

RESEARCH

Open Access



# Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms

Roderick C. Slieker<sup>1</sup>, Maarten van Iterson<sup>1</sup>, René Luijk<sup>1</sup>, Marian Beekman<sup>1</sup>, Daria V. Zhernakova<sup>2</sup>, Matthijs H. Moed<sup>1</sup>, Hailiang Mei<sup>3</sup>, Michiel van Galen<sup>4</sup>, Patrick Deelen<sup>2</sup>, Marc Jan Bonder<sup>2</sup>, Alexandra Zhernakova<sup>2</sup>, André G. Uitterlinden<sup>5</sup>, Ettje F. Tigchelaar<sup>2</sup>, Coen D. A. Stehouwer<sup>6</sup>, Casper G. Schalkwijk<sup>6</sup>, Carla J. H. van der Kallen<sup>6</sup>, Albert Hofman<sup>7</sup>, Diana van Heemst<sup>8</sup>, Eco J. de Geus<sup>9</sup>, Jenny van Dongen<sup>9</sup>, Joris Deelen<sup>1</sup>, Leonard H. van den Berg<sup>10</sup>, Joyce van Meurs<sup>5</sup>, Rick Jansen<sup>11</sup>, Peter A. C. 't Hoen<sup>4</sup>, Lude Franke<sup>2</sup>, Cisca Wijmenga<sup>2</sup>, Jan H. Veldink<sup>10</sup>, Morris A. Swertz<sup>12</sup>, Marleen M. J. van Greevenbroek<sup>6</sup>, Cornelia M. van Duijn<sup>13</sup>, Dorret I. Boomsma<sup>9</sup>, BIOS consortium, P. Eline Slagboom<sup>1</sup> and Bastiaan T. Heijmans<sup>1\*</sup>

## Abstract

**Background:** Epigenetic change is a hallmark of ageing but its link to ageing mechanisms in humans remains poorly understood. While DNA methylation at many CpG sites closely tracks chronological age, DNA methylation changes relevant to biological age are expected to gradually dissociate from chronological age, mirroring the increased heterogeneity in health status at older ages.

**Results:** Here, we report on the large-scale identification of 6366 age-related variably methylated positions (aVMPs) identified in 3295 whole blood DNA methylation profiles, 2044 of which have a matching RNA-seq gene expression profile. aVMPs are enriched at polycomb repressed regions and, accordingly, methylation at those positions is associated with the expression of genes encoding components of polycomb repressive complex 2 (PRC2) *in trans*. Further analysis revealed *trans*-associations for 1816 aVMPs with an additional 854 genes. These *trans*-associated aVMPs are characterized by either an age-related gain of methylation at CpG islands marked by PRC2 or a loss of methylation at enhancers. This distinct pattern extends to other tissues and multiple cancer types. Finally, genes associated with aVMPs *in trans* whose expression is variably upregulated with age (733 genes) play a key role in DNA repair and apoptosis, whereas downregulated aVMP-associated genes (121 genes) are mapped to defined pathways in cellular metabolism.

**Conclusions:** Our results link age-related changes in DNA methylation to fundamental mechanisms that are thought to drive human ageing.

**Keywords:** DNA methylation, Ageing, 450k, DNA damage, Variability

\* Correspondence: bas.heijmans@lumc.nl

<sup>1</sup>Molecular Epidemiology section, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, The Netherlands  
Full list of author information is available at the end of the article

## Background

Studies of model organisms such as yeast, nematodes, and mice have shown that the accumulation of cellular damage is a fundamental cause of ageing across species [1–3]. Epigenetic dysregulation is thought to play a key role in this process [4, 5]. Numerous human population studies have now shown that changes in DNA methylation of CpG dinucleotides, a key epigenetic mechanism [6], are strongly associated with chronological age. Although these epigenetic changes are in part a by-product of age-related changes in the cellular composition of the studied tissue [7–9], many age-related differentially methylated positions (aDMPs) observed in blood samples are independent of cell composition [8–19]. Additional studies showed the consistent occurrence of aDMPs in other tissues [20–22] and species [23] and aDMPs have proven to be a useful tool to predict chronological age [24, 25].

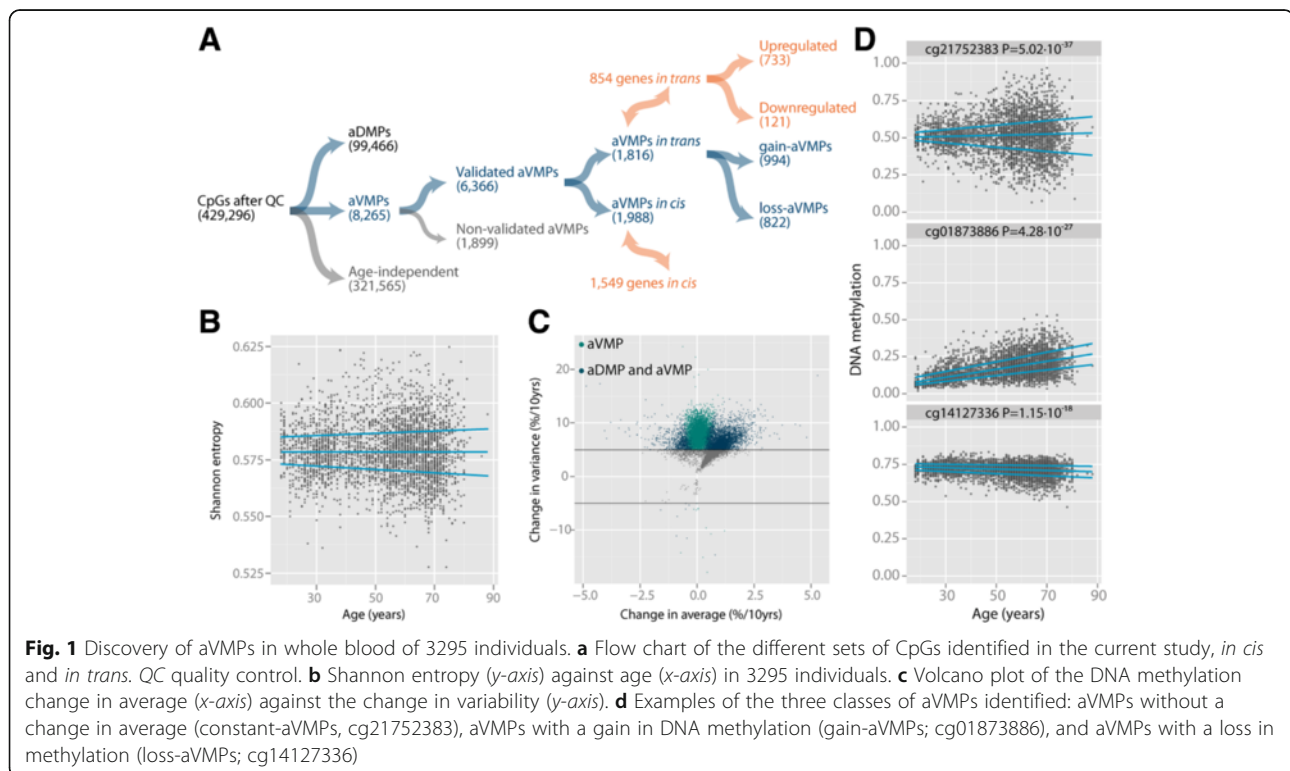
However, aDMPs may not be the most informative marker of the ageing process since they were discovered as close correlates of chronological age instead of biological age [26]. Moreover, only a small proportion of aDMPs are associated with expression changes [12, 17], suggesting that their functional implication may be limited. In contrast, DNA methylation changes that increasingly diverge from chronological age may reflect the increasing inter-individual variation in health that occurs with increasing age. Initial studies, although small or lacking a genome-wide view, indicated that an

increasing variability of DNA methylation with age indeed exists [21, 27, 28]. A recent large twin study showed that genetic factors explain a minor proportion of the variation in DNA methylation in the population, while the influence of environmental and stochastic factors is large and commonly increases with age [19].

In the current study, we charted the occurrence of age-related variably methylated positions (aVMPs) across the genome. We evaluated the methylation at 429,296 CpG sites for increased variability with age in whole blood samples from 3295 individuals. After validation using multiple external datasets and integration of methylome with transcriptome data (n = 2044), we show that aVMPs link to the expression of genes *in trans* that play a central role in fundamental ageing mechanisms.

## Results

To uncover the occurrence of divergence in DNA methylation with age, we studied 3295 whole-blood DNA methylation profiles encompassing 429,296 CpGs of individuals aged 18 to 88 years (Fig. 1a; Additional file 1a, b). To obtain global evidence for an increasing DNA methylomic divergence between individuals with age, we calculated the Shannon entropy on the individual level [25, 29]. With age, we observed a distinct increase in the variability in the Shannon entropy (adjusted for age, sex, and blood cell composition), suggesting that the level of

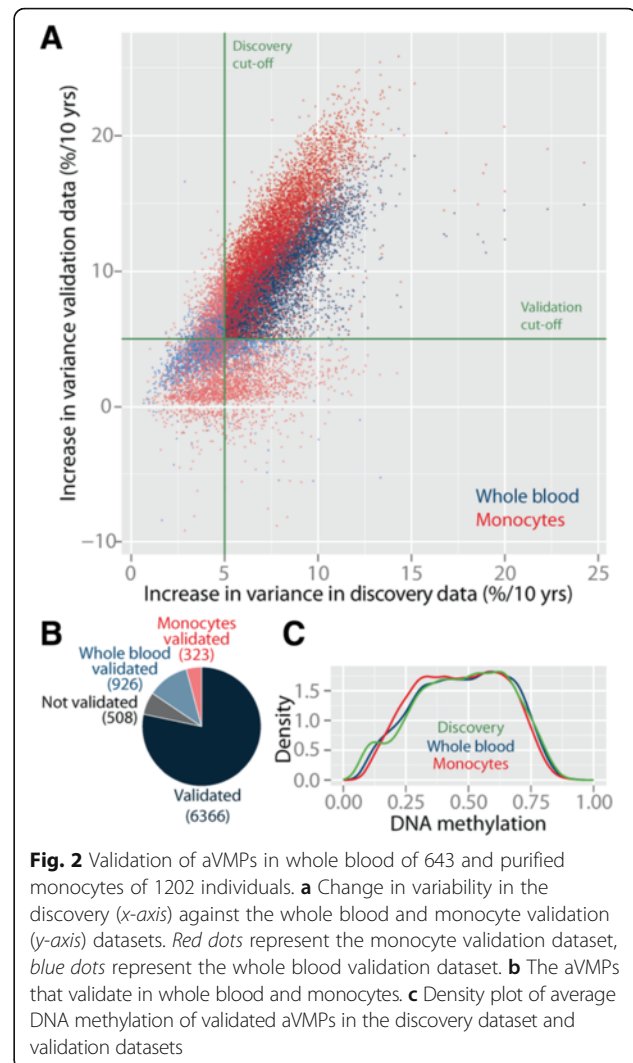


order between individuals becomes more variable at older ages (Fig. 1b).

To map the specific CpGs that change in variability with age (age-related variably methylated positions or aVMPs), we tested for an increase in variability with age, independent of an average change in DNA methylation with age (aDMP effects) and changes in blood cell composition. We identified 8265 aVMPs that showed a robust increase in variability with age ( $P < 10^{-7}$  and  $\geq 5\%$  increase per 10 years; Fig. 1a, c). The number of aVMPs was substantially smaller than aDMPs in our data set: approximately a quarter of the CpGs tested were identified as aDMPs (99,466 CpGs, no effect size cutoff; Fig. 1a; Additional file 1d); the majority of the CpGs did not show an age-related change in DNA methylation (321,565 CpGs; Fig. 1a; Additional file 1c). The number of aVMPs showing a decrease in variability with age was exceedingly small (19 aVMPs,  $P < 10^{-7}$  and  $\geq 5\%$  decrease per 10 years) and these aVMPs were excluded from further analyses (Fig. 1c). Since increases in variability at hypo- or hypermethylated CpGs can only go in one direction (up for DNA methylation levels near 0 and down for those near 1), we observed that 48.2% of the aVMPs were also identified in these data as aDMPs (3980 aVMPs; Fig. 1c). Of note, only 213 aVMPs (2.6%) overlapped with previously reported aDMPs ( $n = 7477$ ) [7, 8, 14, 16, 18, 25] and six aVMPs overlapped with CpGs from the Horvath clock ( $n = 353$ ) [24] (Additional file 1e). Hence, aVMPs represent a new and distinct class of age-related changes in DNA methylation. Clustering analysis showed that aVMPs could be categorized into three groups: aVMPs hypomethylated in young individuals that gained DNA methylation with age (both average and variability; gain-aVMPs), hypermethylated aVMPs that lost DNA methylation with age (both average and variability; loss-aVMPs) and intermediate-methylated aVMPs where gains and losses with age were balanced, resulting in the absence of an aDMP effect (constant-aVMPs) (Fig. 1d).

We validated the aVMPs in two large public datasets consisting of whole blood ( $n = 643$ , age range 20–102, median 65 years) [25] and purified monocytes, a relative homogeneous population of blood cells ( $n = 1202$ , age range 44–83, median 59 years) [17]. This validation step was incorporated to exclude that our results were confounded by the fact that we analyzed multiple population studies of different ages or by age-related changes in the cellular composition of blood. The increase in variability observed in the discovery data was remarkably concordant with the effect size in both validation datasets (Fig. 2a). In total, 6366 aVMPs (78.4%) were validated in both datasets (Fig. 2b; Additional file 2).

Although we corrected for cell heterogeneity using the major blood cell subtypes, age-related changes in the percentage of less frequent cell subtypes could also have introduced aVMPs. To further exclude that the increased



**Fig. 2** Validation of aVMPs in whole blood of 643 and purified monocytes of 1202 individuals. **a** Change in variability in the discovery (x-axis) against the whole blood and monocyte validation (y-axis) datasets. Red dots represent the monocyte validation dataset, blue dots represent the whole blood validation dataset. **b** The aVMPs that validate in whole blood and monocytes. **c** Density plot of average DNA methylation of validated aVMPs in the discovery dataset and validation datasets

variability was driven by changes in blood cell composition, we repeated our aVMP analysis with the inclusion of predicted, more refined blood cell subtype fractions (CD8<sup>+</sup> T cells, CD4<sup>+</sup> T cells, natural killer (NK) cells, B cells, and granulocytes) [30]. Out of the 6366, 6343 (99.6%) were again identified as aVMPs ( $P_{FDR} \leq 0.05$ ). Furthermore, we compared the DNA methylation of whole blood to that of blood cell subtypes and their progenitors using five public data sets on 19 cell types (Additional file 3). The methylation of aVMPs was highly concordant between whole blood and the various cell subtypes, indicating aVMPs cannot be attributed to the outgrowth of a specific cell subtype, but instead represent genuine age-related changes in DNA methylation. Conversely, CpGs differentially methylated between blood cell subtypes [31] had the same, if not smaller, chance of being an aVMP (odds ratio = 0.56;  $P = 0.19$ ).

The validated set of 6366 robust and consistent aVMPs was taken forward for in-depth analysis. aVMPs were

characterized by an intermediate methylation level (Fig. 2c). Among the 6366 aVMPs, a similar number of gain-aVMPs and loss-aVMPs were found (2788 gain-aVMPs, 3578 loss-aVMPs). While some individuals showed differential DNA methylation on many of the identified aVMPs (marked by vertical bands in Additional file 1f), others did not show any difference in DNA methylation. The individual-specific changes in DNA methylation with age also resulted in an age-related loss of correlation between individuals on the 6366 aVMPs (Additional file 1g).

### aVMPs are depleted in active regions and enriched in PcG repressed regions

To characterize the genomic regions harboring aVMPs, we obtained the chromatin state segments of blood cell types (Epigenomics Roadmap [32]), which reflect the biological function of the underlying region in blood cells on the basis of combinations of histone modifications [32]. A pronounced enrichment was found for aVMPs in the segments marking a repressed genome, including *repressed polycomb* (7.2-fold enrichment,  $P < 0.0001$ ), *weak repressed polycomb* (2.3-fold enrichment,  $P < 0.0001$ ) and *heterochromatin* (2.5-fold enrichment,  $P < 0.0001$ ), while aVMPs were depleted for active segments, including *strong transcription* (0.04-fold enrichment,  $P < 0.0001$ ) (Fig. 3a; Additional file 4b). In absolute terms, 4212 aVMPs (66.2 %) mapped to segments marking repressed DNA. These data were supported by a parallel enrichment of aVMPs for repressive histone modifications (Additional file 4a) and for binding sites of the PcG repressive complex 2 (PRC2) protein EZH2 in the ENCODE blood cell line GM12878 (2.1-fold enrichment,  $P < 0.0001$ ).

To explore whether the methylation level of aVMPs was associated with gene expression *in cis* (i.e., gene and aVMP within 500 kb), we studied the relationship between DNA methylation and gene expression using 2044 individuals for whom both DNA methylation and gene expression data (RNA-seq) were available. Despite predominantly mapping to repressed regions, 1988 *cis*-aVMPs out of the 6366 aVMPs (31.2 %; Fig. 1a; Additional file 5) were associated with gene expression of 1549 genes *in cis* (see Additional file 4c for *HOXC4* and *PCDHB6* examples). Gene Ontology (GO) analysis of the *cis*-associated genes showed enrichment for processes involved in *neuron differentiation* and *neuron development* ( $P < 0.0001$ ).

Since age-related divergence was found to be over-represented near developmental genes, we also mapped aVMPs to chromatin state segments of a human embryonic stem cell line (H1-hESC). Interestingly, repressed genomic regions in blood harboring an aVMP were often a (bivalent) transcription start site or (bivalent) enhancer in the H1-hESC line (Fig. 3b), highlighting the developmental history of regions accumulating aVMPs.

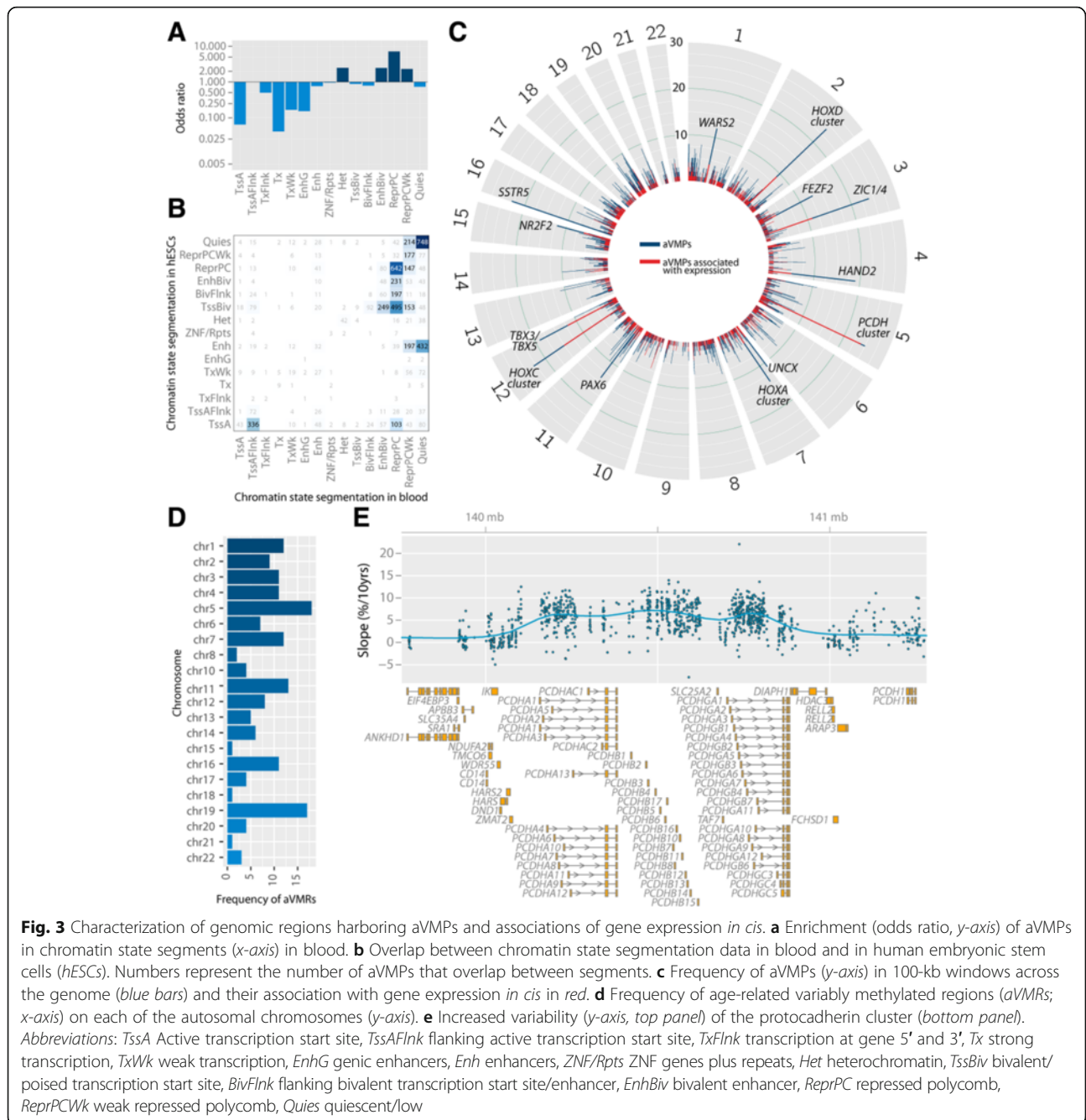
In line with the enrichment for developmental genes, we found a large number of aVMPs near (neuro)developmental genes (Fig. 3c). To investigate whether aVMPs clustered into regions, we identified age-related variably methylated regions (aVMRs) [33]. This resulted in 160 aVMRs encompassing 527 aVMPs (8.3 %; Additional file 6). aVMRs were particularly frequent on chromosomes 5 and 19 (Fig. 3d), the former of which could be attributed, in part, to eight aVMRs (totalling 26 aVMPs) that mapped to the protocadherin gene cluster (Fig. 3e).

### aVMPs are associated with expression of genes *in trans*

The enrichment of aVMPs in polycomb-repressed regions and PRC2 binding sites suggests a role for PcG proteins in the occurrence of aVMPs. Indeed, methylation at the majority of aVMPs was positively associated with the expression of components of the PRC2, including *EED*, *SUZ12*, and *EZH2*, particularly when an aVMP mapped to an EZH2 binding site (Additional file 7a). Further analysis revealed that *trans*-associations were not limited to PRC2 components. Of the total number of 6366 aVMPs, 1816 (28.5 %) were associated with expression of 854 coding genes *in trans* (4.6 % of genes tested; Fig. 4a; Additional file 8), i.e., the aVMP and gene on different chromosomes or on the same chromosome but 5 Mb apart. The association between aVMP methylation and gene expression implies an increase in variance in gene expression in conjunction with DNA methylation. Of note, the number of associations between *trans*-genes and *trans*-aVMPs was high. For example, the expression of *TPRG1* was associated with 1296 correlated aVMPs and, conversely, the aVMP *cg13246235* located near *PHACTR1* was associated with 853 associated *trans*-genes (Fig. 4b).

*Trans*-aVMPs encompassed two classes (instead of three in the complete set of aVMPs): 994 gain-aVMPs that, apart from increasing in variability with age, gained methylation and were hypomethylated in young individuals; and 822 loss-aVMPs that lost DNA methylation with age and were hypermethylated in young individuals (Figs. 1a and 4c). The genomic context of gain- and loss-aVMPs was markedly different. Gain-aVMPs were strongly enriched for CpG islands (CGIs) compared with loss-aVMPs (22.9-fold,  $P < 0.0001$ ), while being depleted for non-CGI regions (Fig. 4d; 0.05-fold,  $P < 0.0001$ ). Conversely, loss-aVMPs were highly enriched for non-CGI regions (19.0-fold,  $P < 0.0001$ ; Fig. 4d). These results are in line with the fact that CGIs are commonly hypomethylated and hence can only gain DNA methylation, while non-CGI regions tend to be hypermethylated and preferentially lose DNA methylation. Furthermore, loss-aVMPs were generally found in active regions (*transcription flanking*, 4.3-fold enrichment,  $P < 0.01$ ), but also in *weak*

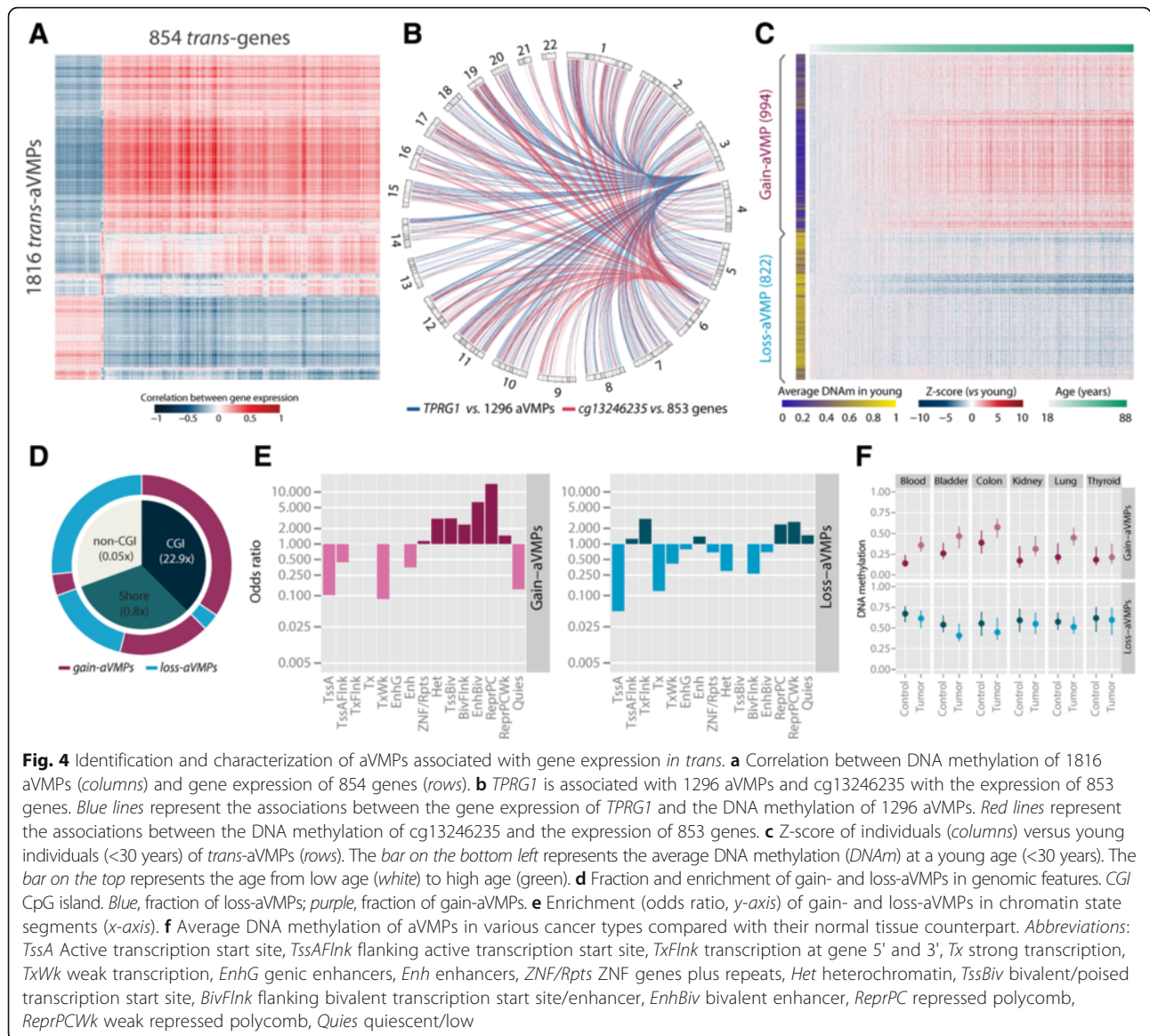




**Fig. 3** Characterization of genomic regions harboring aVMPs and associations of gene expression *in cis*. **a** Enrichment (odds ratio, *y*-axis) of aVMPs in chromatin state segments (*x*-axis) in blood. **b** Overlap between chromatin state segmentation data in blood and in human embryonic stem cells (*hESCs*). Numbers represent the number of aVMPs that overlap between segments. **c** Frequency of aVMPs (*y*-axis) in 100-kb windows across the genome (*blue bars*) and their association with gene expression *in cis* in *red*. **d** Frequency of age-related variably methylated regions (aVMPs; *x*-axis) on each of the autosomal chromosomes (*y*-axis). **e** Increased variability (*y*-axis, *top panel*) of the protocadherin cluster (*bottom panel*). Abbreviations: *TssA* Active transcription start site, *TssAFlnk* flanking active transcription start site, *TxFlnk* transcription at gene 5' and 3', *Tx* strong transcription, *TxWk* weak transcription, *EnhG* generic enhancers, *Enh* enhancers, *ZNF/Rpts* ZNF genes plus repeats, *Het* heterochromatin, *TssBiv* bivalent/poised transcription start site, *BivFlnk* flanking bivalent transcription start site/enhancer, *EnhBiv* bivalent enhancer, *ReprPC* repressed polycomb, *ReprPCWk* weak repressed polycomb, *Quies* quiescent/low

*polycomb repressed* regions (3.7-fold enrichment,  $P < 0.0001$ ) (Fig. 4e). Gain-aVMPs were overrepresented in repressed and bivalent regions and particularly in *PcG repressed* regions (11.5-fold enriched,  $P < 0.0001$ ; Fig. 4e). The enrichment analysis yielded similar results when gain- and loss-aVMPs were evaluated separately: both types were enriched for polycomb repressed regions and gain-aVMPs were enriched for bivalent domains. Bivalent domains have been associated with an enrichment for linear age-related DNA methylation changes [8].

Next, we investigated whether the age-related gain and loss of methylation at the *trans*-aVMPs extended beyond blood using publically available datasets. Gain-aVMPs showed a similar or even larger change in average methylation with age in, for example, colon (91.2 % concordant direction,  $P < 0.0001$ ), lung (89.6 %,  $P < 0.0001$ ), skin (87.0–93.3 %,  $P < 0.0001$ ) and SAT (89.0 %,  $P < 0.0001$ ) (Additional file 9). For loss-aVMPs, evidence for age-related changes in the same direction was found in lung (75.0 %,  $P < 0.0001$ ), colon (70.9 %,  $P$

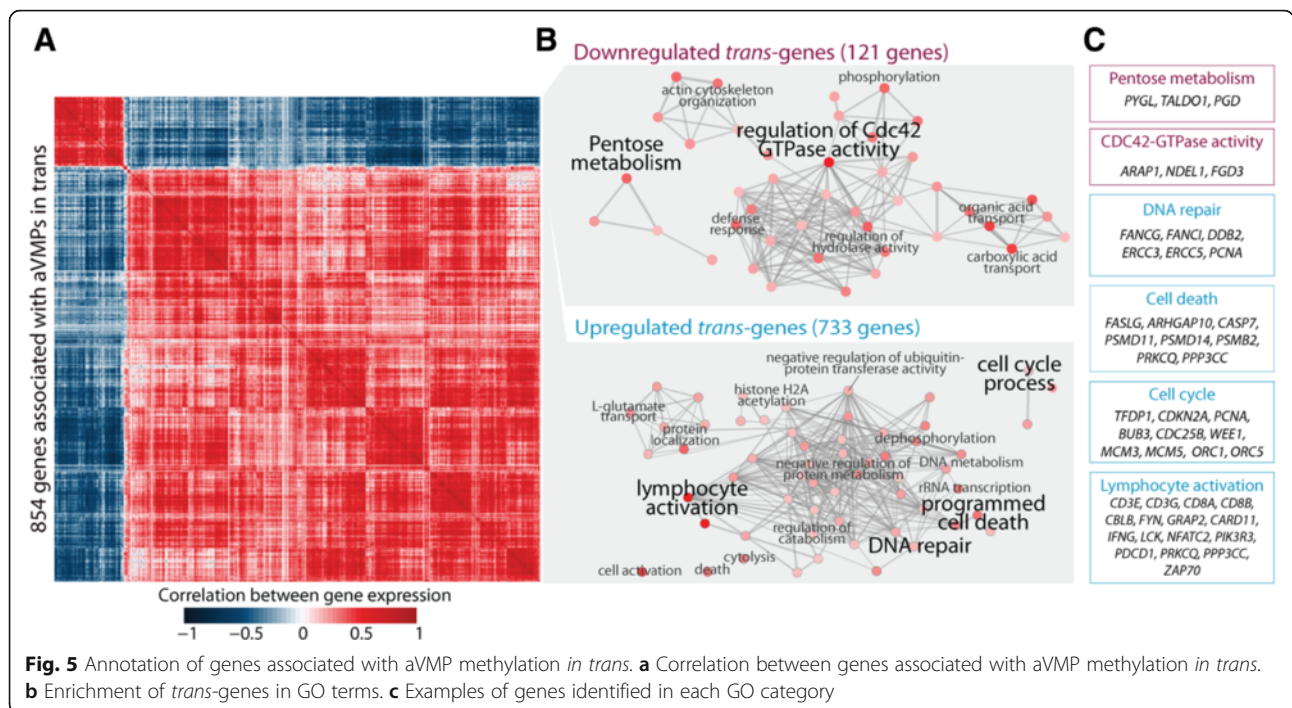


< 0.0001), and skin (56.3–66.1 %,  $P < 0.01$ ). Within skin [34], age-related DNA methylation changes were most pronounced in the epidermis (compared with dermis), with the strongest effect sizes in sun-exposed epidermis samples. In view of the assumed commonalities between ageing and cancer, we investigated whether the changes also extend to different tumors given that gain of DNA methylation at CGIs is a hallmark of tumor biology [35]. Gain-aVMPs were found to show relative hypermethylation and loss-aVMPs hypomethylation, respectively, across various cancer types, including blood, bladder, colon, and lung cancer ( $P < 0.0001$ ; Fig. 4f), which suggests a striking similarity between age-affected and cancer methylomes, in line with previous observations for aDMPs [36]. Taken together, aVMPs are largely driven by tissue-independent

factors, but these changes may be accelerated by external influences (like sun exposure) and tumorigenesis.

**aVMP-associated *trans*-genes linked to DNA damage and apoptosis**

The expression of genes associated with aVMPs *in trans* unambiguously clustered into two highly correlated gene sets (Fig. 5a). An age-affected methylome was associated with down-regulated expression of *trans*-genes in a cluster of 121 genes and with up-regulated expression of a larger cluster of 733 genes. GO analysis showed that down-regulated genes are involved in various intracellular metabolic pathways, including pentose metabolism and regulation of *CDC42* GTPase activity (Fig. 5b), of which the former remained significant when using a



permutation-based enrichment test (Additional file 10). Key genes in the pentose metabolism process whose expression was associated with aVMP methylation included *PYGL*, *TALDO1*, and *PGD* (Fig. 5b, c). Up-regulated *trans*-genes were intimately involved in apoptosis, cell cycle, DNA repair, and lymphocyte activation (Fig. 5b) and these enrichments were also observed using the permutation-based enrichment test (20,000 permutations; Additional file 10). The process unifying these pathways might be the cell damage response encompassing the upregulation of DNA repair, cell cycle changes, and upregulated apoptosis. In each category, key genes were identified, including the *ERCC* genes (DNA repair), *CDKN2A* (encoding the *INK4A/ARF* locus), *BUB3* (checkpoint), and *CASP7* (apoptosis) (Fig. 5c). Of note, a small proportion of the identified genes were previously found to be correlated with chronological age in 14,983 samples (163 genes, 19.1 %; Additional file 7b) [37], illustrating that the *trans*-genes we identified here represent, to a large extent, a different phenomenon. In summary, aVMP DNA methylation is associated with the downregulation of expression of genes in intra-cellular metabolism pathways and the upregulation of expression of genes in ageing pathways.

### Discussion

In a large-scale analysis we discovered and validated 6366 age-related variably methylated positions (aVMPs). These aVMPs were found to occur in three classes: gain-aVMPs that increase in variance and average DNA methylation with age, loss-aVMPs that increase in variance but decrease

in average, and constant-aVMPs that increase in variance but with a constant average. aVMPs accrue in repressed regions that are characterized by both the PcG-deposited histone mark H3K27me3 and the binding site of EZH2, a component of PRC2. While aVMPs were commonly associated with the expression of (neuro)developmental genes *in cis*, they were linked to transcriptional activity of genes *in trans* that have a key role in well-established ageing pathways such as intracellular metabolism, apoptosis, and DNA damage response. Of interest, tumors were found to accumulate DNA methylation changes at CpG sites of aVMPs, thus supporting the long-standing notion that ageing and cancer are in part driven by common mechanisms [2].

Our data show that the genomic regions accumulating variability in ageing populations are highly specific and reproducible. Hence, although the increase in variability may have a stochastic component, the regions affected by this phenomenon are well-defined and not stochastic. aVMPs were discovered in whole blood samples (corrected for blood cell composition) and validated in independent whole blood and purified blood cell type (i.e., monocytes) samples. The results are thus unlikely to be driven by age-related changes in blood cell composition. Furthermore, we compared the DNA methylation of aVMPs to the DNA methylation of 19 blood cell subtypes and showed that the DNA methylation is highly concordant and, conversely, CpGs known to be differentially methylated between blood cell subtypes were not overrepresented among aVMPs. To our knowledge, our



analyses together represent the most comprehensive validation of whole blood-discovered differential DNA methylation to date. Further support for the involvement of a cell type-independent phenomenon is the observation that aVMP methylation was associated with *in cis* and *in trans* expression of genes that function in developmental and ageing-related cellular processes, respectively, instead of immune pathways. To definitively confirm the cross-tissue occurrence of aVMPs, genome-wide DNA methylation data sets for internal tissues are required that have a similarly large size as those currently available for blood samples. Finally, aVMPs associated with gene expression *in trans* discovered in blood displayed a similar pattern of gain and loss of methylation with age across a series of tissues.

Importantly, our data indicate that aVMPs constitute a class of CpGs displaying age-related changes in the level of DNA methylation that is distinct from aDMPs. Known aDMPs rarely were aVMPs (2.6 % of 7477 previously reported aDMPs [7, 8, 14, 16, 18, 25]) and, in contrast to aDMPs, aVMPs showed striking associations with gene expression. Hence, aVMPs are not driven by changes in mean towards a methylation fraction of 0.5 that, in principle, could lead to an increased variance. This mathematical effect that cannot, however, be completely ruled out for all individual aVMPs.

aVMPs preferentially occur in repressed regions marked by PcG repression, namely the PcG binding site of EZH2 and repressive histone marks (H3K27me3). The integration of methylome with transcriptome data revealed that, in contrast to aDMP methylation, aVMP methylation is commonly associated with gene expression. Genes associated with aVMPs *in cis* frequently had a (neuro)developmental role exemplified by the *HOX* gene clusters, the protocadherin gene cluster, *TBX* genes, and *ZIC* genes. aVMPs reported here share the overrepresentation at PRC2-controlled regions with a subset of previously reported aDMPs [7, 8, 11, 12, 17, 36, 38], regions undergoing differential methylation observed in long-term cultured human senescent mesenchymal stem cells [39], various cancer types [7, 36], and, finally, regions displaying differential methylation *in vitro* after oxidative stress-induced DNA damage [40]. The latter study is of particular relevance to the interpretation of our results since it showed that oxidative DNA damage leads to PRC2 recruitment to sites with DNA damage [40] and results in translocation of DNA methyltransferases and EZH2 from CpG-poor regions to CG-rich regions, which in turn leads to hypermethylation of CGIs and hypomethylation at CpG-poor regions [40, 41]. We observed that aVMP methylation is frequently and strongly associated with the expression of PRC2 components *in trans*. Moreover, aVMPs were characterized by a gain of methylation at CGIs and a loss of methylation at CpG-poor regions. Our

data highlight the potentially important role of altered PcG regulation in ageing.

Intriguingly, associations of aVMP methylation with gene expression *in trans* were not limited to PRC2 genes but extended to genes known to play a role in ageing. In older individuals who had an aged DNA methylation profile as compared with young individuals, we observed a downregulation of genes involved in metabolism. The upregulation of ageing pathways, as observed in old individuals with an aged methylome, has been reported previously in hematopoietic stem cells in mice and humans, for which macromolecular or DNA damage may be the driving force [42, 43]. Of note, many of the *trans*-genes we identified are involved in the DNA damage response and are frequently mutated in various cancers, including *CDKN2A*, *DNMT3A*, and *TP53* [44]. Hence, genomic stress, due either to hyperproliferation or DNA damage, may drive upregulation of well-established ageing pathways, downregulation of intra-cellular metabolism, and altered regulation by PcG proteins associated with increased variability of DNA methylation.

## Conclusions

In contrast to aDMPs, aVMPs show a striking variability in DNA methylation at higher ages. Two individuals of the same age may display highly distinct methylation patterns across aVMPs, where one of them may have a DNA methylation profile at aVMPs that is similar to that of young individuals. Therefore, aVMPs fulfill a primary prerequisite for a biomarker of biological age [26]. Further studies are required to establish whether aVMP-based methylation profiles mark health status and predict mortality. Finally, our study shows that large-scale integrative genomics studies are an effective approach toward the identification of fundamental processes involved in ageing and are complementary to experimental work in model organisms.

## Methods

### Data

DNA methylation data and RNA-seq data were generated within the Biobank-based Integrative Omics Studies Consortium (BIOS Consortium; Additional file 11) [45, 46]. Discovery data generated within the BIOS consortium are available from the European Genome-phenome Archive (EGA) under accession number EGAC00001000277. The data comprise six Dutch biobanks: Cohort on Diabetes and Atherosclerosis Maastricht (CODAM) [47], LifeLines (LL) [48], Leiden Longevity Study (LLS) [49], Netherlands Twin Registry (NTR) [50], Rotterdam Study (RS) [51], and the Prospective ALS Study Netherlands (PAN) [52]. A random co-twin per twin pair from the Netherlands Twin Registry was included in the current dataset to restrict our analysis to unrelated individuals. Briefly, 500 ng of genomic DNA



was bisulfite converted using the EZ DNA Methylation kit (Zymo Research, Irvine, CA, USA) and hybridized on Illumina Infinium 450 arrays according to the manufacturer's protocols and signal intensities measured using an Illumina iScan BeadChip scanner. Sample identity of DNA methylation and expression data was confirmed with genotype data using *MixupMapper* [53]. Quality control (QC) on the DNA methylation data was performed using the R package *MethylAid* [54]. Out of the 3391 samples, 3296 samples passed QC. Data were normalized using functional normalization implemented in the R package *minfi* using five principal components [55]. Ambiguously mapped probes [56], probes with a high detection  $P$  value ( $>0.01$ ), probes with a low bead count ( $<3$  beads), and probes with a low success rate (missing in  $>95\%$  of the samples) were set to missing. Probes mapping to chromosomes X and Y were excluded from all analyses. Residual batch effects were removed using Combat as implemented in the R package SVA, with biobank as batch and gender and age as outcome variables [57].

RNA-seq comprised 2044 expression profiles for which also DNA methylation data were available. RNA libraries were prepared using the Illumina's Truseq version 2 library preparation kit and paired-end sequenced of  $2 \times 50$  bp using Illumina's HiSeq2000. Using Illumina's CASAVA, read sets per sample were generated and only reads passing the Chastity Filter were used for further processing. Using *FastQC20* (v0.10.1) [58] initial QC was performed. Adapters were removed using *cutadapt21* (v1.1) [59]. Using *Sickle22* (v1.2) [60], low quality ends of the reads were trimmed (minimum length 25, minimum quality 20). Reads that passed QC were aligned to human genome build *hg19* using STAR23 (v2.3.125) [61] and gene quantifications were based on Ensembl version 71. Gene counts were normalized for GC content and gene length using the R package *cqn* [62]. Associations between gene expression and DNA methylation were performed using *voom*-transformed values [63]. For graphical purposes normalized counts were transformed to RPKM values. Generation of methylome and transcriptome data was performed by the Human Genotyping facility (HugeF) of the ErasmusMC (the Netherlands, <http://www.glimdna.org/>).

Cell count data (neutrophils, lymphocytes, monocytes, eosinophils, and basophils) were available for the majority of the samples ( $>60\%$ ). A prediction model for blood cell composition was fitted using a subset of the data for which cell counts were available using a multivariate partial-least-squares model (including age and gender) on the normalized DNA methylation data. Using the fitted model, the cell fractions were imputed for all samples and only these fractions were used in all analyses. Documentation and code are available in the R package *wbccPredictor* (available from <https://github.com/mvanterson/wbccPredictor>). Cell composition in the whole

blood validation [25] dataset was also imputed using the described method. For the monocyte validation dataset, data on purity of the isolated cells was obtained from the Gene Expression Omnibus (GEO). To further confirm our results with the blood cell subtypes described by Houseman et al. [30], we used the implementation in the R package *minfi* [64] to impute the cell fractions of CD8<sup>+</sup> T cells, CD4<sup>+</sup> T cells, NK cells, B cells, and granulocytes.

Validation of aVMPs was performed in two external datasets, whole blood and monocyte datasets (Additional file 12). For the whole blood data, IDAT files used in Hannum et al. [25] were kindly provided by the authors. Data underwent the same quality and normalization procedure as used above. After quality control, 643 samples were used in subsequent samples. For the monocyte data, normalized data were obtained from the GEO [65] (accession number GSE56046 [17]). Raw DNA methylation data (Level 1; IDAT files) of healthy tissues and cancerous tissues were obtained from The Cancer Genome Atlas (TCGA Research Network, <http://cancergenome.nih.gov>; Additional file 12) and underwent QC and normalization equal to the discovery data. Normalized DNA methylation data for subcutaneous fat were obtained from the ArrayExpress [66] (accession number E-MTAB-1866 [67]). Normalized dermis and epidermis DNA methylation data were downloaded from GEO under accession number GSE51954 [34]. Normalized data for cytogenetic normal acute myeloid leukemia (CN-AML) and healthy bone marrow CD34+ cells were downloaded from GEO under accession number GSE58477 [68].

### Shannon entropy and aVMPs

All analyses were performed on the normalized whole dataset consisting of all biobanks together. To calculate the Shannon entropy, DNA methylation data were corrected for age, cell composition, and gender. Shannon entropy was calculated using a previously described method for DNA methylation data [25].

aVMPs were identified by using the Breusch–Pagan test for heteroscedasticity [69]. First, average change in age, blood cell composition, and gender were regressed out. Next, squared residuals were tested for an association with age with correction for blood cell composition and gender. aVMPs were defined as CpGs that showed a significant association between squared residuals and age with a Bonferroni corrected  $P$  value  $\leq 0.05$ . Moreover, aVMPs were only selected if the increased variability with age was larger than 5 % per 10 years. aVMPs were further subdivided into three classes based on a clustering analysis: gain-aVMPs, loss-aVMPs, and constant-aVMPs. Gain-aVMPs were defined as CpGs that were hypomethylated at young age but where the change in average DNA methylation was positive (Fig. 1d

middle panel). Loss-aVMPs were defined as CpGs that were hypermethylated at young age with a negative change in average DNA methylation with age (Fig. 1d bottom panel). Constant-aVMPs were defined as CpGs that were intermediately methylated and did not show a change in average but only in variance (Fig. 1d top panel).

For the association between DNA methylation of aVMPs and gene expression *in cis*, we only considered genes within 500 kb of the aVMP and we adjusted for gender and blood cell composition. The *P* value was corrected for multiple testing using the Bonferroni method for the number of genes tested *in cis* ( $P_{\text{bonf}} \leq 0.05$ ). Next, the most significant associated gene (if any) was selected. Finally, the *P* value of each of these aVMP-gene pairs was corrected using the false discovery rate (FDR) method ( $P_{\text{FDR}} \leq 0.05$ ) and only significant pairs were used [70].

*In trans*, we only evaluated aVMP and gene pairs that were >5 Mb away from each other on the same or on a different chromosome. *P* values for *trans*-associations were FDR corrected ( $P_{\text{FDR}} \leq 0.05$  considered significant). Genes whose gene expression was associated with less than 5 % of all aVMPs ( $\leq 318$  aVMPs, 6366) *in trans* were discarded.

Z-scores (compared to young) were calculated as a measure of directionality and magnitude of aberrant methylation of an aVMP within an individual:

$$Z_{i,j} = \frac{(M_{i,j} - \overline{M_{y,i}})}{\sigma_{y,i}}$$

where  $Z_{i,j}$  is the Z-score of the  $j^{\text{th}}$  individual and  $i^{\text{th}}$  aVMP,  $M_{i,j}$  the DNA methylation of the  $j^{\text{th}}$  individual and  $i^{\text{th}}$  aVMP,  $\overline{M_{y,i}}$  the average DNA methylation in young individuals (<30 years) of the  $i^{\text{th}}$  aVMP, and  $\sigma_{y,i}$  the standard deviation in young individuals of the  $i^{\text{th}}$  aVMP. All heatmaps were clustered based on Euclidian distance.

aVMPs were clustered to aVMRs using a method described before using default settings [33]. Genes were linked to aVMRs based on the nearest protein coding gene (Ensembl).

#### Annotations and integration with external data

Chromatin state segments were downloaded from the Epigenomics Roadmap for different blood cell subtypes [32]. CpGs were annotated to different segments based on the most frequent occurring feature in the various blood cell subtypes. Transcription factor binding sites (ChIP-seq) were obtained from the ENCODE project for all tissue types available [71]. CGI-centric annotations were obtained from previous work [33]. Enrichments were expressed as odds ratio on a log2 scale (if applicable).

GO enrichments were performed using the default settings of DAVID. To reduce redundancy in the GO enrichment terms, we used REVIGO with default settings [72]. Networks were plotted using Cytoscape version 3.2.1 [73]. The enrichment for gene sets was further verified with the gene set enrichment function (gseGO, biological processes) in the R package *clusterProfiler* [74] based on the fold change of the association between DNA methylation and gene expression *in trans* and 20,000 permutations.

To compare aVMPs in healthy tissues other than blood a linear model was fitted between DNA methylation and age for colon (TCGA), lung (TCGA), kidney (TCGA), bladder (TCGA), thyroid (TCGA), subcutaneous fat (GEO) [67] and skin (GEO) [34]. To compare healthy and cancerous tissues, data for the *trans*-aVMPs were obtained in CN-AML versus bone marrow CD34<sup>+</sup> cells (GEO) [68], bladder urothelial carcinoma versus healthy bladder (TCGA), colon adenocarcinoma versus healthy colon (TCGA), kidney renal clear squamous cell carcinoma versus healthy kidney (TCGA), lung adenocarcinoma versus healthy lung (TCGA), and thyroid carcinoma versus healthy thyroid (TCGA).

Analyses were performed using *R* statistics, version 3.1.2. Figures were produced using *ggplot2* [75], the Perl version of Circos (v0.67-7) [76] and Gviz [77].

#### Additional files

**Additional file 1:** Data characteristics and characteristics of aVMPs. a Sample description per biobank. b Distribution of age in the samples used. c Example of an age-independent CpG site where the DNA methylation does not change in age or variance with age near *NOC2L*. d Example of an aDMP, with an average change in DNA methylation but not variance near *BM11*. e Overlap between aVMPs and frequently identified aDMPs and CpGs in Horvath's ageing clock [24]. f Heatmap of Z-score of individuals (columns) versus young individuals (<30 years) of all 6366 aVMPs (rows). g Correlation between individuals (*y-axis*) for each of the age groups in our study (*x-axis*). The blue curve is a loess smoothed curve. (TIF 4010 kb)

**Additional file 2:** List of validated aVMPs and their genomic location (hg19). (XLSX 234 kb)

**Additional file 3:** Comparison of whole blood DNA methylation (discovery data) to blood sub-cell type DNA methylation. The two plots below each blood cell type are based on aVMPs (*left*) and as a control the 600 CpGs from Jaffe and Irizarry [31] (*right*), of which it is known that the DNA methylation between cell types is known to be different. Percentages were obtained from [https://www.stemcell.com/media/files/wallchart/WA10006-Frequencies\\_Cell\\_Types\\_Human\\_Peripheral\\_Blood.pdf](https://www.stemcell.com/media/files/wallchart/WA10006-Frequencies_Cell_Types_Human_Peripheral_Blood.pdf). (TIF 6010 kb)

**Additional file 4:** Enrichment of aVMPs for histone modifications and relation with expression *in cis*. a Enrichment (odds ratio, *y-axis*) for histone modifications of aVMPs (*x-axis*). b Enrichment (odds ratio, *y-axis*) of gain- and loss-aVMPs in chromatin state segments (*x-axis*). c Two examples of associations between aVMPs and gene expression of *PCDHB6* and *HOXC4*. (TIF 1170 kb)

**Additional file 5:** Relation between DNA methylation and gene expression *in cis*. (XLSX 119 kb)

**Additional file 6:** Identified aVMRs with nearest protein coding gene. (XLSX 19 kb)

**Additional file 7:** a Correlation between DNA methylation and gene expression of PcG complex proteins. b Overlap between genes associated *in cis* and *in trans* with the transcriptome clock [37]. (TIF 531 kb)

**Additional file 8:** Relationship between DNA methylation and gene expression *in trans*. (XLSX 41 kb)

**Additional file 9:** Slope of average DNA methylation of aVMPS in various healthy tissues. (TIF 559 kb)

**Additional file 10:** Enrichment of identified *trans*-genes in GO terms based on 20 k permutations. (XLSX 104 kb)

**Additional file 11:** BIOS Consortium (Biobank-based Integrative Omics Study). (PDF 68 kb)

**Additional file 12:** Description of external datasets used. (PDF 79 kb)

### Acknowledgements

Samples were contributed by LifeLines (<http://lifelines.nl/lifelines-research/general>), the Leiden Longevity Study (<http://www.leidenlangleven.nl>), the Netherlands Twin Registry (<http://www.tweelingenregister.org>), the Rotterdam Study, (<http://www.erasmus-epidemiology.nl/research/ergo.htm>), the CODAM study (<http://www.carimmaastricht.nl/>) and the PAN study (<http://www.alsonderzoek.nl/>). All analyses were carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. Members of the BIOS Consortium are listed in Additional file 11.

### Funding

This work was done within the framework of the Biobank-Based Integrative Omics Studies (BIOS) Consortium funded by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO 184.021.007 and within the European Union's Seventh Framework Program IDEAL (FP7/2007-2011) under grant agreement number 259679.

### Availability of data and materials

DNA methylation and gene expression data, generated within the BIOS Consortium, are available from the European Genome-phenome Archive (EGA) under accession number EGAC00001000277. The results shown here are in part based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>.

### Authors' contributions

Conceptualization, RCS and BTH; methodology, RCS, BTH, MI, RL; investigation, RCS, BTH; writing (original draft), RCS, BTH; resources, MB, DVZ, MHM, HM, MvG, MJB, AZ, AGU, EFT, CDAS, CGS, CJHvdK, BAH, DvH, EldeG, JvD, JD, LvdB, JvM, RJ, LF, CW, JHV, MMJvG, GMvD, DIB, ES. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Ethics approval and consent to participate

Institutional review boards of the participating centers approved this study (CODAM, Medical Ethical Committee of the Maastricht University; LL, Ethics committee of the University Medical Centre Groningen; LLS, Ethical committee of the Leiden University Medical Center; PAN, Institutional review board of the University Medical Centre Utrecht; NTR, Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre; RS, Institutional review board (Medical Ethics Committee) of the Erasmus Medical Center). Informed consent was given by all participants. Experimental procedures in this study comply with the Declaration of Helsinki.

### Author details

<sup>1</sup>Molecular Epidemiology section, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, The Netherlands. <sup>2</sup>Department of Genetics, University Medical Centre Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands. <sup>3</sup>Sequence Analysis Support Core, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, The Netherlands. <sup>4</sup>Department of Human Genetics, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, The Netherlands. <sup>5</sup>Department of Internal Medicine, Erasmus University Medical Center, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands. <sup>6</sup>Department of Internal Medicine and School for Cardiovascular Diseases (CARIM), Maastricht University Medical

Center, Universiteitssingel 50, 6229 ER Maastricht, The Netherlands.

<sup>7</sup>Department of Epidemiology, Erasmus University Medical Center, Dr. Molewaterplein 50, 3015 GE Rotterdam, The Netherlands. <sup>8</sup>Department of Gerontology and Geriatrics, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands. <sup>9</sup>Department of Biological Psychology, VU University Amsterdam, Neuroscience Campus Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. <sup>10</sup>Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands. <sup>11</sup>Department of Psychiatry, VU University Medical Center, Neuroscience Campus Amsterdam, A.J. Ernststraat 1187, 1081 HL Amsterdam, The Netherlands. <sup>12</sup>Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands. <sup>13</sup>Department of Genetic Epidemiology, Erasmus University Medical Center, Dr. Molewaterplein 50, 3015 GE Rotterdam, The Netherlands.

Received: 26 April 2016 Accepted: 1 September 2016

Published online: 22 September 2016

### References

1. Kenyon CJ. The genetics of ageing. *Nature*. 2010;464:504–12.
2. Vijg J, Campisi J. Puzzles, promises and a cure for ageing. *Nature*. 2008;454:1065.
3. Kirkwood TB. Understanding the odd science of aging. *Cell*. 2005;120:437–47.
4. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. *Cell*. 2013;153:1194–217.
5. Benayoun BA, Pollina EA, Brunet A. Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nat Rev Mol Cell Biol*. 2015;16:593–610.
6. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*. 2008;9:465–76.
7. Xu Z, Taylor JA. Genome-wide age-related DNA methylation changes in blood and other tissues relate to histone modification, expression and cancer. *Carcinogenesis*. 2014;35:356–64.
8. Rakan V, Down TA, Maslau S, Andrew T, Yang T-P, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res*. 2010;20:434–9.
9. 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, e1004996.
10. Johansson Å, Enroth S, Gyllenstein U. Continuous aging of the human DNA methylome throughout the human lifespan. *PLoS One*. 2013;8:e67378.
11. McClay JL, Aberg KA, Clark SL, Nerella S, Kumar G, Xie LY, Hudson AD, Harada A, Hultman CM, Magnusson PK. 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;23:1175–85.
12. Steegenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ, Teschendorff AE, Butcher LM, Beck S, Müller M. Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age*. 2014;36:1523–40.
13. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Di Blasio AM, Gentilini D, Vitale G, Collino S. Methylation of ELAVL2 gene as a new epigenetic marker of age. *Aging Cell*. 2012;11:1132–4.
14. Florath I, Butterbach K, Müller 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;23:1186–201.
15. Marttila S, Kananen L, Häyrynen S, Jylhävä J, Nevalainen T, Hervonen A, Jylhä M, Nykter M, Hurme M. Ageing-associated changes in the human DNA methylome: genomic locations and effects on gene expression. *BMC Genomics*. 2015;16:179.
16. Bell JT, Tsai P-C, Yang T-P, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet*. 2012;8:e1002629.
17. Reynolds LM, Taylor JR, Ding J, Lohman K, Johnson C, Siscovick D, Burke G, Post W, Shea S, Jacobs Jr DR, et al. Age-related variations in the methylome associated with gene expression in human monocytes and T cells. *Nat Commun*. 2014;5:1–8.
18. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, Diez J, Sanchez-Mut JV, Setien F, Carmona FJ. Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A*. 2012;109:10522–7.

19. van Dongen J, Nivard MG, Willemsen G, Hottenga J-J, Helmer Q, Dolan CV, Ehli EA, Davies GE, van Ijzerman M, Breeze CE. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nature Commun.* 2016;7:1–13.
20. Hernandez DG, Nalls MA, Gibbs JR, Arepalli S, van der Brug M, Chong S, Moore M, Longo DL, Cookson MR, Traynor BJ. Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum Mol Genet.* 2011;20:1164–72.
21. Fernández AF, Bayón GF, Urduingio RG, Torano EG, García MG, Carella A, Petrus-Reurer S, Ferrero C, Martínez-Cambor P, Cubillo I. H3K4me1 marks DNA regions hypomethylated during aging in human stem and differentiated cells. *Genome Res.* 2015;25:27–40.
22. Day K, Waite LL, Thalacker-Mercer A, West A, Bamman MM, Brooks JD, Myers RM, Absher D. Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. *Genome Biol.* 2013;14:R102.
23. Maegawa S, Hinkal G, Kim HS, Shen L, Zhang L, Zhang J, Zhang N, Liang S, Donehower LA, Issa J-PJ. Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res.* 2010;20:332–40.
24. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013;14:R115.
25. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, Klotzle B, Bibikova M, Fan JB, Gao Y, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell.* 2013;49:359–67.
26. Deelen J, Beekman M, Capri M, Franceschi C, Slagboom PE. Identifying the genomic determinants of aging and longevity in human population studies: progress and challenges. *Bioessays.* 2013;35:386–96.
27. Talens RP, Christensen K, Putter H, Willemsen G, Christiansen L, Kremer D, Suchiman HED, Slagboom PE, Boomsma DI, Heijmans BT. Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell.* 2012;11:694–703.
28. Phipson B, Oshlack A. DiffVar: a new method for detecting differential variability with application to methylation in cancer and aging. *Genome Biol.* 2014;15:465.
29. Shannon CE, Weaver W. *The mathematical theory of communication.* Champaign: University of Illinois Press; 1959.
30. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* 2012;13:1.
31. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* 2014;15:R31.
32. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518:317–30.
33. Slieker RC, Bos SD, Goeman JJ, Bovée J, Talens RP, van der Breggen R, Suchiman H, Lameijer E-W, Putter H, van den Akker EB. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics Chromatin.* 2013;6:26.
34. Vandiver AR, Irizarry RA, Hansen KD, Garza LA, Runarsson A, Li X, Chien AL, Wang TS, Leung SG, Kang S. Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin. *Genome Biol.* 2015;16:80.
35. Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev Cell.* 2010;19:698–711.
36. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, Campan M, Noushmehr H, Bell CG, Maxwell AP. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res.* 2010;20:440–6.
37. Peters MJ, Joehanes R, Pilling LC, Schurmann C, Conneely KN, Powell J, Reinmaa E, Sutphin GL, Zernakova A, Schramm K. The transcriptional landscape of age in human peripheral blood. *Nat Commun.* 2015;6:1–14.
38. Tserel L, Kolde R, Limbach M, Tretyakov K, Kasela S, Kisand K, Saare M, Vilo J, Metspalu A, Milani L. Age-related profiling of DNA methylation in CD8+ T cells reveals changes in immune response and transcriptional regulator genes. *Sci Rep.* 2015;5:1–11.
39. Schellenberg A, Lin Q, Schuler H, Koch CM, Jousset S, Denecke B, Walenda G, Pallua N, Suschek CV, Zenke M. Replicative senescence of mesenchymal stem cells causes DNA-methylation changes which correlate with repressive histone marks. *Aging (Albany NY).* 2011;3:873.
40. O'Hagan HM, Wang W, Sen S, Shields CD, Lee SS, Zhang YW, Clements EG, Cai Y, Van Neste L, Easwaran H. Oxidative damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members to promoter CpG Islands. *Cancer Cell.* 2011;20:606–19.
41. Ding N, Bonham EM, Hannon BE, Amick TR, Baylin SB, O'Hagan HM. Mismatch repair proteins recruit DNA methyltransferase 1 to sites of oxidative DNA damage. *J Mol Cell Biol.* 2016;8:244–25.
42. Rube C, Fricke A, Widmann TA, Furst T, Madry H, Pfreundschuh M, Rube C. Accumulation of DNA damage in hematopoietic stem and progenitor cells during human aging. *PLoS One.* 2011;6, e17487.
43. Rossi DJ, Bryder D, Seita J, Nussenzweig A, Hoeijmakers J, Weissman IL. Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. *Nature.* 2007;447:725–9.
44. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, McMichael JF, Schmidt HK, Yellapantula V, Miller CA. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med.* 2014;20:1472–8.
45. Zernakova D, Deelen P, Vermaat M, van Ijzerman M, van Galen M, Arindarto W, van t Hof P, Mei H, van Dijk F, Westra H-J, et al. Hypothesis-free identification of modulators of genetic risk factors. *BioRxiv.* 2015.
46. Bonder MJ, Luijk R, Zernakova D, Moed M, Deelen P, Vermaat M, van Ijzerman M, van Dijk F, van Galen M, Bot J, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *BioRxiv.* 2015.
47. van Greevenbroek MM, Jacobs M, van der Kallen CJ, Vermeulen VM, Jansen EH, Schalkwijk CG, Ferreira I, Feskens EJ, Stehouwer CD. The cross-sectional association between insulin resistance and circulating complement C3 is partly explained by plasma alanine aminotransferase, independent of central obesity and general inflammation (the CODAM study). *Eur J Clin Invest.* 2011;41:372–9.
48. Tigchelaar EF, Zernakova A, Dekens JA, Hermes G, Baranska A, Mujagic Z, Swertz MA, Muñoz AM, Deelen P, Cénit MC. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open.* 2015;5:e006772.
49. Westendorp RG, Van Heemst D, Rozing MP, Frölich M, Mooijaart SP, Blauw GJ, Beekman M, Heijmans BT, De Craen AJ, Slagboom PE. Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. *J Am Geriatr Soc.* 2009;57:1634–7.
50. Boomsma DI, Vink JM, Van Beijsterveldt TC, de Geus EJ, Beem AL, Mulder EJ, Derks EM, Riese H, Willemsen GA, Bartels M. Netherlands Twin Register: a focus on longitudinal research. *Twin Res.* 2002;5:401–6.
51. Hofman A, Brusselle GG, Murad SD, van Duijn CM, Franco OH, Goedegebure A, Ikram MA, Klaver CC, Nijsten TE, Peeters RP. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol.* 2015;30:661–708.
52. Huisman MH, de Jong SW, van Doormaal PT, Weinreich SS, Schelhaas HJ, van der Kooij AJ, de Visser M, Veldink JH, van den Berg LH. Population based epidemiology of amyotrophic lateral sclerosis using capture–recapture methodology. *J Neurol Neurosurg Psychiatry.* 2011;82:1165–70.
53. Westra H-J, Jansen RC, Fehrmann RS, te Meerman GJ, Van Heel D, Wijmenga C, Franke L. MixupMapper: correcting sample mix-ups in genome-wide datasets increases power to detect small genetic effects. *Bioinformatics.* 2011;27:2104–11.
54. van Ijzerman M, Tobi EW, Slieker RC, den Hollander W, Luijk R, Slagboom PE, Heijmans BT. MethylAid: visual and interactive quality control of large Illumina 450 k datasets. *Bioinformatics.* 2014;30:3435–7.
55. Fortin J-P, Labbe A, Lemire M, Zanke B, Hudson T, Fertig E, Greenwood C, Hansen K. Functional normalization of 450 k methylation array data improves replication in large cancer studies. *Genome Biol.* 2014;15:503.
56. Chen Y-A, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics.* 2013;8:203–9.
57. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007;8:118–27.
58. Andrews S. FastQC: a quality control tool for high throughput sequence data. Reference Source; 2010.
59. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17:10–2.
60. Joshi N, Fass J. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33). 2011. <https://github.com/najoshi/sickle>.



61. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
62. Hansen KD, Irizarry RA, Zhijun W. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics*. 2012;13:204–16.
63. Law CW, Chen Y, Shi W, Smyth GK. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;15:1.
64. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30:1363–69.
65. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41:D991–5.
66. Kolesnikov N, Hastings E, Keays M, Melnichuk O, Tang YA, Williams E, Dylag M, Kurbatova N, Brandizi M, Burdett T, et al. ArrayExpress update—simplifying data submissions. *Nucleic Acids Res*. 2015;43:D1113–6.
67. Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, Bell JT, Yang T-P, Meduri E, Barrett A. Mapping cis-and trans-regulatory effects across multiple tissues in twins. *Nat Genet*. 2012;44:1084–9.
68. Qu Y, Lennartsson A, Gaidzik VI, Deneberg S, Karimi M, Bengtzen S, Höglund M, Bullinger L, Döhner K, Lehmann S. Differential methylation in CN-AML preferentially targets non-CGI regions and is dictated by DNMT3A mutational status and associated with predominant hypomethylation of HOX genes. *Epigenetics*. 2014;9:1108–19.
69. Breusch TS, Pagan AR. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*. 1979;47:1287–94.
70. Luijk R, Goeman JJ, Slagboom EP, Heijmans BT, van Zwet EW. An alternative approach to multiple testing for methylation QTL mapping reduces the proportion of falsely identified CpGs. *Bioinformatics*. 2015;31:340–5.
71. ENCODE. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
72. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIKO summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011;6:e21800.
73. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498–504.
74. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OmicS*. 2012;16:284–7.
75. Wickham H. ggplot2: elegant graphics for data analysis. Berlin: Springer Science & Business Media; 2009.
76. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19:1639–45.
77. Hahne F, Durinck S, Ivanek R, Mueller A, Lianoglou S. Gviz: plotting data and annotation information along genomic coordinates. R package version 1.2.1. Bioconductor; 2012. <https://bioconductor.org/packages/release/bioc/html/Gviz.html>.

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](http://www.biomedcentral.com/submit)

