



Variability in DNA methylation at the serotonin transporter gene promoter: epigenetic mechanism or cell-type artifact?

Sarah R. Moore¹ · Michael S. Kobor ¹

Received: 12 April 2017 / Revised: 10 April 2018 / Accepted: 25 May 2018 / Published online: 6 August 2018
© Macmillan Publishers Limited, part of Springer Nature 2018

Swartz et al. [1] report that an epigenetic signature links socioeconomic status (SES) to amygdala function and depression, built on an impressive prospective design. Here, we discuss the methods for quantifying blood DNA methylation (DNAm) employed in this study, focusing on their use of principal component analysis (PCA) to collapse 20 target CpG sites and extract one PC without prior correction of blood cell-type heterogeneity in DNAm profiles. Since the primary driver of variability in DNAm within whole blood is cell type [2–5], we argue that the index used in this analysis is likely reflective of interindividual differences in cell-type proportions rather than the DNAm mechanism inferred by the authors. We demonstrate this position analytically using publicly available DNAm data.

CpG sites located in the proximal promoter of the serotonin transporter gene (*SLC6A4*) were targeted by Swartz et al. [1] using bisulfite pyrosequencing. CpG sites are densely located at many gene promoters, referred to as “CpG islands”, which are typically unmethylated [6]. Consequently, there tends to be a low dynamic range of DNAm across individuals for promoter sites containing CpG islands. These invariable measurements are partly attributed to the way that DNAm is quantified: a single CpG site is present two times per cell, one on each of a chromosome pair, and thus DNA can be 0, 50 (rarely) or 100% methylated. An average across thousands or millions of cells present in a typical biological sample results in DNAm values ranging between 0 and 100% methylated, with CpG island sites typically demonstrating values very close to zero. This is indeed the case with the 20 CpGs used by the authors (from a prior publication), which demonstrate a

mean percent methylation across participants ranging between 0.88 and 3.57% methylated [7] (Supplementary Table 1). Considering that a 5% difference in DNAm is a commonly used threshold for a biologically meaningful effect [8] and that 5% is the error for pyrosequencing [9], 0.88–3.57% would be considered a low range.

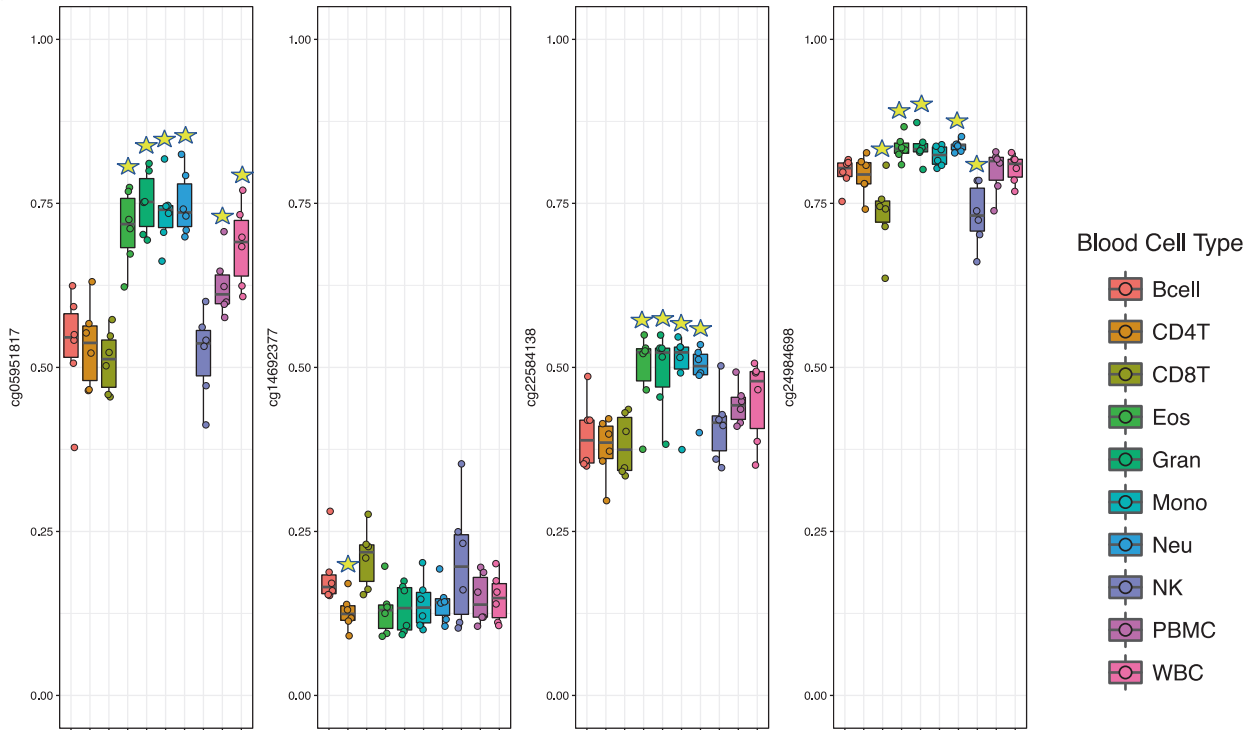
Although the authors address the limitations of making inferences about DNAm in the brain using blood, the important influence of cell-type differences in DNAm variability is overlooked. DNAm plays an essential role in the differentiation of tissues and cell types, resulting in highly cell-type-specific DNAm patterns [10, 11]. Cell-type heterogeneity within blood is generally the largest contributor to DNAm variation [3–5, 12], which includes seven cell types (and subcategories within these types) with distinct DNAm profiles. Variation attributable to cell type in blood exceeds interindividual variability explained by age, ethnicity, or exposures, and thus emerges in top PCs [3–5, 12]. For instance, one group reported that the top two DNAm PCs from blood are highly significantly correlated with cell-type proportions in five publicly available DNAm data sets [5]. This large contribution of cell type is so robust that the top PC is statistically leveraged by some methods for cell-type correction [13], which has been argued to be valid for blood samples but not necessarily for other more homogeneous or disease-related tissues [2].

Figure 1: Below, we use two publicly available DNAm data sets [12, 14] to argue what has been shown before: that the primary driver of variability in DNAm data in blood, including at the *SLC6A4* promoter, is due to cell-type proportions. First, for descriptive purposes, we show DNAm levels in isolated white blood cell types for the four most variable *SLC6A4* CpGs (one of which belongs to a CpG island, showing very low DNAm levels). DNAm at each of these sites is highly associated with blood cell type. Because these cell-type-specific patterns were measured on an older array technology, there is limited representation and no overlap with the region investigated by Swartz et al. However, we also show data drawn from a second more

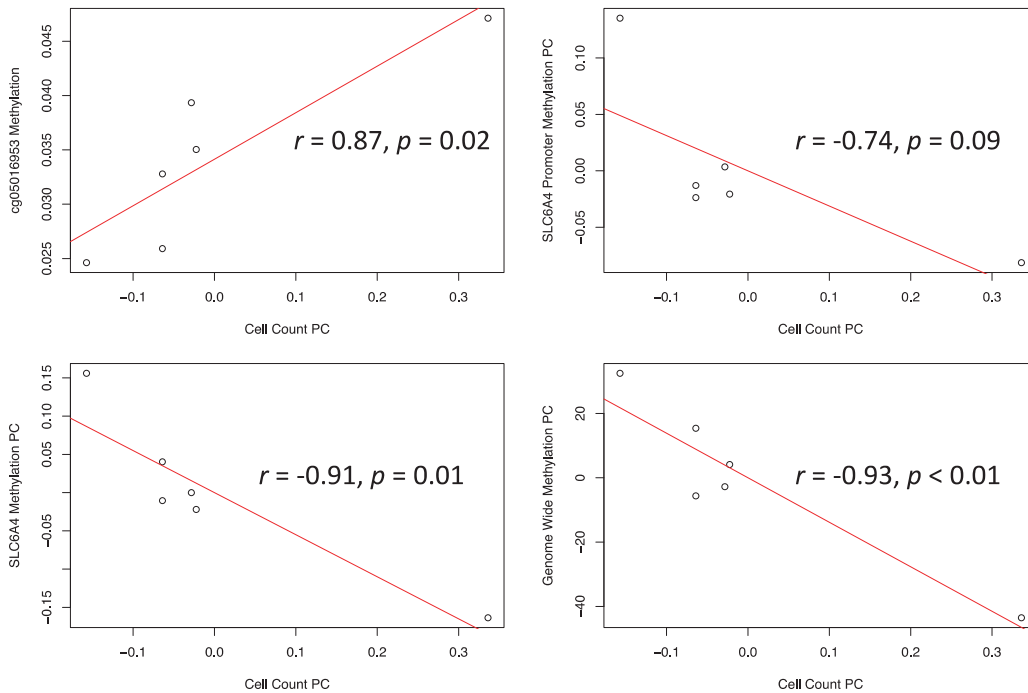
✉ Sarah R. Moore
smoore@cmmt.ubc.ca

¹ Centre for Molecular Medicine and Therapeutics, BC Children’s Hospital, Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

a. *SLC6A4* Variable CpG β Values By Blood Cell Types



b. Associations between First PC of Cell Count Predictions and DNA methylation



recent data set using the Illumina EPIC array [15]. With greater coverage, this array captures one CpG included in the PC used by Swartz et al. Although this study ran the array on whole blood, reference-based methods can be applied to array data to bioinformatically estimate cell-type

proportions [8]. We correlated the estimates of cell-type variability derived for each individual with (1) DNAm at the Swartz CpG site, (2) within 1000 bp of this region, (3) across *SLC6A4*, and (4) across all CpG sites assayed by the EPIC array, and show that in all cases, cell-type proportions

◀ **Fig. 1 a, b** Associations between cell types and DNAm levels. **a** *SLC6A4* variable CpG β values by blood cell types. **b** Associations between the first PC of cell count predictions and DNA methylation. *Note:* **a** B cell = CD19+ B cells; CD4T = CD4+ T cells; CD8T = CD8+ T cells; Eos = eosinophils; Gran = granulocytes; Mono = CD14+ monocytes; Neu = neutrophils; NK = CD56+ natural killer cells; PBMC = peripheral blood mononuclear cells; WB = whole blood. DNAm levels in isolated white blood cell types for the four most variable *SLC6A4* CpGs drawn from Reinius et al., 2012 Infinium HumanMethylation450k array data [12]. Four CpGs from the *SLC6A4* region with the highest standard deviations are shown. The second CpG is classified as located within a CpG island by the UCSC Annotation. For all four examples, β values are significantly associated with cell-type category (significant cell types with $p < 0.05$ are indicated by a “*” symbol). *Note:* **b** Data on DNA methylation from whole-blood samples and relations to cell-type proportions drawn from Guastafierro et al., 2017 Infinium HumanMethylationEPIC array data [14]. The first PC of cell counts was based on estimated counts of blood cell types produced by the “estimateCellCounts” function in the minfi package in R. Top left: β values of one CpG overlapping with the region investigated by Swartz et al. (Genome Build 37 location is 17:28562813). Top right: first PC from *SLC6A4* promoter region (17:28561783:28563929, which is 1000 bp around the area included in the PC used by Swartz et al., and includes 12 CpGs). Bottom left: the first PC from all 31 *SLC6A4* CpGs. Bottom right: the second PC from genome-wide PCA. The second rather than the first PC is plotted because the first/zeroth PC of DNA methylation quantified by array technology represents the variation in mean DNA methylation values of probes across the genome [17]. This is opposed to quantitative bisulfite pyrosequencing used by Swartz et al., or targeted gene regions, which capture a small number of highly correlated CpGs, opposed to epigenome-wide variability patterns

and DNAm are extremely, highly correlated. Although not conclusive, this analysis strongly suggests that the variable probes contributing to the first PC used in the authors’ analysis do indeed reflect cell-type proportions rather than interindividual variability in DNAm.

This observation warrants an adjustment of the interpretation that DNAm status of the *SLC6A4* promoter is predicted by SES, and in turn predicts amygdala reactivity, at least until the finding is replicated in the context of cell-type correction or cell-type confounding is explicitly tested in the author’s data. Indeed, blood cell-type proportions can themselves be related to environmental exposures [16]. We previously documented that after correcting for cell type in blood samples, a connection between DNAm and current SES was no longer present [17]. Moreover, it was recently reported that the ratio of inflammatory to antiviral white blood cell types, calculated bioinformatically using DNAm profiles, mediated the association between SES and chronic illness. Specifically, a higher ratio of monocytes and natural killer cells (i.e., markers of chronic inflammation) to T and B cells (adaptive immune system and antiviral cells) accounted for the relationship between low SES and chronic disease [18]. These reported interindividual cell-type differences in immune cells are consistent with a broad

literature linking circulating immune cells to stress levels as an adaptive response of the body to threat [19].

It is thus essential to determine whether findings from the target article reflect DNAm variability, cell-type proportion differences related to SES, or a combination of the two. Regardless, these findings provide an intriguing example of the complexity of the biological processes potentially affected by SES.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Swartz JR, Hariri AR, Williamson DE. An epigenetic mechanism links socioeconomic status to changes in depression-related brain function in high-risk adolescents. *Mol Psychiatry*. 2017;22:209–14 <http://www.nature.com/doi/10.1038/mp.2016.82>
- Zheng SC, Beck S, Jaffe AE, Koestler DC, Hansen KD, Houseman AE, et al. Correcting for cell-type heterogeneity in epigenome-wide association studies: revisiting previous analyses. *Nat Methods*. 2017;14:216–7. <http://www.nature.com/articles/nmeth.4187>
- Farré P, Jones MJ, Meaney MJ, Emberly E, Turecki G, Kobor MS. Concordant and discordant DNA methylation signatures of aging in human blood and brain. *Epigenetics Chromatin*. 2015;8:19 <http://www.epigeneticsandchromatin.com/content/8/1/19>
- Koestler DC, Christensen B, Karagas MR, Marsit CJ, Langevin SM, Kelsey KT, et al. Blood-based profiles of DNA methylation predict the underlying distribution of cell types: a validation analysis. *Epigenetics*. 2013;8:816–26. <http://www.ncbi.nlm.nih.gov/pubmed/23903776>
- Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 2014;15:R31 <http://www.ncbi.nlm.nih.gov/pubmed/24495553>
- Illingworth RS, Bird AP. CpG islands – a rough guide. *FEBS Lett*. 2009;583:1713–20. <https://www.sciencedirect.com/science/article/pii/S0014579309002919>
- Nikolova YS, Koenen KC, Galea S, Wang C-M, Seney ML, Sibille E, et al. Beyond genotype: serotonin transporter epigenetic modification predicts human brain function. *Nat Neurosci*. 2014;17:1153–5. <https://doi.org/10.1038/nn.3778>
- Jones MJ, Moore SR, Kobor MS. Principles and challenges of applying epigenetic epidemiology to psychology. *Annu Rev Psychol*. 2018;69:459–85. <http://www.annualreviews.org/doi/10.1146/annurev-psych-122414-033653>
- Mikeska T, Felsberg J, Hewitt CA, Dobrovic A. *Analysing DNA methylation using bisulphite pyrosequencing*. Vol. 791, p. 33–53 (Humana Press: Clifton, NJ), 2011. Available from: http://link.springer.com/10.1007/978-1-61779-316-5_4
- Yang X, Shao X, Gao L, Zhang S. Systematic DNA methylation analysis of multiple cell lines reveals common and specific patterns within and across tissues of origin. *Hum Mol Genet*. 2015;24:4374–84. <https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddv172>
- Mohn F, Schübeler D. Genetics and epigenetics: stability and plasticity during cellular differentiation. *Trends Genet*. 2009;25:129–36. <http://www.ncbi.nlm.nih.gov/pubmed/19185382>
- Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén S-E, Greco D, et al. Differential DNA methylation in purified human

- blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS ONE*. 2012;7:e41361. <http://dx.plos.org/10.1371/journal.pone.0041361>
13. Rahmani E, Zaitlen N, Baran Y, Eng C, Hu D, Galanter J, et al. Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. *Nat Methods*. 2016;13:443–5. <http://www.nature.com/doi/10.1038/nmeth.3809>
 14. Guastafierro T, Bacalini MG, Marcocchia A, Gentilini D, Pisoni S, Di Blasio AM, et al. Genome-wide DNA methylation analysis in blood cells from patients with Werner syndrome. *Clin Epigenetics*. 2017;9:92 <http://www.ncbi.nlm.nih.gov/pubmed/28861129>
 15. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics*. 2016;8:389–99. <http://www.ncbi.nlm.nih.gov/pubmed/26673039>
 16. Esposito EA, Jones MJ, Doom JR, MacIsaac JL, Gunnar MR, Kobor MS. Differential DNA methylation in peripheral blood mononuclear cells in adolescents exposed to significant early but not later childhood adversity. *Dev Psychopathol*. 2016;28:1385–99.
 17. Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, et al. Factors underlying variable DNA methylation in a human community cohort. *Proc Natl Acad Sci USA*. 2012;109:17253–60.
 18. Simons RL, Lei M-K, Beach SRH, Barr AB, Cutrona CE, Gibbons FX, et al. An index of the ratio of inflammatory to antiviral cell types mediates the effects of social adversity and age on chronic illness. *Soc Sci Med*. 2017;185:158–65.
 19. Steptoe A, Hamer M, Chida Y. The effects of acute psychological stress on circulating inflammatory factors in humans: a review and meta-analysis. *Brain Behav Immun*. 2007;21:901–12.