health Perspect*. 2014;122(2):131–7.
- Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet*. 2011;88(4):450–7.
- Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, et al. Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A*. 2012; 109(26):10522–7.
- Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):R115.
- Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, et al. Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*. 2012;11(6):1132–4.
- Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013;49(2):359–67.
- Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, et al. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res*. 2010;20(4):434–9.
- Marttila S, Kananen L, Hayrynen S, Jylhava J, Nevalainen T, Hervonen A, et al. Ageing-associated changes in the human DNA methylome: genomic locations and effects on gene expression. *BMC Genomics*. 2015;16(1):179. Epub 2015 Mar 14.
- Yuan T, Jiao Y, de Jong S, Ophoff RA, Beck S, Teschendorff AE. An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging. *PLoS Genet*. 2015;11(2), e1004996.
- Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res*. 2010; 20(4):440–6.
- Raitakari OT, Juonala M, Ronnemaa T, Keltikangas-Jarvinen L, Rasanen L, Pietikainen M, et al. Cohort profile: the cardiovascular risk in Young Finns study. *Int J Epidemiol*. 2008;37(6):1220–6.
- Steegenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ, et al. Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age (Dordr)*. 2014;36(3):9648. Epub 2014 May 2.
- Alisch RS, Barwick BG, Chopra P, Myrick LK, Satten GA, Conneely KN, et al. Age-associated DNA methylation in pediatric populations. *Genome Res*. 2012;22(4):623–32.
- Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. 2009;10:48.
- Zambelli F, Pesole G, Pavesi G. Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res*. 2009;37(Web Server issue):W247–52.
- Florath I, Butterbach K, Muller H, Bewerunge-Hudler M, Brenner H. Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum Mol Genet*. 2014;23(5):1186–201.
- Johansson A, Enroth S, Gyllenstein U. Continuous aging of the human DNA methylome throughout the human lifespan. *PLoS One*. 2013;8(6), e67378.
- Gubbels Bupp MR. Sex, the aging immune system, and chronic disease. *Cell Immunol*. 2015.
- Oertelt-Prigione S. The influence of sex and gender on the immune response. *Autoimmun Rev*. 2012;11(6-7):A479–85.
- McClay JL, Aberg KA, Clark SL, Nerella S, Kumar G, Xie LY, et al. A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects. *Hum Mol Genet*. 2014;23(5):1175–85.
- Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, et al. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol*. 2014;15(2):R24.
- Hernandez DG, Nalls MA, Gibbs JR, Arepalli S, van der Brug M, Chong S, et al. Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum Mol Genet*. 2011;20(6):1164–72.
- Reynolds LM, Taylor JR, Ding J, Lohman K, Johnson C, Siscovick D, et al. Age-related variations in the methylome associated with gene expression in human monocytes and T cells. *Nat Commun*. 2014;5:5366.
- West J, Beck S, Wang X, Teschendorff AE. An integrative network algorithm identifies age-associated differential methylation interactome hotspots targeting stem-cell differentiation pathways. *Sci Rep*. 2013;3: 1630.
- Lukic S, Nicolas JC, Levine AJ. The diversity of zinc-finger genes on human chromosome 19 provides an evolutionary mechanism for defense against inherited endogenous retroviruses. *Cell Death Differ*. 2014;21(3):381–7.
- Taruscio D, Mantovani A. Factors regulating endogenous retroviral sequences in human and mouse. *Cytogenet Genome Res*. 2004;105(2-4): 351–62.
- Nuotio J, Oikonen M, Magnussen CG, Jokinen E, Laitinen T, Hutri-Kahonen N, et al. Cardiovascular risk factors in 2011 and secular trends since 2007: the cardiovascular risk in Young Finns study. *Scand J Public Health*. 2014; 42(7):563–71.
- Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, et al. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res*. 2006;16(3):383–93.
- Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, et al. Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics*. 2009;1(1):177–200.
- Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011; 98(4):288–95.
- Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013;29(2):189–96.
- Marabita F, Almgren M, Lindholm ME, Ruhmann S, Fagerstrom-Billai F, Jagodic M, et al. An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics*. 2013;8(3):333–46.
- Pidsley R, Wong CC Y, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 2013;14:293.
- Cribari-Neto F, Zeileis A. Beta regression in R. *JSS*. 2010;34:1–24 [http://www.jstatsoft.org/v34/i02/].

42. Triche T, Jr. InfiniumMethylation.hg19: Annotation package for Illumina Infinium DNA methylation probes. R package version 2.2.0.
43. Eden E, Lipson D, Yogev S, Yakhini Z. Discovering motifs in ranked lists of DNA sequences. *PLoS Comput Biol.* 2007;3(3), e39.
44. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One.* 2011;6(7), e21800.

Submit your next manuscript to BioMed Central
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

