

1 **Newborn Differential DNA Methylation and Subcortical Brain Volumes**
2 **as Early Signs of Severe Delay in Neurodevelopment**

3
4 Anke Hüls¹, Catherine J Wedderburn^{2,3,4}, Nynke A Groenewold^{4,5}, Nicole Gladish^{6,7,8}, Meaghan Jones⁹,
5 Nastassja Koen^{4,5,10}, Julia L MacIsaac^{6,7,8}, David TS Lin^{6,7,8}, Katia E Ramadori^{6,7,8}, Michael P Epstein¹¹,
6 Kirsten A. Donald^{2,4}, Michael S Kobor^{6,7,8}, Heather J Zar^{2,12}, Dan J Stein^{4,5,10}

7
8 ¹Department of Epidemiology and Gangarosa Department of Environmental Health, Rollins School of
9 Public Health, Emory University, Atlanta, Georgia, USA

10 ²Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital, University
11 of Cape Town, SA

12 ³Department of Clinical Research, London School of Hygiene & Tropical Medicine, London, UK

13 ⁴Neuroscience Institute, University of Cape Town, Cape Town, South Africa

14 ⁵Department of Psychiatry & Mental Health, University of Cape Town, Cape Town, South Africa

15 ⁶Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

16 ⁷BC Children's Hospital Research Institute, Vancouver, BC, Canada

17 ⁸Centre for Molecular Medicine and Therapeutics, Vancouver, BC, Canada

18 ⁹Department of Biochemistry and Medical Genetics, University of Manitoba, and Children's Hospital
19 Research, Institute of Manitoba, Winnipeg, Canada

20 ¹⁰SU/UCT MRC Unit on Risk and Resilience in Mental Disorders, University of Cape Town, Cape Town,
21 South Africa

22 ¹¹Department of Human Genetics, School of Medicine, Emory University, Atlanta, Georgia, USA

23 ¹²SAMRC Unit on Child & Adolescent Health, University of Cape Town, Cape Town, South Africa

24

25

26 **Corresponding author:**

27 Anke Huels, PhD

28 Rollins School of Public Health, Emory University

29 1518 Clifton Road, Atlanta, GA 30322

30 E-mail: anke.huels@emory.edu

31

32 **Funding:**

33 The Drakenstein Child Health Study was funded by the Bill & Melinda Gates Foundation (OPP 1017641),
34 Discovery Foundation, Medical Research Council South Africa, National Research Foundation South
35 Africa, CIDRI Clinical Fellowship and Wellcome Trust (204755/2/16/z). AH was supported by a research
36 fellowship from the Deutsche Forschungsgemeinschaft (DFG; HU 2731/1-1) and by the HERCULES
37 Center (NIEHS P30ES019776). MPE was supported by NIH grant R01 GM117946. NAG was supported
38 by a Claude Leon Fellowship. DJS, HJZ, KAD and NK are supported by the SA Medical Research Council
39 (SAMRC). PDS is supported by the National Health and Medical Research Council, Australia. CJW is
40 supported by the Wellcome Trust through a Research Training Fellowship [203525/Z/16/Z]. Support for
41 the neuroimaging was also received by KAD from an ABMRF young investigator grant, the South African
42 Medical Association and the Harry Crossley Foundation. The funders had no role in the study design, data
43 collection and analysis, decision to publish, or preparation of manuscript.

44

45 **Acknowledgments:**

46 The authors thank the study and clinical staff at Paarl Hospital, Mbekweni and TC Newman clinics, as well
47 as the CEO of Paarl Hospital, and the Western Cape Health Department for their support of the study. The
48 authors thank the families and children who participated in this study.

49

50 **Competing financial interests declaration:**

51 All authors declare they have no actual or potential competing financial interest.

52

53

54 **Abstract**

55 *Objective:* The first two years of life are a critical period of rapid brain development. Since early
56 neurodevelopment is influenced by prenatal risk factors and genetics, neonatal biomarkers can potentially
57 provide the opportunity to detect early signs of neurodevelopmental delay. We analyzed associations
58 between DNA methylation levels from cord blood, neonatal MRI imaging data, and neurodevelopmental
59 delay at two years of age.

60 *Methods:* Neurodevelopment was assessed in 161 children from the South African Drakenstein Child
61 Health Study at two years of age using the Bayley Scales of Infant and Toddler Development. We performed
62 an epigenome-wide association study of neurodevelopmental delay using DNA methylation levels from
63 cord blood. A mediation analysis was conducted to analyze if associations between differential methylation
64 and neurodevelopmental delay were mediated by changes in neonatal brain volumes.

65 *Results:* We found epigenome-wide significant associations between severe neurodevelopmental delay and
66 differential methylation in the *SPTBN4* locus (cg26971411, p-value=3.10x10⁻⁰⁸), an intergenic region on
67 chromosome 11 (cg00490349, p-value=2.41x10⁻⁰⁸) and in the *NBPF8* locus (FDR p-value for the
68 region=9.06x10⁻⁰⁵). While these associations were not mediated by neonatal brain volume, neonatal caudate
69 volumes were independently associated with severe neurodevelopmental delay, particularly with language
70 development (p-value=0.0443) and delay in motor function (p-value=0.0082).

71 *Conclusion:* Differential methylation levels from cord blood and increased neonatal caudate volumes were
72 associated with severe neurodevelopmental delay at two years of age. If confirmed by future studies, our
73 findings could have implications for early detection of developmental delay and timely implementation of
74 interventions, which are essential for positive child development.

75

76

77

78

79 **Introduction**

80 The first two years of life are a critical period of rapid growth and brain development. Child
81 cognitive development is influenced by genetic and environmental factors which interact to determine how
82 the brain develops and functions ¹. Multiple factors have been shown to affect neurodevelopment; these
83 include poverty, maternal education, maternal physical and psychological health, substance use, and
84 nutrition ²⁻⁶. As a result, low- and middle-income countries in general and in particular sub-Saharan Africa
85 have the highest proportion of young children at risk of developmental delay ³. New research in early human
86 development shows that epigenetic, immunological, physiological, and psychological adaptations to the
87 environment occur from conception, and that these adaptations affect development throughout the life
88 course ⁶, which highlights the importance of early identification of developmental delay and timely
89 implementation of interventions ⁴. Studies have shown that children who have received early intervention
90 services experience improvements in cognitive and academic performance and engage less in risky
91 behaviors such as alcohol, tobacco, and drug use, and high-risk sexual activity ⁷⁻⁹. However, screening for
92 neurodevelopmental delay is only conducted at nine months of age at the earliest ¹⁰. Therefore, exploring
93 neonatal biomarkers that can potentially be used to detect early signs of severe neurodevelopmental delay
94 is a nascent area of research.

95 DNA methylation levels from cord blood could potentially be used as such neonatal biomarkers.
96 DNA methylation can be altered by psychological, environmental and genetic factors and is one of the most
97 studied modifications of the genome. Multiple studies show that epigenetic modifications are associated
98 with the risk of cognitive disability or impairment ¹¹. However, the role of epigenetic variation on cognitive
99 development of infants remains largely unknown. Neuropsychiatric outcomes that are known to be
100 associated with altered DNA methylation in newborns include attention-deficit/hyperactivity disorder
101 (ADHD) ¹² and amygdala:hippocampal volume ratio, which is a marker for anxiety and aggression ¹³. Some
102 evidence that altered methylation levels in neonates can also be used as a biomarker for neurodevelopmental
103 delay comes from a study of 238 Mexican-American children ¹⁴. However, this study only found suggestive
104 (non-significant) associations between two methylation sites and working memory and processing speed.

105 Over the last decades, structural magnetic resonance imaging (MRI) studies in children have
106 increasingly been used to ascertain relations between brain volume and various measures of atypical
107 behavioral development ^{15,16}. Changes in brain volume been associated neurodevelopmental disorders such
108 as autism ^{17,18}, ADHD ¹⁹, and specific language impairment ²⁰. However, most of these studies have been
109 conducted in children older than two years of age and information regarding structural and functional brain
110 development in the first two years of life is very limited ²¹. More recent studies used MRI brain scans from
111 infants at 6 or 7 months of age and neonates to predict cognitive development, suggesting involvement of

112 the maturation of subcortical brain regions ^{16,22,23}. These early changes in brain volume can potentially be
113 used as an early biomarker for neurodevelopmental delay. Furthermore, given their association with
114 prenatal environmental, psychological risk factors and genetics ⁶, we argue that early changes in brain
115 volume can potentially mediate the association between DNA methylation and neurodevelopmental delay.

116 In this study we will estimate associations between DNA methylation levels from cord blood,
117 subcortical volumes derived from neonatal MRI imaging data and neurodevelopmental delay at two years
118 of age in the South African Drakenstein Child Health Study (DCHS).

119

120 **Methods**

121 *Study design and study population*

122 The DCHS, a population-based birth cohort, has been described previously ²⁴⁻²⁶. Mothers were
123 enrolled prenatally in their second trimester and followed through pregnancy at two primary care clinics
124 serving two distinct populations (predominantly black African ancestry or predominantly mixed ancestry).
125 Mother-child pairs were followed from birth and infants enrolled in the DCHS are followed until at least
126 five years of age ²⁴. All births occurred at a single, central facility, Paarl Hospital. The sample included in
127 the present study were 161 children who had all measurements available including cognitive Bayley scores
128 at 2 years of age, methylation data from cord blood and genotyping data.

129 Ethical approval for human subjects' research was obtained from the Human Research Ethics
130 Committee of the Faculty of Health Sciences of University of Cape Town (HREC UCT REF 401/2009;
131 HREC UCT REF 525/2012). Informed consent was signed by the mothers on behalf of herself and her
132 infant for participation in this study.

133

134 *Assessment of neurodevelopmental delay*

135 At 24 months, children were assessed using the Bayley Scales of Infant and Toddler Development,
136 third edition (BSID-III) ²⁷. The BSID-III is a gold standard assessment of child development used
137 internationally, and validated in South Africa ^{28,29}. The assessment generates scores for cognitive, language
138 and motor development. Trained assessors administered the BSID-III using direct observation to score
139 children on their cognitive, language and motor development ³⁰. BSID specialized software was used to
140 produce normed and age-adjusted scores calculated using data from a US-reference group. Composite
141 scores for cognitive, motor and language scales were scaled to have a mean of 100 and standard deviation

142 of 15. The scores were categorized into severe delay if they were two standard deviations from the BSID-
143 III reference mean, and mild delay if they were one standard deviation from the mean ²⁷.

144

145 *DNA methylation*

146 DNA was isolated from cord blood samples that were collected at time of delivery ³¹. DNA
147 methylation was assessed with the Illumina Infinium HumanMethylation450 BeadChips (n=156) and the
148 MethylationEPIC BeadChips (n=160). Pre-processing and statistics were done using R 3.5.1 ³². Raw iDat
149 files were imported to RStudio where intensity values were converted into beta values. The 450K and EPIC
150 datasets were then combined using the minfi package ^{33,34} resulting in 316 samples and 453,158 probes.
151 Background subtraction, color correction and normalization were performed using the preprocessFunnorm
152 function ³⁵. After sample and probe filtering, 273 samples and 409,033 probes remained for the downstream
153 analyses (see supplementary methods for details). Batch effects were removed using ComBat from the R
154 package sva ³⁶. Cord blood cell type composition was predicted using the most recent cord blood reference
155 data set ³⁷ and the IDOL algorithm and probe selection ³⁸ based on the previous methods ³⁹.

156

157 *Genotype data*

158 Offspring samples were selected for genotyping analysis based on a number of criteria relevant to
159 the DCHS as a whole – including (but not limited to) maternal psychosocial risk/stressors and/or availability
160 of offspring lung function data. DNA was isolated from cord blood samples that were collected at time of
161 delivery ³¹. Genome-wide genotyping was performed in 270 newborns using the Illumina Infinium
162 PsychArray (n=119) and the Illumina Infinium Global Screening Array, GSA (n=151). After quality control
163 (QC) SNPs were imputed on the 1000 Genomes reference panel (Phase III) using the Michigan Imputation
164 Server ⁴⁰. Imputed genotypes reaching an $R^2 \geq 0.3$ in both arrays were used in analyses. Principal components
165 were calculated using PLINK v1.90b4 64-bit ^{41,42}.

166

167 *MRI imaging data from neonates*

168 Neonatal imaging was performed on a subgroup of newborns from the DCHS at the Cape
169 Universities Brain Imaging Centre, Tygerberg Hospital ². Newborns were excluded if they were found to
170 have medical comorbidities or had neonatal intensive care admission at birth, were premature (<36 weeks'
171 gestation), had an APGAR core of <7 at 5 minutes, or their mothers used illicit drugs in pregnancy.
172 Structural T2-weighted images were acquired during natural sleep on a Siemens Magnetom 3T Allegra

173 MRI scanner (Erlangen, Germany) using a head coil with a wet clay inlay ^{43,44}. Statistical parametric
174 mapping software SPM8 (www.fil.ion.ucl.ac.uk/spm/software/spm8) was used to process the T2-images,
175 using a well-established neonatal brain template ⁴⁵ for image normalization to standard space and
176 segmentation of images into three tissue types using probabilistic maps as priors. Volumetric data was
177 extracted for the following subcortical regions: caudate, pallidum, putamen, thalamus, amygdala and
178 hippocampus. In addition, total grey matter, total white matter, and total CSF volumes were extracted and
179 summed to obtain a measure of intracranial volume. More details regarding image acquisition and
180 processing can be found elsewhere ⁴⁶. Since an appreciable number of scans needed to be discarded largely
181 due to movement artifacts ⁴⁶, only 51 out of the 161 samples could be included in the analysis of neonatal
182 MRI imaging data.

183

184 *Statistical analysis*

185 To identify DNA methylation patterns in cord blood that are associated with severe delay in
186 neurodevelopment at two years of age, we conducted an epigenome-wide association study (EWAS) on
187 single CpG sites as well as an analysis of differentially methylated regions (DMRs). For the EWAS, we ran
188 a multivariate robust linear regression model with empirical Bayes from the R package limma (version
189 3.40.6) ⁴⁷ using severe neurodevelopmental delay at the age of two years as the independent variable and
190 each CpG methylation as a dependent variable, adjusting for sex, preterm birth, maternal smoking,
191 household income, the first five genetic PCs to control for population stratification and the first three PCs
192 from the estimated cell type proportions (after centered log-ratio transformation), which explain >90% of
193 the cell type heterogeneity ⁴⁸. We applied a Bonferroni threshold to correct for multiple testing based on
194 the number of tested CpG sites (threshold: $0.05/403933=1.24 \times 10^{-07}$). We conducted the following
195 sensitivity analyses: 1) We validated our findings by using mild delay in neurodevelopment and the
196 continuous Bayley scores as independent variables, 2) We confirmed our associations using linear
197 regression with p-values obtained from normal theory (lm() function in R) as well as from a permutation
198 test. To identify plausible pathways associated with severe delay in neurodevelopment, we performed an
199 over-representation analysis based on the CpGs with p-values < 0.001 for the association with severe delay
200 in neurodevelopment. We used the R Bioconductor package missMethyl (version 1.18.0 gometh function),
201 which performs one-sided hypergeometric tests taking into account and correcting for any bias derived
202 from the use of differing numbers of probes per gene interrogated by the array ⁴⁹. Differentially methylated
203 regions (DMRs) in severe neurodevelopmental delay were identified using DMRcate, that identifies DMRs
204 from tunable kernel smoothing process of association signals ⁵⁰. Input files were our single-CpG EWAS
205 results on severe neurodevelopmental delay including regression coefficients, standard deviations and

206 uncorrected p-values. DMRs were defined based on the following criteria: a) a DMR should contain more
207 than one probe; b) regional information can be combined from probes within 1,000 bp; c) the region showed
208 FDR corrected p-value < 0.05.

209 Finally, we analyzed if associations between differential methylation and neurodevelopmental
210 delay were mediated by neonatal brain volume as follows: (1) We analyzed associations between CpG sites,
211 that were associated with neurodevelopmental delay, and MRI imaging data from neonates (total grey
212 matter, total white matter and subcortical brain volumes) using linear regression models adjusted for the
213 same covariates as the EWAS analyses plus age at MRI scan, child sex and intracranial volume. (2) We
214 analyzed associations between MRI imaging data and severe delay in neurodevelopment using linear
215 regression models adjusted for age at MRI scan, child sex and intracranial volume. These associations were
216 validated in two sensitivity analyses: First, associations were additionally adjusted for preterm birth,
217 maternal smoking, household income and the first five genetic PCs to control for population stratification
218 and second, the findings were validated by using mild neurodevelopmental delay and the continuous Bayley
219 scores as independent variables. Lastly, if the associations in (1) and (2) were significant, a formal
220 mediation analysis was conducted⁵¹.

221

222 **Results**

223 *Description of Study Participants*

224 The study sample included 161 children with DNA methylation and genotype data, BSID-III scores
225 at two years of age as well as information on all relevant covariates (Table 1). Of these, neonatal MRI
226 imaging data were available in 51 children. Almost half of our study sample was female, half of them were
227 of African ancestry and the other half were of mixed ancestry, and a quarter of the population was exposed
228 to maternal smoking during pregnancy. Cognitive development at two years of age was severely delayed
229 in 4 to 12% of the study sample (depending on the tested domain). Most of the cases had a severe delay in
230 language development (8% in the whole study sample and 12% in the subsample with imaging data),
231 followed by severe neurodevelopmental delay (7-8%) and severe delay in motor function (4%). MRI scans
232 were conducted in neonates at three weeks of age.

233

234 *Differential Methylation as an Early Sign of Neurodevelopmental Delay*

235 Differentially methylated CpG sites in the *SPTBN4* locus (cg26971411) and in an intergenic region
236 on chromosome 11 (cg00490349) were associated with severe neurodevelopmental delay at the epigenome-

237 wide significance level (Bonferroni-adjustment) after adjusting for sex, preterm birth, maternal smoking,
238 household income, the first five genetic PCs and cell type proportions (Table 2, Figure 1, Figure S1).
239 Differential methylation in cg26971411 was significantly associated with severe delay in motor function
240 (beta = -0.024, p-value = 3.10×10^{-08}) and nominally significant for cognitive development (beta = -0.013,
241 p-value = 3.93×10^{-05}) and language development (beta = -0.014, p-value = 2.57×10^{-06}). Differential
242 methylation in cg00490349 was significantly associated with severe delay in language development (beta
243 = -0.036, p-value = 2.41×10^{-08}) and nominally significant for cognitive development (beta = -0.031, p-
244 value = 1.94×10^{-05}) and motor function (beta = -0.050, p-value = 2.73×10^{-07}). The significant associations
245 were confirmed in a sensitivity analysis using linear regression models as well as permutation tests (Table
246 S1). Associations with differential methylation in cg26971411 and cg00490349 were weaker, but still
247 nominally significant, for mild neurodevelopmental delay (p-values < 0.05, Table S2). Associations with
248 the continuous Bayley scores were only nominally significant for cognitive function, but not for language
249 development or motor function (Table S2).

250 In addition to the two significant single CpG sites, we identified one significant DMR from our
251 EWAS results on severe delay in language development that is located in the *NBPF8* locus (Figure 1, Table
252 2, minimum FDR p-value for the region = 9.06×10^{-05}).

253 No significantly enriched pathway was found among the most significant CpG sites (p-values <
254 0.001) from the EWAS of severe neurodevelopmental delay (Tables S3-S5).

255

256 *Changes in Neonatal Subcortical Brain Volumes Associated with Neurodevelopmental delay*

257 Increased caudate volumes were associated with severe neurodevelopmental delay at two years of
258 age (Table 3, Figure 2), particularly for language development (beta = 165.30, p-value = 0.0443) and delay
259 in motor function (beta = 365.36, p-value = 0.0082). The association between caudate volumes and delay
260 in motor function was also significant for mild delay in motor function (beta = 156.62, p-value = 0.0148)
261 and when using the continuous Bayley score (beta = -4.94, p-value = 0.0150) (Table S6). Furthermore, the
262 associations were robust towards additional adjustment for preterm birth, maternal smoking, household
263 income and the first five genetic PCs to control for population stratification (beta = 349.97, p-value =
264 0.0144, Table S7). Associations with language development were only significant for a severe delay (Table
265 S6) and not significant after including additional covariates (Table S8). There were no associations with
266 other subcortical brain volumes or with total grey or white matter (Table 3).

267 However, since we did not find associations between methylation in the significant CpG sites from
268 the EWAS of neurodevelopmental delay (cg26971411 and cg00490349) and changes in brain volumes
269 (Table S9), our data did not support a mediating effect of altered brain volumes on the association between
270 methylation and neurodevelopmental delay.

271

272 **Discussion**

273 In this study of infants from a poor peri-urban community in South Africa, we showed that
274 differential methylation levels from cord blood as well as larger neonatal caudate volumes were associated
275 with severe neurodevelopmental delay at two years of age. These results suggest that methylation levels
276 and brain imaging data from newborns can potentially be useful as neonatal biomarkers for
277 neurodevelopmental delay and highlight the need to understand the biological pathways how prenatal
278 environmental, psychological risk factors and genetics affect neurodevelopment. If confirmed by future
279 studies, our findings could have implications for early detection of neurodevelopmental delay and
280 consequently timely implementation of interventions ⁴, which are essential for a positive cognitive
281 functioning throughout the life course ⁷⁻⁹.

282

283 *Differential Methylation in SPTBN4 and NBPF8 as an Early Sign of Neurodevelopmental Delay*

284 Differential methylation in three loci (*SPTBN4*, *NBPF8*, and an intergenic region on chromosome
285 11) was significantly associated with severe neurodevelopmental delay at two years of age. Differential
286 methylation in the *SPTBN4* locus (cg26971411) showed the strongest association with severe delay in
287 motor function. *SPTBN4* (Spectrin Beta, Non-Erythrocytic 4) is a protein coding gene, which is, according
288 to the Human Protein Atlas, primary expressed in brain tissue and particularly in the cerebellum, which is
289 the part of the brain that functions as a co-processor of movement in concert with the cortex and basal
290 ganglia ⁵². In line, mutations and loss-of-function variants in *SPTBN4* were reported in association with
291 arthrogyrosis, which is a neuromuscular condition ⁵³, and with a severe neurological syndrome that
292 includes congenital hypotonia, intellectual disability, and motor axonal and auditory neuropathy ⁵⁴. To the
293 best of our knowledge, only one study has identified differential DNA methylation in *SPTBN4* in
294 association with neurocognitive outcomes (Alzheimer's disease) and this study was conducted in mice ⁵⁵.
295 Therefore, our study extends the current literature by highlighting that not only genetic variants or
296 methylation levels from brain tissue, but also methylation levels from cord blood are linked to cognitive

297 outcomes by providing to our knowledge the first report of an association between neonatal differential
298 methylation in *SPTBN4* and neurodevelopment.

299 The second locus for which we found an association with severe neurodevelopmental delay was
300 *NBPF8*. *NBPF8* is a member of the neuroblastoma breakpoint family (NBPF). Members of this gene family
301 are characterized by tandemly repeated copies of DUF1220 protein domains. Copy-number variations in
302 the 1q21.1 region, where most DUF1220 sequences map, have been implicated in an increasing number of
303 human diseases, including autism, schizophrenia, microcephaly/macrocephaly, and neuroblastoma ⁵⁶.
304 However, to the best of our knowledge, this is the first study showing differential methylation in this region
305 in association with cognitive outcomes.

306

307 *Increased Neonatal Caudate Volumes Associated with Severe Neurodevelopmental Delay*

308 While we did not find indications that altered neonatal brain volumes mediate the association
309 between methylation and neurodevelopmental delay, we showed that increased neonatal caudate volumes
310 were associated with neurodevelopmental delay at two years of age, particularly in motor function. This
311 finding suggests the potential use of brain imaging data as neonatal biomarker for neurodevelopmental
312 delay in addition to methylation. The caudate nucleus is one of the structures that make up the corpus
313 striatum, which is a component of the basal ganglia. The caudate nucleus plays a prominent role in motor
314 processes, and caudate nucleus dysfunction has been found in Parkinson's disease, Huntington's chorea,
315 dyskinesias, obsessive-compulsive disorder and other movement and cognitive disorders ⁵⁷. The human
316 brain undergoes a rapid change in the first year of life. The growth rates of subcortical grey-matter structures
317 are similar to cortical grey-matter growth rates, with the amygdala, thalamus, caudate, putamen, and
318 pallidum growing about 105% in the first year and roughly 15% in the second ⁵⁸. Our study showed that
319 increased neonatal caudate volumes were associated with severe neurodevelopmental delay at two years of
320 age, suggesting that risk factors for neurodevelopmental delay impact brain development in utero, resulting
321 in deviations in maturation that can already be detected soon after birth. Here, we assume that any extreme
322 deviation from the normal trajectory can result in adverse cognitive outcomes.

323

324 *Strengths & Limitations*

325 Our study has a number of limitations. First, our findings were based on a relatively small sample
326 size, particularly the analyses of neonatal brain imaging data, because an appreciable number of scans
327 needed to be discarded largely due to movement artifacts. Furthermore, although cortical brain regions are

328 expected to contribute to neurodevelopmental delay ¹⁶, cortical segmentation in the present dataset was
329 hindered by low rates of white matter myelination and the absence of T1-weighted images that have been
330 shown to enhance segmentation quality. While the infants whose scans were included and excluded did not
331 differ significantly on background variables, the loss of this large amount of data may have impacted our
332 results and limited the statistical power to detect mediation effects. Another limitation of our analyses was
333 the small number of cases of severe neurodevelopmental delay in our sample. However, to reduce the risk
334 of false positive findings due to the imbalanced study design, we validated our findings by using different
335 modelling approaches in our EWAS analyses (limma, linear regression, permutation tests, DMR analysis)
336 and different variables to measure neurodevelopmental delay in all our analyses (severe
337 neurodevelopmental delay, mild neurodevelopmental delay, continuous Bayley scores).

338 Despite these limitations, this is a very well characterized group of infants recruited as part of a
339 population-based prospective study design, and the selection of the infant age-group is an additional
340 strength. For this study, 2 to 4 weeks old infants were chosen for imaging because cerebral changes are
341 particularly intense during the last weeks of gestation and the first postnatal months. Furthermore, imaging
342 during the early postnatal period may more accurately reflect the effects of prenatal environmental,
343 psychological risk factors and genetics on brain structure.

344

345 *Conclusions*

346 We have presented evidence that differential neonatal methylation in *SPTBN4*, *NBPF8*, and an
347 intergenic region on chromosome 11 as well as increased neonatal caudate volumes were associated with
348 severe neurodevelopmental delay at two years of age. If confirmed by future studies, our findings could
349 have implications for early detection of developmental delay and timely implementation of interventions,
350 which are essential for positive child development.

351

352 **References**

- 353 1. Tucker-Drob EM, Briley DA. Continuity of genetic and environmental influences on cognition
354 across the life span: a meta-analysis of longitudinal twin and adoption studies. *Psychol Bull.*
355 2014;140(4):949-979. doi:10.1037/a0035893
- 356 2. Donald KA, Hoogenhout M, Du Plooy CP, et al. Drakenstein Child Health Study (DCHS):
357 Investigating determinants of early child development and cognition. *BMJ Paediatr Open.*
358 2018;2(1):1-9. doi:10.1136/bmjpo-2018-000282
- 359 3. Black MM, Walker SP, Fernald LCH, et al. Early childhood development coming of age: science
360 through the life course. *Lancet.* 2017;389(10064):77-90. doi:10.1016/S0140-6736(16)31389-7
- 361 4. Daelmans B, Darmstadt GL, Lombardi J, et al. Early childhood development: the foundation of
362 sustainable development. *Lancet.* 2017;389(10064):9-11. doi:10.1016/S0140-6736(16)31659-2
- 363 5. Chan M, Lake A, Hansen K. The early years: silent emergency or unique opportunity? *Lancet.*
364 2017;389(10064):11-13. doi:10.1016/S0140-6736(16)31701-9
- 365 6. Britto PR, Lye SJ, Proulx K, et al. Nurturing care: promoting early childhood development.
366 *Lancet.* 2017;389(10064):91-102. doi:10.1016/S0140-6736(16)31390-3
- 367 7. Roberts MY, Kaiser AP. Early intervention for toddlers with language delays: a randomized
368 controlled trial. *Pediatrics.* 2015;135(4):686-693. doi:10.1542/peds.2014-2134
- 369 8. McCormick MC, Brooks-Gunn J, Buka SL, et al. Early intervention in low birth weight premature
370 infants: results at 18 years of age for the Infant Health and Development Program. *Pediatrics.*
371 2006;117(3):771-780. doi:10.1542/peds.2005-1316
- 372 9. Spittle A, Orton J, Anderson PJ, Boyd R, Doyle LW. Early developmental intervention
373 programmes provided post hospital discharge to prevent motor and cognitive impairment in
374 preterm infants. *Cochrane Database Syst Rev.* 2015;24(11):CD005495.
375 doi:10.1002/14651858.CD005495.pub4
- 376 10. Vitrikas K, Savard D, Bucaj M. Developmental Delay: When and How to Screen. *Am Fam*
377 *Physician.* 2017;96(1):36-43.
- 378 11. Roubroeks JAY, Smith RG, van den Hove DLA, Lunnon K. Epigenetics and DNA methylomic
379 profiling in Alzheimer's disease and other neurodegenerative diseases. *J Neurochem.*
380 2017;143(2):158-170. doi:10.1111/jnc.14148

- 381 12. Walton E, Pingault JB, Cecil CAM, et al. Epigenetic profiling of ADHD symptoms trajectories: A
382 prospective, methylome-wide study. *Mol Psychiatry*. 2017;22(2):250-256.
383 doi:10.1038/mp.2016.85
- 384 13. Walton E, Cecil CAM, Suderman M, et al. Longitudinal epigenetic predictors of
385 amygdala:hippocampus volume ratio. *J Child Psychol Psychiatry*. 2017;58(12):1341-1350.
386 doi:10.1111/jcpp.12740
- 387 14. Huen K, Solomon O, Kogut K, Eskenazi B, Holland N. PON1 DNA methylation and
388 neurobehavior in Mexican-American children with prenatal organophosphate exposure. *Environ*
389 *Int*. 2018;121(August):31-40. doi:10.1016/j.envint.2018.08.044
- 390 15. Paterson SJ, Heim S, Friedman JT, Choudhury N, Benasich AA. Development of structure and
391 function in the infant brain: implications for cognition, language and social behaviour. *Neurosci*
392 *Biobehav Rev*. 2006;30(8):1087-1105. doi:10.1016/j.neubiorev.2006.05.001
- 393 16. Spann MN, Bansal R, Rosen TS, Peterson BS. Morphological features of the neonatal brain
394 support development of subsequent cognitive, language, and motor abilities. *Hum Brain Mapp*.
395 2014;35(9):4459-4474. doi:10.1002/hbm.22487
- 396 17. Schumann CM, Hamstra J, Goodlin-Jones BL, et al. The amygdala is enlarged in children but not
397 adolescents with autism; the hippocampus is enlarged at all ages. *J Neurosci*. 2004;24(28):6392-
398 6401. doi:10.1523/JNEUROSCI.1297-04.2004
- 399 18. Mosconi MW, Cody-Hazlett H, Poe MD, Gerig G, Gimpel-Smith R, Piven J. Longitudinal study
400 of amygdala volume and joint attention in 2- to 4-year-old children with autism. *Arch Gen*
401 *Psychiatry*. 2009;66(5):509-516. doi:10.1001/archgenpsychiatry.2009.19
- 402 19. Castellanos FX, Lee PP, Sharp W, et al. Developmental trajectories of brain volume abnormalities
403 in children and adolescents with attention-deficit/hyperactivity disorder. *JAMA*.
404 2002;288(14):1740-1748. doi:10.1001/jama.288.14.1740
- 405 20. Gauger LM, Lombardino LJ, Leonard CM. Brain morphology in children with specific language
406 impairment. *J Speech, Lang Hear Res*. 1997;40(6):1272-1284. doi:10.1044/jslhr.4006.1272
- 407 21. Knickmeyer RC, Gouttard S, Kang C, et al. A structural MRI study of human brain development
408 from birth to 2 years. *J Neurosci*. 2008;28(47):12176-12182. doi:10.1523/JNEUROSCI.3479-
409 08.2008
- 410 22. Ortiz-Mantilla S, Choe M sun, Flax J, Grant PE, Benasich AA. Associations between the size of

- 411 the amygdala in infancy and language abilities during the preschool years in normally developing
412 children. *Neuroimage*. 2010;49(3):2791-2799. doi:10.1016/j.neuroimage.2009.10.029
- 413 23. Deniz Can D, Richards T, Kuhl PK. Early gray-matter and white-matter concentration in infancy
414 predict later language skills: A whole brain voxel-based morphometry study. *Brain Lang*.
415 2013;124(1):34-44. doi:10.1016/j.bandl.2012.10.007
- 416 24. Zar HJ, Barnett W, Myer L, Stein DJ, Nicol MP. Investigating the early-life determinants of
417 illness in Africa: The Drakenstein Child Health Study. *Thorax*. 2015;70(6):592-594.
418 doi:10.1136/thoraxjnl-2014-206242
- 419 25. Gray DM, Turkovic L, Willemsse L, et al. Lung function in African infants in the Drakenstein child
420 health study impact of lower respiratory tract illness. *Am J Respir Crit Care Med*.
421 2017;195(2):212-220. doi:10.1164/rccm.201601-0188OC
- 422 26. Vanker A, Barnett W, Nduru PM, Gie RP, Sly PD, Zar HJ. Home environment and indoor air
423 pollution exposure in an African birth cohort study. *Sci Total Environ*. 2015;536:362-367.
424 doi:10.1016/j.scitotenv.2015.06.136
- 425 27. Bayley N. *Bayley Scales of Infant and Toddler Development, Technical Manual. 3rd Ed.*
426 Bloomington; 2006.
- 427 28. Ballot DE, Potterton J, Chirwa T, Hilburn N, Cooper PA. Developmental outcome of very low
428 birth weight infants in a developing country. *BMC Pediatr*. 2012;12:11. doi:10.1186/1471-2431-
429 12-11
- 430 29. Rademeyer V, Jacklin L. A study to evaluate the performance of black South African urban infants
431 on the Bayley Scales of Infant Development III. *South African J Child Heal*. 2013;7(2):54-59.
- 432 30. Donald KA, Wedderburn CJ, Barnett W, et al. Risk and protective factors for child development:
433 An observational South African birth cohort. *PLoS Med*. 2019;16(9).
- 434 31. Morin AM, Gatev E, McEwen LM, et al. Maternal blood contamination of collected cord blood
435 can be identified using DNA methylation at three CpGs. *Clin Epigenetics*. 2017;9(1):1-9.
436 doi:10.1186/s13148-017-0370-2
- 437 32. R Core Team. R: A language and environment for statistical computing. 2018. [https://www.r-](https://www.r-project.org/)
438 [project.org/](https://www.r-project.org/).
- 439 33. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: A flexible and comprehensive Bioconductor

- 440 package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*.
441 2014;30(10):1363-1369. doi:10.1093/bioinformatics/btu049
- 442 34. Fortin JP, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina
443 HumanMethylationEPIC array with minfi. *Bioinformatics*. 2017;33(4):558-560.
444 doi:10.1093/bioinformatics/btw691
- 445 35. Fortin J-P, Labbe A, Lemire M, et al. Functional normalization of 450k methylation array data
446 improves replication in large cancer studies. *Genome Biol*. 2014;15(12):503. doi:10.1186/s13059-
447 014-0503-2
- 448 36. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch
449 effects and other unwanted variation in high-throughput experiments. *Bioinformatics*.
450 2012;28(6):882-883. doi:10.1093/bioinformatics/bts034
- 451 37. Gervin K, Salas LA, Bakulski KM, et al. Systematic evaluation and validation of reference and
452 library selection methods for deconvolution of cord blood DNA methylation data. *Clin*
453 *Epigenetics*. 2019;11(1):1-15. doi:10.1186/s13148-019-0717-y
- 454 38. Koestler DC, Jones MJ, Usset J, et al. Improving cell mixture deconvolution by identifying
455 optimal DNA methylation libraries (IDOL). *BMC Bioinformatics*. 2016;17(1):1-21.
456 doi:10.1186/s12859-016-0943-7
- 457 39. Houseman EA, Molitor J, Marsit CJ. Reference-free cell mixture adjustments in analysis of DNA
458 methylation data. *Bioinformatics*. 2014;30(10):1431-1439. doi:10.1093/bioinformatics/btu029
- 459 40. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat*
460 *Genet*. 2016;48(10):1284-1287. doi:10.1038/ng.3656
- 461 41. Purcell S, Chang C. PLINK. 2017. www.cog-genomics.org/plink/1.9/.
- 462 42. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK:
463 rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. doi:10.1186/s13742-
464 015-0047-8
- 465 43. Donald KA, Fouche JP, Roos A, et al. Alcohol exposure in utero is associated with decreased gray
466 matter volume in neonates. *Metab Brain Dis*. 2016;31(1):81-91. doi:10.1007/s11011-015-9771-0
- 467 44. Donald KA, Roos A, Fouche J-P, et al. A study of the effects of prenatal alcohol exposure on
468 white matter microstructural integrity at birth. *Acta Neuropsychiatr*. 2015;27(4):197-205.

- 469 doi:10.1017/neu.2015.35
- 470 45. Shi F, Yap PT, Wu G, et al. Infant brain atlases from neonates to 1- and 2-year-olds. *PLoS One*.
471 2011;6(4). doi:10.1371/journal.pone.0018746
- 472 46. Donald KA, Fouche JP, Roos A, et al. Alcohol exposure in utero is associated with decreased gray
473 matter volume in neonates. *Metab Brain Dis*. 2016;31(1):81-91. doi:10.1007/s11011-015-9771-0
- 474 47. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-
475 sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007
- 476 48. Pawlowsky-Glahn V, Egozcue JJ. Compositional data and their analysis: An introduction. *Geol*
477 *Soc Spec Publ*. 2006;264:1-10. doi:10.1144/GSL.SP.2006.264.01.01
- 478 49. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from
479 Illumina's HumanMethylation450 platform. *Bioinformatics*. 2016;32(2):286-288.
480 doi:10.1093/bioinformatics/btv560
- 481 50. Peters TJ, Buckley MJ, Statham AL, et al. De novo identification of differentially methylated
482 regions in the human genome. *Epigenetics Chromatin*. 2015;8(1):6. doi:10.1186/1756-8935-8-6
- 483 51. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and
484 causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros.
485 *Psychol Methods*. 2013;18(2):137-150. doi:10.1037/a0031034
- 486 52. Koziol LF, Budding D, Andreasen N, et al. Consensus paper: The cerebellum's role in movement
487 and cognition. *Cerebellum*. 2014;13(1):151-177. doi:10.1007/s12311-013-0511-x
- 488 53. Pehlivan D, Bayram Y, Gunes N, et al. The Genomics of Arthrogyriposis, a Complex Trait:
489 Candidate Genes and Further Evidence for Oligogenic Inheritance. *Am J Hum Genet*.
490 2019;105(1):132-150. doi:10.1016/j.ajhg.2019.05.015
- 491 54. Wang CC, Ortiz-González XR, Yum SW, et al. β IV Spectrinopathies Cause Profound Intellectual
492 Disability, Congenital Hypotonia, and Motor Axonal Neuropathy. *Am J Hum Genet*.
493 2018;102(6):1158-1168. doi:10.1016/j.ajhg.2018.04.012
- 494 55. Sanchez-Mut J V., Aso E, Panayotis N, et al. DNA methylation map of mouse and human brain
495 identifies target genes in Alzheimer's disease. *Brain*. 2013;136(10):3018-3027.
496 doi:10.1093/brain/awt237
- 497 56. Dumas L, Sikela JM. DUF1220 domains, cognitive disease, and human brain evolution. *Cold*

- 498 *Spring Harb Symp Quant Biol.* 2009;74:375-382. doi:10.1101/sqb.2009.74.025
- 499 57. Schultz W. Reward functions of the basal ganglia. *J Neural Transm.* 2016;123(7):679-693.
500 doi:10.1007/s00702-016-1510-0
- 501 58. Gilmore JH, Knickmeyer RC, Gao W. Imaging structural and functional brain development in
502 early childhood. *Nat Rev Neurosci.* 2018;19(3):123-137. doi:10.1038/nrn.2018.1
- 503
- 504

505 **Tables**

506

Table 1. Study characteristics of the whole study population as well as the subsample with MRI imaging data.

	Whole study sample	Children with MRI scan as neonates
N	161	51
Female, n (%)	71 (44.1%)	24 (47.1%)
Ethnicity of child black, n (%)	87 (54.0%)	22 (43.1%)
Ethnicity of child colored, n (%)	74 (46.0%)	29 (56.9%)
Preterm birth (<37 weeks gestation), n (%)	19 (11.8%)	4 (7.8%)
Average household income		
< R1,000/month, n (%)	62 (38.5%)	16 (31.4%)
R1,000 - R5,000/month, n (%)	74 (46.0%)	22 (43.1%)
> R5,000/month, n (%)	25 (15.5%)	13 (25.5%)
Maternal smoking at enrolment, n (%)	43 (26.7%)	13 (25.5%)
Cognition severely delayed, n (%)	11 (6.8%)	4 (7.8%)
Language severely delayed, n (%)	13 (8.1%)	6 (11.8%)
Motor function severely delayed, n (%)	6 (3.7%)	2 (3.9%)
Age at MRI scan (in weeks), mean ± sd	n/a	3.1 ± 0.9
Intracranial volume (mm ³), mean ± sd	n/a	425906.8 ± 4235.7
Total grey matter (mm ³), mean ± sd	n/a	235978.2 ± 12843.8
Total white matter (mm ³), mean ± sd	n/a	122128.8 ± 11222.7
Caudate volume (mm ³), mean ± sd	n/a	1877.1 ± 192.2
Pallidum volume (mm ³), mean ± sd	n/a	715.6 ± 32.4
Putamen volume (mm ³), mean ± sd	n/a	2847.0 ± 72.3
Thalamus volume (mm ³), mean ± sd	n/a	3216.3 ± 107.7
Amygdala volume (mm ³), mean ± sd	n/a	538.7 ± 25.1
Hippocampus volume (mm ³), mean ± sd	n/a	1584.8 ± 144.4

507

508

509

510

Table 2. Significant single CpG sites and differentially methylated regions (DMRs) in relation to severe neurodevelopmental delay.

A. Single CpGs				Cognition		Language		Motor function	
CpG	chr	position	Gene	beta	p-value^a	beta	p-value^a	beta	p-value^a
cg00490349	11	45740105	N/A	-0.031	1.94E-05	-0.036	2.41E-08	-0.050	2.73E-07
cg26971411	19	41016501	<i>SPTBN4</i>	-0.013	3.93E-05	-0.014	2.57E-06	-0.024	3.10E-08

B. DMRs				Cognition		Language		Motor function	
CpGs	chr	position (start - end)	Genes^b	Max beta^c	p-value^d	Max beta^c	p-value^d	Max beta^c	p-value^d
9.06E-05	1	147826191 147826387	- <i>NBPF8</i>	N/A	N/A	-0.099	9.06E-05	N/A	N/A

Adjusted for sex, preterm birth, maternal smoking, household income, the first five genetic PCs to control for population stratification and the first three PCs from cell type proportions (after centered log-ratio transformation). ^a Bonferroni-threshold (0.05/403933=1.24e-07) was used to correct for multiple testing in the single CpG analyses. Significant p-values in **bold**; ^b Annotated genes in the region ^c Maximum beta for the region; ^d Minimum FDR p-value for the region. Significant p-values in **bold**.

511

512

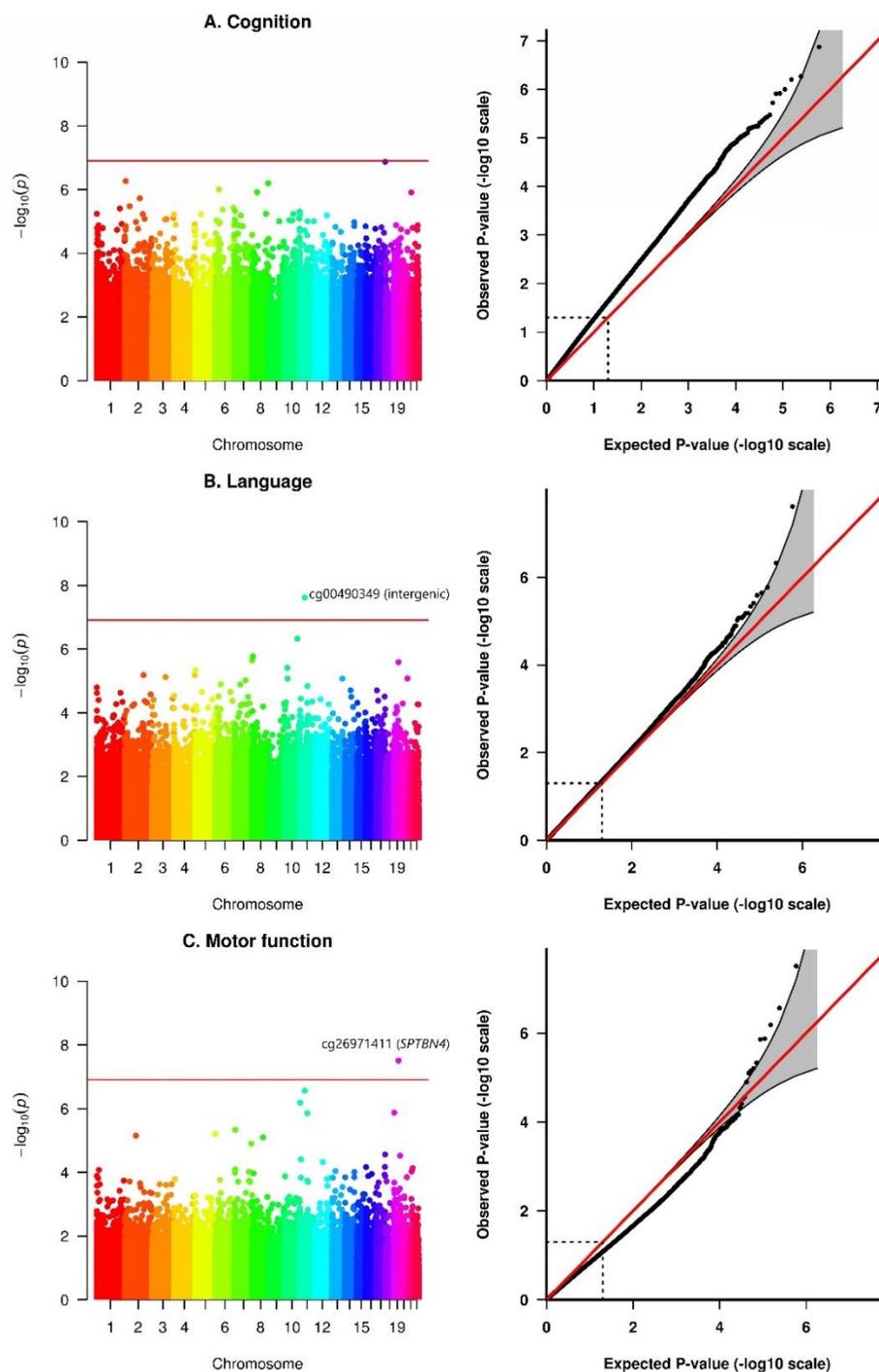
513

Table 3. Association between MRI imaging data (both hemispheres combined) from neonates and severe neurodevelopmental delay at two years of age.

MRI	Cognition		Language		Motor function	
	beta	p-value	beta	p-value	beta	p-value
Total grey matter	8790.70	0.1788	4878.53	0.3611	4605.74	0.6151
Total white matter	-5404.95	0.3679	527.50	0.9143	12192.75	0.1408
Caudate volume	117.61	0.2520	165.30	0.0443	365.36	0.0082
Pallidum volume	9.03	0.5850	-15.72	0.2394	0.56	0.9806
Putamen volume	27.69	0.4555	-18.37	0.5429	34.08	0.5089
Thalamus volume	-2.12	0.9694	-70.01	0.1149	26.07	0.7347
Amygdala volume	11.57	0.3855	3.77	0.7289	29.38	0.1092
Hippocampus volume	105.80	0.1510	64.80	0.2814	171.76	0.0919

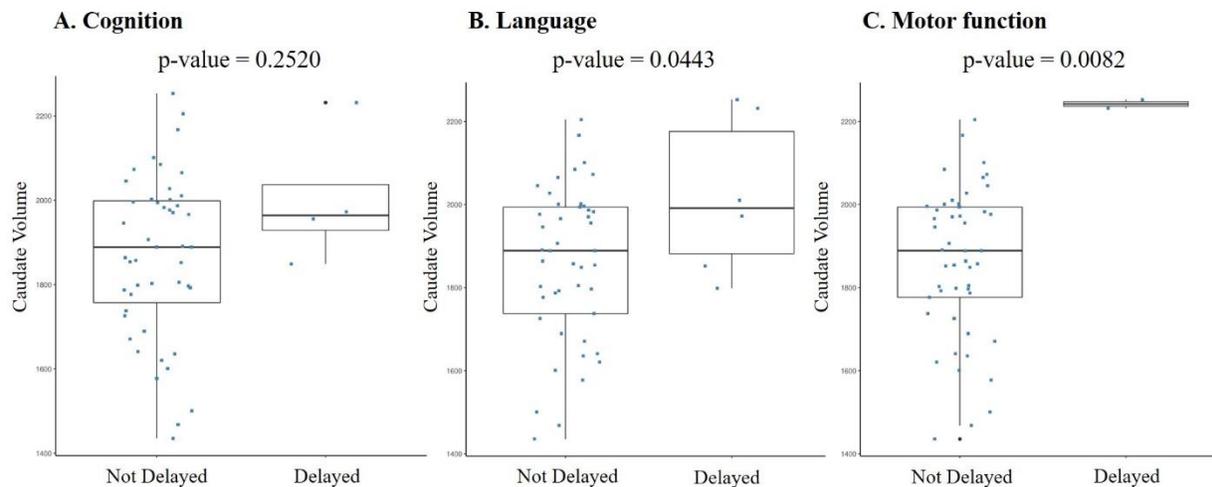
Adjusted for age at scan, child sex and intracranial volume.

515 **Figures**



516

517 **Figure 1.** Manhattan and QQ-Plots from the EWAS on severe neurodevelopmental delay in different
518 cognitive domains (A. Cognition, B. Language, C. Motor function). Adjusted for sex, preterm birth,
519 maternal smoking, household income, the first five genetic PCs to control for population stratification and
520 the first three PCs from cell type proportions (after centered log-ratio transformation). Bonferroni-
521 threshold: $0.05/403933 = 1.24 \times 10^{-7}$



522

523 **Figure 2.** Association between caudate volume in neonates and severe neurodevelopmental delay. Adjusted
524 for age at scan, child sex and intracranial volume.

525

526