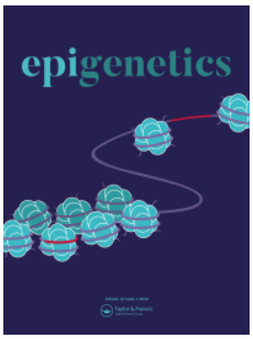

Epigenetics

ISSN: 1559-2294 (Print) 1559-2308 (Online) Journal homepage: www.tandfonline.com/journals/kepi20

Genome-wide DNA methylation profiles and ribosomal DNA copy number at birth

Kathrin Barth, Rossella Alfano, Michelle Plusquin, Congrong Wang, Tim S. Nawrot & Dries S. Martens

To cite this article: Kathrin Barth, Rossella Alfano, Michelle Plusquin, Congrong Wang, Tim S. Nawrot & Dries S. Martens (2026) Genome-wide DNA methylation profiles and ribosomal DNA copy number at birth, Epigenetics, 21:1, 2633818, DOI: 10.1080/15592294.2026.2633818

To link to this article: https://doi.org/10.1080/15592294.2026.2633818

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

View supplementary material

Published online: 03 Mar 2026.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=kepi20

EPIGENETICS
2026, VOL. 21, NO. 1, 2633818
https://doi.org/10.1080/15592294.2026.2633818

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Genome-wide DNA methylation profiles and ribosomal DNA copy number at birth

Kathrin Barth[a], Rossella Alfano[a], Michelle Plusquin[a], Congrong Wang[a], Tim S. Nawrot[a,b] and Dries S. Martens[a]

[a]Centre for Environmental Sciences, Hasselt University, Hasselt, Belgium; [b]Research Unit Environment and Health, KU Leuven Department of Public Health and Primary Care, University of Leuven, Leuven, Belgium

ABSTRACT

Ribosomal DNA copy number (rDNAcn) and DNA methylation are important modulators of the human genome, both studied in relation to overall cellular function, biological ageing, and disease development. Despite the overlapping roles, their relationship remains poorly understood, especially in the early stages of life, characterized by rapid growth and high cellular demands. Even though previous studies have associated rDNA methylation with cancer and ageing, no study to date has examined the interplay between rDNAcn and whole-genome DNA methylation. In an epigenome-wide association study of 45S rDNAcn variation in 194 newborns, we show strong positive associations between rDNAcn and single DNA methylated CpGs, measured with the Illumina EPIC array. Out of the 122 Bonferroni-significant CpGs, 63.5% were also Bonferroni-significant in a replication cohort of 167 newborns, in which a second EWAS was conducted using DNA methylation data from the Illumina 450K array. The identified CpGs were dispersed over the autosomes and were not functionally related to the rDNA-forming nucleolar-associated domains. The top CpGs were annotated to genes (*GFI1*, *USP46*, *ABHD14B*, *CHL1*, *CGREF1*) that are functionally linked to cancer and cellular proliferation. In downstream analyses, the 122 rDNAcn-related CpGs revealed 31 differentially methylated regions and 253 nominally significant correlations with cord blood gene transcripts in an eQTM analysis. Pathway enrichment analyses showed an overrepresentation of the following pathways: 'RNA Polymerase III transcription' (R-HSA-76071, R-HSA-76046, R-HSA-74158, R-HSA-749476, R-HSA-73780, R-HSA-73980, R-HSA-76066, R-HSA-76061, hsa03020), 'cytosolic sensors of pathogen-associated DNA' (R-HSA-1834949), 'RNA polymerase II transcribes snRNA genes' (R-HSA-6807505), and 'translation initiation' (R-HSA-72613, R-HSA-72737). Our findings reveal a close link between rDNAcn variation and DNA methylation in early life. Disruptions in this interplay may influence cellular functions critical for early development, potentially shaping health and disease trajectories later in life.

ARTICLE HISTORY
Received 10 October 2025
Revised 4 February 2026
Accepted 14 February 2026

KEYWORDS
Ribosomal DNA; whole-genome DNA methylation; early life aging; EWAS; birth cohort

Introduction

Ribosomal DNA (rDNA) is organized in a tandem repeat structure, which is highly conserved across all kingdoms. Of the 100 to 600 rDNA repeats that can be found in the human diploid genome [1], only about 50% are actively transcribed [2]. Still, the rDNA transcription makes up the majority of all the RNA production in a cell [3].

In humans, rDNA is encoded from the 45S and the 5S array. The 45S array is localized on the short arms of the acrocentric chromosomes 13, 14, 15, 21, and 22. It is transcribed by RNA Polymerase I and processed into 18S, 5.8S, and 28S rRNAs. The 5S rDNA array is found on chromosome 1, where it is transcribed by RNA Polymerase III. The 5S rRNA, combined with the 45S rRNAs and cytoplasmic ribosomal proteins (cRP), builds the human ribosomes responsible for protein synthesis [4]. Furthermore, rDNA clusters encompass nucleolus organizer regions (NORs), which aggregate to form the nucleolus. This membrane-

CONTACT Dries S. Martens dries.martens@uhasselt.be Centre for Environmental Sciences, Hasselt University, Agoralaan Building D, Diepenbeek BE-3590, Belgium

Supplemental data for this article can be accessed online at https://doi.org/10.1080/15592294.2026.2633818.

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

less organelle, located within the cell's nucleus, is the primary site of ribosome biogenesis. It has been shown that several NORs are affiliated with other genomic regions, known as nucleolar-associated domains (NADs) [5,6]. Even though the NADs are spread across all chromosomes, they are typically associated with repeated DNA sequences, low gene density, and reduced global gene expression [5].

Variations in the structure, sequence, and copy number of the rDNA arrays are frequently reported [7,8]. The direct implications of those changes have, however, only barely been studied. The regulation of the rDNAcn in humans is only partly understood. Repeats can get lost over time due to deleterious recombination events, double-strand-break repair, or the formation of unintended secondary structures [9]. An increase in the copy number was observed in yeast, where the rDNA amplification mechanism has been described [10].

In recent years, accumulating evidence supports that rDNA not only plays an important role in ribosome biogenesis but also that variation in the rDNA arrays can induce genetic and phenotypic diversity within populations [11,12]. rDNAcn is connected to global gene expression [11], ageing [13,14], and health phenotypes [15], including body mass [16], renal function [17], and different types of human cancer [18]. Additionally, it was found that rDNA can trigger genomic instability [19] and alter genome organization [20]. Therefore, it is argued that rDNA could play a role in cellular senescence [21] and induce replication gaps [22].

DNA methylation is one of the major processes that regulates gene expression. A broad range of studies reveal insights into molecular underpinnings and areas of influence. Hypermethylation of a gene's promoter region suppresses gene expression [23], while gene body methylation is connected to an upregulation of gene transcription [24]. Besides this, DNA methylation is listed as a hallmark of ageing [25] and cancer [26]. It is associated with genomic instability [27] and developmental processes [28].

While DNA methylation has been the focus of many studies, rDNA appears to be a rather understudied genomic regulator. Despite the considerable overlap in functionalities, their relation remains largely unclear. Recent evidence suggests a link between rDNAcn and gene expression [11], potentially mediated by DNA methylation. Therefore, here we examined the associations or interconnections between rDNAcn and genome-wide DNA methylation. The large biological implications of both of these genome regulators may be related to important phenotypic differences in health and disease. The Developmental Origins of Health and Disease (DOHaD) hypothesis stipulates that diseases in later life may be programmed in the early stages of life. Consequently, investigating key omic regulators such as rDNAcn and genome-wide methylation may provide insights into the starting point of the molecular programming of health and disease, potentially shaping later health phenotypes. Therefore, we investigated the interconnection between rDNAcn and genome-wide DNA methylation in newborns. We conducted an epigenome-wide association study (EWAS) to identify CpGs associated with the number of rDNA copies in cord blood samples of two independent newborn populations. By analysing gene-annotated CpGs, we evaluated the overrepresentation of biological pathways to uncover insights into molecular underpinnings and regulatory processes. Our results provide the first evidence for a positive interconnection between DNA methylation and rDNAcn in cord blood. Furthermore, we were able to identify biomolecular pathways that may contribute, as a cause or consequence, to this relation.

Methods

Study population

This study included participants from the ongoing ENVIR*ON*AGE (ENVIRonmental influence *ON* early life AGEing) birth cohort. General study procedures have been described before [29]. In brief, mother-newborn pairs were recruited at birth at the East-Limburg Hospital (Genk, Belgium). Inclusion criteria were the mothers' ability to fill out a Dutch questionnaire and that they did not have a planned C-section or a multiple birth. The ENVIR*ON*AGE study protocol is approved by the local ethical committees of Hasselt University (Diepenbeek, Belgium, reference no. B371201216090 and B37120107805) and East-Limburg Hospital (Genk, Belgium, reference no. EudraCT B37120107805) and has been carried out according to the Helsinki declaration. Mothers provided written informed consent. Between February 2010 and June 2015, 930 participants were

recruited. This study included a subset of 361 participants for whom rDNAcn measurements were available. Of these, 194 had genome-wide DNA methylation measured via the Illumina HumanMethylationEPIC Bead-Chip array and constituted the EPIC discovery dataset. In addition, 167 newborns had methylation measured via the Illumina HumanMethylation450K Bead-Chip array and constituted the 450K replication dataset (for a flow chart, see Supplementary Figure S1). To assess the representativeness of the population, we compared the analysed fraction of the ENVIR*ON*AGE population with all births from Flanders, Belgium, between 1999–2009. Overall, the ENVIR*ON*AGE population is representative of the gestational segment at large in Flanders, Belgium, with a slightly higher degree of newborns classified as European and a higher educational level of the mothers (Supplementary Table S1).

Measures

At birth, clinical, anthropometrical, and lifestyle data were collected for mothers and newborns via medical records and questionnaire data. Gestational age (in weeks) was estimated based on the last menstrual cycle and the first fetal ultrasound examination. Birth weight was measured in grams. Maternal height and weight were measured without shoes, wearing light clothes at the first antenatal visit (week 7–9 of gestation) to the nearest centimetre and 0.1 kg. Pre-pregnancy BMI was weight in kilograms divided by height in metres squared. Maternal educational achievement was categorized using the International Standard Classification of Education (for further information, see Education https://uis.unesco.org/en/topic/international-standard-classification-education-isced). It was low when no diploma was obtained, middle when a secondary diploma was obtained, and high when a college or university diploma was obtained. Maternal smoking was coded as nonsmokers, stopped before pregnancy, and continued smokers. Newborn ethnicity was coded as European if at least two grandparents of the child were of European origin; otherwise, they were considered non-European.

Sample collection, DNA, and RNA extraction

Immediately after birth, umbilical cord blood was collected in BD Vacutainer® plastic whole blood tubes with spray-coated K2EDTA (BD, Franklin Lakes, NJ, USA). Blood samples were centrifuged within one hour at 3200 rpm for 15 min to obtain the buffy coat layer. Buffy coats were aliquoted and stored at −80°C upon analysis. Cord blood genomic DNA (gDNA) was extracted from buffy coats using the QIAamp DNA Mini Kit (Qiagen, Inc., Venlo, the Netherlands). gDNA quantity and purity were assessed by a Nanodrop 1000 spectrophotometer (Isogen, Life Science, Belgium). gDNA passed the purity control test when the A260/280 was greater than 1.80 and A260/230 greater than 2.0. Samples were re-extracted when poor purity was detected. gDNA integrity was visually assessed and confirmed by agarose gel electrophoresis (Supplementary Figure S2). Total RNA was extracted from whole blood collected in Tempus tubes (ThermoFisher Scientific, Waltham, MA, USA) using the Tempus Spin RNA Isolation kit (Life Technologies, Paisley, UK) according to the manufacturer's instructions. RNA yields were evaluated using the Nanodrop 1000 spectrophotometer, and the quality of the samples was checked with an Agilent 2100 Bioanalyzer (Agilent Technologies, Amstelveen, the Netherlands). Samples with an RNA Integrity Number that was higher than 6 were stored at – 80°C until further processing.

DNA methylation measurement and data processing

Illumina HumanMethylationEPIC Bead-Chip arrays (EPIC, Illumina Inc., San Diego, USA) were used to measure DNA methylation of the EPIC discovery dataset at the GenomeScan laboratory (Leiden, The Netherlands). For the replication dataset, DNA methylation measurements were performed with the Infinium HumanMethylation450 BeadChip arrays (450K, Illumina Inc., San Diego, USA), at the International Agency for Research on Cancer (Lyon, France). Full details of sample preprocessing, measurements, and data processing, including quality control, are given in Supplementary Text S1 and S2. After the processing, the EPIC array yielded *M*-values for 787,264 CpG probes, while the 450K array yielded *M*-values for 459,781 CpG probes.

rDNA content assessment

rDNA content was measured using a multiplex droplet digital PCR method as described by Xu *et al.*, 2017 [18]. Primer-probe sequences were obtained directly after contacting the corresponding authors. First, a 20x primer-probe stock was prepared for the rDNA (45s rDNA, for exact genomic location of primers see Supplementary Figure S3) and the single-copy gene (*TBP*, TATA-box binding protein) assays using a primer:probe ratio of 2 (fluorescent labels and primer-probes were obtained from Bio-Rad, and sequences are provided in Supplementary Table S2). gDNA samples were normalized to ensure a uniform DNA input of 1 ng for each ddPCR reaction. This was verified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies, Europe). Reaction master mixtures (20 µl) contained 20x rDNA and *TBP* primer-probe stocks (1 µl), 2 units HaeIII restriction enzyme (Thermo-Fisher), 10 µl ddPCR™ supermix for probes (no dUTP) (Bio-Rad), and 1ng of template gDNA. Samples were first incubated for 15 minutes at room temperature, allowing DNA digestion. A total of 70 µl droplet generation oil for probes (Bio-Rad) was added, and droplets were generated using the QX200™ Droplet Generator (Bio-Rad). After droplet generation, a PCR reaction was carried out in 96-well semi-skirted ddPCR Plates (Bio-Rad) using the C100 Touch Thermal Cycler (Bio-Rad). The following cycling conditions were applied: 95°C for 10 min, followed by 40 cycles at 96°C for 30s and 57°C for 56s, followed by 98°C for 10 min. All samples were analysed in triplicate, and on each plate, a non-template control was added, and 2 inter-run calibrators were included to account for plate-to-plate variation. Finally, after PCR amplification, PCR plates containing droplets with amplified DNA were read using the QX200™ Droplet Reader and Quantasoft™ software (Bio-Rad). rDNAcn was calculated by dividing the mean of rDNA copies by the mean of *TBP* copies. The reliability of the assay is shown by an intra-assay intraclass correlation coefficient (ICC) of 0.97 (95%CI: 0.963–0.975) and an inter-assay ICC of 0.94 (95%CI: 0.891–0.987).

Micro-array gene-expression analysis

An aliquot of 0.2 µg total RNA was reverse-transcribed into cDNA, labelled with cyanine-3 following the Agilent one-colour Quick-Amp labelling protocol (Agilent Technologies), and hybridized onto Agilent Whole Human Genome 8 × 60 K microarrays. Microarray signals were detected using the Agilent DNA G2505C Microarray Scanner (Agilent Technologies). The Agilent Feature Extraction Software (Version 10.7.3.1, Agilent Technologies, Amstelveen, The Netherlands) was used to convert the scanned images into TXT files. After an in-house developed quality control pipeline (Supplementary Text S3) in R software, the final dataset contained 28,555 probes. Transcriptomic measurements were available for a subset of 64 out of the 194 participants comprising the EPIC discovery dataset. Those were included in the subsequent eQTM analysis.

Statistical analysis

All statistical analyses were performed using the R software version 4.5.0. The main statistical approach is applied to the EPIC discovery dataset. In the first step, an EWAS was performed using methylation at individual CpG sites (measured with Illumina EPIC Bead-Chip array) as an outcome and cord blood rDNAcn as the predictor. In the second step, results from the EWAS were used to identify differentially methylated regions (DMRs). In the third step, the methylation level of the CpGs identified in step 1 was correlated with gene expression data in an eQTM analysis. As a fourth step, genes annotated to the CpGs identified in step 1, DMRs identified in step 2, and correlated gene expression identified in step 3 were used as input for pathway overrepresentation analyses. Finally, the EWAS, DMR, and pathway analyses were replicated using the 450K replication dataset containing DNA methylation data measured with the Illumina 450K Bead-Chip array. An overview of the statistical approach is provided in Figure 1

Epigenome-wide association analysis of cord blood rDNAcn

Robust linear regression models (rlm function from the MASS package [30]) were used to associate individual CpG sites (M values, measured with Illumina EPIC Bead-Chip array) and cord blood rDNAcn of 194 newborns. Models were adjusted for *a priori* selected potential confounders known to be linked with

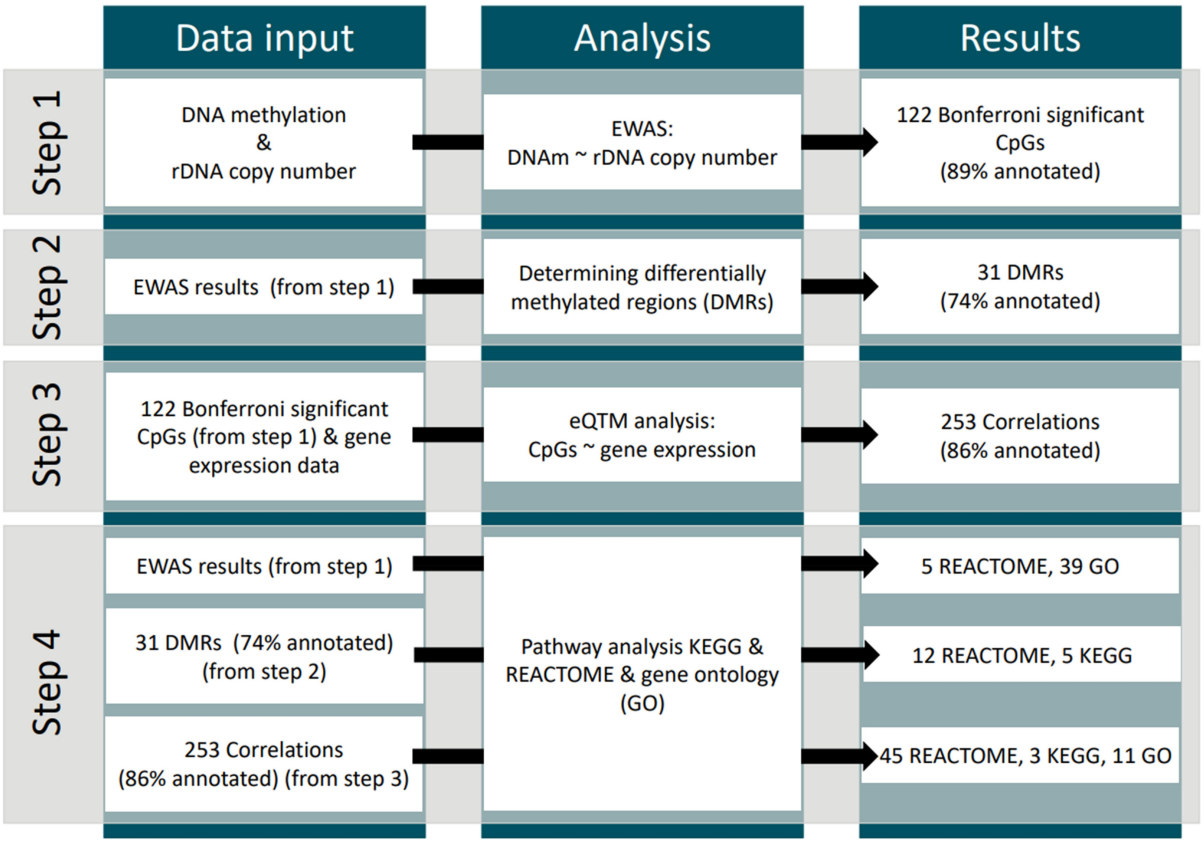

Figure 1. Overview of the statistical analyses and results. In the first step, an EWAS of cord blood rDNAcn was performed in an EPIC discovery dataset (measured with the Illumina EPIC Bead-Chip array). In a second step, the EWAS results were used in a differentially methylated region analysis. In step three, the methylation level of the Bonferroni-significant CpGs from the EWAS was correlated with gene-expression profiles. In a fourth step, genes annotated to the Bonferroni-significant CpGs (from step one), DMRs (from step two), and transcripts (from step three) were independently examined in a pathway overrepresentation analysis for further interpretability of the results. Finally, the EWAS, DMR, and pathway analyses were replicated using a 450K replication dataset (DNA methylation measured with the Illumina 450K Bead-Chip array).

cord blood DNA methylation, namely newborn sex, gestational age, birth weight, maternal age, pre-pregnancy BMI, estimated blood cell counts (CD8+ T cells, CD4+ T cells, natural killer cells, B cells, and monocytes), and the technical factor storage time. We applied Bonferroni correction for multiple testing and considered a 0.05 threshold for significance. Prior to EWAS, we tested the model's multicollinearity by calculating the variance inflation factor (VIF) with the R package car. In the initial model, including all aforementioned covariates and all estimated blood cell counts, granulocytes showed a high VIF of 18.8. Therefore, this variable was removed from the final model. Quality control of the EWAS results was performed by visual inspection of residual plots and scatter plots from the top results, as well as the calculation of p-value inflation (λ). Due to the small sample size, we used the R package bacon [31] to handle p-value inflation. A comparison of the quantile-quantile plots (QQ-plots) and the λ values of the original p-values with the bacon-corrected p-values (Supplementary Figure S4) showed no substantial improvement by the correction, which suggested minimal inflation. We therefore used the original p-values for further analysis.

Gene annotation was performed using IlluminaHuman-MethylationEPICanno.ilm10b4.hg19. Manhattan and volcano plots were created using ggplot2. Ideograms were created to show the genomic locations of the associated CpGs using karyoploteR [32]. To evaluate whether there is a chromosome-specific association between rDNAcn and DNA methylation, we calculated the proportion of significant CpGs per chromosome and, for each chromosome, tested whether this proportion is significantly different compared to the proportion in the rest of the genome, using a Fisher's Exact Test for count data.

To test the robustness of the EWAS findings, we used a minimally adjusted model (only adjusting for storage time and estimated blood cell counts). Additionally, model robustness was assessed in several sensitivity analyses by adjusting models for factors that have been well-established predictors of the cord blood methylome. First, we additionally adjusted for maternal smoking and maternal education, both factors associated with cord blood DNA methylation [33,34]. Second, we performed a restricted analysis using only European participants, as clear genome-wide methylation differences have been observed in populations from different ethnic backgrounds [35,36]. Finally, as rDNAcn is linked to haematological profiles [17], we also evaluated whether cord blood rDNAcn is associated with cord blood haematological data (platelets, neutrophils, lymphocytes, monocytes, eosinophils, basophils) using a subset of 150 participants, for whom measured blood cell type counts were available. Within this subgroup, we assessed whether our main results from the EPIC discovery analysis remained significant after replacing the estimated with the measured cell counts (neutrophils, lymphocytes, monocytes, eosinophils, basophils).

Differentially methylated region analysis

DMRs associated with the cord blood rDNAcn were identified using DMRcate [37] and ENmix-combp [38]. The analysis was performed with the results obtained in the EWAS. DMRcate was set to lambda = 500, C = 5. The package is powered by limma and smooths t-test statistics with a Gaussian kernel; *p*-values were adjusted for multiple testing using FDR correction. For combp, the maximal distance to combine adjacent DMRs was set to 750 base pairs. The bin size was 310, and the seed was 0.05. The package uses *p*-values from neighbouring CpG sites within a specific region; the default method of *p*-value adjustment for multiple testing is 1-step Siddak correction. The minimum number of CpGs constituting a DMR was two. Ideograms, for illustration of the results, were created.

Expression quantitative trait methylation (eQTM) analysis

To improve the biological interpretation of the rDNAcn-associated CpGs, we correlated the methylation level of the identified CpGs from the EWAS with available cord blood gene expression data in an eQTM analysis. For 64 participants, both methylation and micro-array gene-expression data were available. Pearson correlations were calculated between the methylation level of the Bonferroni-significant CpGs found in the EWAS and the expression levels of transcripts located on the same chromosome, with starting sites within 1Mb from the CpG locus. Partial correlations were calculated, adjusted for estimated blood cell counts (CD8+ T cells, CD4+ T cells, natural killer cells, B cells, and monocytes). *p*-values were not corrected for multiple testing because the statistical power was constrained by the limited sample size, and the eQTM analysis was conducted for exploratory purposes. Correlations having a nominal *p*-value below 0.05 were considered significant. Network plots showing the CpGs that are correlated with more than one transcript were created using the packages ggraph [39] and igraph [40].

Pathway overrepresentation analysis

A pathway overrepresentation analysis was performed using the Reactome, KEGG, and Gene Ontology (GO) term databases. Three separate pathway analyses using different inputs were performed: i) The results of the EWAS were examined with the package methylGSA to account for CpG number bias [41], ii). Genes annotated to the DMRs, and iii) genes annotated to transcripts that significantly correlated with the methylation level of the Bonferroni-significant CpGs in the eQTM analysis were analysed using ReactomePA and clusterProfiler [42,43]. The *p*-value threshold for the overrepresentation analyses was 0.05 (after FDR correction). Dotplots were used to show identified pathways having FDR-corrected *p*-values smaller than 0.05.

Additional analyses

(1) As a complementary analysis, we evaluated the association of global DNA, gene promoter, gene body, and intergenic methylation with rDNAcn using linear regression models. Global DNA

methylation was calculated as the study participants' arithmetic mean of the *M*-values ($n = 787{,}264$ probes). The same was done for promoter ($n = 180{,}705$ probes), gene body ($n = 332{,}964$ probes), and intergenic regions ($n = 84{,}958$ probes). Models were adjusted for the same covariates as in the EWAS.

(2) We evaluated whether previously identified NADs were over- or underrepresented in rDNAcn-associated methylation marks. The genomic locations of previously identified NADs were obtained from Bersaglieri *et al.*, 2022 [44]. An overlap of the Bonferroni-significant CpGs or the identified DMRs with NADs was assessed by i) calculating the percentage of significant CpGs and DMRs that overlap with the NADs and by ii) creating ideograms that show the positions of significant CpGs or DMRs and NADs. For the overlap between the CpGs and the NADs, we additionally performed a Fisher's exact test.

(3) The webtool eFORGE v2.0 [45,46]) was used to test for enrichment of Bonferroni-significant CpGs in DNase I hypersensitive sites (DHSs) across a range of tissues and cell types, using the default settings of the eFORGE web interface.

Replication of EWAS, DMR, and pathway overrepresentation analysis using the 450K replication dataset

The EWAS (main model), the DMR analysis, and the corresponding pathway analyses were replicated with the 450K replication dataset, containing DNA methylation data from 459,781 CpGs measured with the Illumina 450K Bead-Chip array. This replication analysis included data obtained from 167 newborns. Participants from the 450K replication dataset were independent of the EPIC discovery dataset.

Results

Study population

Comparative characteristics of the participants in the EPIC discovery dataset and the 450K replication dataset are presented in Table 1. The included newborns had an average (±SD) birth weight of 3394 ± 487 grams and a gestational age of 39.1 ± 1.6 weeks. Among them, 51.2% were boys, and most newborns were of European descent (92.2%). Non-European newborns were predominantly from Turkish (50%) and Moroccan (36%)

Table 1. Comparative characteristics of participants from the EPIC discovery dataset and the 450K replication dataset from the ENVIR*ON*AGE study.

Characteristic	EPIC discovery dataset ($n = 194$)	450K replication dataset ($n = 167$)	*P*-value
Newborn			
Gestational age, weeks	39.1 ± 1.6	39.1 ± 1.6	>0.99
Birth weight, grams	3386.5 ± 473.1	3403 ± 505.5	0.75
Sex			>0.99
Boys, n	99 (51%)	86 (51.5%)	
Girls, n	95 (49%)	81 (48.5%)	
Ethnicity			0.051
European, n	184 (94.8%)	149 (89.2%)	
Non-European, n	10 (5.2%)	18 (10.8%)	
rDNAcn	478.6 ± 108.4	450.2 ± 97.6	0.0088
Maternal			
Age, years	30.1 ± 4.2	29.4 ± 4.6	0.15
Pre-pregnancy BMI, kg/m^2	24 ± 3.9	24.6 ± 4.5	0.13
Education at delivery			0.02
Low, n	18 (9.3%)	25 (15%)	
Middle, n	53 (27.3%)	60 (35.9%)	
High, n	123 (63.4%)	82 (49.1%)	
Smoking			0.83
Never smoked, n	128 (66%)	105 (62.9%)	
Stopped smoking, n	41 (21.1%)	39 (23.3%)	
Continued smoking, n	25 (12.9%)	23 (13.8%)	

Values are reported as a mean ± standard deviation or as a number (%). Maternal education was categorized using the International Standard Classification of Education (for further information, see Education https://uis.unesco.org/en/topic/international-standard-classification-education-isced) and was low when no diploma was obtained, middle when a secondary diploma was obtained, and high when a college or university diploma was obtained. The reported *p*-values are from Fisher's exact test for the categorical variables and from the large sample Z-test for the continuous variables.

origin; only 14% were of other origins, including Russia, Algeria, and Jordan. Mothers were on average 29.8 ± 4.4 years and had a mean pre-pregnancy BMI of 24.3 ± 4.4 kg/m^2. However, 34.6% were overweight, of whom 10.5% were obese. Most mothers obtained a college or university diploma (56.8%), whereas 11.9% did not obtain a diploma. In total, 13.3% of the mothers continued smoking during pregnancy, 22.2% stopped at the time of pregnancy, and 64.5% never smoked. Newborn rDNAcn was on average 465.42 (range 208.68–857.84) and did not differ between boys and girls (468.4 *vs* 462.3; $p = 0.57$), nor between Europeans and non-Europeans (467.2 vs 444.5; $p = 0.14$). Compared to the EPIC discovery dataset, participants in the 450K replication dataset showed no differences in gestational age, birth weight, sex, maternal pre-pregnancy BMI, maternal age, and smoking status. However, the 450K replication dataset had a slightly higher proportion of non-European newborns (5.2% *vs* 10.8%; $p = 0.051$), a lower proportion of highly educated mothers (63.4% *vs* 49.1%; $p = 0.02$), and participants had a slightly lower cord blood rDNAcn (478.6 *vs* 450.2; $p = 0.0088$) (Table 1, Figure 2).

Cord blood rDNAcn epigenome-wide associations

In the EPIC discovery analysis, cord blood rDNAcn was associated with 122 Bonferroni-significant single CpG probes (120 positive and 2 negative associations). An overview of associated CpGs and annotated genes is provided in Supplementary Table S3, and results are visualized in a Manhattan (Figure 3A) and a volcano plot (Supplementary Figure S5A). The asymmetrical shape of the volcano plot, with a clear skew towards positive associations, reflects the predominance of CpGs showing a positive relationship with rDNAcn. Of the 122 CpGs, 109 were annotated to one or more genes, resulting in a set of 125 genes, with only two CpGs annotated to the same gene (Supplementary Table S3). The top 10 ranked CpGs based on *p*-values were annotated to the following genes: *GFI1* (growth factor independent 1 transcriptional repressor, $p_{adj} = 2.43^{-19}$), *USP46* (ubiquitin specific peptidase 46, $p_{adj} = 1.01^{-18}$), *ABHD14B* (abhydrolase domain containing 14B, $p_{adj} = 1.36^{-18}$), *ABHD14A* (abhydrolase domain containing 14A, $p_{adj} = 1.36^{-18}$), *CHL1* (cell adhesion molecule L1 like, $p_{adj} = 7.07^{-18}$), *CGREF1* (cell growth regulator with EF-hand domain 1, $p_{adj} = 7.43^{-18}$), *CRCP* (CGRP receptor component, $p_{adj} = 3.21^{-17}$), *CCNI* (cyclin I, $p_{adj} = 1.11^{-16}$), *AGBL3* (AGBL carboxypeptidase 3, $p_{adj} = 3.07^{-16}$), *EIF4G1* (eukaryotic translation initiation factor 4 gamma 1, $p_{adj} = 1.24^{-15}$) and *PROM1* (prominin 1, $p_{adj} = 2.50^{-15}$) (for scatter plots see Figure 4). We assessed whether the CpGs were positioned within promoter or gene body regions to gain insight into potential effects on gene expression. Among the top

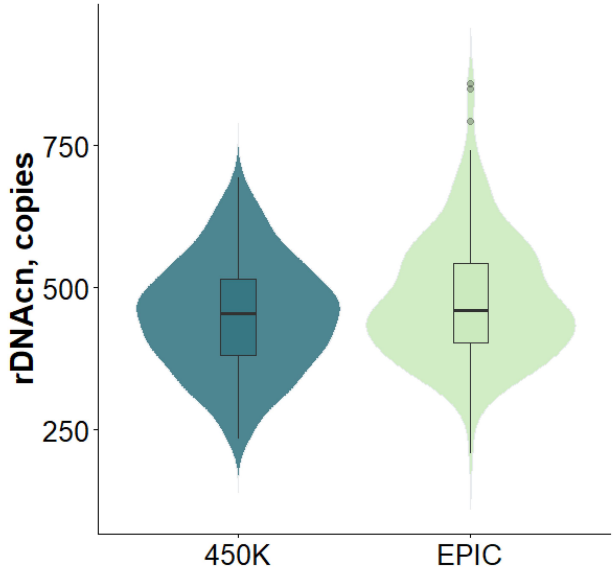

Figure 2. Violin plots of rDNAcn in the EPIC discovery and 450K replication dataset. The plot shows the distribution of rDNAcn in the EPIC discovery and the 450K replication dataset, with violin widths reflecting the kernel density of the data. Overlaid boxplots summarize the distribution: The central line indicates the median, and the lower and upper hinges correspond to the first and third quartiles (25th and 75th percentiles), respectively. Whiskers extend to the most extreme values within 1.5 times the interquartile range from the hinges, while observations beyond this range are shown as individual outliers.

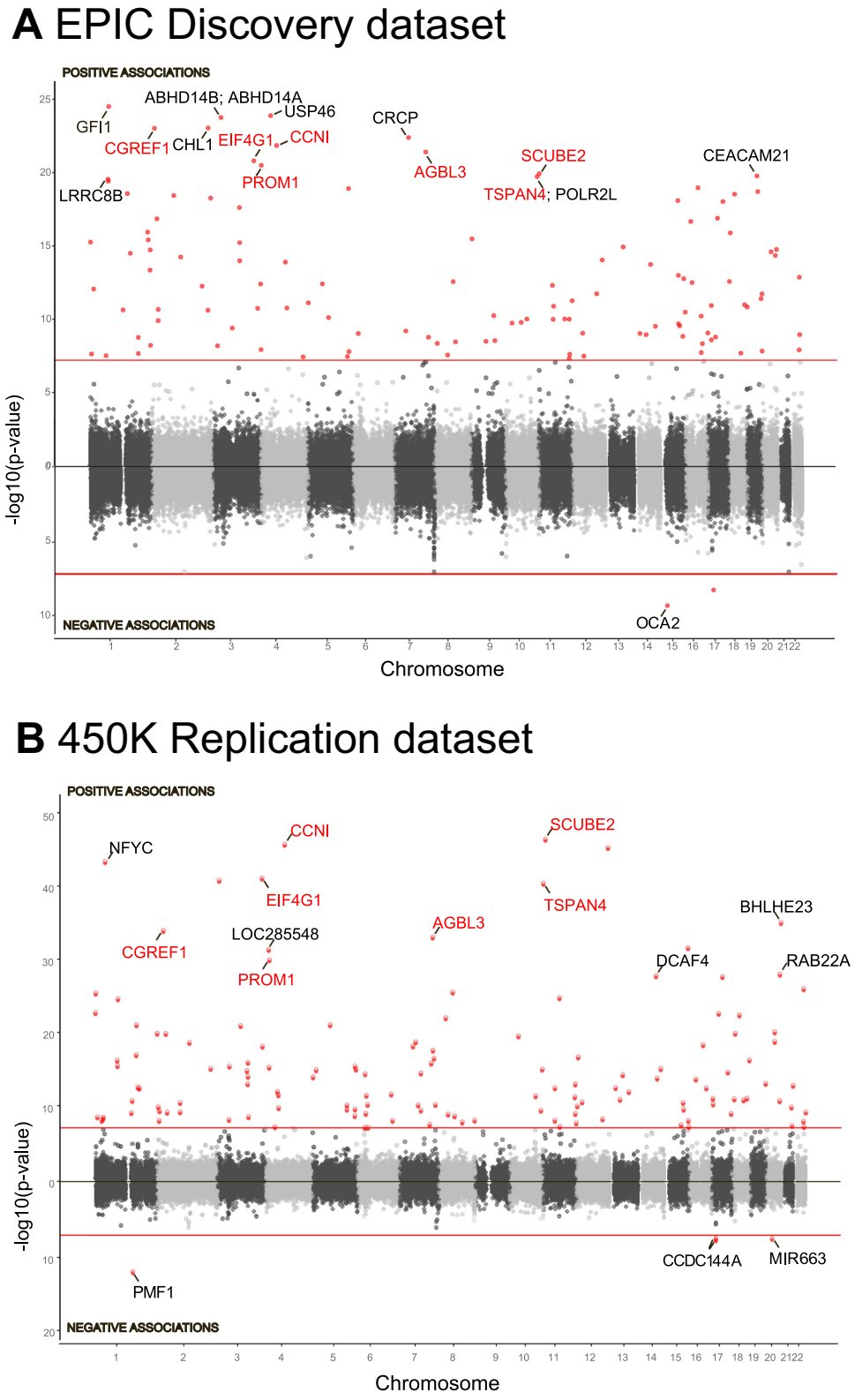

Figure 3. Manhattan plots showing the EWAS results of cord blood rDNAcn in the EPIC discovery dataset (A) and the 450K replication dataset (B). Manhattan plots show the -$\log_{10}$ transformed *p*-value on the y-axis of the associated CpGs across all chromosomes on the x-axis. CpGs with positive associations to rDNAcn are shown above the middle line, while CpGs with negative associations to rDNAcn are shown below. The top 15 positively and negatively associated Bonferroni-significant CpGs were annotated. The genes (within those top 15) that were found in the EPIC discovery dataset, as well as in the 450K replication dataset, are highlighted in red. The *p*-values were obtained from robust regression models [30] adjusted for newborn sex, gestational age, birth weight, maternal age, pre-pregnancy BMI, estimated blood cell counts (CD8+ T cells, CD4+ T cells, natural killer cells, B cells, and monocytes) and the technical factor storage time, using the EPIC discovery and the 450K replication dataset. The horizontal red line represents the Bonferroni level of significance.

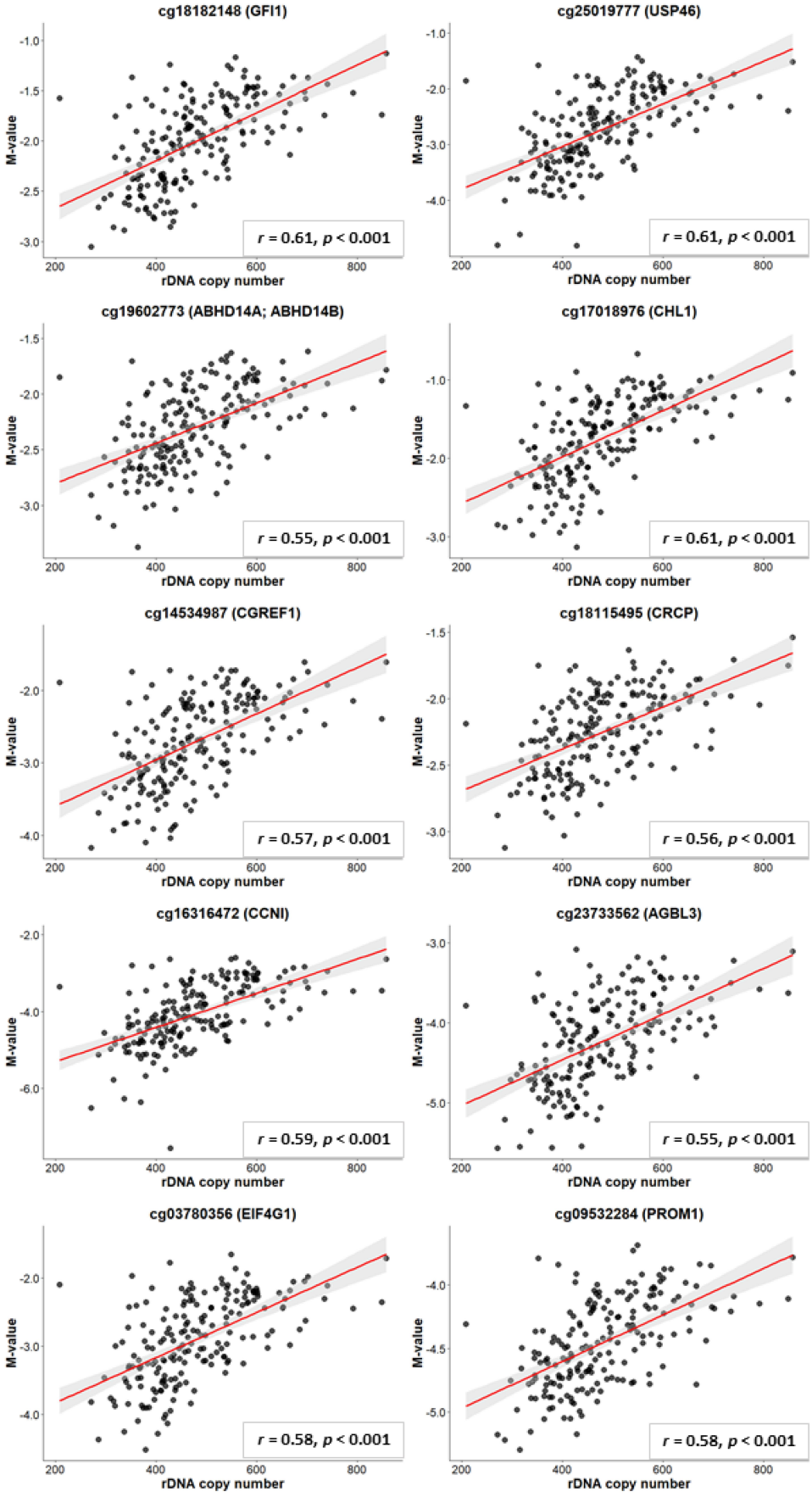

Figure 4. Scatter plots from the top 10 CpGs identified in the cord blood rDNAcn using the EPIC discovery dataset. Scatter plots show the rDNAcn on the x-axis and CpG methylation (*M*-values) on the y-axis. Linear regression line (red) and 95% confidence interval (grey shade) are shown. Pearson correlation coefficients (r) and corresponding *p*-values are displayed in the lower right corner of each plot.

10 ranked CpGs, those annotated to genes *GFI1*, *ABHD14A*, *CHL1, CGREF1, CCNI, AGBL3, EIF4G1*, and *PROM1* were located in the promoter region. For *USP46*, *ABHD14B*, and *CRCP*, the specific location of the methylation sites could not be mapped due to missing functional annotations for the corresponding genomic regions. Overall, among the 125 genes to which CpGs were annotated, 79 showed promoter methylation, while only 16 exhibited gene body methylation, and 30 genes could not be classified. The associated CpGs were dispersed throughout the autosomes, as shown in the ideogram (Figure 5). We examined whether the number of significant CpGs is equally distributed over the chromosomes. Although no significant CpGs were detected on chromosomes 6, 10, and 21, the ratio between the significant CpGs and the total number of CpGs was found to be significantly lower only on chromosome 6 ($p = 0.005$), as its overall CpG content is comparatively high (Figure 6). A QQ-plot of the *p*-values is shown in Supplementary Figure S3, with an inflation factor $\lambda = 0.97$. In our sensitivity analyses (minimally adjusted model, models additionally adjusted for maternal smoking or maternal education, and restricting the analysis to European participants), 108 out of the 122 Bonferroni-significant CpGs remained significant across all analyses (Supplementary Figure S6). rDNAcn was not associated with any of the measured cell-type proportions (see Supplementary Table S4). Moreover, all 122 CpG sites that were significant after Bonferroni correction in the original analysis remained significant when measured cell-type proportions were used in the model instead of estimated cell counts (see Supplementary Table S5).

Using the 450K replication dataset, we observed 96 out of the 122 Bonferroni-significant CpGs from the EPIC discovery analysis to be present on the 450K array. In the 450K replication analysis, all 96 were found significant based on the nominal *p*-value, and all were consistent in the direction of association as observed in the EPIC discovery dataset. After Bonferroni correction for 459,781 tests, in total 61 CpGs (63.5%) remained significant (for a comparison of the results, see Supplementary Table S6 and S7). Overall, when using the 450K replication dataset, we observed 118 Bonferroni-significant CpGs (114 positive and 4 negative associations) associated with rDNAcn at birth. An overview of these associated CpGs and annotated genes is provided in Supplementary Table S8. Results are visualized in a Manhattan (Figure 3B) and a volcano plot (Supplementary Figure S5B). The genomic location and overlap between the Bonferroni-significant CpGs from the EPIC discovery and 450K replication analysis are shown in an ideogram (Figure 7) and in a Venn diagram (Figure 8).

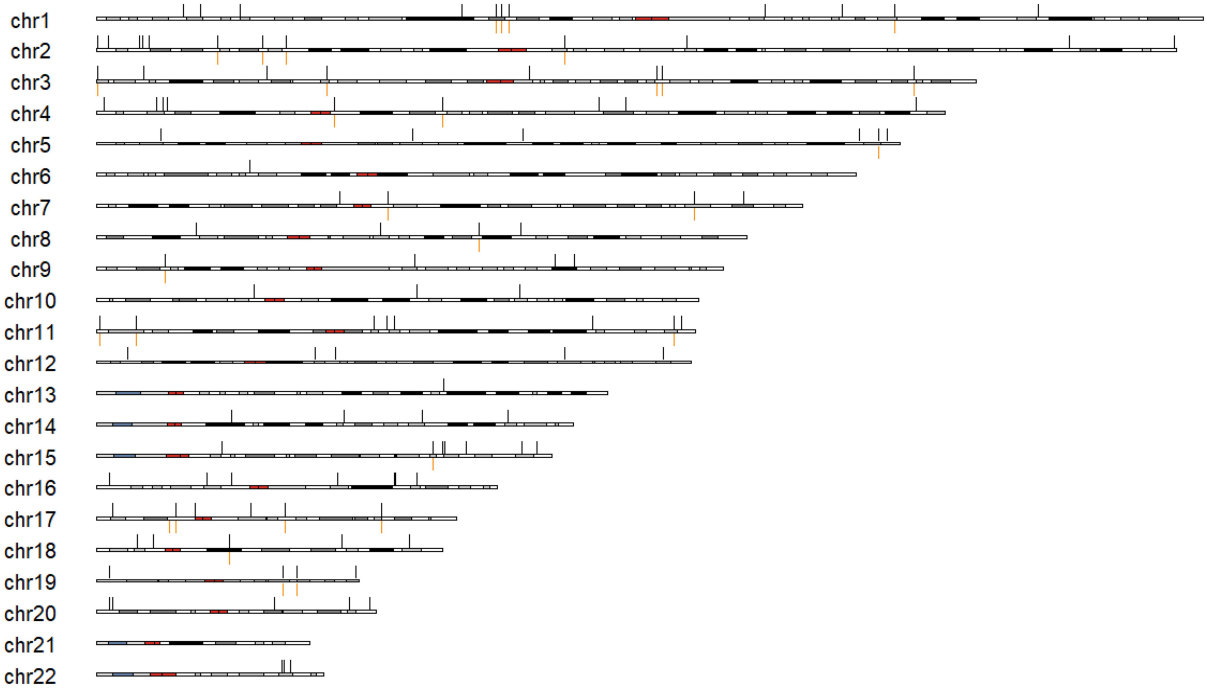

Figure 5. Ideogram of autosomes showing cord blood rDNAcn-associated CpGs and DMRs from the EPIC discovery analysis. EWAS-derived CpGs (black) and DMRs (orange) are indicated above and below each chromosome structure, respectively.

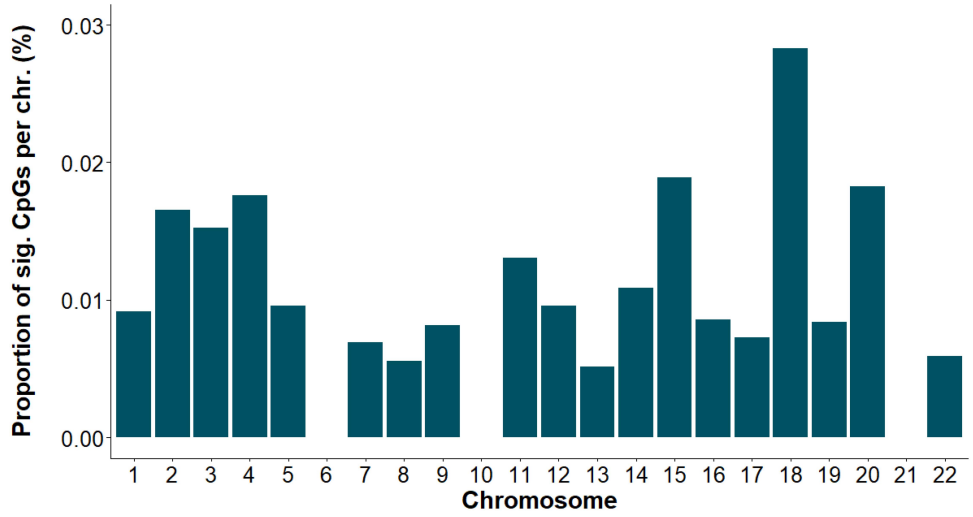

Figure 6. Bar plot showing the proportion of Bonferroni-significant CpGs per chromosome. The bar plot shows the proportion of Bonferroni-significant CpGs per autosome (y-axis), calculated as the number of significant CpGs on each chromosome divided by the total number of CpGs on that chromosome. Chromosome numbers are displayed along the x-axis.

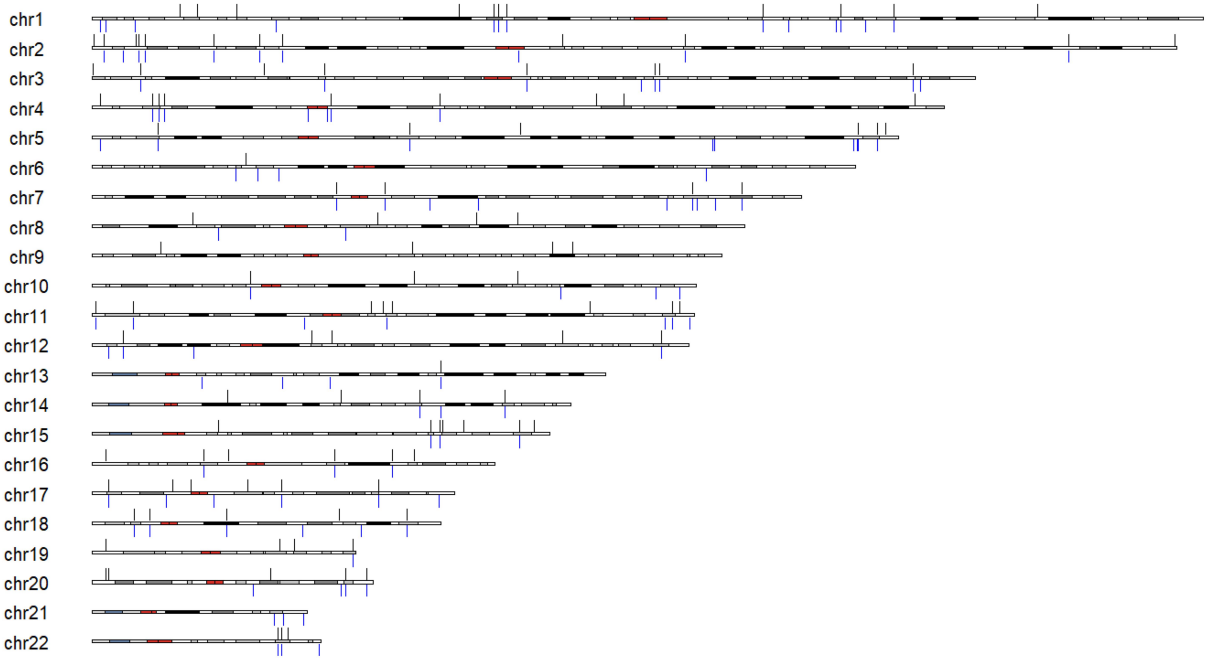

Figure 7. Ideogram of autosomes showing cord blood rDNAcn-associated CpGs from the EPIC discovery and 450K replication analysis. EWAS-derived CpGs from the EPIC discovery analysis (black) and from the 450K replication analysis (blue) are indicated above and below each chromosome structure, respectively.

Differentially methylated regions

In the EPIC discovery dataset, 31 FDR-significant DMRs were observed (Supplementary Table S9) in relation to cord blood rDNAcn when using DMRcate. When using combp, we did not identify any DMRs. Of the 31 identified DMRs, 23 were annotated, resulting in 30 genes overlapping with the DMRs. The 5 most significant DMRs are related to *USP46* (ubiquitin specific peptidase 46, $p_{adj} = 5.68^{-23}$), *ABHD14A* (abhydrolase domain containing 14A, $p_{adj} = 5.68^{-23}$), *ABHD14B* (abhydrolase domain containing 14B, $p_{adj} = 5.68^{-23}$), *RP11-155D18.14* (uncharacterized, $p_{adj} = 5.68^{-23}$), *RP11-155D18.12* (uncharacterized, $p_{adj} = 5.68^{-23}$), *CHL1* (cell adhesion molecule L1 like, $p_{adj} = 2.20^{-22}$), *CHL1-AS2* (CHL1 antisense RNA 2,

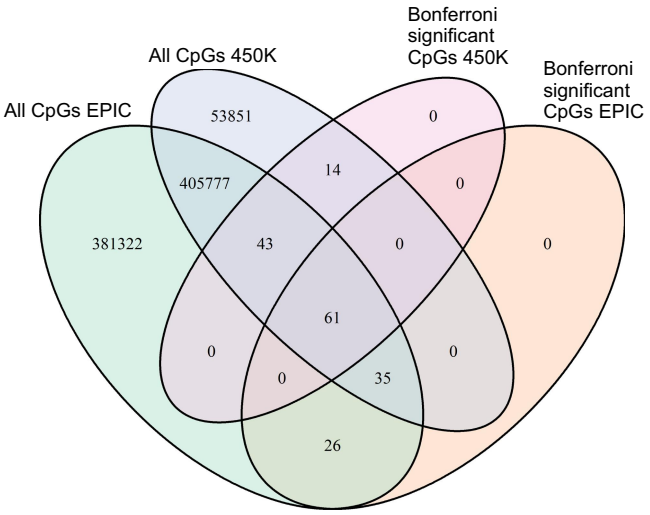

Figure 8. Overlap of CpGs from the EPIC discovery analysis and the 450K replication analysis. The Venn diagram shows the overlap of CpGs included in the EWAS analyses performed with the EPIC and 450K arrays, as well as the number of Bonferroni-significant CpGs observed in each dataset.

$p_{adj} = 2.20^{-22}$), *CGREF1* (cell growth regulator with EF-hand domain 1, $p_{adj} = 2.77^{-22}$). Those DMRs spanned between 2 and 8 CpGs. The locations of the identified DMRs are shown in the ideogram (Figure 5). From the 31 DMRs, only 1 is located far from a differentially methylated CpG. This is likely because the observed differential methylation was driven by strong, localized CpG-level effects rather than weak, spatially diffuse signals. Such a pattern favours detection by DMRcate over combp, which furthermore may also explain why no DMRs were identified using combp.

In the 450K replication dataset, a total of 24 DMRs (using DMRcate) were observed, of which 18 were annotated, resulting in 22 overlapping genes (Supplementary Table S10). 11 out of the 24 DMRs from the 450K replication dataset overlapped with the DMRs found in the EPIC discovery analysis (Figure 9).

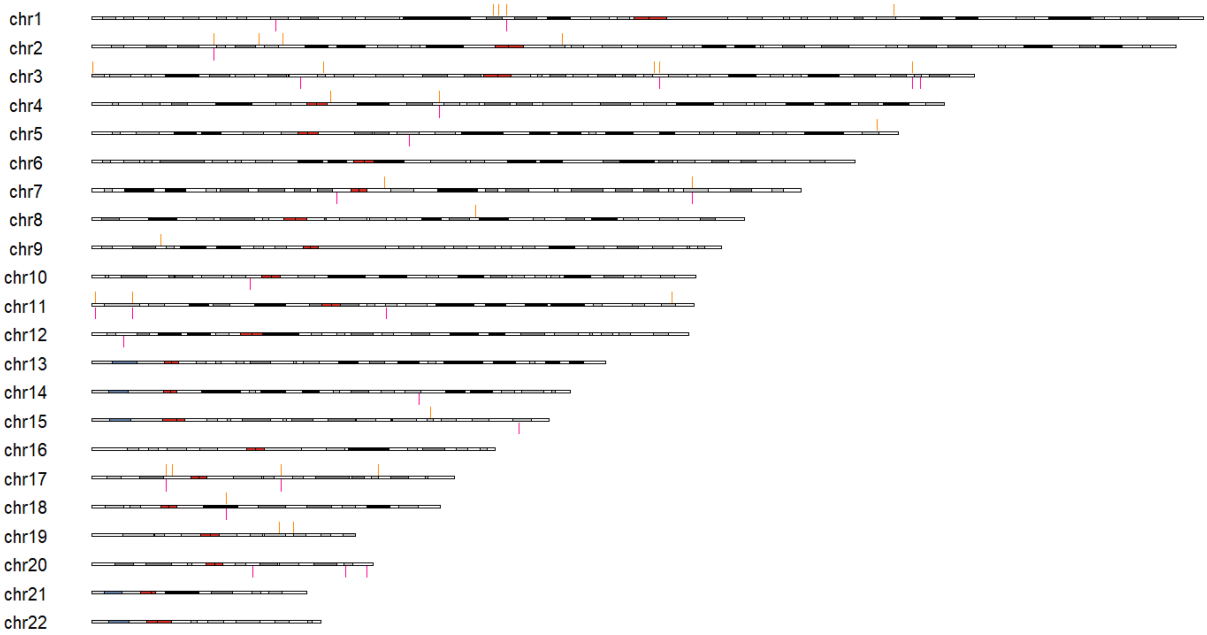

Figure 9. Ideogram of autosomes showing cord blood rDNAcn-associated DMRs from the EPIC discovery and 450K replication analysis. DMRs from the EPIC discovery analysis (orange) and DMRs from the 450K replication analysis (pink) are indicated above and below each chromosome structure, respectively.

Downstream analyses using eQTM analysis

64 newborns from the EPIC discovery dataset were available for an eQTM analysis. A total of 253 nominally significant correlations (73 positive and 180 negative) were observed between the methylation level of the EWAS-identified CpGs and gene expression (Supplementary Table S11). This resulted in 192 unique genes whose expression correlated with the methylation level of the significant CpGs. The top 5 negative correlations included: cg06297958 and *CENPT* (centromere protein T; $r = -0.434$, $p < 0.001$), cg18182148 and *GLMN* (glomulin, FKBP associated protein; $r = -0.427$, $p < 0.001$), cg01255913 and *LOC202181* (SUMO-interacting motifs containing 1 pseudogene; $r = -0.41$, $p < 0.001$), cg11817993 and *GPR68* (G protein-coupled receptor 68; $r = -0.4$, $p = 0.001$), cg09463466 and *LOC91450* (uncharacterized; $r = -0.399$, $p = 0.001$). The top 5 positive correlations included: cg02660524 and *PFKM* (phosphofructokinase, muscle; $r = 0.422$, $p < 0.001$), cg23733562 and *TMEM140* (transmembrane protein 140; $r = 0.381$, $p = 0.002$), cg06297958 and *PARD6A* (par-6 family cell polarity regulator alpha; $r = 0.354$, $p = 0.004$), cg02660524 and *TUBA1A* (tubulin alpha 1a; $r = 0.352$, $p = 0.004$), cg05696584 and *MRPS27* (mitochondrial ribosomal protein S27, $r = 0.348$, $p = 0.005$).

The 253 nominally significant correlations involved 239 unique transcript probes; 14 were associated with two CpGs each. In contrast, the 253 nominally significant correlations included 89 unique CpGs. From those, one CpG (cg00231422) was associated with 11 transcript probes, two CpGs (cg02660524, cg27138088) were associated with the expression of 10 transcript probes each, one CpG (cg17681277) was associated with 8 transcript probes, one CpG (cg06297958) was associated with 7 transcript probes, 4 CpGs (cg01255913, cg03727280, cg23806621, cg25371950) with 6 transcript probes each, and 7 CpGs (cg04228104, cg05035456, cg07049977, cg11817993, cg21855135, cg25552435, cg26228123) with 5 transcript probes each. Collectively, these 16 CpGs accounted for 105 of the 253 nominally significant correlations (Figure 10 and Supplementary Figure S7).

Pathway overrepresentation analysis

The EWAS results from the EPIC discovery analysis revealed 44 overrepresented pathways. 'Cytosolic sulfonation of small molecules' (R-HSA-156584) and 'RNA Polymerase III Transcription' (R-HSA-76071, R-HSA-76046, R-HSA-74158, R-HSA-749476) were found within the Reactome database. The GO terms show, among others, pathways related to regulation of translation (GO:0032056, GO:0008190, GO:0036490), kidney development (GO:0072160, GO:0061004, GO:0072048, GO:0072070), and regulation of vitamin metabolism (GO:0030656, GO:0042368, GO:0042362, GO:0009110) (for full details, see Supplementary Table S12). When using the 25 genes annotated to the DMRs, a total of 13 overrepresented pathways were identified, of which 11 were significant ($p_{adj} < 0.05$), and 2 remained only borderline significant ($p_{adj} = 0.06$) after FDR correction. Those pathways were mainly connected to 'RNA Polymerase III Transcription' (R-HSA-76071, R-HSA-76046, R-HSA-74158, R-HSA-749476, R-HSA-73780, R-HSA-73980, R-HSA-76066, R-HSA-76061). Additionally, an overrepresentation of 'Cytosolic sensors of pathogen-associated DNA' (R-HSA-1834949), 'RNA polymerase II transcribes snRNA genes' (R-HSA-6807505), 'RNA polymerase' (hsa03020), 'Eukaryotic Translation Initiation' (R-HSA-72613), and 'Cap-dependent Translation Initiation' (R-HSA-72737) was observed (for full details, see Supplementary Table S13). Figure 11 shows the overrepresented pathways (Reactome and KEGG) from both analyses (EWAS and DMRs).

Running the pathway overrepresentation analysis using the EWAS results from the 450K replication dataset, we were able to partly reproduce the findings from the EPIC discovery analysis (see Supplementary Table S14). Using the 22 genes annotated to the DMRs from the 450K replication dataset supported all pathways identified in the genes annotated to the DMRs from the EPIC discovery analysis (for full details, see Supplementary Table S15).

Finally, a pathway overrepresentation analysis conducted using all correlated gene transcripts from the eQTM analysis revealed 59 overrepresented pathways (Supplementary Table S16). These pathways were mainly related to hormone and peptide regulation, including 'Positive regulation of insulin secretion' (GO:0032024) or 'Regulation of peptide transport' (GO:0090087). Furthermore, we found pathways connected to intracellular transport, e.g., 'Kinesins' (R-HSA-983189) and 'Intra-Golgi and

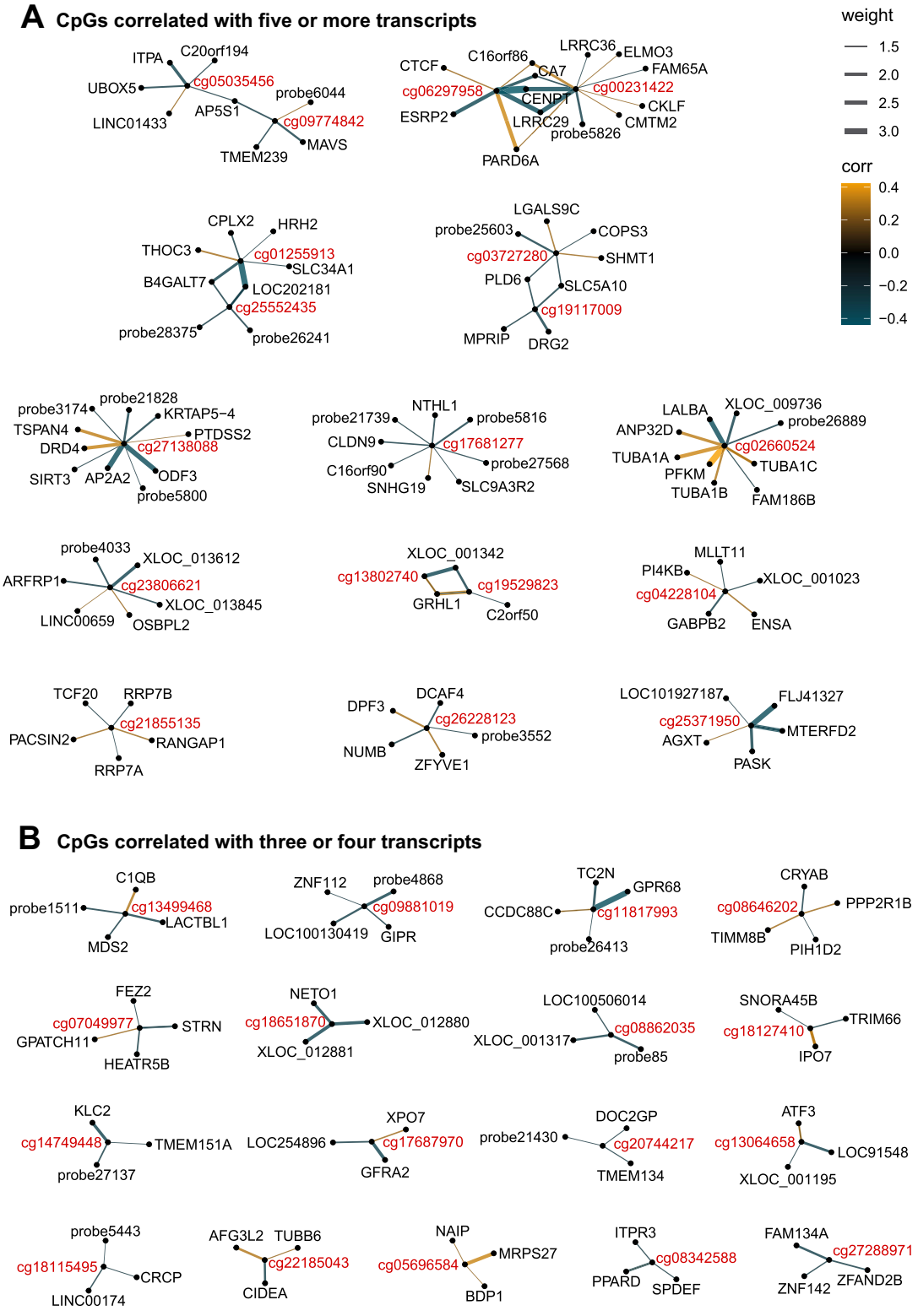

Figure 10. Network plot from the eQTM analysis showing the CpGs correlated with five or more gene transcripts (A) and CpGs correlated with three or four gene transcripts (B). The network plot visualizes the correlations from the eQTM analysis between the methylation level of the Bonferroni-significant CpGs found in the EWAS and the expression levels of transcripts (Pearson correlations adjusted for estimated blood cell counts (CD8+ T cells, CD4+ T cells, natural killer cells, B cells, and monocytes)). The colour of the lines encodes the correlation coefficient, whereas line thickness corresponds to −log10 (*p*-values), such that thicker lines indicate stronger statistical significance. Only CpGs correlated with three or more gene transcripts (probe ID shown if there is no gene symbol available for the probe) are shown in this plot. The network plot of CpGs correlated with two gene transcripts can be found in Supplementary Figure S7.

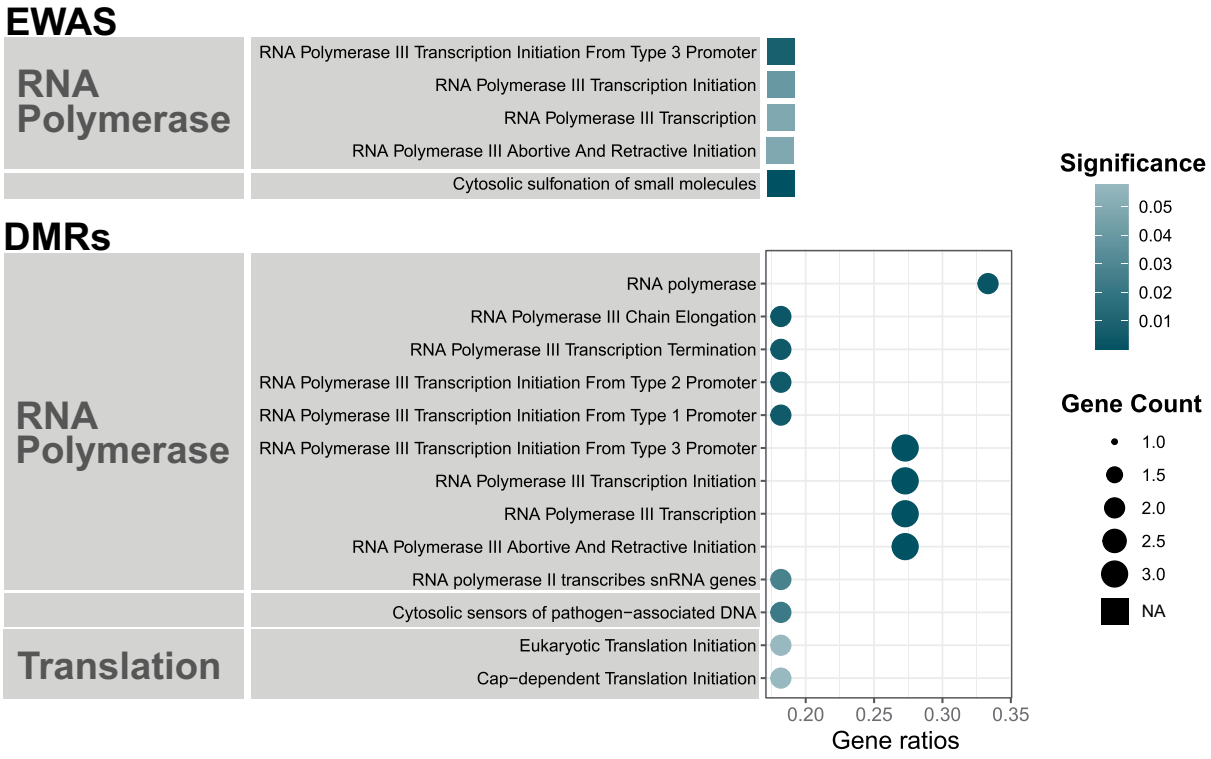

Figure 11. Pathway analysis from the EWAS results and from annotated genes to cord blood rDNAcn-associated DMRs from the EPIC discovery analysis. The dot size shows the number of genes (count) appearing in a pathway. The gene ratio (ratio of input genes that map to this pathway) is represented on the x-axis. Significance is indicated by FDR-corrected *p*-values. The databases used are the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome. The overrepresentation analysis from the EWAS results was conducted with the package methylGSA, and therefore, gene counts and gene ratios were not available.

retrograde Golgi-to-ER traffic' (R-HSA-6811442), pathways connected to mitosis and 'Separation of Sister Chromatids' (R-HSA-2467813), as well as signalling pathways such as 'Signaling by Hedgehog' (R-HSA-5358351) and 'PKR-mediated signaling' (R-HSA-9833482). All those pathways were not identified when using CpG and DMR annotated genes. Overall, the majority of pathways are driven by the genes *TUBA1A*, *TUBA1B*, *TUBA1C*, and *TUBB6*, all of which encode tubulins (Supplementary Table S16).

Global, promoter, gene body, and intergenic DNA methylation and rDNAcn

Linear regression analysis did not reveal an association between global, promoter, gene body, or intergenic DNA methylation and rDNAcn (all $p > 0.19$) within the EPIC discovery dataset (Supplementary Table S17).

Overlap of NADs with rDNAcn-associated CpGs and eForge analysis

Previously reported genomic locations of NADs showed that there was a non-significant overlap of 4.9% of the Bonferroni-significant CpGs, and a 6.5% overlap with the DMRs (for ideograms, see Supplementary Figures S8 and S9). Using the web-based tool eFORGE, we identified 32 tissues and cell types in which the Bonferroni-significant CpGs were enriched in DNase I hypersensitive sites. Of these, 15 corresponded to fetal or embryonic stem cells. The most significant enrichment was observed in fetal kidney DNase I hypersensitive sites (full results see Supplementary Table S18).

Discussion

To date, studies have evaluated methylation on the rDNA array, revealing associations of rDNA methylation patterns with cancer and ageing in humans [14,47,48]. However, no study has examined the interplay between rDNAcn and genome-wide DNA methylation. We performed an EWAS of cord blood rDNAcn in two independent newborn populations. Using the EPIC discovery dataset, we found 122 significant single CpGs related to rDNAcn, of which 98.4% were positively associated. Furthermore, 31 DMRs were identified. In downstream analyses, we could relate the methylation level of the identified CpGs to cord blood gene expression and observed 253 nominally significant correlations with 177 annotated genes. Pathway analyses revealed a consistent enrichment of RNA Polymerase III Transcription pathways even when different gene input approaches were used. Using the 450K replication dataset in an EWAS, DMR, and downstream pathway enrichment analyses, we were able to find consistent evidence for those findings.

A relation between genome-wide DNA methylation and rDNAcn in cord blood has not been described before, and it is currently unknown whether methylation affects rDNAcn, rDNAcn affects methylation, or if there is a third factor that influences rDNAcn and DNA methylation at the same time, but in an independent manner. In our study, we aimed to highlight possible biological responses to differences in rDNAcn. Therefore, we made the following assumption: rDNAcn acts as the predictor variable, and DNA methylation was the outcome variable. However, we can only speculate if a higher rDNAcn leads to differences in the methylation degree of the identified CpGs or *vice versa*. Overall, currently, mechanistic evidence that directly connects rDNAcn and DNA methylation is missing. Nevertheless, both potential dependencies as well as the possibility of effects related to a third confounder are discussed.

First, the following observations might indicate a role of rDNAcn on DNA methylation. In a study by Gibbons *et al.*, 2014 [11], it was shown that a strong association exists between gene expression and rDNAcn, and they speculate that differential recruitment of chromatin-modifying proteins and epigenetic changes could be the mechanism behind their observations [11]. Furthermore, they argue that variations in rDNAcn alter the required amounts of chromatin regulators and DNA-binding proteins needed for rRNA transcription. This might result in concentration changes of chromatin-modifying proteins throughout the rest of the genome, which affects genome-wide chromatin environments and transcription rates. In addition to that, it is known that rDNA clusters may play a role in global epigenetic gene regulation [49].

Second, based on some studies, we can speculate on how changes in DNA methylation might impact the rDNAcn. DNA methylation is known to affect gene expression. Therefore, regulatory genes of rDNAcn that are controlled by methylation can serve as a link between DNA methylation and rDNAcn. Currently, the mechanistic underpinnings of rDNA maintenance in humans remain largely unexplored, and knowledge in other multicellular organisms is limited. However, evidence suggests that the genes *mus-101* and *mei-41*, both involved in the homologous recombination repair, are required for germline rDNA maintenance and the rapid expansion of rDNA copies in *Drosophila* [50]. Consequently, the expression of *mus-101* or *mei-41* homologs and other genes involved in homologous recombination repair could potentially influence rDNAcn in humans.

Third, the considerable overlap in functionalities between rDNAcn and DNA methylation may suggest the presence of a third factor that influences rDNAcn and DNA methylation at the same time, but in an independent manner. However, a review of the current literature does not provide clear evidence for this. Although several attributes (including genomic instability, cancer, or body mass [16,18,19,51–53]) have been associated with rDNAcn and DNA methylation, the directionality of these associations often remains unclear, or those attributes are known to be the consequence rather than the cause of rDNAcn- or DNA methylation changes. Therefore, to the best of our knowledge, no common factors have been identified yet that directly cause changes in both rDNAcn and DNA methylation. Nonetheless, higher-order biological processes, including environmental exposures or developmental processes, could induce alterations in rDNAcn as well as DNA methylation, which have not yet been shown.

Our results indicate that a higher rDNAcn is associated with higher DNA methylation (120 out of the 122 identified CpGs showed a higher methylation with increasing rDNAcn). This is in line with recent studies that showed strong positive correlations between rDNAcn and rDNA methylation in multiple human tissues and in mice [54,55]. Out of these 122 CpGs, 79 CpGs are found in the promoter regions of the

annotated genes, which mostly leads to a reduction in gene expression. Therefore, for the genes identified in the EWAS, our results indicate a trend of reduced expression with increasing rDNAcn in cord blood.

The top significant CpGs from the EPIC discovery analysis were related to the genes *GFI1* (growth factor independent 1 transcriptional repressor), *USP46* (ubiquitin specific peptidase 46), *ABHD14A* (abhydrolase domain containing 14A), *ABHD14B* (abhydrolase domain containing 14B), *CHL1* (cell adhesion molecule L1 like) and *CGREF1* (cell growth regulator with EF-hand domain 1). *GFI1* preserves haematopoietic stem cell integrity [56], plays a crucial role in haematopoiesis [57], and its expression is associated with acute myeloid leukaemia [58]. *USP46* is part of a large family of cysteine proteases that generate free ubiquitin and cleave ubiquitin from ubiquitylated substrates [59]. The gene plays a role in early embryogenesis [60] and has been connected to cancer, with some studies reporting suppression of progression [61] while others indicate it promotes metastasis [62]. *ABHD14B* was identified as a lysine deacetylase in mammals [63] and is potentially involved in metabolic activities in cancer cells [64]. *ABHD14A*, however, remains functionally uncharacterized. The gene *CHL1* is mainly known for its connection to human cancers. It is involved in breast cancer tumorigenesis and progression [65] and acts as a tumour suppressor in neuroblastoma [66]. Furthermore, it plays an important role in the development of the nervous system [67]. *CGREF1* is connected to cell proliferation [68] and cell growth [69]. It is associated with lipogenesis [70] and its transcript has been identified as a prognostic marker for osteosarcoma [71].

Taken together, the top genes we identified as associated with rDNAcn in newborns share a common functional link to (cancer) cell proliferation. This forms a bridge to rDNA, as studies have shown a strong connection between rDNA and cell proliferation. In mammalian cells, rDNA transcription fluctuates significantly throughout cell cycle progression [72], and it is regulated by cellular growth factors, nutrient availability, and the energy household of the cell [73]. At the same time, changes in ribosome biogenesis can hinder cell proliferation [74]. The 45S rDNAcn is negatively correlated with cell proliferation, while 5S rDNAcn is positively correlated with it [75]. When evaluating the functions of our top genes, it appeared conspicuous that several of them are associated with the mTOR pathway [61,76–78]. This signalling network coordinates eukaryotic cell growth and is responsive to growth factors, amino acids, cellular energy levels, oxygen availability, and stress. mTOR signalling is suggested to play an important role in regulating fetal growth by adjusting the growth trajectory of the foetus to the maternal supply of oxygen, nutrients, and growth factors [79]. Furthermore, mTOR has been shown to influence rDNA transcription [73] and may also be connected to rDNAcn [18]. Therefore, we speculate that mTOR could be the mechanistic connection between our top genes and rDNAcn. Besides the role in proliferation, most top genes are also functionally connected to cell growth and development, which might be in line with observations in this very early stage of life. It may also implicate potential long-term effects, as alterations in the development of the neuronal and immune systems are likely to be relevant later in life.

The most significant CpG overall was cg01804284. It emerged as the top hit in the 450K replication dataset; however, this probe is not present on the EPIC array. It is annotated to *NFYC* (nuclear transcription factor Y subunit gamma), a gene known to be connected to developmental processes [80,81] and cancer [82,83], which aligns with the findings reported above.

Despite the relatively large number of highly significant CpGs, we did not find any associations between the rDNAcn and the mean genomic CpG methylation level. Therefore, we currently have no evidence that rDNAcn does exert a global influence on overall genomic CpG methylation levels, but our results might indicate more localized or gene-specific effects.

Our downstream pathway analyses were performed in three different ways. First, the results of the EWAS were used as an input, showing enrichment of 'Cytosolic sulfonation of small molecules,' RNA Polymerase III Transcription, as well as pathways related to regulation of translation, kidney development, and regulation of vitamin metabolism. Second, we performed a pathway analysis using the annotated genes of the significant CpGs and identified DMRs as input. This revealed a strong connection to RNA Polymerase III, 'Cytosolic sensors of pathogen-associated DNA,' 'RNA polymerase II transcribes snRNA genes,' 'RNA polymerase,' 'Eukaryotic Translation Initiation,' and 'Cap-dependent Translation Initiation.' Third, we performed a pathway analysis using the genes whose expression was correlated with the methylation level of the CpGs, and we observed an overrepresentation in pathways related to hormone and peptide regulation, intracellular transport, and mitosis. Although the key findings from our pathway overrepresentation

analyses have not been reported in previous studies, several of the identified pathways can be corroborated by the literature. A study conducted in 2014 by Gibbons and his colleagues directly investigated the correlations between rDNAcn and gene expression in adult humans. They reported 15 GO terms that were positively associated with rDNA dosage and 20 GO terms with negative correlations [11]. A higher rDNAcn was connected to a higher gene expression of genes associated with the nucleolus, ribosomes, spliceosomal complex, RNA metabolism, and response to virus, while a lower gene expression was mainly found for genes localized at the endoplasmatic reticulum [11]. Our findings overlap to some extent with those found by Gibbons. The pathway analysis conducted with the DMRs from the EWAS revealed a connection to RNA metabolism. Our third pathway analysis using genes from the eQTM analysis showed an enrichment of pathways connected to intracellular protein transport, as in line with Gibbons *et al.*, 2014 [11]. Besides that, we identified pathways that are related to kidney development. Consistent with this finding, a recent study reported an association between rDNAcn and the estimated glomerular filtration rate, which is known to be a marker for kidney function [17]. The pathway analysis using genes from the eQTM analysis showed an enrichment of pathways connected to mitosis. As previously discussed, there is a well-established link between rDNA and cell proliferation. The specific connection between rDNAcn and mitosis, described by Ide *et al.*, 2010 [84], underlines the importance of rDNAcn for replicative functioning. Their study in yeast indicated that the separation of the sister chromatids is more sensitive to damage once the rDNAcn is decreased. They conclude that additional rDNA repeats promote condensin association and sister-chromatid cohesion, which enhances recombinational repair.

We are unaware of any direct association with rDNAcn for the remaining findings. Therefore, in the following, we speculate on possible links between some of those pathways and the RNA machinery and highlight some mechanistic interplays.

The most prominent result of the pathway analyses was RNA Polymerase III transcription activity. RNA Polymerase III mostly transcribes tRNAs and 5S rRNAs, which are integral components of the ribosomes. In this study, however, we assessed the copy number of 45S rDNA, which is transcribed by RNA Polymerase I rather than RNA polymerase III. Therefore, it is important to point out that RNA polymerase III is also required for the repair of DNA double-strand breaks [85]. These occur significantly more often in rDNA compared to the rest of the genome [86]. Based on this, we speculate that an increased 45S rDNAcn may be associated with a higher incidence of double-strand breaks, thereby increasing the demand for RNA polymerase III – mediated repair processes. Such a mechanism could indirectly lead to altered RNA Polymerase III transcription activity in people with higher 45S rDNAcn. However, direct experimental evidence supporting this hypothesis is currently lacking.

Furthermore, there was an overrepresentation of 'RNA polymerase II transcribes snRNA genes.' Besides small nuclear RNAs, which play an important role in splicing, RNA polymerase II transcribes messenger RNA and microRNA. In humans, there is no direct link between rDNAcn and RNA polymerase II. However, it was recently shown that RNA polymerase II plays a role in rDNAcn expansion in *Drosophila* [87].

Several results from the overrepresentation analysis seem to be connected to the RNA machinery via translation initiation. The (Cap-dependent) translation initiation pathway was overrepresented. It is described as the assembly of the elongation-competent 80S ribosomes, where the initiation codon is base-paired with the anticodon loop of the initiator tRNA in the ribosome (for further information, see [88]).

'Cytosolic sensors of pathogen-associated DNA' are known to be different receptors capable of initiating diverse immune responses upon the detection of pathogen-derived molecules or DNA from invading organisms [89]. Chromosomal-derived cytoplasmic DNA can trigger DNA-sensing receptors as well [90]. As rDNA sequences are enriched within the cytoplasmic DNA [91], they may induce pro-inflammatory responses [92].

It should be noted that most pathways identified based on the eQTM analysis are related to the *TUBA1A*, *TUBA1B*, *TUBA1C*, and *TUBB6* genes, all of which encode tubulins. Tubulins constitute the primary structural components of microtubules, which are essential elements of the cytoskeleton, playing a critical role in maintaining cellular architecture and facilitating intracellular transport.

Also, it is important to point out that the results obtained from the first and second pathway analysis are considered to provide more robust and biologically meaningful results. The third analysis, based on genes identified through the eQTM analysis, was intended to be exploratory and used a subset of the data with less stringent statistical criteria. Its primary goal was to validate findings and generate directions for future research. As such, the pathways derived from the eQTM analysis should be interpreted with caution.

Strengths and limitations

To our knowledge, this is the first study showing an association between rDNAcn and genome-wide DNA methylation in the human genome. In the main analysis, we used the highly reliable EPIC array [93], which covers a wide range of genome regions, and a robust and sensitive ddPCR technique for rDNAcn measurement. The observed associations remained consistent in a comprehensive series of tests, including several adjustment sets and sensitivity analyses. In addition, we replicated the results using a fully independent dataset, which contained methylation data measured with the 450K array.

Nevertheless, it is important to point out some limitations of the study. First, as discussed above, we cannot pursue the direction of the dependencies. Second, the limited sample size in the eQTM analysis represents a constraint, as it may limit the robustness of the conclusions and warrants caution in the interpretation of the results. Third, a recent study reported maternal cell contamination of approximately 4% in umbilical cord blood samples [94]. Given this low level, we do not expect it to substantially affect our results. Nevertheless, we note it as a limitation, as it could potentially influence our analyses. Finally, rDNAcn measurements obtained by ddPCR do not capture sequence variation within rDNA arrays; given emerging evidence that such variation may be biologically relevant, this represents an additional limitation of the present study.

Conclusions

In this study, we showed an association between rDNAcn variation and genome-wide DNA methylation at birth. An increase in cord blood rDNAcn was related to higher DNA methylation, and we identified 122 CpGs to be significantly associated with the number of rDNA repeats at birth. Like rDNA, the top genes annotated to the identified CpGs can functionally be linked to (cancer) cell proliferation, cell growth, and development. A pathway overrepresentation analysis uncovered potentially relevant biological pathways and, most intriguingly, a possible connection of the association to RNA Polymerase III transcription. Finally, we showed here that rDNAcn may serve as a key genomic regulator in early life, which highlights the need for further studies to explore the functional role of rDNAcn variation and its associated epigenomic effects in development, health, and disease.

Author contributions

CRediT: **Kathrin Barth:** Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Rossella Alfano:** Data curation, Formal analysis, Methodology, Software, Writing – review & editing; **Michelle Plusquin:** Writing – review & editing; **Congrong Wang:** Data curation, Formal analysis, Methodology, Software, Writing – review & editing; **Tim S. Nawrot:** Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing; **Dries S. Martens:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Tim Nawrot, who holds funding from Methusalem. Dries Martens, who holds a postdoctoral grant from the Research Foundation Flanders [FWO grant 12X9623N], and Rossella Alfano, who holds a postdoctoral grant from the Research Foundation Flanders [FWO grant 1296523N].

ORCID

Kathrin Barth http://orcid.org/0009-0008-3499-8303
Rossella Alfano http://orcid.org/0000-0001-9374-4727
Michelle Plusquin http://orcid.org/0000-0002-7271-9722
Congrong Wang http://orcid.org/0000-0002-2488-4918
Tim S. Nawrot http://orcid.org/0000-0002-3583-3593
Dries S. Martens http://orcid.org/0000-0001-7893-3642

Data availability statement

The raw data underlying the findings can be accessed online at: https://figshare.com/s/2a14b192a72ce42ec47c. However, personal data is subject to ethical restrictions, but is available from the corresponding author upon request and after approval by the legal entity of the local university committees.

Contributors

TSN coordinates together with MP and DSM the ENVIRONAGE birth cohort. DSM designed the research hypotheses together with KB. DSM performed the measurements of the rDNAcn. KB performed the statistical analyses, guided by CW, RA, and DSM. KB and DSM prepared the first draft of the manuscript. KB, DSM, CW, and RA were involved in data interpretation. All authors contributed to the critical revision of the manuscript. They read and approved the final manuscript.

References

[1] Hall AN, Turner TN, Queitsch C. Thousands of high - quality sequencing samples fail to show meaningful correlation between 5S and 45S ribosomal DNA arrays in humans. Sci Rep. 2021;11(1):1–12. doi: 10.1038/s41598-020-80049-y
[2] Sanij E, Poortinga G, Sharkey K, et al. UBF levels determine the number of active ribosomal RNA genes in mammals. J Cell Biol. 2008;183(7):1259–1274. doi: 10.1083/jcb.200805146
[3] Warner JR. The economics of ribosome biosynthesis in yeast. Trends Biochem Sci. 1999;24(11):437–440. doi: 10.1016/S0968-0004(99)01460-7
[4] Thomson E, Ferreira-Cerca S, Hurt E. Eukaryotic ribosome biogenesis at a glance. J Cell Sci. 2013;126(21):4815–4821. doi: 10.1242/jcs.111948
[5] Németh A, Conesa A, Santoyo-Lopez J, et al. Initial genomics of the human nucleolus. PLOS Genet. 2010;6(3):e1000889. doi: 10.1371/journal.pgen.1000889
[6] Van Koningsbruggen S, Gierliński M, Schofield P, et al. High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. Mol Biol Cell. 2010;21(21):3735–3748. doi: 10.1091/mbc.E10-06-0508
[7] Caburet S, Conti C, Schurra C, et al. Human ribosomal RNA gene arrays display a broad range of palindromic structures. Genome Res. 2005;15(8):1079–1085. doi: 10.1101/gr.3970105
[8] Kim JH, Dilthey AT, Nagaraja R, et al. Variation in human chromosome 21 ribosomal RNA genes characterized by TAR cloning and long-read sequencing. Nucleic Acids Res. 2018;46(13):6712–6725. doi: 10.1093/nar/gky442
[9] Kobayashi T. Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. Cell Mol Life Sci. 2011;68(8):1395–1403. doi: 10.1007/s00018-010-0613-2
[10] Hori Y, Engel C, Kobayashi T. Regulation of ribosomal RNA gene copy number, transcription and nucleolus organization in eukaryotes. Nat Rev Mol Cell Biol. 2023;24(6):414–429. doi: 10.1038/s41580-022-00573-9
[11] Gibbons JG, Branco AT, Yu S, et al. Ribosomal DNA copy number is coupled with gene expression variation and mitochondrial abundance in humans. Nat Commun. 2014;5(1):5. doi: 10.1038/ncomms5850
[12] Stults DM, Killen MW, Pierce HH, et al. Genomic architecture and inheritance of human ribosomal RNA gene clusters. Genome Res. 2008;18(1):13–18. doi: 10.1101/gr.6858507
[13] Malinovskaya EM, Ershova ES, Golimbet VE, et al. Copy number of human ribosomal genes with aging: unchanged mean, but narrowed range and decreased variance in elderly group. Front Genet. 2018;9(AUG):1–11. doi: 10.3389/fgene.2018.00306
[14] Shao F, Liu X, Zhang X, et al. Methylation of 45S ribosomal DNA (rDNA) is associated with cancer and aging in humans. Int J Genomics. 2021;2021:1–9. doi: 10.1155/2021/8818007
[15] Porokhovnik LN, Lyapunova NA. Dosage effects of human ribosomal genes (rDNA) in health and disease. Chromosome Res. 2019;27(1–2):5–17. doi: 10.1007/s10577-018-9587-y
[16] Law PP, Mikheeva LA, Rodriguez-Algarra F, et al. Ribosomal DNA copy number is associated with body mass in humans and other mammals. Nat Commun. 2024;15(1):15(1. doi: 10.1038/s41467-024-49397-5

[17] Rodriguez-Algarra F, Evans DM, Rakyan VK. Ribosomal DNA copy number variation associates with hematological profiles and renal function in the UK Biobank. Cell Genomics. 2024;4(6):100562. doi: 10.1016/j.xgen.2024.100562
[18] Xu B, Li H, Perry JM, et al. Ribosomal DNA copy number loss and sequence variation in cancer. PLOS Genet. 2017;13(6):1–25. doi: 10.1371/journal.pgen.1006771
[19] Salim D, Gerton JL. Ribosomal DNA instability and genome adaptability. Chromosome Res. 2019;27(1–2):73–-87. doi: 10.1007/s10577-018-9599-7
[20] Potapova TA, Gerton JL. Ribosomal DNA and the nucleolus in the context of genome organization. Chromosome Res. 2019;27(1–2):109–127. doi: 10.1007/s10577-018-9600-5
[21] Kobayashi T. A new role of the rDNA and nucleolus in the nucleus—rDNA instability maintains genome integrity. Bioessays. 2008;30(3):267–272. doi: 10.1002/bies.20723
[22] Foss EJ, Lao U, Dalrymple E, et al. SIR2 suppresses replication gaps and genome instability by balancing replication between repetitive and unique sequences. Proc Natl Acad Sci USA. 2017;114(3):552–557. doi: 10.1073/pnas.1614781114
[23] Tate PH, Bird AP. Effects of DNA methylation on DNA-binding proteins and gene expression. Curr Opin Genet Dev. 1993;3(2):226–231. doi: 10.1016/0959-437X(93)90027-M
[24] Jones PA. The DNA methylation paradox. Trends Genet. 1999;15(1):34–37. doi: 10.1016/S0168-9525(98)01636-9
[25] López-Otín C, Blasco MA, Partridge L, et al. Hallmarks of aging: an expanding universe. Cell. 2023;186(2):243–278. doi: 10.1016/j.cell.2022.11.001
[26] Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022;12(1):31–46. doi: 10.1158/2159-8290.CD-21-1059
[27] Zhou D, Robertson KD. Role of DNA methylation in genome stability. In: Kovalchuk I, Kovalchuk O, editors. Genome stability: from virus to human application. Lethbridge (AB): Elsevier Inc; 2016. p. 409–424. doi: 10.1016/B978-0-12-803309-8.00024-0
[28] Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013;14(3):204–220. doi: 10.1038/nrg3354
[29] Janssen BG, Madhloum N, Gyselaers W, et al. Cohort profile: the ENVIRonmental influence on early AGEing (ENVIRONAGE): a birth cohort study. Int J Epidemiol. 2017;46(5):1386–1387M. doi: 10.1093/ije/dyx033
[30] Venables WN, Ripley BD. Modern applied statistics with S. Fourth Edi ed. (NY): Springer; 2002.
[31] van Iterson M, van Zwet EW, Heijmans BT, et al. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. Genome Biol. 2017;18(1):1–13. doi: 10.1186/s13059-016-1131-9
[32] Gel B, Serra E. KaryoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics. 2017;33(19):3088–3090. doi: 10.1093/bioinformatics/btx346
[33] Alfano R, Guida F, Galobardes B, et al. Socioeconomic position during pregnancy and DNA methylation signatures at three stages across early life: Epigenome-wide association studies in the ALSPAC birth cohort. Int J Epidemiol. 2019;48(1):30–44. doi: 10.1093/ije/dyy259
[34] Joubert BR, Felix JF, Yousefi P, et al. DNA methylation in newborns and maternal smoking in pregnancy: genome-wide consortium meta-analysis. Am J Hum Genet. 2016;98(4):680–696. doi: 10.1016/j.ajhg.2016.02.019
[35] Adkins RM, Krushkal J, Tylavsky FA, et al. Racial differences in gene-specific DNA methylation levels are present at birth. Birth Defects Res Part Clin Mol Teratol. 2011;91(8):728–736. doi: 10.1002/bdra.20770
[36] Li B, Aouizerat BE, Cheng Y, et al. Incorporating local ancestry improves identification of ancestry-associated methylation signatures and meQtls in African Americans. Commun Biol. 2022;5(1). doi: 10.1038/s42003-022-03353-5
[37] Peters TJ, Buckley MJ, Chen Y, et al. Calling differentially methylated regions from whole genome bisulphite sequencing with DMRcate. Nucleic Acids Res. 2021;49(19):e109–e109. doi: 10.1093/nar/gkab637
[38] Xu Z, Xie C, Taylor JA, et al. IpDMR: identification of differentially methylated regions with interval P-values. Bioinformatics. 2021;37(5):711–713. doi: 10.1093/bioinformatics/btaa732
[39] Pedersen TL. Ggraph: an implementation of grammar of graphics for graphs and networks. 2025. doi: 10.32614/CRAN.package.ggraph
[40] Csárdi G, Nepusz T, Traag V, et al. Igraph: network analysis and visualization in R. 2025. doi: 10.5281/zenodo.7682609
[41] Ren X, Kuan PF. Genome analysis methylGSA: a Bioconductor package and shiny app for DNA methylation data length bias adjustment in gene set testing. Bioinformatics. 2019;35(11):1958–1959. doi: 10.1093/bioinformatics/bty892
[42] Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innov. 2021;2(3):100141. doi: 10.1016/j.xinn.2021.100141
[43] Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst. 2016;12(2):477–479. doi: 10.1039/C5MB00663E
[44] Bersaglieri C, Kresoja-Rakic J, Gupta S, et al. Genome-wide maps of nucleolus interactions reveal distinct layers of repressive chromatin domains. Nat Commun. 2022;13(1):1483. doi: 10.1038/s41467-022-29146-2

[45] Breeze CE, Paul DS, Van Dongen J, et al. eFORGE : a tool for identifying cell type-specific signal in epigenomic data resource eFORGE : a tool for identifying cell type-specific signal in epigenomic data. Cell Reports. 2016;17 (8):2137–50. doi: 10.1016/j.celrep.2016.10.059
[46] Breeze CE, Reynolds AP, Van Dongen J, et al. Genome analysis eFORGE v2. 0: updated analysis of cell type-specific signal in epigenomic data. Bioinformatics. 2019;35(22):4767–4769. doi: 10.1093/bioinformatics/btz456
[47] Wang M, Lemos B. Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. Genome Res. 2019;29(3):325–333. doi: 10.1101/gr.241745.118
[48] Yang F, Guo X, Bao Y, et al. The role of ribosomal DNA methylation in embryonic development, aging and diseases. Epigenet Chromatin. 2024;17(23):1–11.
[49] Tchurikov NA, Kravatsky YV. The role of rDNA clusters in global epigenetic gene regulation. Front Genet. 2021;12(August):1–11. doi: 10.3389/fgene.2021.730633
[50] Lu KL, Nelson JO, Watase GJ, et al. Transgenerational dynamics of rDNA copy number in drosophila male germline stem cells. Elife. 2018;7:1–20. doi: 10.7554/eLife.32421
[51] Chen L, Ganz PA, Sehl ME. DNA methylation, aging, and cancer risk: a mini-review. Front Bioinform. 2022;2 (June):1–7. doi: 10.3389/fbinf.2022.847629
[52] Dick KJ, Nelson CP, Tsaprouni L, et al. DNA methylation and body-mass index: a genome-wide analysis. Lancet. 2014;383(9933):1990–1998. doi: 10.1016/S0140-6736(13)62674-4
[53] Rizwana R, Hahn PJ. CpG methylation reduces genomic instability. J Cell Sci. 1999;112(24):4513–4519. doi: 10.1242/JCS.112.24.4513
[54] Razzaq A, Bejaoui Y, Alam T, et al. Ribosomal DNA copy number variation is coupled with DNA methylation changes at the 45S rDNA locus. Epigenetics. 2023;18(1). doi: 10.1080/15592294.2023.2229203
[55] Rodriguez-Algarra F, Seaborne RAE, Danson AF, et al. Genetic variation at mouse and human ribosomal DNA influences associated epigenetic states. Genome Biol. 2022;23(1):1–17. doi: 10.1186/s13059-022-02617-x
[56] Zeng H, Yücel R, Kosan C, et al. Transcription factor Gfi1 regulates self-renewal and engraftment of hematopoietic stem cells. Embo J. 2004;23(20):4116–4125. doi: 10.1038/sj.emboj.7600419
[57] Van Der Meer LT, Jansen JH, Van Der Reijden BA. Gfi1 and Gfi1b: key regulators of hematopoiesis. Leukemia. 2010;24(11):1834–1843. doi: 10.1038/leu.2010.195
[58] Hönes JM, Thivakaran A, Botezatu L, et al. Enforced GFI1 expression impedes human and murine leukemic cell growth. Sci Rep. 2017;7(1):1–13. doi: 10.1038/s41598-017-15866-9
[59] Quesada V, Díaz-Perales A, Gutiérrez-Fernández A, et al. Cloning and enzymatic analysis of 22 novel human ubiquitin-specific proteases. Biochem Biophys Res Commun. 2004;314(1):54–62. doi: 10.1016/j.bbrc.2003.12.050
[60] Joo HY, Jones A, Yang C, et al. Regulation of histone H2A and H2B deubiquitination and xenopus development by USP12 and USP46. J Biol Chem. 2011;286(9):7190–7201. doi: 10.1074/jbc.M110.158311
[61] Li X, Stevens PD, Yang H, et al. The deubiquitination enzyme USP46 functions as a tumor suppressor by controlling PHLPP-dependent attenuation of akt signaling in colon cancer. Oncogene. 2013;32(4):471–478. doi: 10.1038/onc.2012.66
[62] Tian M, Zhu R, Ding F, et al. Ubiquitin-specific peptidase 46 promotes tumor metastasis through stabilizing ENO1 in human esophageal squamous cell carcinoma. Exp Cell Res. 2020;395(1):112188. doi: 10.1016/j.yexcr.2020.112188
[63] Rajendran A, Vaidya K, Mendoza J, et al. Functional annotation of ABHD14B, an orphan serine hydrolase enzyme. Biochemistry. 2020;59(2):183–196. doi: 10.1021/acs.biochem.9b00703
[64] Namani A, Cui QQ, Wu Y, et al. NRF2-regulated metabolic gene signature as a prognostic biomarker in non-small cell lung cancer. Oncotarget. 2017;8(41):69847–69862. doi: 10.18632/oncotarget.19349
[65] He LH, Ma Q, Shi YH, et al. CHL1 is involved in human breast tumorigenesis and progression. Biochem Biophys Res Commun. 2013;438(2):433–438. doi: 10.1016/j.bbrc.2013.07.093
[66] Ognibene M, Pagnan G, Marimpietri D, et al. CHL1 gene acts as a tumor suppressor in human neuroblastoma. Oncotarget. 2018;9(40):25903–25921. doi: 10.18632/oncotarget.25403
[67] Katic J, Loers G, Kleene R, et al. Interaction of the cell adhesion molecule CHL1 with vitronectin, integrins, and the plasminogen activator inhibitor-2 promotes CHL1-induced neurite outgrowth and neuronal migration. J Neurosci. 2014;34(44):14606–14623. doi: 10.1523/JNEUROSCI.3280-13.2014
[68] Deng W, Wang L, Xiong Y, et al. The novel secretory protein CGREF1 inhibits the activation of AP-1 transcriptional activity and cell proliferation. Int J Biochem Cell Biol. 2015;65:32–39. doi: 10.1016/j.biocel.2015.05.019
[69] Madden SL, Galella EA, Riley D, et al. Induction of cell growth regulatory genes by p53. Cancer Res. 1996;56 (23):5384–5390.
[70] Chan P, Cheung PHH, Kang XZ, et al. Cgref1 is a CREB-H-regulated hepatokine that promotes hepatic de novo lipogenesis by mediating epididymal fat insulin resistance. Int J Biol Sci. 2025;21(6):2568–2588. doi: 10.7150/ijbs.97008
[71] Zhihao Z, Cheng J, Xiaoshuang Z, et al. Cancer-associated fibroblast infiltration in osteosarcoma: the discrepancy in subtypes pathways and immunosuppression. Front Pharmacol. 2023;14(June):1–13. doi: 10.3389/fphar.2023.1136960

[72] Grummt I. Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. Genes Dev. 2003;17(14):1691–1702. doi: 10.1101/gad.1098503R
[73] Kusnadi EP, Hannan KM, Hicks RJ, et al. Regulation of rDNA transcription in response to growth factors, nutrients and energy. Gene. 2015;556(1):27–34. doi: 10.1016/j.gene.2014.11.010
[74] Donati G, Montanaro L, Derenzini M. Ribosome biogenesis and control of cell proliferation: p53 is not alone. Cancer Res. 2012;72(7):1602–1607. doi: 10.1158/0008-5472.CAN-11-3992
[75] Wang M, Lemos B. Ribosomal DNA copy number amplification and loss in human cancers is linked to tumor genetic context, nucleolus activity, and proliferation. PLOS Genet. 2017;13(9):1–24. doi: 10.1371/journal.pgen.1006994
[76] Huang Z, Gao Y, Sun Y, et al. NB-3 signaling mediates the cross-talk between post-traumatic spinal axons and scar-forming cells. Embo J. 2016;35(16):1745–1765. doi: 10.15252/embj.201593460
[77] Kurebayashi Y, Nagai S, Ikejiri A, et al. PI3K-Akt-mTORC1-S6K1/2 axis controls Th17 differentiation by regulating Gfi1 expression and nuclear translocation of RORγ. Cell Rep. 2012;1(4):360–373. doi: 10.1016/j.celrep.2012.02.007
[78] You H, Zhang H, Jin X, et al. Dysregulation of ubiquitination modification in renal cell carcinoma. Front Genet. 2024;15(December):1–18. doi: 10.3389/fgene.2024.1453191
[79] Gupta MB, Jansson T. Novel roles of mechanistic target of rapamycin signaling in regulating fetal growth†. Biol Reprod. 2019;100(4):872–884. doi: 10.1093/biolre/ioy249
[80] Nylén C, Aoi W, Abdelmoez AM, et al. IL6 and LIF mRNA expression in skeletal muscle is regulated by AMPK and the transcription factors NFYC, ZBTB14, and SP1. Am J Physiol-Endocrinol Metab. 2025;4(5):995–1004. doi: 10.1152/ajpendo.00398.2017
[81] Wu D, Liang Z, Li Z, et al. Transcription factor-mediated gene regulatory networks in the formation of oocytes. Reprod Fertil Dev. 2025;37(12). doi: 10.1071/RD25054
[82] Takenoyama M, Baurain J-F, Yasuda M. A point mutation in the NFYC gene generates an antigenic peptide recognized by autologous cytolytic T lymphocytes on a human squamous cell lung carcinoma. Int J Cancer. 2006;118(8):1992–1997. doi: 10.1002/ijc.21594
[83] Tong Y, Merino D, Nimmervoll B, et al. Cross-species Genomics identifies TAF12, NFYC, and RAD54L as choroid plexus carcinoma oncogenes. Cancer Cell. 2015;27(5):712–727. doi: 10.1016/j.ccell.2015.04.005
[84] Ide S, Miyazaki T, Maki H, et al. Abundance of ribosomal RNA gene copies maintains genome integrity. Science. 2010;327(5966):693–696. doi: 10.1126/science.1179044
[85] Liu S, Hua Y, Wang J, et al. RNA polymerase III is required for the repair of DNA double-strand breaks by homologous recombination. Cell. 2021;184(5):1314–1329.e10. doi: 10.1016/j.cell.2021.01.048
[86] Tchurikov NA, Fedoseeva DM, Sosin DV, et al. Hot spots of DNA double-strand breaks and genomic contacts of human rDNA units are involved in epigenetic regulation. J Mol Cell Biol. 2015;7(4):366–382. doi: 10.1093/jmcb/mju038
[87] Watase GJ, Yamashita YM. RNA polymerase II-mediated rDNA transcription mediates rDNA copy number expansion in Drosophila. PLOS Genet. 2024;20(5 May):1–21. doi: 10.1371/journal.pgen.1011136
[88] Jackson RJ, Hellen CUT, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol. 2010;11(2):113–127. doi: 10.1038/nrm2838
[89] Barber GN. Cytoplasmic DNA innate immune pathways. Immunol Rev. 2011;243(1):99–108. doi: 10.1111/j.1600-065X.2011.01051.x
[90] Hu MM, Shu HB. Innate immune response to cytoplasmic DNA: mechanisms and diseases. Annu Rev Immunol. 2020;38(1):79–98. doi: 10.1146/annurev-immunol-070119-115052
[91] Cheng J, Torkamani A, Peng Y, et al. Plasma membrane associated transcription of cytoplasmic DNA. Proc Natl Acad Sci USA. 2012;109(27):10827–10831. doi: 10.1073/pnas.1208716109
[92] Malinovskaya EM, Ershova ES, Okorokova NA, et al. Ribosomal DNA as DAMPs signal for MCF7 cancer cells. Front Oncol. 2019;9(MAY). doi: 10.3389/fonc.2019.00445
[93] Pidsley R, Zotenko E, Peters TJ, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. 2016;17(1):1–17. doi: 10.1186/s13059-016-1066-1
[94] Smeekens SP, Leferink M, Yntema HG, et al. Maternal cell contamination in postnatal umbilical cord blood samples implies a low risk for genetic misdiagnoses. Prenat Diagn. 2024;44(11):1304–1309. doi: 10.1002/pd.6595