



OPEN

# Plasma cell-free DNA methylation marks for episodic memory impairment: a pilot twin study

M. Konki<sup>1,2</sup>✉, N. Lindgren<sup>3,4</sup>, M. Kyläniemi<sup>1</sup>, R. Venho<sup>1</sup>, E. Laajala<sup>1</sup>, B. Ghimire<sup>5</sup>, R. Lahesmaa<sup>1</sup>, J. Kaprio<sup>5,6</sup>, J. O. Rinne<sup>4</sup> & R. J. Lund<sup>1</sup>

Decline in episodic memory performance usually causes the first clinical symptoms of Alzheimer's disease. At present, Alzheimer's disease can only be diagnosed at a very late stage when neurodegeneration and cognitive impairment is already irreversible. New early disease markers are needed for earlier and more efficient Alzheimer's disease intervention. To identify early disease markers, we implemented a genome-wide bisulphite sequencing method for the analysis of plasma cell-free DNA methylation profiles and compared differences associated with episodic memory performance in Finnish twin pairs. A noticeable amount of cell-free DNA was present in plasma, however, the amounts as well as the genomic coverage of these fragments varied substantially between individuals. We found no significant markers associated with episodic memory performance in the twins' plasma cell-free DNA methylation profiles. Furthermore, our results indicate that due to the low genomic coverage of cell-free DNA fragments and the variety in these fragments between individuals, the implemented genome-wide bisulphite sequencing method is not optimal for comparing cell-free DNA methylation differences between large groups of individuals.

Episodic memory (EM) is the memory type that starts to decline early in Alzheimer's disease (AD) causing the usual first symptoms of dementia, like problems in remembering recent events and setting these events to the timeline. Alzheimer's disease is the most common cause of dementia, inflicting tens of millions of patients worldwide<sup>1</sup>. Currently, AD can be clinically diagnosed only at a very late stage when neurodegeneration is far-progressed and patients' cognitive impairment is already irreversible. While the initial triggers for AD are not completely known, the disease progression may start already 20–30 years before it can be diagnosed<sup>2</sup>. New early diagnostic or prognostic markers would be valuable enabling early intervention studies and evaluation of their effectiveness prior to the stage when severe neuronal damage has occurred in the disease process. Others and we have previously identified DNA methylation marks associated with AD or dementia in specific brain regions and in peripheral whole blood<sup>3–5</sup>. Plasma circulating cell-free DNA (cfDNA) methylation marks have emerged as a potential method for monitoring of cancers, rejection of organ transplantation and in prenatal diagnostics<sup>6–8</sup>. Whether these marks could be utilized to detect dementia is not clear. As cfDNA has been detected to leak into plasma from dying cells in cancers, cfDNA could be leaking to the blood circulation from the degenerating neuronal cells as well<sup>9</sup>. In this pilot study, our goal was to test and optimize the method for plasma cfDNA methylation sequencing and determine whether cfDNA methylation marks associated with EM performance can be detected in Finnish twin study participants.

## Results

**Optimization of the sample preparation method for cell-free DNA methylome analysis.** To optimize the sample preparation method for cfDNA methylome analysis, we first examined the impact of anti-coagulant type on the cfDNA yields. For this purpose, whole blood was collected into EDTA (n=3) and CPT (n=4) tubes. Plasma was fractionated and cfDNAs were isolated from each sample by using 1,000 µl volume of plasma. The cfDNA yields varied between 2.05 and 17.08 ng. No statistically significant differences were

<sup>1</sup>Turku Bioscience Centre, University of Turku and Åbo Akademi University, 20520 Turku, Finland. <sup>2</sup>Turku Doctoral Programme of Molecular Medicine, University of Turku, 20014 Turku, Finland. <sup>3</sup>Drug Research Doctoral Program, University of Turku, 20014 Turku, Finland. <sup>4</sup>Turku PET Centre, University of Turku, 20520 Turku, Finland. <sup>5</sup>Institute for Molecular Medicine Finland, University of Helsinki, 00014 Helsinki, Finland. <sup>6</sup>Department of Public Health, University of Helsinki, 00271 Helsinki, Finland. ✉email: miilko@utu.fi

Twin pair ID	Zygosity	Gender	EM status	EM z-score	APOE genotype	Amount of DNA isolated (ng)
MZ1	1	Male	Impaired	- 1.75	33	4.9
MZ1	1	Male	Impaired	- 2.67	33	7.5
MZ2	1	Male	Normal	1.23	33	10.7
MZ2	1	Male	Normal	- 0.11	33	11.6
DZ1	2	Female	Impaired	- 1.19	33	30.9
DZ1	2	Female	Normal	0.56	33	6.8
DZ2	2	Male	Normal	- 0.52	33	8.7
DZ2	2	Male	Normal	- 0.52	33	5.6
DZ3	2	Male	Normal	0.30	34	5.5
DZ3	2	Male	Impaired	- 3.08	34	19.5
DZ4	2	Female	Normal	0.59	33	11.4
DZ4	2	Female	Impaired	- 2.80	34	8.0
DZ5	2	Female	Impaired	- 2.13	44	3.9
DZ5	2	Female	Normal	1.12	34	5.4

**Table 1.** Description of the seven Finnish twin pairs included in the plasma cell-free DNA methylation analysis. Their mean age was 74.1 years (range 73–76). Zygosity (1: monozygotic, 2: dizygotic), EM episodic memory.

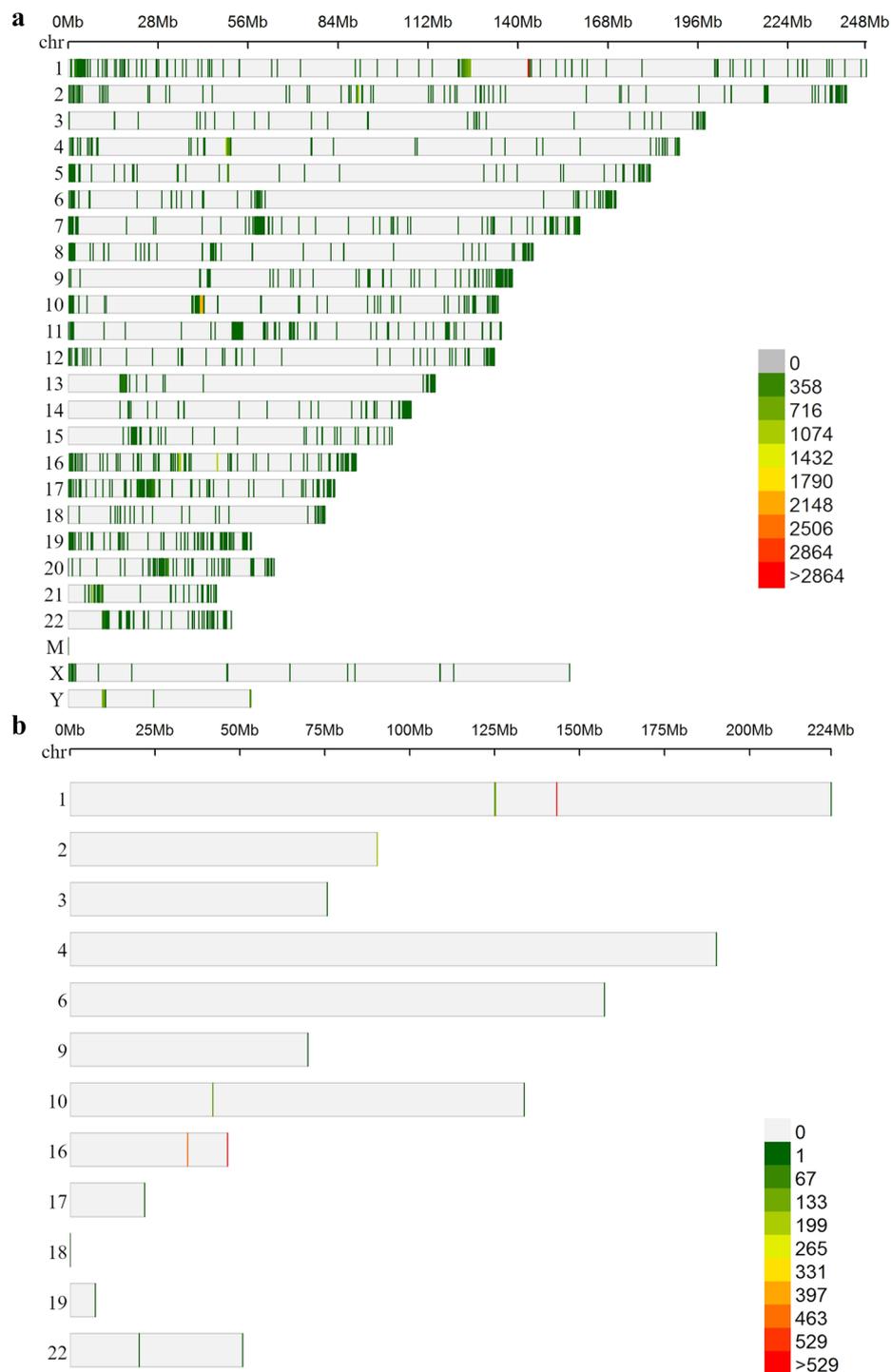
detected in the cfDNA yields between EDTA or CPT plasma (paired t-test p-value 0.20, Supplementary Fig. S1). We then examined impact of plasma volume on the cfDNA yields. For this purpose, cfDNAs were isolated from 1,000, 750 and 500  $\mu$ l of EDTA plasma with Qiagen QIAamp Circulating Nucleic Acid kit that is optimized for up to 5 ml plasma volumes. When using 500  $\mu$ l of EDTA plasma as a starting volume, two of three replicates did not yield any cfDNA indicating that this starting volume is not robust for cfDNA isolation. When using 750 and 1,000  $\mu$ l of EDTA plasma cfDNAs were successfully isolated from all samples. There were no differences (t-test p-value > 0.20) in the cfDNA yields between different starting volumes of plasma (Supplementary Fig. S1), although, there was more variation in the isolated cfDNA yields with the lower plasma volumes. Our results indicated that further increase in the plasma volume would not increase the yield.

Next, we determined whether bisulphite sequencing libraries can be prepared from these samples and whether the type of the anticoagulant and starting amount of cfDNA affect the number of detected CpG sites. Library preparation from all the samples was successful. The libraries were prepared from 2.0–10 ng of cfDNA isolated from EDTA plasma and 1.0–17 ng of cDNA isolated from CPT plasma. According to the results, there were no differences in the number of CpG sites detected in samples prepared from EDTA plasma when compared to the CPT plasma (t-test p-value 0.7). Furthermore, the correlation between the number of detected CpG sites and the amount of cfDNA used as starting material for library preparation was not statistically significant (Pearson's correlation 0.28, p-value 0.44, Supplementary Fig. S2).

### Pilot analysis of plasma cfDNA markers associated with episodic memory performance in Finnish twin pairs.

To identify plasma cell-free markers for EM performance, we then repeated the analysis by using samples collected from Finnish twin cohort. Our first null hypothesis was that the plasma cell-free DNA yields do not correlate with the EM performance. To test this hypothesis, cell-free DNAs were isolated from 1 ml of plasma collected from seven Finnish twin pairs (see materials and methods section for details). Of these twin pairs four were discordant for EM impairment, two pairs were cognitively preserved and in one pair both siblings were affected (Table 1). The cfDNA yields from the 14 individuals varied between 3.9 and 30.90 ng, however, did not correlate with the Logical Memory delayed recall ( $r = -0.16$ ,  $p = 0.58$ ) or delayed word list recall ( $r = -0.17$ ,  $p = 0.57$ ) scores. Therefore, the null hypothesis was not rejected.

Our second null hypothesis was that cell-free DNA methylation marks associated with EM performance cannot be detected in plasma. To test this hypothesis, we characterised plasma cfDNA methylomes of the seven twin pairs (14 individuals) with whole genome bisulphite sequencing and analysed whether plasma cfDNA methylation is associated with EM performance. The number of CpG sites detected in each individual with  $10 \times$  sequencing coverage varied between 4,023 and 13,414. A total number of 2,089 CpG sites were detected in all 14 individuals with  $10 \times$  sequencing coverage and were included in the differential methylation analysis (Fig. 1). Of these sites 99.52% were located in intergenic regions of the genome. Differential methylation was analysed using a binomial mixed model implemented in R package PQLseq<sup>10</sup>. The EM z-score was set as the test variable, gender, age, zygosity and APOE genotype were included as fixed effects and twin pair ID was included as random effect in the model. APOE genotype was added as a fixed effect variable since it is a significant risk factor for AD and our previous results indicate that it can correlate with DNA methylation in peripheral blood<sup>3</sup>. The model converged in 2002 of the 2089 CpG sites. To control for multiple testing, p-values were adjusted by combining the significance levels of nearby CpG sites, using RADMeth<sup>11</sup>, and corrected using FDR. After correction, we detected no significant methylation differences associated with EM performance (adjusted FDR < 0.05) (Fig. 2, Supplementary Table S1). Therefore, we did not reject the null hypothesis.

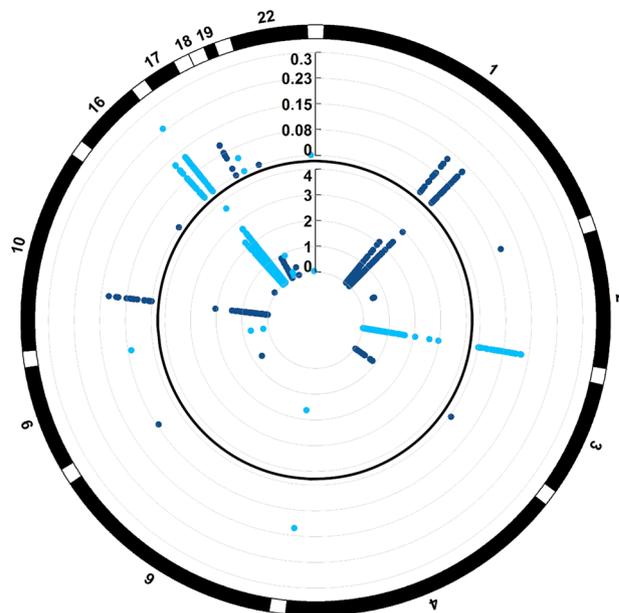


**Figure 1.** CpG sites detected with plasma cell-free DNA bisulphite sequencing. **(a)** CpG sites detected with  $\geq 10\times$  sequencing coverage in one or more twin study participants. **(b)** CpG sites detected with  $\geq 10\times$  sequencing coverage in all of the twin participants. The CpG sites in figure **(b)** were included in the differential methylation analysis. Color scaling: number of sites detected within one million base pair windows.

As a conclusion, no differentially methylated sites associated with EM performance were detected in plasma circulating cfDNA with our sample size and analysis methods.

## Discussion

In this study, we implemented a method for whole genome bisulphite sequencing of plasma cfDNA. The method was piloted by comparing plasma cfDNA methylation differences associated with EM performance in Finnish twin pairs. To our knowledge, there is no previous published studies on genome-wide characterisation of plasma



**Figure 2.** Methylation differences associated with episodic memory performance in the Finnish twins. The Manhattan plots show multiple test corrected  $\log_{10}(p)$ -values in the outer circle and unadjusted  $\log_{10}(p)$ -values in the inner circle for CpG sites that were included in the analysis. Generalised linear mixed effects model (PQLseq) was utilised to test for plasma cfDNA methylation marks associated with EM performance within the seven twin pair sample set.

cfDNA methylation markers for EM impairment even though several groups have identified DNA methylation differences associated with neurodegenerative diseases in brain tissue and peripheral blood<sup>3–5</sup>. However, plasma cell-free DNA methylation markers have been characterized for other human diseases, like cancers, and biological conditions, like prenatal diagnostics and organ transplant monitoring using similar sequencing-based methods as well as methylation arrays<sup>6–8</sup>. In addition, a recent study that utilised targeted bisulphite pyrosequencing reported neuronal tissue specific methylation pattern in LIM homeobox 2 (*LHX2*) gene in plasma cfDNA associated with AD<sup>12</sup>.

The cfDNA amounts, isolated from 1 ml twin plasma samples varied from less than four nanograms to more than thirty nanograms. Our results suggest that there is noticeable amount of cfDNA circulating in peripheral blood, however, the amount may vary a lot between individuals. Furthermore, plasma cfDNA amounts did not correlate with EM performance, in contrast to the findings Pai and his group discovered in their study<sup>12</sup>. However, Pai's group quantified circulating cfDNA in AD patients who have more severe neurodegeneration and cognitive impairment than our study participants. Thus, it is possible that plasma cfDNA levels increase as neurodegeneration progresses further. We detected between 4,023 and 13,414 CpG sites with  $10\times$  sequencing coverage in each twin sample, which indicates that also the genomic coverage of the cfDNA fragments varies substantially between individuals and therefore the method is not optimal to compare large groups of individuals. A vast majority of the detected sites located in the intergenic regions of the genome, especially in centromeres and telomeres, which suggests that these functional genomic regions are highly represented in plasma cfDNA. We did not detect any significant differences in methylation of the cfDNA fragments associated with EM impairment. Since our study material consisted of samples from individuals with relatively mild cognitive deficits, it is possible that DNA methylation markers would be detectable at a later stage when neurodegeneration is more severely progressed.

The low number of twin pairs that we were able to obtain for this study was a limitation in the pilot analysis of plasma cell-free DNA methylation differences associated with EM performance. However, we detected no promising marker candidates and our results indicate that both cfDNA amounts and genomic coverage of these fragments vary remarkably between individuals. After all, only 2,089 of the 28 million CpG sites in human genome were detected in all of our study participants with sufficient sequencing coverage. Thus, we think that no reliable methylation marks associated with mild EM impairment can be identified in plasma cfDNA with the implemented genome-wide bisulphite sequencing method that we applied even if the number of twin pairs could be increased.

## Materials and methods

**Description of the study participants.** A total number of 559 same-sex twin pairs (born 1938–1944) from the older Finnish Twin Cohort participated in a telephone-based cognitive screening. According to primary information about their cognitive performance 45 individuals; nine monozygotic twin pairs and nine dizygotic twin pairs as well as eight non-twin controls, were invited for further assessment at the Turku PET Centre, Finland, where they underwent blood sampling and neurophysiological testing in 2014–2017<sup>13–15</sup>. From the 18 twin pairs we selected the seven first ones with plasma sample available at the time of the study; two

monozygotic and five dizygotic twin pairs, who were examined at PET Centre before July 2016, for a cell-free DNA methylation characterization pilot (Table 1). The participants were administered a neuropsychological test battery that included the delayed word list recall from the Consortium to Establish a Registry for Alzheimer's disease Neuropsychological Battery (CERAD-NB) and Logical Memory delayed recall from the Wechsler Memory Scale-Revised (WMS-R). The test performances were transformed into standard deviation (SD) units based on age-appropriate Finnish norms<sup>16,17</sup>. Following the Jak/Bondi neuropsychological criteria, participants were defined to have EM impairment if they had  $\leq -1$  SD or poorer performance in both delayed verbal recall tests<sup>18</sup>. According to these criteria, four twin pairs were discordant for EM impairment (Table 1). The EM z-score values in Table 1 are the combined means from CERAD-NB and WMS-R test scores. In addition to the twin samples, we collected test plasma samples into EDTA and CPT tubes from eight non-related controls in total for testing and optimizing the sample preparation protocol. (Supplementary Table S2).

The study was approved by the Ethics Committee of Hospital District of Southwest Finland. The study was conducted according to good clinical and scientific practices and following the ethical principles of the Declaration of Helsinki. Informed consent was obtained from all study participants.

**DNA isolation.** DNA was isolated from plasma with Qiagen's QIAamp Circulating Nucleic Acids kit. For library preparation testing DNA was isolated from 500  $\mu$ l, 750  $\mu$ l or 1,000  $\mu$ l of plasma. Plasma was extracted from EDTA or CPT blood samples within two hours after sampling, snap-frozen in dry ice and stored in  $-80$  °C before DNA isolation. Twin samples were stored in CPT tubes and 1,000  $\mu$ l of plasma was used in DNA isolation per sample. Isolated DNA quality was analysed with Agilent Bioanalyzer 2100 High-Sensitivity DNA assay and DNA was quantified with Invitrogen Qubit 2.0 dsDNA HS assay.

**Library preparation.** Bisulphite sequencing library preparation test was started with 2 to 10 ng of plasma DNA per sample. Library preparation conditions for the test samples can be seen in more detail in Supplementary Table S2. Library preparation started with end-filling/A-tailing reaction with Klenow fragment (New England Biolabs) and nucleotide mix containing  $10\times$  concentration of ATP compared to other nucleotides, and unmethylated lambda DNA. The reaction mix was incubated 30 min in 30 °C and 30 min in 37 °C. In the next step Illumina TruSeq sequencing adapters were ligated to the DNA fragments with T4 DNA ligase (New England Biolabs) and overnight incubation in 16 °C. After adapter ligation, libraries were purified with AMPure XP beads in 2.5X bead concentration. Bisulphite conversion was carried out with Invitrogen MethylCode Bisulfite Conversion kit, according to the manufacturer's instructions. 25 ng of carrier DNA (sheared *E. Coli* DNA) was added to each library sample before conversion. Libraries were amplified in an 18 cycle PCR and purified two times with AMPure XP beads, 1.2X and 1.5X bead concentration. Library concentrations were measured with Qubit 2.0 HS DNA assay and quality analysed with Bioanalyzer 2,100 High Sensitivity DNA assay. The twin sample bisulphite sequencing libraries were prepared from 3 ng of plasma DNA.

**Sequencing and data analysis.** Libraries were sequenced with Illumina HiSeq 2500/3000  $1\times 50$  bp chemistry. The trimmed reads (Trim Galore 0.4.1<sup>19</sup>) were mapped to hg38 reference genome with Bismark 0.14.5<sup>20</sup>. Centromere regions, haplotype and sex chromosomes and repetitive regions that had very high sequencing coverage were filtered out from the data set. Differential methylation was tested at individual CpG sites with PQLseq<sup>10</sup> (version 1.1) software. PQLseq was utilised to identify plasma cfDNA methylation markers associated with EM performance in the whole seven twin pair sample set. PQLseq implements a generalised linear mixed effects model to test for differential methylation associated with a given trait. All seven pairs (14 twins) were included in the mixed model analysis, however, only CpG sites that were detected with at least  $10\times$  sequencing coverage in all individuals were included.  $10\times$  sequencing coverage means that a CpG site is covered by ten unique sequencing reads after mapping, i.e. the site is sequenced ten times in the library sample. EM z-score was set as the predictor variable. Gender, age, twin zygosity and APOE genotype were included as fixed effects and twin pair ID as a random effect in the model. The p-values obtained from PQLseq analysis were adjusted by combining the p-values of nearby CpG sites and FDR corrected for multiple testing using methPipe/RADMeth (version 3.4.3)<sup>11</sup>. The threshold for differential methylation was adjusted FDR  $< 0.05$ .

### Data availability

All the data supporting the findings in this study are available within the article or upon request from the corresponding author.

Received: 16 March 2020; Accepted: 5 August 2020

Published online: 25 August 2020

### References

1. Winblad, B. *et al.* Defeating Alzheimer's disease and other dementias: a priority for European science and society. *Lancet. Neurol.* **15**, 455–532 (2016).
2. Villemagne, V. L. *et al.* Amyloid  $\beta$  deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *Lancet Neurol.* **12**, 357–367 (2013).
3. Konki, M. *et al.* Peripheral blood DNA methylation differences in twin pairs discordant for Alzheimer's disease. *Clin. Epigenet.* **11**, 130 (2019).
4. De Jager, P. L. *et al.* Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* **17**, 1156–1163 (2014).
5. Smith, A. R. *et al.* A cross-brain regions study of ANK1 DNA methylation in different neurodegenerative diseases. *Neurobiol. Aging* **74**, 70–76 (2019).

6. Guo, S. *et al.* Identification of methylation haplotype blocks AIDS in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat. Genet.* **49**, 635–642 (2017).
7. Sun, K. *et al.* Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc. Natl. Acad. Sci. USA* **112**, E5503–E5512 (2015).
8. Lun, F. M. *et al.* Noninvasive prenatal methylomic analysis by genomewide bisulfite sequencing of maternal plasma DNA. *Clin. Chem.* **59**, 1583–1594 (2013).
9. Delgado, P. O. *et al.* Characterization of cell-free circulating DNA in plasma in patients with prostate cancer. *Tumor Biol.* **34**, 983–986 (2013).
10. Sun, S. *et al.* Heritability estimation and differential analysis of count data with generalized linear mixed models in genomic sequencing studies. *Bioinformatics* **35**, 487–496 (2019).
11. Song, Q. *et al.* A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS ONE* **8**, e81148 (2013).
12. Pai, M. C., Kuo, Y. M., Wang, I. F., Chiang, P. M. & Tsai, K. J. The role of methylated circulating nucleic acids as a potential biomarker in Alzheimer's disease. *Mol. Neurobiol.* **56**, 2440–2449 (2019).
13. Järvenpää, T. *et al.* Characteristics of two telephone screens for cognitive impairment. *Dement. Geriatr. Cogn. Disord.* **13**, 149–155 (2002).
14. Kaprio, J. *et al.* The older Finnish twin cohort: 45 years of follow-up. *Twin Res. Hum. Genet.* **22**, 240–254 (2019).
15. Lindgren, N., Rinne, J. O., Palviainen, T., Kaprio, J. & Vuoksima, E. Prevalence and correlates of dementia and mild cognitive impairment classified with different versions of the modified Telephone Interview for Cognitive Status (TICS-m). *Int. J. Geriatr. Psychiatry* **34**, 1883–1891 (2019).
16. Ylikoski, R. *The Relationship of Neuropsychological Functioning with Demographic Characteristics, Brain Imaging Findings, and Health in Elderly Individuals* (Raija Ylikoski, Helsinki, 2000).
17. Sotaniemi, M. *et al.* CERAD-neuropsychological battery in screening mild Alzheimer's disease. *Acta Neurol. Scand.* **125**, 16–23 (2012).
18. Jak, A. J. *et al.* Quantification of five neuropsychological approaches to defining mild cognitive impairment. *Am. J. Geriatr. Psychiatry* **17**, 368–375 (2009).
19. Krueger, F. *Babraham Bioinformatics: Trim Galore!* (2012). [https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). (Accessed 4 Sept 2019)
20. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).

## Acknowledgements

We acknowledge Finnish Functional Genomics Centre, Turku Bioscience Centre, University of Turku, Åbo Akademi, Biocenter Finland and the Finnish IT Center for Science (CSC) for the research infrastructure and support. The study was funded by State Research Funding, Hospital District of Southwest Finland, The Academy of Finland, Sigrid Jusélius Foundation, Finnish Cultural Foundation, Varsinais-Suomi Regional Fund, Yrjö Jahns-son foundation, Jenny and Antti Wihuri foundation and a private donation from Tirkkonen family.

## Author contributions

N.L., J.K. and J.R. were responsible for characterizing and recruiting the twin pairs enrolled in the study and collection of the blood samples. R.V. and M.Ko. were responsible for testing and optimizing the sample processing and bisulphite sequencing library preparation protocols under supervision of R.Lu. M.Ko. processed the twin blood samples and prepared the twin bisulphite sequencing libraries under supervision of R.Lu., B.G., M.Ko., M.Ky. and E.L. performed the bioinformatics analysis of the bisulphite sequencing data under supervision of R.Lu. and R.La. All authors contributed in preparation of the manuscript and revised and accepted the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-71239-9>.

**Correspondence** and requests for materials should be addressed to M.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020