

1 **Epigenetic scores indicate differences in the proteome of preterm infants**

2 Katie Mckinnon¹, Eleanor L.S. Conole², Kadi Vaher^{1,4}, Robert F. Hillary³, Danni A. Gadd³, Justyna
3 Binkowska¹, Gemma Sullivan⁴, Anna J. Stevenson³, Amy Corrigan¹, Lee Murphy⁵, Heather C.
4 Whalley⁴, Hilary Richardson⁶, Riccardo E. Marioni³, Simon R. Cox², James P. Boardman^{1,4}

5

6 1. Centre for Reproductive Health, Institute for Regeneration and Repair, University of Edinburgh,
7 Edinburgh, UK.

8 2. Lothian Birth Cohorts, Department of Psychology, University of Edinburgh, Edinburgh, UK.

9 3. Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of
10 Edinburgh, Edinburgh, UK.

11 4. Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK.

12 5. Edinburgh Clinical Research Facility, University of Edinburgh, Edinburgh, UK.

13 6. School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Edinburgh,
14 UK.

15

16 Corresponding author: Professor James P Boardman

17 Institute for Regenerative Medicine, 4-5 Little France Drive, Edinburgh, EH16 4UU

18 James.Boardman@ed.ac.uk

19

20

21

22

23

24

25

26

27

28

29

30

1 **Abstract**

2 *Background:* Epigenetic scores (EpiScores), reflecting DNA methylation (DNAm)-based surrogates for
3 complex traits, have been developed for multiple circulating proteins. EpiScores for pro-inflammatory
4 proteins, such as C-reactive protein (DNAm CRP), are associated with brain health and cognition in
5 adults and with inflammatory comorbidities of preterm birth in neonates. Social disadvantage can
6 become embedded in child development through inflammation, and deprivation is over-represented in
7 preterm infants. We tested the hypotheses that preterm birth and socioeconomic status (SES) are
8 associated with alterations in a set of EpiScores enriched for inflammation-associated proteins.

9
10 *Results:* 104 protein EpiScores were derived from saliva samples of 332 neonates born at gestational
11 age (GA) 22.14 to 42.14 weeks. Saliva sampling was between 36.57 and 47.14 weeks. Forty-three
12 (41%) EpiScores were associated with low GA at birth (standardised estimates |0.14 to 0.88|,
13 Bonferroni-adjusted p -value $<8.3 \times 10^{-3}$). These included EpiScores for chemokines, growth factors,
14 proteins involved in neurogenesis and vascular development, cell membrane proteins and receptors,
15 and other immune proteins. Three EpiScores were associated with SES, or the interaction between
16 birth GA and SES: afamin, intercellular adhesion molecule 5 and hepatocyte growth factor-like protein
17 (standardised estimates |0.06 to 0.13|, Bonferroni-adjusted p -value $<8.3 \times 10^{-3}$). In a preterm sub-group
18 ($n=217$, median [range] GA 29.29 weeks [22.14 to 33.0 weeks]), SES-EpiScore associations did not
19 remain statistically significant after adjustment for sepsis, bronchopulmonary dysplasia, necrotising
20 enterocolitis, and histological chorioamnionitis.

21
22 *Conclusions:* Low birth GA is substantially associated with a set of EpiScores. The set was enriched
23 for inflammatory proteins, providing new insights into immune dysregulation in preterm infants. SES
24 had fewer associations with EpiScores; these tended to have small effect sizes and were not
25 statistically significant after adjusting for inflammatory comorbidities. This suggests that inflammation
26 is unlikely to be the primary axis through which SES becomes embedded in the development of
27 preterm infants in the neonatal period.

28
29 **Keywords:** Epigenetic, Inflammation, Neonatal, Preterm birth, Socioeconomic status.

30

1 **1.0 Background**

2 Preterm birth (delivery at less than 37 weeks of gestation) affects around 10% of births worldwide and
3 is closely associated with increased likelihood of cerebral palsy, neurocognitive impairment,
4 behavioural, social and communication difficulties, and mental and cardiometabolic health diagnoses
5 across the life course [1–5]. These adverse outcomes can be explained, in part, by deleterious effects
6 of early exposure to extrauterine life on brain and cardiac development, and they are often
7 accompanied by changes in blood proteins, including those reflecting the perinatal innate and
8 adaptive immune response [6–8].

9
10 Socioeconomic status (SES) is also associated with the adverse neurodevelopmental and health
11 outcomes listed above [9–12], and social deprivation is consistently over-represented among preterm
12 children and their families [13,14]. In a meta-analysis of 43 studies (n=111,156 individuals), low SES
13 associated with increased inflammatory markers of disease risk (C-reactive protein [CRP] and
14 interleukin-6 [IL6]), which suggests that pro-inflammatory pathways may be important mechanisms for
15 translating social inequalities into health disparities [15]. However, only four studies included
16 participants under 10 years of age, leaving uncertainty about SES-inflammation correlations in early
17 life [16–19].

18
19 Although protein levels are commonly used as biomarkers of exposure and disease risk, they are
20 limited because they are often phasic in the systemic circulation, rely on venepuncture, and may not
21 capture baseline status or chronicity. For example, inflammation is often measured using acute-phase
22 inflammatory proteins such as CRP [20,21], but it is not always reliable [22], particularly in neonates,
23 and a single-time-point measure may not reflect baseline inflammation or capture chronic
24 inflammation [23]. These challenges have been addressed by the development of DNA methylation
25 (DNAm) markers of protein expression (EpiScores), which are derived from a linear weighted sum of
26 DNAm sites that are correlated with protein levels. Several EpiScores are associated with magnetic
27 resonance imaging (MRI) measures of brain health, cognition, child mental health, stroke, ischaemic
28 heart disease Alzheimer's disease, lung cancer [24–31]. In neonates, DNAm CRP is associated with
29 birth gestational age (GA), perinatal inflammatory processes, and MRI features of encephalopathy of
30 prematurity [32]. Childhood SES is associated with differential DNAm in inflammation-related genes

1 [33,34] and at CpG sites that correlate with an inflammation index [35]. Adult SES and social mobility
2 are associated with variations in DNAm in inflammation-related genes [33,36]. Importantly, SES-
3 related DNAm variations are associated with differences in gene expression so may have functional
4 consequences [33,36].

5
6 We investigated relationships between preterm birth, SES, and 104 EpiScores enriched for
7 inflammation-related proteins [26–28,31,37]. We tested the following hypotheses: first, low GA is
8 associated with differences in EpiScores; second, SES is correlated with EpiScores, and interacts
9 with birth GA, but the relationship is attenuated by inflammatory disease burden in preterm infants.

10 11 **2.0 Methods**

12 *2.1 Participants*

13 Participants were preterm infants (born ≤ 33 weeks' gestation) and term-born infants born at the Royal
14 Infirmary of Edinburgh, UK. These infants were recruited to a longitudinal cohort study designed to
15 investigate the effect of preterm birth on brain development and outcomes with multimodal data
16 collection [38]. Infants were recruited between February 2012 and December 2021.

17
18 Exclusion criteria were congenital malformation, chromosomal abnormality, congenital infection, cystic
19 periventricular leukomalacia, haemorrhagic parenchymal infarction, and post-haemorrhagic
20 ventricular dilatation. These criteria mean the cohort is representative of the majority of survivors of
21 modern intensive care practices [38].

22
23 Final participants included were 217 preterm infants (born ≤ 33 weeks' GA) and 115 term-born infants,
24 with median birth GA of 29.29 weeks and 39.71 weeks respectively. Their demographic
25 characteristics are shown in Table 1. The three SES measures (Scottish Index of Multiple Deprivation
26 (SIMD 2016) [39], maternal education and maternal occupation) differed between the preterm and
27 term groups (Cohen's d effect sizes 0.52-0.68). Ethnicity did not differ between groups and is
28 representative of the Edinburgh area [40].

29
30 *[Table 1 near here]*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

2.2 DNA methylation

Saliva samples for DNAm were collected at term-equivalent age using Oragene OG-575 Assisted Collection kits (DNA Genotek, ON, Canada) and DNA was extracted using prepIT.L2P reagent (DNA Genotek, ON, Canada). Saliva sampling was used due to accessibility and the non-invasiveness of the method; DNAm patterns measured via saliva samples correlate with brain and other tissue DNAm patterns [41,42]. We chose to sample at the term-corrected gestation timepoint to include the allostatic load of both prenatal and early postnatal exposures.

DNA was bisulphite converted and methylation levels were measured using Illumina HumanMethylationEPIC BeadChip (Illumina, San Diego, CA, USA) at the Edinburgh Clinical Research Facility (Edinburgh, UK). The arrays were imaged on the Illumina iScan or HiScan platform, and genotypes were called automatically using GenomeStudio Analysis software version 2011.1 (Illumina). DNAm was processed in four batches.

Raw intensity (.idat) files were read into the R environment using minfi. wateRmelon and minfi were used for preprocessing, quality control and normalisation [43]. The pfilter function in wateRmelon was used to exclude samples with 1% of sites with a detection p -value >0.05 , sites with beadcount <3 in 5% of samples, and sites with 1% of samples with detection p -value >0.05 . Cross hybridising probes, probes targeting single nucleotide polymorphisms with overall minor allele frequency ≥ 0.05 , and control probes were also removed. Samples were removed if there was a mismatch between predicted sex (minfi) and recorded sex ($n=3$), or if samples did not meet preprocessing quality control criteria ($n=29$). Data were danet normalised, which includes background correction and dye bias correction [43]. Saliva contains different cells types, including buccal epithelial cells. Epithelial cell proportions were estimated with epigenetic dissection of intra-sample heterogeneity with the reduced partial correlation method implemented in the R package EpiDISH [44]. Probes located on sex chromosomes were removed before analysis. The cohort includes twins ($n=32$); these were randomly removed leaving one participant per twin pair. This left a final sample size of $n=332$.

2.3 EpiScore calculation

1 The 104 protein EpiScores included 100 EpiScores from Gadd *et al* [26], a study enriched for
2 inflammatory-related proteins, excluding those where the required CpGs were not available, owing to
3 differences in assay platform CpG coverage relative to Gadd *et al* [26]. For duplicate proteins, those
4 developed using Olink platform-identified proteins (antibody-based assays) were prioritised over those
5 from SOMAscan platforms (aptamer-based assays), due to specificity and reproducibility, and the
6 variable correlation between the two methods [45–48] (for details see Supplementary eMethods,
7 Additional File 1). In addition, we included EpiScores for growth and differentiation factor 15 (GDF15)
8 and N-terminal-pro B-type natriuretic peptide (NTproBNP) [37], and CRP. The CRP EpiScore used
9 was Barker *et al*'s seven-CpG variation of Ligthart *et al*'s CRP EpiScore [27,31], as this is known to
10 correlate with birth GA, perinatal pro-inflammatory exposures, and neonatal brain development [32].

11

12 For each individual, EpiScores were obtained by multiplying the methylation proportion at a given
13 CpG by the effect size from previous studies. This was performed using the MethylDetectR platform
14 [49] for those inflammatory proteins currently included and using R for those not currently included
15 (CRP, GDF15, IL6, NTproBNP). All CpG sites and coefficients required to calculate the 104
16 EpiScores are in Supplementary Table 1 (Additional File 2).

17

18 2.4 Statistics

19 The predictor variables were SES and birth GA. SES was operationalised in three ways:
20 neighbourhood-level SES using the Scottish Index of Multiple Deprivation (SIMD) [39], and two
21 measures of family-level SES, which were maternal education (highest educational qualification) and
22 maternal occupation (current or most recent occupation). For further details, see Supplementary
23 eMethods, Additional File 1. Birth GA was a continuous variable to maximise statistical power [50,51].
24 We adjusted for GA at saliva sampling, DNAm batch, infant sex, and birthweight z-score.

25

26 All statistical analyses were performed in R (version 4.3.1) and were preregistered [52].

27

28 Principal component analysis (PCA) was used to determine the significance threshold for controlling
29 type 1 error in analyses of multiple EpiScores [53]. We began with a correlation analysis, which
30 showed correlation coefficients between EpiScores of |0.01 to 0.93| (Figure 1A). To determine the

1 number of statistical “families” among the 104 EpiScores, PCA was performed. This yielded two
2 principal components with eigenvalues >1, our pre-specified threshold, which explained 59.5% and
3 17.2% of variance, respectively (Figure 1B). Standardised component loadings are provided in
4 Supplementary Table 3 (Additional File 1). In all subsequent analyses, we corrected for multiple
5 comparisons across EpiScores and SES measures using a Bonferroni-adjusted p -value threshold of
6 8.3×10^{-3} . This is $0.05/(2 \times 3)$, with two reflecting the two principal components for EpiScores and three
7 reflecting the number of SES measures used.

8

9 *[Figure 1 near here]*

10

11 We constructed generalised linear regression models for each EpiScore as outcome measure to
12 assess associations between GA, each of the three SES measures (separate models for each of
13 SIMD, maternal education, and maternal occupation), and the product interaction term SES*birth GA
14 (removing the term if not significant), and adjusting for GA at sampling, sex, and batch.

15

16 For the preterm sub-group, we then additionally adjusted for perinatal inflammatory exposures known
17 to be associated with the CRP EpiScore as, to our knowledge, this is the only DNAm proxy of an
18 inflammatory protein that has been studied in this context [32]. These were histological
19 chorioamnionitis (HCA), sepsis, bronchopulmonary dysplasia (BPD) and necrotising enterocolitis
20 (NEC). For definitions see Supplementary eMethods (Additional File 1) and for frequencies see
21 Supplementary Table 2 (Additional File 1).

22

23 **3.0 Results**

24 *3.1 Associations between gestational age and EpiScores*

25 Gestational age associated with 43 of the 104 EpiScores after adjustment for SIMD, maternal
26 education, or maternal occupation (Figure 2A-C). The proteins represented by the 43 EpiScores are
27 listed in Table 2 categorised by functional annotation adapted from the STRING database[54], and
28 their broader roles in immune processes and inflammation, and the pathogenesis of neonatal
29 diseases, where known, are described in Supplementary Table 4 (Additional File 1).

1 29 (67%) EpiScores negatively associated with birth GA (standardised estimates |0.14-0.76|, adjusted
2 p -value $<8.3 \times 10^{-3}$), and 14 (33%) EpiScores positively associated with birth GA (standardised
3 estimates 0.14-0.88, adjusted p -value $<8.3 \times 10^{-3}$).
4 33 EpiScores associated with low GA irrespective of SES measure used in the model. The results for
5 all 104 EpiScores are provided in Supplementary Figures 1-9 (Additional File 1).

6

7 [Table 2 near here]

8 [Figure 2 near here]

9

10 *3.2 Associations between SES, EpiScores, and the effect of inflammatory comorbidities of preterm*
11 *birth.*

12 Three out of 104 EpiScores associated with SES measures or the interaction between SES and GA
13 (Figure 3). There was a small effect size association between higher afamin EpiScore and higher
14 maternal occupation (standardised $\beta=0.06$, 95% confidence interval (CI) 0.02-0.11, $p=0.0082$), and
15 DNAm afamin associated with the birth GA*maternal education interaction term such that afamin
16 positively correlated with birth GA among babies with mothers without university education, and
17 negatively correlated with birth GA among babies with mothers with university education
18 (undergraduate or postgraduate) (standardised $\beta=-0.12$, 95% CI -0.20 to -0.04, $p=0.0041$,
19 Supplementary Figure 10, Additional File 1).

20

21 Higher intercellular adhesion molecule 5 (ICAM5) EpiScore associated with higher SIMD
22 (standardised $\beta=0.13$, 95% CI 0.03-0.23, $p=0.0079$). Hepatocyte growth factor-like protein (HGFI)
23 associated with the birth GA*maternal occupation interaction term (standardised $\beta=-0.10$, 95% CI -
24 0.16- -0.04, $p=0.0021$, Supplementary Figure 11, Additional File 1), such that HGFI EpiScore
25 positively correlated with birth GA among babies with mothers who were unemployed, homemakers,
26 in full time education, or in unskilled or manual occupations, but negatively correlated with birth GA
27 among babies with mothers in partly skilled, non-manual skilled or professional occupations.

28

29 In the planned analysis of preterm-born babies only, when controlling for inflammatory exposures
30 (sepsis, HCA, NEC, and BPD), a greater proportion of R^2 was explained ($R^2=0.053$ -0.28 in unadjusted

1 models, $R^2=0.029-0.437$ in adjusted models), but the EpiScores no longer met our statistical threshold
2 (p -values >0.045 with adjusted p -value threshold $<8.3 \times 10^{-3}$, see Supplementary Table 5, Additional
3 File 1).

4

5 *[Figure 3 near here]*

6

7 **4.0 Discussion**

8 In this study, we identified several associations between a set of EpiScores enriched for inflammatory
9 proteins and low GA at birth. Few EpiScores associated with SES within the whole sample and these
10 associations were partially attenuated in preterm infants who experienced inflammatory comorbidities.

11 This is the first study to assess the impact of preterm birth and social status using epigenetic
12 signatures designed to reflect the circulating proteome.

13

14 *4.1 Associations between birth GA and EpiScores*

15 43 EpiScores associated with preterm birth when sampled at term equivalent age. The EpiScores
16 reflect chemokines, growth factors, proteins required for neurogenesis and vascular development, cell
17 membrane proteins and receptors, and immune response proteins (Table 2). As well as having
18 specific immunoregulatory roles, in the neonatal setting or relevant animal models, these proteins are
19 associated with several comorbidities and developmental consequences of preterm birth. These
20 include lung development and disease such as BPD [55–67], in utero and postnatal growth failure
21 [68–71], HCA [7,72,73], patent ductus arteriosus [74–77], retinopathy of prematurity [78–82], NEC
22 [83–85], hyperglycaemia [86], sepsis [85,87–90], brain injury [32,91–94], and neurodevelopmental
23 outcomes [95–99].

24

25 *4.2 Associations between SES measures and EpiScores*

26 SES appears to play a much smaller role in the patterning of EpiScores compared to birth GA. SES
27 measures, or interactions between birth GA and SES measures, were associated with only three of
28 the 104 EpiScores studied: afamin, ICAM5 and HGFI. Afamin and ICAM5 positively associated with
29 maternal occupation and SIMD, respectively. Afamin and HGFI both associated with the interaction
30 term between SES and birth GA. Afamin is involved in vitamin E transport [100], ICAM5 has a role in

1 microglial regulation [101], and HGFI is a macrophage-stimulating protein [102]. The relationships
2 between these proteins and SES have not previously been investigated, although afamin is
3 associated with the development of metabolic syndrome [103], which varies with SES [104].
4
5 Among preterm infants, no SES-EpiScore associations survived adjustment for inflammatory
6 exposures, which suggests that the weak effects of SES on the neonatal proteome that we observed
7 in a small number of EpiScores are at least partially accounted for by inflammatory pathologies in
8 early life. Taken together, the results suggest immune dysregulation, proxied by EpiScores, may not
9 be the primary axis through which SES becomes embedded in the development of preterm infants
10 during neonatal intensive care.

11
12 SES has been consistently associated with inflammation in adulthood, including in relation to
13 childhood deprivation [15,105], but less is known about the relationship between SES and
14 inflammation in the neonatal period. A longitudinal study by Leviton *et al* [106], with five sampling
15 timepoints during the first month of life after preterm birth, showed that maternal eligibility for Medicaid
16 associated with levels of 14 inflammatory proteins (IL6R, TNFR1, TNFR2, IL8, ICAM1, VCAM1, TSH,
17 EPO, bFGF, IGF1, VEGF, PIGF, Ang-1, Ang-2). However, only three were significant at more than
18 one timepoint during the month after preterm birth (TSH, bFGF, Ang-1), and none associated at all
19 five measurement timepoints. These studies, taken together with our results, suggest that the impact
20 of SES on immune regulation is relatively modest and inconsistent in the newborn period but accrues
21 through to adulthood. Further research is required to understand how and when SES becomes
22 embedded in child development and whether early life events such as preterm birth modify that
23 process; EpiScores could be a powerful tool for investigating the temporal dynamics of social
24 determinants of child health.

25

26 *4.3 Strengths and limitations*

27 Strengths of this study include the large sample of term and preterm neonates; to the best of our
28 knowledge, this is the first examination of multiple DNA methylation-based estimators of circulating
29 proteins in a neonatal sample. We derived EpiScores from minimally-invasive sampling (buccal cells
30 from saliva) which overcomes the ethical challenge of venepuncture for research in children. The

1 EpiScores, serving as proxies of inflammatory proteins and sampled at term-equivalent age in
2 preterm infants, were selected for their potential to capture chronic, cumulative inflammation
3 associated with preterm birth and neonatal intensive care exposures [23–25]. We adjusted for
4 variables associated with DNAm, and additionally for inflammatory exposures to increase the clinical
5 validity of our results.

6

7 The study has some limitations. The EpiScores used were developed in adult cohorts [26–28,37] and
8 have not been validated in neonatal populations with neonatal protein levels, although we have
9 previously established that neonatal DNAmCRP scores correlate with cumulative inflammatory
10 exposures [32]. Further studies that evaluate serial blood protein levels with EpiScores could be
11 informative but would need to rely on small volume samples taken during time of venepuncture for
12 clinical reasons, given the practical and ethical challenges of serial venepuncture for research
13 purposes in neonates. The EpiScores were also trained using blood samples [26–28,37], whereas we
14 have projected these scores into saliva samples. However, previous studies have successfully used
15 similar cross-tissue techniques [32,107,108], and in neonates saliva provides a non-invasive and
16 accessible sample method. Not all inflammatory-related proteins are represented, as we were limited
17 by available EpiScores, so we may have underestimated the full complexity of the relationship
18 between birth GA, SES, and inflammation. Longitudinal investigations are imperative for elucidating
19 whether the DNA methylation signatures associated with gestational age identified in this study exert
20 a causal influence on the inflammation-associated mechanisms in preterm birth. It remains crucial to
21 discern whether these signatures represent a direct downstream consequence of GA itself or are
22 induced by specific factors correlated with GA, yet not necessarily driven by chronic inflammation.
23 Mendelian randomization studies, integrating genomic and epigenomic determinants, are a promising
24 methodological approach to disentangle the directionality of these intricate relationships. The study
25 population is comparable to other neonatal populations in high-income, majority-white settings, but
26 these results may not generalise to settings with different socioeconomic or ethnicity profiles. We
27 studied several measures of SES but were not able to include all that could be relevant, such as
28 household income.

29

30 **5.0 Conclusion**

1 We identified 43 EpiScores enriched for inflammatory proteins that associated with low birth GA.
2 These 43 proteins offer novel insights into the physiological response to preterm birth and warrant
3 further study to explore their role in the relationship between preterm birth, inflammation, and longer-
4 term outcomes. We found only three EpiScores associated with SES in the neonatal period, none of
5 which survived adjustment for perinatal pro-inflammatory exposures, suggesting that inflammation is
6 unlikely to be the primary axis through which SES becomes embedded in the development of preterm
7 infants in the neonatal period.

8

9 **6.0 Declarations**

10 *6.1 Ethics approval and consent to participate*

11 Ethical approval for the study was obtained from the National Research Ethics Service, South-East
12 Scotland Research Ethics Committee (REC 11/55/0061, 13/SS/0143, 16/SS/0154), and NHS Lothian
13 Research and Development (2016/0255). Written informed consent was obtained from parents.

14

15 *6.2 Consent for publication*

16 Not applicable.

17

18 *6.3 Availability of data and materials*

19 DNA methylation data are available to researchers subject to the terms of the Data Access Policy:
20 [https://www.ed.ac.uk/centre-reproductive-health/tebc/about-tebc/for-researchers/data-access-](https://www.ed.ac.uk/centre-reproductive-health/tebc/about-tebc/for-researchers/data-access-collaboration)
21 [collaboration](https://www.ed.ac.uk/centre-reproductive-health/tebc/about-tebc/for-researchers/data-access-collaboration).

22

23 *6.4 Competing interests*

24 LM has received speaker and consultancy fees from Illumina. REM is a scientific advisor to the
25 Epigenetic Clock Development Foundation and to Optima Partners. All other authors declare that they
26 have no competing interests.

27

28 *6.5 Funding*

29 This work was supported by Theirworld (www.theirworld.org) and a UKRI MRC Programme Grant
30 MR/X003434/1. KM receives salary from NHS Scotland. GS is funded by an MRC Clinician Scientist

1 Fellowship MR/X019535/1. SRC is supported by a Sir Henry Dale Fellowship jointly funded by the
2 Wellcome Trust and the Royal Society (221890/Z/20/Z).

3

4 *6.6 Authors' contributions*

5 KM: Conceptualisation, methodology, formal analysis, writing (original draft), writing (review/editing),
6 visualization. ELSC: Formal analysis, writing (review/editing). KV: Formal analysis, writing
7 (review/editing), visualization. RH: Methodology, formal analysis, writing (review/editing). DG:
8 Methodology, formal analysis, writing (review/editing). JB: Formal analysis, writing (review/editing).
9 GS: Investigation, writing (review/editing). AJS: Methodology, writing (review/editing). AC:
10 Investigation, resources, data curation, writing (review/editing). LM: Investigation, resources, writing
11 (review/editing). HCW: Methodology, writing (review/editing). HR: Conceptualisation, writing (original
12 draft), writing (review/editing), supervision. REM: Conceptualisation, methodology, writing (original
13 draft), writing (review/editing), supervision. SRC: Conceptualisation, methodology, writing (original
14 draft), writing (review/editing), supervision. JPB: Conceptualisation, methodology, writing (original
15 draft), writing (review/editing), supervision, project administration, funding acquisition.

16

17 *6.7 Acknowledgements*

18 The authors are grateful to the families who consented to take part in the study.

19

20

21

22

23

24

25

26

27

28

29

30

1 **Figure legends**

2 *Figure 1. Determining significance threshold*

3 Principal component analysis was used to determine the adjusted statistical significance threshold,
4 given multiple statistical comparisons. Figure 1A shows a correlation matrix of 104 EpiScores,
5 showing correlation coefficient as red for positive and blue negative associations when significant
6 ($p < 0.05$). Figure 1B shows a scree plot of principal components, with the eigenvalues for each
7 component. Standardised component loadings for principal components one and two are provided in
8 Supplementary Table 3 (Additional File 1).

9
10 *Figure 2. EpiScores associated with gestational age in regression models adjusted for socioeconomic*
11 *status*

12 EpiScores associated with gestational age in regression models with (A) Scottish Index of Multiple
13 Deprivation, (B) maternal education, and (C) maternal occupation. Figure 2A (n=331) shows 39
14 associations, Figure 2B (n=323) shows 35 associations, and Figure 2C (n=328) shows 39
15 associations. Points and bars represent standardised beta and 95% confidence intervals, with red
16 indicating positive and blue negative associations. Covariates included in all models: age at sample,
17 birthweight z-score, sex, and methylation processing batch. Bonferroni-adjusted p -value $< 8.3 \times 10^{-3}$.
18 CCL11 – C-C chemokine 11, CCL18 – C-C chemokine 18, CCL21 – C-C chemokine 21, CCL22 – C-
19 C chemokine 22, CCL25 – C-C chemokine 25, CD5L – CD5 antigen-like protein, CD6 – T-cell
20 differentiation antigen, CD163 – scavenger receptor cysteine-rich type 1 protein M130, CI –
21 confidence interval, CRP – C-reactive protein, CRTAM – Cytotoxic and regulatory T-cell molecule,
22 CXCL9 – C-X-C motif chemokine 9, CXCL10 – C-X-C motif chemokine 10, FAP – Fibroblast
23 Activation Protein alpha, FCGR3B – Low affinity immunoglobulin gamma Fc region receptor III-B,
24 FcRL2 – Fc receptor-like protein 2, FGF21 – Fibroblast growth factor 21, GDF15 –
25 Growth/differentiation factor 15, GHR – Growth hormone receptor, HCII – Heparin cofactor II, HGF –
26 Hepatocyte growth factor alpha chain, IGFBP4 – Insulin-like growth factor-binding protein 4, MMP9 –
27 Matrix metalloproteinase-9, NCAM1 – Neural cell adhesion molecule 1, PIGR – Polymeric
28 immunoglobulin receptor, SCGF – stem cell growth factor, SIMD – Scottish Index of Multiple
29 Deprivation, SKR3 – Serine/threonine-protein kinase receptor R3, SLITRK5 – SLIT and NTRK-like

1 protein 5, VCAM1 – Vascular cell adhesion protein 1, VEGFA – Vascular endothelial growth factor A,
2 WFIKKN2 – WAP Kazal immunoglobulin Kunitz and NTR domain-containing protein 2.

3

4 *Figure 3. EpiScores associated with socioeconomic status or an interaction between socioeconomic*
5 *status and birth gestational age*

6 EpiScores associated with socioeconomic status (Scottish Index of Multiple Deprivation, maternal
7 education, or maternal occupation), or with an interaction between socioeconomic status and birth
8 gestational age. Regression models with gestational age at birth, gestational age at sample,
9 birthweight z-score, sex, and methylation processing batch. Sample sizes: for the Scottish Index of
10 Multiple Deprivation n=331, for maternal education n=323, and for maternal occupation n=328. 3/104
11 EpiScores were significant (Bonferroni-adjusted p -value $<8.3 \times 10^{-3}$). Points and bars represent
12 standardised beta and 95% confidence intervals, with red indicating positive and blue negative
13 associations.

14 CI – confidence interval, GA – gestational age, HGFI – Hepatocyte growth factor-like protein alpha
15 chain, ICAM5 – Intercellular adhesion molecule 5, SIMD – Scottish Index of Multiple Deprivation.

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

1 **Tables**

2 *Table 1. Participant characteristics*

Demographic measure		Preterm (n=217)	Term (n=115)
Sex	Male	114/217 (52.5%)	64/115 (55.7%)
	Female	103/217 (47.5%)	51/115 (44.3%)
Birth GA (weeks) - median (range)		29.29 (22.14-33.0)	39.71 (37.0-42.14)
Birthweight (g) - median (range)		1200 (370-2510)	3450 (2346-4670)
Birthweight z-score - median (range)		0.10 (-3.13-2.07)	0.43 (-2.3-2.96)
SIMD rank - median (range) ^A		3720 (6-6966)	5344 (267-6967)
Maternal ethnicity	African	1/217 (0.5%)	0/115 (0%)
	Bangladeshi	0/217 (0%)	1/115 (0.9%)
	Caribbean	0/217 (0%)	0/115 (0%)
	Chinese	0/217 (0%)	1/115 (0.9%)
	Indian	3/217 (2.4%)	1/115 (0.9%)
	Pakistani	4/214 (1.8%)	1/115 (0.9%)
	White	195/217 (89.9%)	106/115 (92.2%)
	White/Asian	2/217 (0.9%)	0/115 (0%)
	White/Black African	0/217 (0%)	1/115 (0.9%)
	White/Black Caribbean	1/217 (0.5%)	1/115 (0.9%)
	Other Asian ^B	1/217 (0.5%)	2/115 (1.7%)
	Other ethnic group ^C	6/217 (2.8%)	0/115 (0%)
	Other mixed ethnic background ^D	4/217 (1.8%)	1/115 (0.9%)
Maternal education	None	7/208 (3.4%)	0/115 (0%)
	Basic high school qualification (1-4)	5/208 (2.4%)	2/115 (1.7%)
	Basic high school qualification (>/=5)	8/208 (3.8%)	1/115 (0.9%)
	Advanced high school	32/208 (15.4%)	3/115 (2.6%)

	qualification		
	College qualification	46/208 (22.1%)	8/115 (7.0%)
	University undergraduate	61/208 (29.3%)	50/115 (43.5%)
	University postgraduate	49/208 (23.6%)	51/115 (44.3%)
Maternal occupation	Unemployed	12/213 (5.6%)	1/115 (0.9%)
	Homemaker	10/213 (4.7%)	1/115 (0.9%)
	Still in full time education	9/213 (4.2%)	1/115 (0.9%)
	Sheltered employment	1/213 (0.5%)	0/115 (0%)
	Unskilled	11/213 (5.2%)	3/115 (2.6%)
	Partly skilled	7/213 (3.3%)	2/115 (1.7%)
	Manual skilled	25/213 (11.7%)	9/115 (7.8%)
	Non-manual skilled	44/213 (20.7%)	11/115 (9.6%)
	Professional	94/213 (44.1%)	87/115 (75.7%)
Sample GA (weeks) - median (range)		40.57 (36.57-45.86)	42.0 (39.86-47.14)
Batch	1	92/217 (42.4%)	46/115 (40.0%)
	2	62/217 (28.6%)	56/115 (48.7%)
	3	32/217 (14.7%)	1/115 (0.9%)
	4	31/217 (14.3%)	12/115 (10.4%)

1 A: Preterm n=216, term n=115

2 B: "Other Asian" includes Sri Lankan (n=1), Malay (n=1)

3 C: "Other ethnic group" includes Arab (n=1), Iraqi (n=1), white Bulgarian (n=1), Fijian (n=1), Japanese
4 (n=1), Hong Kong (n=1).

5 D: "Other mixed ethnic background" includes Pakistani/Scottish (n=2), British/Arab (n=1), Sri
6 Lankan/Black (n=1), Sri Lankan/Indian (n=1).

7 GA = gestational age, SIMD = Scottish index of multiple deprivation.

8

9

10

11

1 *Table 2. Protein EpiScores associated with birth gestational age*

Chemokines	Growth factors	Neurogenesis	Vascular development	Cell membrane proteins / receptors	Other immune response
CCL11	FGF21	NCAM1	SKR3	Afamin	Complement C9
CCL18	GDF15	Semaphorin 3E	VCAM1	CD5L	CRP
CCL21	GHR	SLITRK5	VEGFA	CD6	CRTAM
CCL22	HGF			CD48	FAP
CCL25	IGFBP4			CD163	HCII
CXCL9	SCGF alpha			CD209	L-selectin
CXCL10	SCGF beta			Contactin-4	Lactotransferrin
	WFIKKN2			FcRL2	MMP9
				FcGR3B	S100A9
				PIGR	Sialoadhesin
				Thrombopoietin receptor	Trypsin-2

2 43 EpiScores associated with birth gestational age in regression models adjusted for socioeconomic status. Roles adapted from the STRING database[54].

3 See Supplementary Table 4, Additional File 1, for further details of the functional roles of each protein, including roles in immunity and inflammation, and in
4 preterm infants specifically.

5 CCL11 – C-C chemokine 11, CCL18 – C-C chemokine 18, CCL21 – C-C chemokine 21, CCL22 – C-C chemokine 22, CCL25 – C-C chemokine 25, CD5L –
6 CD5 antigen-like protein, CD6 – T-cell differentiation antigen, CD163 – scavenger receptor cysteine-rich type 1 protein M130, CRP – C-reactive protein,
7 CRTAM – Cytotoxic and regulatory T-cell molecule, CXCL9 – C-X-C motif chemokine 9, CXCL10 – C-X-C motif chemokine 10, FAP – Fibroblast Activation
8 Protein alpha, FCGR3B – Low affinity immunoglobulin gamma Fc region receptor III-B, FcRL2 – Fc receptor-like protein 2, FGF21 – Fibroblast growth factor

1 21, GDF15 – Growth/differentiation factor 15, GHR – Growth hormone receptor, HCII – Heparin cofactor II, HGF – Hepatocyte growth factor alpha chain,
2 IGFBP4 – Insulin-like growth factor-binding protein 4, MMP9 – Matrix metalloproteinase-9, NCAM1 – Neural cell adhesion molecule 1, PIGR – Polymeric
3 immunoglobulin receptor, SCGF – stem cell growth factor, SIMD – Scottish Index of Multiple Deprivation, SKR3 – Serine/threonine-protein kinase receptor
4 R3, SLITRK5 – SLIT and NTRK-like protein 5, VCAM1 – Vascular cell adhesion protein 1, VEGFA – Vascular endothelial growth factor A, WFIKKN2 – WAP
5 Kazal immunoglobulin Kunitz and NTR domain-containing protein 2.

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

1 **References**

- 2 1. Ohuma EO, Moller A-B, Bradley E, Chakwera S, Hussain-Alkhateeb L, Lewin A, et al. National,
3 regional, and global estimates of preterm birth in 2020, with trends from 2010: a systematic analysis.
4 *Lancet*. 2023;402:1261–71.
- 5 2. Johnson S, Marlow N. Early and long-term outcome of infants born extremely preterm. *Archives of*
6 *Disease in Childhood*. 2017;102:97–102.
- 7 3. Agrawal S, Rao SC, Bulsara MK, Patole SK. Prevalence of Autism Spectrum Disorder in Preterm
8 Infants: A Meta-analysis. *Pediatrics*. 2018;142:1–14.
- 9 4. Twilhaar ES, Wade RM, deKieviet JF, vanGoudoever JB, vanElburg RM, Oosterlaan J. Cognitive
10 Outcomes of Children Born Extremely or Very Preterm Since the 1990s and Associated Risk Factors:
11 A Meta-analysis and Meta-regression. *Jama Pediatr*. 2018;172:361.
- 12 5. Crump C. An overview of adult health outcomes after preterm birth. *Early Hum Dev*.
13 2020;150:105187.
- 14 6. Hagberg H, Mallard C, Ferriero DM, Vannucci SJ, Levison SW, Vexler ZS, et al. The role of
15 inflammation in perinatal brain injury. *Nat Rev Neurol*. 2015;11:192–208.
- 16 7. Sullivan G, Galdi P, Blesa-Cábez M, Borbye-Lorenzen N, Stoye DQ, Lamb GJ, et al. Interleukin-8
17 dysregulation is implicated in brain dysmaturation following preterm birth. *Brain Behav Immun*.
18 2020;90:311–8.
- 19 8. Raisi-Estabragh Z, Cooper J, Bethell MS, McCracken C, Lewandowski AJ, Leeson P, et al. Lower
20 birth weight is linked to poorer cardiovascular health in middle-aged population-based adults. *Heart*.
21 2023;109:535–41.
- 22 9. Farah MJ. The Neuroscience of Socioeconomic Status: Correlates, Causes, and Consequences.
23 *Neuron*. 2017;96:56–71.
- 24 10. Kivimäki M, Batty GD, Pentti J, Shipley MJ, Sipilä PN, Nyberg ST, et al. Association between
25 socioeconomic status and the development of mental and physical health conditions in adulthood: a
26 multi-cohort study. *Lancet Public Heal*. 2020;5:e140–9.
- 27 11. Duncan GJ, Ziol-Guest KM, Kalil A. Early-Childhood Poverty and Adult Attainment, Behavior, and
28 Health. *Child Dev*. 2010;81:306–25.
- 29 12. Pillas D, Marmot M, Naicker K, Goldblatt P, Morrison J, Pikhart H. Social inequalities in early
30 childhood health and development: a European-wide systematic review. *Pediatr Res*. 2014;76:418–
31 24.
- 32 13. Thomson K, Moffat M, Arisa O, Jesurasa A, Richmond C, Odeniyi A, et al. Socioeconomic
33 inequalities and adverse pregnancy outcomes in the UK and Republic of Ireland: a systematic review
34 and meta-analysis. *Bmj Open*. 2021;11:e042753.
- 35 14. Ruiz M, Goldblatt P, Morrison J, Kukla L, Švancara J, Riitta-Järvelin M, et al. Mother's education
36 and the risk of preterm and small for gestational age birth: a DRIVERS meta-analysis of 12 European
37 cohorts. *J Epidemiol Commun H*. 2015;69:826–33.
- 38 15. Muscatell KA, Brosso SN, Humphreys KL. Socioeconomic status and inflammation: a meta-
39 analysis. *Mol Psychiatr*. 2020;25:2189–99.

- 1 16. Broyles ST, Staiano AE, Drazba KT, Gupta AK, Sothorn M, Katzmarzyk PT. Elevated C-Reactive
2 Protein in Children from Risky Neighborhoods: Evidence for a Stress Pathway Linking Neighborhoods
3 and Inflammation in Children. *Plos One*. 2012;7:e45419.
- 4 17. Dowd JB, Zajacova A, Aiello AE. Predictors of Inflammation in U.S. Children Aged 3–16 Years.
5 *Am J Prev Med*. 2010;39:314–20.
- 6 18. Schmeer KK, Yoon AJ. Home sweet home? Home physical environment and inflammation in
7 children. *Soc Sci Res*. 2016;60:236–48.
- 8 19. Schmeer KK, Yoon A. Socioeconomic status inequalities in low-grade inflammation during
9 childhood. *Arch Dis Child*. 2016;101:1043.
- 10 20. Chiesa C, Natale F, Pascone R, Osborn JF, Pacifico L, Bonci E, et al. C reactive protein and
11 procalcitonin: Reference intervals for preterm and term newborns during the early neonatal period.
12 *Clin Chim Acta*. 2011;412:1053–9.
- 13 21. Brown JVE, Meader N, Cleminson J, McGuire W. C-reactive protein for diagnosing late-onset
14 infection in newborn infants. Group CN, editor. *Cochrane Database of Systematic Reviews*.
15 2019;10:16–78.
- 16 22. Bower JK, Lazo M, Juraschek SP, Selvin E. Within-Person Variability in High-Sensitivity C-
17 Reactive Protein. *Arch Intern Med*. 2012;172:1519–21.
- 18 23. Dammann O, Leviton A. Intermittent or sustained systemic inflammation and the preterm brain.
19 *Pediatr Res*. 2014;75:376–80.
- 20 24. Stevenson AJ, McCartney DL, Harris SE, Taylor AM, Redmond P, Starr JM, et al. Trajectories of
21 inflammatory biomarkers over the eighth decade and their associations with immune cell profiles and
22 epigenetic ageing. *Clin Epigenetics*. 2018;10:159.
- 23 25. Stevenson AJ, McCartney DL, Hillary RF, Campbell A, Morris SW, Birmingham ML, et al.
24 Characterisation of an inflammation-related epigenetic score and its association with cognitive ability.
25 *Clin Epigenetics*. 2020;12:113.
- 26 26. Gadd DA, Hillary RF, McCartney DL, Zaghlool SB, Stevenson AJ, Cheng Y, et al. Epigenetic
27 scores for the circulating proteome as tools for disease prediction. *Elife*. 2022;11:e71802.
- 28 27. Ligthart S, Marzi C, Aslibekyan S, Mendelson MM, Conneely KN, Tanaka T, et al. DNA
29 methylation signatures of chronic low-grade inflammation are associated with complex diseases.
30 *Genome Biol*. 2016;17:255.
- 31 28. Stevenson AJ, Gadd DA, Hillary RF, McCartney DL, Campbell A, Walker RM, et al. Creating and
32 Validating a DNA Methylation-Based Proxy for Interleukin-6. *Journals Gerontology Ser Biological Sci
33 Medical Sci*. 2021;76:2284–92.
- 34 29. Conole ELS, Stevenson AJ, Maniega SM, Harris SE, Green C, Hernández M del CV, et al. DNA
35 Methylation and Protein Markers of Chronic Inflammation and Their Associations With Brain and
36 Cognitive Aging. *Neurology*. 2021;97:e2340–52.
- 37 30. Green C, Shen X, Stevenson AJ, Conole ELS, Harris MA, Barbu MC, et al. Structural brain
38 correlates of serum and epigenetic markers of inflammation in major depressive disorder. *Brain
39 Behav Immun*. 2021;92:39–48.
- 40 31. Barker ED, Cecil CAM, Walton E, Houtepen LC, O'Connor TG, Danese A, et al. Inflammation-
41 related epigenetic risk and child and adolescent mental health: A prospective study from pregnancy to
42 middle adolescence. *Dev Psychopathol*. 2018;30:1145–56.

- 1 32. Conole ELS, Vaheer K, Blesa-Cábez M, Sullivan G, Stevenson AJ, Hall J, et al. Immuno-epigenetic
2 signature derived in saliva associates with the encephalopathy of prematurity and perinatal
3 inflammatory disorders. *Brain Behav Immun.* 2023;110:322–38.
- 4 33. Needham BL, Smith JA, Zhao W, Wang X, Mukherjee B, Kardia SLR, et al. Life course
5 socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation:
6 The multi-ethnic study of atherosclerosis. *Epigenetics.* 2015;10:958–69.
- 7 34. Stringhini S, Polidoro S, Sacerdote C, Kelly RS, Veldhoven K van, Agnoli C, et al. Life-course
8 socioeconomic status and DNA methylation of genes regulating inflammation. *Int J Epidemiol.*
9 2015;44:1320–30.
- 10 35. McDade TW, Ryan C, Jones MJ, Maclsaac JL, Morin AM, Meyer JM, et al. Social and physical
11 environments early in development predict DNA methylation of inflammatory genes in young
12 adulthood. *Proc National Acad Sci.* 2017;114:7611–6.
- 13 36. Smith JA, Zhao W, Wang X, Ratliff SM, Mukherjee B, Kardia SLR, et al. Neighborhood
14 characteristics influence DNA methylation of genes involved in stress response and inflammation: The
15 Multi-Ethnic Study of Atherosclerosis. *Epigenetics.* 2017;12:662–73.
- 16 37. Gadd DA, Smith HM, Mullin D, Chybowska O, Hillary RF, Kimenai DM, et al. DNAm scores for
17 serum GDF15 and NT-proBNP levels associate with a range of traits affecting the body and brain.
18 *medRxiv.* 2023;2023.10.18.23297200.
- 19 38. Boardman JP, Hall J, Thrippleton MJ, Reynolds RM, Bogaert D, Davidson DJ, et al. Impact of
20 preterm birth on brain development and long-term outcome: protocol for a cohort study in Scotland.
21 *Bmj Open.* 2020;10:e035854.
- 22 39. Scottish National Statistics. SIMD - Scottish Index of Multiple Deprivation: SIMD16 technical
23 notes. Edinburgh: Scottish National Statistics; 2016 p. 1–69.
- 24 40. National Records of Scotland. Scotland's Census [Internet]. 2020 [cited 2023 Nov 23]. Available
25 from: <https://www.scotlandscensus.gov.uk/search-the-census/#/>
- 26 41. Lowe R, Gemma C, Beyan H, Hawa MI, Bazeos A, Leslie RD, et al. Buccals are likely to be a
27 more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics.*
28 2013;8:445–54.
- 29 42. Braun PR, Han S, Hing B, Nagahama Y, Gaul LN, Heinzman JT, et al. Genome-wide DNA
30 methylation comparison between live human brain and peripheral tissues within individuals. *Transl
31 Psychiat.* 2019;9:47.
- 32 43. Pidsley R, Wong CCY, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to
33 preprocessing Illumina 450K methylation array data. *Bmc Genomics.* 2013;14:293.
- 34 44. Zheng SC, Webster AP, Dong D, Feber A, Graham DG, Sullivan R, et al. A novel cell-type
35 deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and
36 cervix. *Epigenomics-uk.* 2018;10:925–40.
- 37 45. Katz DH, Robbins JM, Deng S, Tahir UA, Bick AG, Pampana A, et al. Proteomic profiling
38 platforms head to head: Leveraging genetics and clinical traits to compare aptamer- and antibody-
39 based methods. *Sci Adv.* 2022;8:eabm5164.
- 40 46. Haslam DE, Li J, Dillon ST, Gu X, Cao Y, Zeleznik OA, et al. Stability and reproducibility of
41 proteomic profiles in epidemiological studies: comparing the Olink and SOMAscan platforms.
42 *PROTEOMICS.* 2022;22:e2100170.

- 1 47. Raffield LM, Dang H, Pratte KA, Jacobson S, Gillenwater LA, Ampleford E, et al. Comparison of
2 Proteomic Assessment Methods in Multiple Cohort Studies. *PROTEOMICS*. 2020;20:1900278.
- 3 48. Joshi A, Mayr M. In Aptamers They Trust: The caveats of the SOMAscan biomarker discovery
4 platform from SomaLogic. *Circulation*. 2018;138:2482–5.
- 5 49. Hillary RF, Marioni RE. MethylDetectR: a software for methylation-based health profiling.
6 *Wellcome Open Res*. 2021;5:283.
- 7 50. Preacher KJ. Extreme Groups Designs. *The Encyclopedia of Clinical Psychology*. 2015;1–4.
- 8 51. Fisher JE, Guha A, Heller W, Miller GA. Extreme-Groups Designs in Studies of Dimensional
9 Phenomena: Advantages, Caveats, and Recommendations. *J Abnorm Psychol*. 2020;129:14–20.
- 10 52. Mckinnon K, Conole ELS, Vaheer K, Binkowska J, Sullivan G, Hillary R, et al. The relationship
11 between socioeconomic status, preterm birth and systemic inflammation using DNA methylation
12 proxies. *OSF*. 2023;
- 13 53. Gao X, Becker LC, Becker DM, Starmer JD, Province MA. Avoiding the high Bonferroni penalty in
14 genome-wide association studies. *Genet Epidemiol*. 2010;34:100–5.
- 15 54. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING
16 database in 2023: protein–protein association networks and functional enrichment analyses for any
17 sequenced genome of interest. *Nucleic Acids Res*. 2022;51:D638–46.
- 18 55. Kandasamy J, Roane C, Szalai A, Ambalavanan N. Serum eotaxin-1 is increased in extremely-
19 low-birth-weight infants with bronchopulmonary dysplasia or death. *Pediatr Res*. 2015;78:498–504.
- 20 56. Almudares F, Hagan J, Chen X, Devaraj S, Moorthy B, Lingappan K. Growth and differentiation
21 factor 15 (GDF15) levels predict adverse respiratory outcomes in premature neonates. *Pediatr*
22 *Pulmonol*. 2023;58:271–8.
- 23 57. Al-Mudares F, Reddick S, Ren J, Venkatesh A, Zhao C, Lingappan K. Role of Growth
24 Differentiation Factor 15 in Lung Disease and Senescence: Potential Role Across the Lifespan.
25 *Frontiers Medicine*. 2020;7:594137.
- 26 58. Wan Y, Fu J. GDF15 as a key disease target and biomarker: linking chronic lung diseases and
27 ageing. *Mol Cell Biochem*. 2023;1–14.
- 28 59. Zhang Y, Jiang W, Wang L, Lingappan K. Sex-specific differences in the modulation of Growth
29 Differentiation Factor 15 (GDF15) by hyperoxia in vivo and in vitro: Role of Hif-1 α . *Toxicol Appl*
30 *Pharm*. 2017;332:8–14.
- 31 60. Wang G, Wen B, Deng Z, Zhang Y, Kolesnichenko OA, Ustiyanyan V, et al. Endothelial progenitor
32 cells stimulate neonatal lung angiogenesis through FOXF1-mediated activation of BMP9/ACVRL1
33 signaling. *Nat Commun*. 2022;13:2080.
- 34 61. Collaco JM, McGrath-Morrow SA, Griffiths M, Chavez-Valdez R, Parkinson C, Zhu J, et al.
35 Perinatal Inflammatory Biomarkers and Respiratory Disease in Preterm Infants. *J Pediatrics*.
36 2022;246:34-39.e3.
- 37 62. Oak P, Hilgendorff A. The BPD trio? Interaction of dysregulated PDGF, VEGF, and TGF signaling
38 in neonatal chronic lung disease. *Mol Cell Pediatrics*. 2017;4:11.
- 39 63. Course CW, Lewis PA, Kotecha SJ, Cousins M, Hart K, Watkins WJ, et al. Characterizing the
40 urinary proteome of prematurity-associated lung disease in school-aged children. *Respir Res*.
41 2023;24:191.

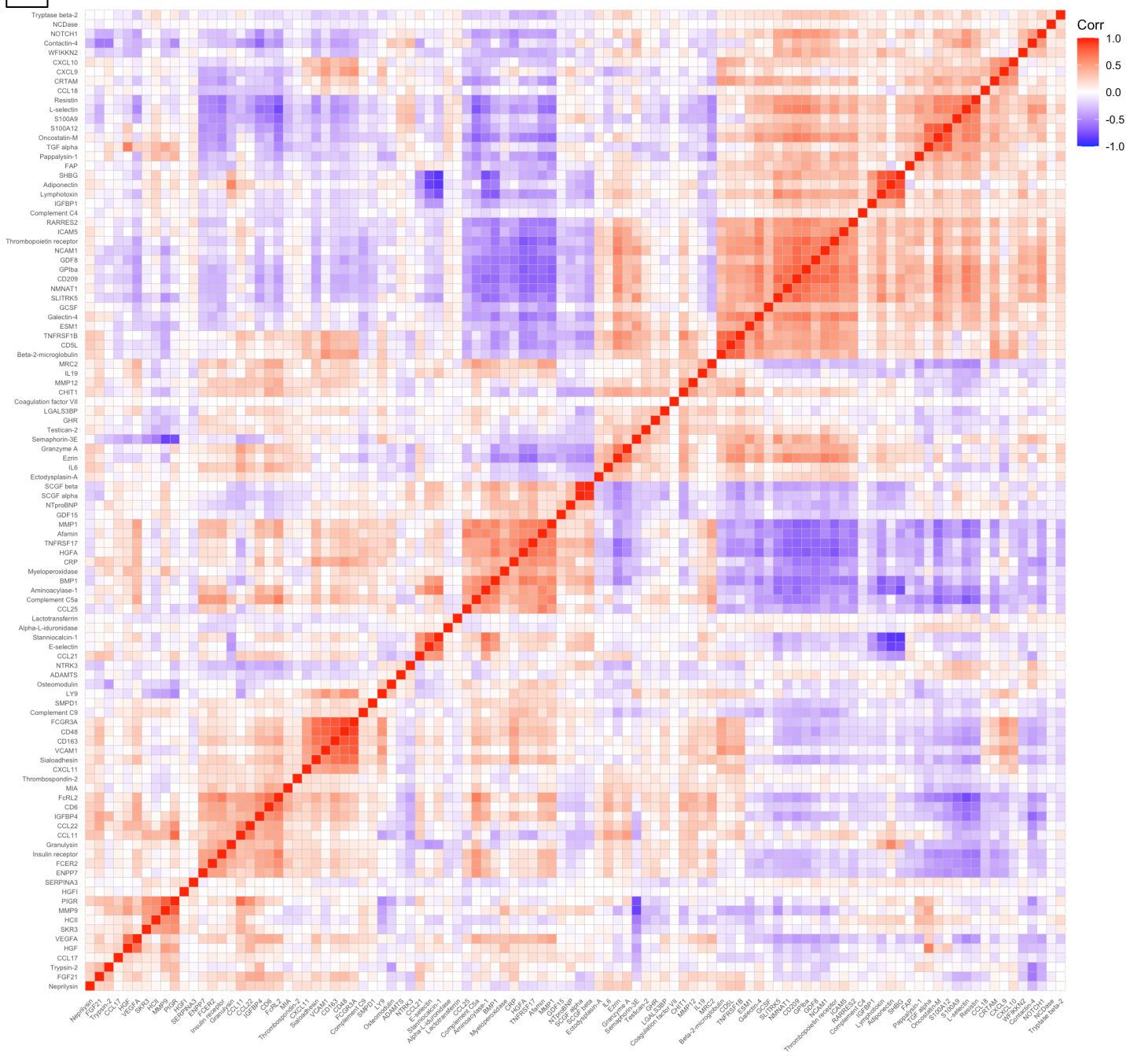
- 1 64. Bose CL, Dammann CEL, Laughon MM. Bronchopulmonary dysplasia and inflammatory
2 biomarkers in the premature neonate. *Archives Dis Child - Fetal Neonatal Ed.* 2008;93:F455.
- 3 65. Zasada M, Suski M, Bokinić R, Szwarc-Duma M, Borszewska-Kornacka MK, Madej J, et al.
4 Comparative two time-point proteome analysis of the plasma from preterm infants with and without
5 bronchopulmonary dysplasia. *Ital J Pediatr.* 2019;45:112.
- 6 66. Lassus P, Heikkilä P, Andersson LC, Boguslawski K von, Andersson S. Lower concentration of
7 pulmonary hepatocyte growth factor is associated with more severe lung disease in preterm infants. *J*
8 *Pediatr.* 2003;143:199–202.
- 9 67. Wagner BD, Babinec AE, Carpenter C, Gonzalez S, O'Brien G, Rollock K, et al. Proteomic
10 Profiles Associated with Early Echocardiogram Evidence of Pulmonary Vascular Disease in Preterm
11 Infants. *Am J Respir Crit Care Med.* 2018;197:394–7.
- 12 68. Guasti L, Silvennoinen S, Bulstrode NW, Ferretti P, Sankilampi U, Dunkel L. Elevated FGF21
13 Leads to Attenuated Postnatal Linear Growth in Preterm Infants Through GH Resistance in
14 Chondrocytes. *J Clin Endocrinol Metab.* 2014;99:E2198–206.
- 15 69. Spencer R, Maksym K, Hecher K, Maršál K, Figueras F, Ambler G, et al. Ultrasound and
16 biochemical predictors of pregnancy outcome at diagnosis of early-onset fetal growth restriction.
17 *Medrxiv.* 2023;2023.01.27.23285087.
- 18 70. Qiu Q, Bell M, Lu X, Yan X, Rodger M, Walker M, et al. Significance of IGFBP-4 in the
19 Development of Fetal Growth Restriction. *J Clin Endocrinol Metab.* 2012;97:E1429–39.
- 20 71. Voller SB, Chock S, Ernst LM, Su E, Liu X, Farrow KN, et al. Cord blood biomarkers of vascular
21 endothelial growth (VEGF and sFlt-1) and postnatal growth: A preterm birth cohort study. *Early Hum*
22 *Dev.* 2014;90:195–200.
- 23 72. Kim MJ, Romero R, Kim CJ, Tarca AL, Chhauy S, LaJeunesse C, et al. Villitis of Unknown
24 Etiology Is Associated with a Distinct Pattern of Chemokine Up-Regulation in the Feto-Maternal and
25 Placental Compartments: Implications for Conjoint Maternal Allograft Rejection and Maternal Anti-
26 Fetal Graft-versus-Host Disease. *J Immunol.* 2009;182:3919–27.
- 27 73. Romero R, Chaemsaitong P, Chaiyasit N, Docheva N, Dong Z, Kim CJ, et al. CXCL10 and IL-6:
28 Markers of two different forms of intra-amniotic inflammation in preterm labor. *Am J Reprod Immunol.*
29 2017;78:e12685.
- 30 74. Sellmer A, Henriksen TB, Palmfeldt J, Bech BH, Astono J, Bennike TB, et al. The Patent Ductus
31 Arteriosus in Extremely Preterm Neonates Is More than a Hemodynamic Challenge: New Molecular
32 Insights. *Biomol.* 2022;12:1179.
- 33 75. Olsson KW, Larsson A, Jonzon A, Sindelar R. Exploration of potential biochemical markers for
34 persistence of patent ductus arteriosus in preterm infants at 22–27 weeks' gestation. *Pediatr Res.*
35 2019;86:333–8.
- 36 76. Waleh N, Seidner S, McCurnin D, Giavedoni L, Hodara V, Goelz S, et al. Anatomic Closure of the
37 Premature Patent Ductus Arteriosus: The Role of CD14+/CD163+ Mononuclear Cells and VEGF in
38 Neointimal Mound Formation. *Pediatr Res.* 2011;70:332–8.
- 39 77. Xu C, Su X, Chen Y, Xu Y, Wang Z, Mo X. Proteomics analysis of plasma protein changes in
40 patent ductus arteriosus patients. *Ital J Pediatr.* 2020;46:64.
- 41 78. Cheng Y, Zhu X, Linghu D, Xu Y, Liang J. Serum levels of cytokines in infants treated with
42 conbercept for retinopathy of prematurity. *Sci Rep-uk.* 2020;10:12695.

- 1 79. Sato T, Kusaka S, Shimojo H, Fujikado T. Simultaneous Analyses of Vitreous Levels of 27
2 Cytokines in Eyes with Retinopathy of Prematurity. *Ophthalmology*. 2009;116:2165–9.
- 3 80. Rathi S, Jalali S, Patnaik S, Shahulhameed S, Musada GR, Balakrishnan D, et al. Abnormal
4 Complement Activation and Inflammation in the Pathogenesis of Retinopathy of Prematurity. *Front*
5 *Immunol*. 2017;8:1868.
- 6 81. The Royal College of Ophthalmologists. Treating Retinopathy of Prematurity in the UK. The Royal
7 College of Ophthalmologists; 2022.
- 8 82. Rivera JC, Holm M, Austeng D, Morken TS, Zhou T (Ellen), Beaudry-Richard A, et al. Retinopathy
9 of prematurity: inflammation, choroidal degeneration, and novel promising therapeutic strategies. *J*
10 *Neuroinflamm*. 2017;14:165.
- 11 83. Klerk DH, Plösch T, Verkaik-Schakel RN, Hulscher JBF, Kooi EMW, Bos AF. DNA Methylation of
12 TLR4, VEGFA, and DEFA5 Is Associated With Necrotizing Enterocolitis in Preterm Infants. *Frontiers*
13 *Pediatrics*. 2021;9:630817.
- 14 84. Olaloye OO, Liu P, Toothaker JM, McCourt BT, McCourt CC, Xiao J, et al. CD16+CD163+
15 monocytes traffic to sites of inflammation during necrotizing enterocolitis in premature infants. *J Exp*
16 *Med*. 2021;218:e20200344.
- 17 85. Pammi M, Suresh G. Enteral lactoferrin supplementation for prevention of sepsis and necrotizing
18 enterocolitis in preterm infants. *Cochrane Db Syst Rev*. 2020;3:CD007137.
- 19 86. Satrom KM, Ennis K, Sweis BM, Matveeva TM, Chen J, Hanson L, et al. Neonatal hyperglycemia
20 induces CXCL10/CXCR3 signaling and microglial activation and impairs long-term synaptogenesis in
21 the hippocampus and alters behavior in rats. *J Neuroinflamm*. 2018;15:82.
- 22 87. Groselj-Grenc M, Ihan A, Derganc M. Neutrophil and Monocyte CD64 and CD163 Expression in
23 Critically Ill Neonates and Children with Sepsis: Comparison of Fluorescence Intensities and
24 Calculated Indexes. *Mediat Inflamm*. 2008;2008:202646.
- 25 88. Kingsmore SF, Kennedy N, Halliday HL, Velkinburgh JCV, Zhong S, Gabriel V, et al. Identification
26 of Diagnostic Biomarkers for Infection in Premature Neonates. *Mol Cell Proteomics*. 2008;7:1863–75.
- 27 89. Zilow EP, Hauck W, Linderkamp O, Zilow G. Alternative Pathway Activation of the Complement
28 System in Preterm Infants with Early Onset Infection. *Pediatr Res*. 1997;41:334–9.
- 29 90. Mohammed A, Okwor I, Shan L, Onyilagha C, Uzonna JE, Gounni AS. Semaphorin 3E Regulates
30 the Response of Macrophages to Lipopolysaccharide-Induced Systemic Inflammation. *J Immunol*.
31 2020;204:128–36.
- 32 91. Leviton A, Allred EN, Fichorova RN, O'Shea TM, Fordham LA, Kuban KKC, et al. Circulating
33 biomarkers in extremely preterm infants associated with ultrasound indicators of brain damage. *Eur J*
34 *Paediatr Neuro*. 2018;22:440–50.
- 35 92. Schultz SJ, Aly H, Hasanen BM, Khashaba MT, Lear SC, Bendon RW, et al. Complement
36 component 9 activation, consumption, and neuronal deposition in the post-hypoxic–ischemic central
37 nervous system of human newborn infants. *Neurosci Lett*. 2005;378:1–6.
- 38 93. Morales DM, Townsend RR, Malone JP, Ewersmann CA, Macy EM, Inder TE, et al. Alterations in
39 Protein Regulators of Neurodevelopment in the Cerebrospinal Fluid of Infants with Posthemorrhagic
40 Hydrocephalus of Prematurity. *Mol Cell Proteom*. 2012;11:M111.011973.
- 41 94. Ochoa TJ, Sizonenko SV. Lactoferrin and prematurity: a promising milk protein? *Biochem Cell*
42 *Biol*. 2017;95:22–30.

- 1 95. Che X, Hornig M, Bresnahan M, Stoltenberg C, Magnus P, Surén P, et al. Maternal mid-
2 gestational and child cord blood immune signatures are strongly associated with offspring risk of ASD.
3 *Mol Psychiatr.* 2022;27:1527–41.
- 4 96. Leviton A, Joseph RM, Fichorova RN, Allred EN, Taylor HG, O’Shea TM, et al. Executive
5 Dysfunction Early Postnatal Biomarkers among Children Born Extremely Preterm. *J Neuroimmune*
6 *Pharm.* 2019;14:188–99.
- 7 97. Allred EN, Dammann O, Fichorova RN, Hooper SR, Hunter SJ, Joseph RM, et al. Systemic
8 Inflammation during the First Postnatal Month and the Risk of Attention Deficit Hyperactivity Disorder
9 Characteristics among 10 year-old Children Born Extremely Preterm. *J Neuroimmune Pharm.*
10 2017;12:531–43.
- 11 98. Aly H, Khashaba M, Nada A, Hasanen B, McCarter R, Schultz S, et al. The Role of Complement
12 in Neurodevelopmental Impairment following Neonatal Hypoxic-Ischemic Encephalopathy. *Am J*
13 *Perinatol.* 2009;26:659–65.
- 14 99. Limbrick DD, Morales DM, Shannon CN, Wellons JC, Kulkarni AV, Alvey JS, et al. Cerebrospinal
15 fluid NCAM-1 concentration is associated with neurodevelopmental outcome in post-hemorrhagic
16 hydrocephalus of prematurity. *PLoS ONE.* 2021;16:e0247749.
- 17 100. Voegelé AF, Jerković L, Wellenzohn B, Eller P, Kronenberg F, Liedl KR, et al. Characterization
18 of the Vitamin E-Binding Properties of Human Plasma Afamin. *Biochemistry.* 2002;41:14532–8.
- 19 101. Paetau S, Rolova T, Ning L, Gahmberg CG. Neuronal ICAM-5 Inhibits Microglia Adhesion and
20 Phagocytosis and Promotes an Anti-inflammatory Response in LPS Stimulated Microglia. *Front Mol*
21 *Neurosci.* 2017;10:431.
- 22 102. Skeel A, Yoshimura T, Showalter SD, Tanaka S, Appella E, Leonard EJ. Macrophage stimulating
23 protein: purification, partial amino acid sequence, and cellular activity. *J Exp Med.* 1991;173:1227–34.
- 24 103. Kronenberg F, Kollerits B, Kiechl S, Lamina C, Kedenko L, Meisinger C, et al. Plasma
25 Concentrations of Afamin Are Associated With the Prevalence and Development of Metabolic
26 Syndrome. *Circ: Cardiovasc Genet.* 2018;7:822–9.
- 27 104. Montez JK, Bromberger JT, Harlow SD, Kravitz HM, Matthews KA. Life-Course Socioeconomic
28 Status and Metabolic Syndrome Among Midlife Women. *J Gerontol: Ser B.* 2016;71:1097–107.
- 29 105. Castagné R, Delpierre C, Kelly-Irving M, Campanella G, Guida F, Krogh V, et al. A life course
30 approach to explore the biological embedding of socioeconomic position and social mobility through
31 circulating inflammatory markers. *Sci Rep-uk.* 2016;6:25170.
- 32 106. Leviton A, Allred EN, Dammann O, Joseph RM, Fichorova RN, O’Shea TM, et al. Socioeconomic
33 status and early blood concentrations of inflammation-related and neurotrophic proteins among
34 extremely preterm newborns. *Plos One.* 2019;14:e0214154.
- 35 107. Suarez A, Lahti J, Lahti-Pulkkinen M, Girchenko P, Czamara D, Arloth J, et al. A polyepigenetic
36 glucocorticoid exposure score at birth and childhood mental and behavioral disorders. *Neurobiology*
37 *Stress.* 2020;13:100275.
- 38 108. Blostein FA, Fisher J, Dou J, Schnepfer L, Ware EB, Natterman DA, et al. Polymethylation scores
39 for prenatal maternal smoke exposure persist until age 15 and are detected in saliva in the Fragile
40 Families and Child Wellbeing cohort. *Epigenetics.* 2022;17:2223–40.

41

A



B

