

Methylation status of *VTRNA2-1/nc886* is stable across human populations, monozygotic twin pairs and in majority of somatic tissues

Saara Marttila^{1,2}, Hely Tamminen¹, Sonja Rajić¹, Pashupati P Mishra^{3,4,5}, Terho Lehtimäki^{3,4,5}, Olli Raitakari^{6,7,8}, Mika Kähönen^{4,9}, Laura Kananen^{10,11,12}, Juulia Jylhävä^{11,12}, Sara Hägg¹¹, Thomas Delerue¹³, Annette Peters^{14,15}, Melanie Waldenberger^{13,15}, Marcus E Kleber^{16,17}, Winfried März^{16,18,19,20}, Riitta Luoto^{21,22}, Jani Raitanen^{22,23}, Elina Sillanpää^{24,25}, Eija K Laakkonen²⁴, Aino Heikkinen²⁵, Miina Ollikainen²⁵, Emma Raitoharju^{1,4}

¹Molecular Epidemiology, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

²Gerontology Research Center, Tampere University, Finland

³Department of Clinical Chemistry, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

⁴Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

⁵Fimlab Laboratories, Tampere, Finland

⁶Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland

⁷Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland

⁸Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland

⁹Department of Clinical Physiology, Tampere University Hospital, Tampere Finland

¹⁰Faculty of Medicine and Health Technology, and Gerontology Research Center, Tampere University, Tampere, Finland

¹¹Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

¹²Faculty of Social Sciences (Health Sciences), and Gerontology Research Center, Tampere University, Tampere, Finland

¹³Research Unit Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, D-85764, Neuherberg, Bavaria, Germany

¹⁴Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, D-85764, Neuherberg, Bavaria, Germany

¹⁵German Center for Cardiovascular Research (DZHK), Partner Site Munich Heart Alliance, Munich, Germany

¹⁶Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim 68167, Germany

¹⁷SYNLAB MVZ Humangenetik Mannheim, Mannheim, Germany

¹⁸Competence Cluster for Nutrition and Cardiovascular Health (nutriCARD) Halle-Jena-Leipzig, Jena 07743, Germany

¹⁹SYNLAB Academy, SYNLAB Holding Deutschland GmbH, Augsburg 86156, Germany

²⁰Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz 8010, Austria

²¹The Social Insurance Institute of Finland (Kela), Helsinki, Finland

²²The UKK Institute for Health Promotion Research, Tampere, Finland,

²³Faculty of Social Sciences (Health Sciences), Tampere University, Tampere, Finland

²⁴Gerontology Research Center and Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

²⁵Institute for Molecular Medicine Finland, FIMM, HiLIFE, University of Helsinki, Helsinki, Finland

Abstract

Aims and methods: Our aim was to characterise the methylation level of a polymorphically imprinted gene, *VTRNA2-1/nc886*, in human populations and somatic tissues. We utilised 48 datasets, consisting of >30 different tissues and >30 000 individuals.

Results: We show that the *nc886* methylation status is associated with twin status and ethnic background, but the variation between populations is limited. Monozygotic twin pairs present concordant methylation, while ~30% of dizygotic twin pairs present discordant methylation in the *nc886* locus. The methylation levels of *nc886* are uniform across somatic tissues, except in cerebellum and skeletal muscle.

Conclusion: We hypothesize that the *nc886* imprint is established in the oocyte and that after implantation, the methylation status is stable, excluding a few specific tissues.

Keywords: non-coding 886, nc886, VTRNA2-1, polymorphic imprinting, metastable epiallele, imprinting, DNA methylation, Developmental Origins of Health and Disease hypothesis, population studies

Introduction

Genomic imprinting can be defined as the expression of a gene only from the maternal or the paternal allele, while the corresponding allele in the other chromosome is silenced via epigenetic mechanisms, including DNA methylation (1). The epigenetic profiles maintaining the imprinted status are established during gametogenesis when first the DNA methylation pattern is erased, after which the parent-of-origin-related DNA methylation pattern is created. For males, the DNA methylation profile of the sperm, including imprints, is completed in the primordial germ cells, before the birth of the male child. In oocytes, on the other hand, *de novo* DNA methylation will begin only after the birth of the female child, during the follicular growth phase, with gene-specific timepoints for imprinting reported (2–4). Canonically imprinted genes, approximately 130 of which exist in humans, retain the parent-of-origin-related expression pattern throughout an individuals' life in all of their somatic tissues, although tissue or developmental stage-specific imprinting can be seen, for example, in the placenta (5–7). The significance of intact genetic imprints is highlighted by the severe disorders caused by imprinting defects (8).

In humans, the locus harbouring non-coding 886 (*nc886*, also known as *VTRNA2-1*) is polymorphically imprinted, with approximately 75% of individuals having a methylated maternal allele, i.e. imprinted *nc886* locus, and the remaining approximately 25% having both maternal and paternal *nc886* allele unmethylated (9–11). According to current literature, this pattern is not due to genetic variation (9,12,13). We have also previously identified individuals who escape this bimodal methylation pattern. We found they present either intermediate methylation levels (methylation beta value 0.20-0.40 i.e. methylation level of 20-40%, in approximately 1-5% of the population) or methylation beta values over 0.60 (methylation level of over 60%), indicating that also the paternal allele has gained methylation in somatic tissues (in approximately 0.1% of the population) (11). The *nc886* locus codes for a 102nt long non-

coding RNA, which might then be cleaved into miRNA-like short non-coding RNAs, although the nature of these RNAs is still widely debated (14,15).

The establishment of the *nc886* imprint has been suggested to be an early event in the zygote, happening between four and six days after fertilization (13). Recently, it was suggested that the methylation pattern of *nc886* is already established in the preconceptional oocyte (10). Early establishment of *nc886* methylation status is supported by the fact that it has been shown to be uniform across analysed somatic tissues (10,16,17). As the methylation pattern of *nc886* had been reported to be concordant in monozygotic twin pairs (MZ) but not in dizygotic twin pairs (DZ), genetic factors were hypothesized to influence the methylation pattern (9), which was later shown not to be the case (10,11,18). Once established, the methylation status of *nc886* is stable from childhood to adolescence (16) and from adolescence to adulthood (11).

Changes in the proportion of individuals with methylated or unmethylated maternal *nc886* allele have been associated with maternal age (9,11), maternal socioeconomic status (11) and maternal alcohol consumption (10), as well as the season of conception in rural Gambia (9,16). Furthermore, the methylation level or status of *nc886* or level of *nc886* RNAs transcribed from the locus have been associated with childhood BMI (19), adiposity and cholesterol levels (11), as well as allergies (20), asthma (21), infections (22) and inflammation (23). Interestingly, the *nc886* methylation status and the level of *nc886* RNAs have also been associated with indicators of glucose metabolism (11,24). These findings suggest that the methylation status of *nc886* is a potential molecular mediator of the Developmental Origins of Health and Disease (DOHaD) hypothesis (also known as the Barker hypothesis) (25).

Detailed understanding of the determinants and functions of the methylation status of *nc886* is still lacking. *In vitro* methods are of limited feasibility, as both carcinogenesis (13) and pluripotency induction (11) have an effect on the methylation pattern at the *nc886* locus.

Currently, no animal models are available to study *nc886*, as most species, including mice and rats, do not harbour the gene, and in species harbouring the *nc886* gene, the locus is not polymorphically imprinted (26). Thus, we wanted to utilise available resources, the numerous existing DNA methylation datasets from humans, to gain insight on the methylation of *nc886*, a unique polymorphically imprinted gene and a potential molecular mediator of DOHaD.

More precisely, we wanted to investigate 1) the prevalence of *nc886* DNA methylation status groups in a large number of human populations (with a total $n > 30\,000$) with divergent historical and geographical origins in order to identify the factors associating with the existing variation, 2) the *nc886* methylation status in MZ and DZ twin pairs to elucidate the contribution of shared gametes vs unique gametes in shared prenatal environment to the establishment of *nc886* methylation pattern, and finally 3) the *nc886* methylation patterns in a larger variety of tissues, including brain regions and placenta, which have been shown to express a multitude of imprinted genes, as well as present dynamic imprinting (5,27,28) to have insights on the potential function of *nc886* and the stability of this polymorphic imprint in human tissues.

Materials and Methods

Datasets

In this study, we utilized 48 DNA methylation datasets, including DILGOM, FTC, ERMA, KORA, LURIC, NELLI, SATSA and YFS as well as 39 datasets available in Gene Expression Omnibus (GEO) (29) consisting of >30 different tissues and >30 000 individuals. (Supplementary Table 1). These datasets were used to study the methylation of *nc886* locus across populations, in twin pairs and across different tissues, with some datasets utilized in multiple settings.

For the population analysis, we included 32 datasets, with the number of individuals ranging from 131 to 2711 (total $n = 30\,347$). In these datasets, the DNA methylation data was available from blood (11,30–48), separated peripheral blood cells (49–51), blood spots (19), umbilical cord buffy coats (52), foetal cord tissue (53) or buccal swabs (52). Association of zygosity (MZ vs DZ pairs) with *nc886* methylation was analysed in 5 datasets (36,47,48,54,55).

To analyse the methylation of *nc886* across different tissues, 17 datasets were utilized. These included a dataset consisting of 30 tissues from a 112-year-old female (56), as well as datasets consisting of different brain regions ((57), GSE134379), adipose tissue (54,58), muscle (46,55–57, GSE142141, GSE171140), liver (58), buccal swabs (52,61), skin (62), sperm (63,64) and placenta (65–68). All utilized datasets are described in detail in Supplementary Table 1.

DILGOM (the Dietary, Lifestyle, and Genetic Determinants of Obesity and Metabolic Syndrome) (44) has been collected as an extension of the FINRISK 2007 survey. FINRISK surveys consist of cross-sectional, population-based studies conducted to monitor the risk of chronic diseases in Finland (69). The data used for the research were obtained from the THL biobank (study number THLBB2021_22).

FTC (the Finnish Twin Cohort) study includes three longitudinal cohorts (70,71) - the Older Twin cohort (71), FinnTwin16 (FT16) (72), and FinnTwin12 (FT12) (73). In this study, we included two subsets of FT16 and FT12 cohorts - a smaller subset of individuals for whom DNA methylation data was available from muscle, adipose tissue and blood (60) and a larger subset with methylation data only available from blood (54). For 49 MZ twin pairs, the information on whether they were dichorionic and diamniotic (DCDA, 22 pairs), monochorionic and diamniotic (MCDA, 9 pairs), or monochorionic and monoamniotic (MCMD, 18 pairs) was available. Participants of FTC were born before 1987 in Finland.

ERMA (Estrogenic Regulation of Muscle Apoptosis) study was a prospective cohort study designed to reveal how hormonal differences over the menopausal stages affect the physiological and psychological functioning of middle-aged women (74). The original ERMA cohort includes 47- to 55-year-old women living in the city of Jyväskylä or neighbouring municipalities in Finland. For this study, a subset of 47 individuals with both whole blood and muscle tissue samples available were included (60).

KORA (Kooperative Gesundheitsforschung in der Region Augsburg/the Cooperative Health Research in the Region Augsburg cohort) is a population-based health survey that collected both clinical and genetic data from the general population in the region of Augsburg and two surrounding counties in Germany (75). Data utilized here is from KORA F4 (collected in 2006-2008) and KORA FF4 (collected in 2013-2014) cohorts.

LURIC (the LUdwigshafen RIsk and Cardiovascular Health study) consists of patients of German ancestry who underwent coronary angiography between 1997 and 2000 at a tertiary care centre in Southwestern Germany (76). For this study, we utilized 2423 individuals with DNA methylation data available.

NELLI (Neuvonta, ELintavat ja LIIkunta neuvolassa) cohort consists of pregnant mothers participating in an intervention study aimed at preventing gestational diabetes from 14 municipalities in the southern part of Finland (77). Included in this study is data from children of these mothers, collected at a 7-year follow-up (61). The children participating in this study were born in 2007/2008 in Finland

SATSA (the Swedish Adoption/Twin Study of Aging) is a population-based study drawn from the Swedish Twin Registry that includes same-sex pairs of twins reared together and pairs separated before the age of 11, collected between 1984 and 2014 (78), 478 of whom had DNA methylation data available.

YFS (Young Finns Study) is a multicentre follow-up study on cardiovascular risk from childhood to adulthood, launched in 1980 (79). DNA methylation data utilized here is from the 30-year follow-up in 2011, including 1714 individuals. The participants in this study were born between 1962 and 1977 in Finland.

DNA methylation data processing

The majority of datasets in GEO were available as processed data, and these datasets were utilized as such. Datasets GSE61454, GSE71678, GSE134379, and GSE157896 were downloaded as raw idat-files, extracted with minfi package function `read.metharray.exp` and quantile normalized with default settings (80). For DILGOM (44), FTC (54,60), ERMA (60), KORA (11,45), LURIC (46), SATSA (47) and YFS (11), the processing of DNA methylation data has been described in detail in previous publications referenced here. For all datasets, information on sample material, Illumina array type (Illumina Infinium MethylationEPIC or Methylation450K BeadChip), and the used processing methods are provided in Supplementary Table 1.

For NELLI, genomic DNA from buccal swabs was extracted with Gentra Buccal Cell Kit (QIAGEN, cat.no: 158845) and stored at -20°C. Aliquots of 1 µg of DNA were subjected to bisulphite conversion, and a 4-µl aliquot of bisulphite-converted DNA was subjected to whole-genome amplification, followed by enzymatic fragmentation and hybridization onto an Illumina Infinium MethylationEPIC BeadChip at Helmholtz Zentrum, Munich, Germany. The arrays were scanned with the iScan reader (Illumina). All the analysed samples had a sum of detection p-values across all the probes less than 0.01. Logged (log2) median of methylated and unmethylated intensities of the analysed samples clustered well visually except for one outlier which was excluded from the analysis. Samples were checked for discrepancies between the reported and the predicted sex and none failed the test. Normalization was done with a stratified quantile normalization method implemented in preprocessQuantile function in minfi R/Bioconductor package (80,81). Probes with a detection p-value of more than 0.01 in 99% of the samples were filtered out. Similarly, cross-reactive probes and probes with SNPs were excluded from the analysis (82,83). All pre-processing steps were performed using functions implemented in the minfi R/Bioconductor package (80).

Clustering of individuals according to nc886 locus methylation level

For all datasets utilised, we retrieved the methylation values for 14 CpGs - cg07158503, cg04515200, cg13581155, cg11978884, cg11608150, cg06478886, cg04481923, cg18678645, cg06536614, cg25340688, cg26896946, cg00124993, cg08745965, and cg18797653, shown to display bimodal DNA methylation pattern in the *nc886* locus, which is indicative of polymorphic imprinting (9,11). For some datasets, the methylation data was not available from all 14 CpGs as, depending on the quality control steps performed, some probes had been

omitted from the data. The number of probes available for each dataset is provided in Supplementary Table 1.

Based on the methylation levels of the 14 CpGs, individuals were clustered into 3 groups by k-means clustering. Based on the median methylation level of the cluster, *nc886* methylation status groups were defined as imprinted (typical methylation β value > 0.40 , indicative of monoallelic methylation), non-methylated (typical methylation β value < 0.15 , indicative of two unmethylated alleles), and intermediately methylated (typical methylation β value $0.15-0.40$). Data was visualized to verify that the clustering had identified the groups as expected. The intermediately methylated individuals could not be detected in all datasets. In the datasets they were identified, the proportion ranged from 2% to 6%.

To verify that the clustering is reproducible across different datasets, we compared the clustering results within one dataset processed with different methods. We established that while the imprinted group could be reliably identified across different normalization methods, there were some inconsistencies between intermediately methylated individuals and non-methylated individuals (Supplementary data 1). Therefore, for analyses where proportions of *nc886* status groups were investigated across datasets, we combined the intermediately methylated and non-methylated clusters, i.e. *nc886* status is described as ‘imprinted’ or as ‘other’, the latter group including both intermediately methylated and non-methylated individuals. Only when comparing individuals within a dataset, namely in the case of MZ twin pairs, we have kept all three status groups in the analyses.

Comparison to established imprinted genes

When analysing different tissues, we also examined the methylation level of six known imprinted genes (*DIRAS3*, *KCNQ10T1*, *MEG3*, *MEST*, *NAP1L5*, *PEG10*, and *ZNF597*) (84).

This was done in tissues not presenting bimodal *nc886* pattern, to ensure imprinted genes, in general, do not display atypical methylation patterns in these tissues.

Statistical analyses

Differences in the proportion of individuals with imprinted *nc886* locus between sexes or in a case-control setting (depression (GSE125105), assisted reproductive technologies (GSE157896), rheumatoid arthritis (GSE42861), gestational diabetes mellitus (GSE141065), schizophrenia (GSE80417, GSE84727), inflammatory bowel disease (GSE87648), childhood abuse (GSE132203), and Parkinson's disease (GSE111629)) were investigated with χ^2 -test, with a threshold for significance set at $p < 0.05$ (Supplementary Table 2). For DZ twin pairs, the mathematically estimated proportion of miss-matched pairs (i.e. one twin imprinted, one intermediately methylated or non-methylated) was calculated. The difference between the observed number of miss-matched pairs and the estimated numbers was then investigated with the χ^2 -test, with a threshold for significance set at $p < 0.05$.

In tissues not presenting bimodal *nc886* methylation pattern, the difference in methylation levels between imprinted and other groups (as clustered according to a tissue presenting the expected methylation pattern from the same individuals) were analysed, with Mann-Whitney U-test with the threshold for significance set at $p < 0.05$. The median methylation levels in the *nc886* locus were correlated between different tissues with Spearman correlation, with the threshold for significance set at $p < 0.05$.

Results

Methylation status of nc886 across population cohorts

We characterized the methylation status of the *nc886* locus in 32 cohorts, including both population-based and case-control settings, consisting of 30 347 individuals in total. In the majority of the cohorts, the participants were described as being of European descent, white or Caucasian, hereafter referred to as white. In the datasets utilized, DNA methylation data was from whole blood, blood cells or blood cell subtypes, buccal swabs or foetal cord tissue (Supplementary Table 2). In these tissues, the methylation level of the *nc886* locus followed the expected bimodal pattern, and thus the individuals could be clustered into *nc886* methylation status groups (Supplementary Figure 1). Across all datasets, the proportion of imprinted individuals (individuals with the methylation level indicative of monoallelic methylation) varied between 65.8% and 83.5%, with an average percentage of 75.3% (Figure 1, Supplementary Table 2). When considering only cohorts consisting of white singletons, the proportion of imprinted individuals varied less, and was between 72.6% and 77.6% (Figure 1, Supplementary Table 2).

The lowest proportion of imprinted individuals was observed in datasets GSE157896 and GSE55763, with 68.2% and 65.8% of imprinted individuals, respectively (Figure 1, Supplementary Table 2). GSE157896 consists of new-borns whose mothers were Singaporean citizens or permanent residents, with self-reported homogenous Chinese, Indian or Malay ancestry (53), while GSE55763 consists of individuals with Indian Asian ancestry living in the UK (85).

In contrast, datasets consisting primarily of African American individuals (GSE117859 and GSE132203), had the third- and fourth-highest proportions of imprinted individuals - 79.1% and 78.7%, respectively. A third cohort consisting primarily of African American individuals

(GSE117860), did not stand out in this regard, with 76.8% of imprinted individuals. As these three datasets consist of other ethnicities in addition to African Americans, we tested whether there is a difference in the proportion of imprinted individuals across ethnic groups. In two of the datasets - GSE117859 and GSE117860, the proportion of imprinted individuals was significantly higher in African Americans than in white individuals (χ^2 -test p-value < 0.05, Supplementary Figure 2).

The highest proportion of imprinted individuals was observed in datasets GSE121633 and GSE56105, with 83.5% and 79.5% of imprinted individuals respectively. Both GSE121633 and GSE56105 consist of twins. Dataset GSE100227, also consisting of twins, has the fifth-highest percentage of imprinted individuals with 78.7%. However, some datasets consisting of twins had an average proportion of imprinted individuals (74.3% in GSE105018, 75.5% in FTC and 74.5% in SATSA, Figure 1 and Supplementary Table 2).

We found no difference (χ^2 -test p-value > 0.05) in the proportion of imprinted individuals between males and females in 26 datasets (Supplementary Table 2). The only exception was the dataset GSE82273, where the proportion of imprinted individuals was 73.3% (total n=505) in males and 80.7% (total n=384) in females (χ^2 -test p-value = 0.009). This dataset consists of individuals born in Norway with a facial cleft, and unaffected controls matched for the time of birth (33,86). In addition, we found no statistically significant differences in the proportion of imprinted individuals in any of the case-control settings reported (χ^2 -test p-value > 0.05, Supplementary Table 2) and no bias was seen according to the array type utilized (EPIC vs 450K).

The methylation of nc886 in MZ and DZ twin pairs

We investigated the concordance of the *nc886* methylation level and status between twin pairs. In five datasets, almost all MZ pairs were concordant regarding the *nc886* locus methylation level, whereas a large proportion of DZ twins were discordant for the *nc886* locus methylation level (Table 1, Supplementary Figures 3-6).

Of the total of 1250 MZ twin pairs investigated, we identified 17 pairs (1.3%), that were clustered to different *nc886* methylation status groups, when individuals were classified as imprinted and other (Table 1). For the datasets in which we could identify the intermediately methylated individuals (GSE61496, GSE105018), we were able to refine the analysis and to include the intermediately methylated and non-methylated groups as well. When considering all three *nc886* status groups (imprinted, intermediately methylated, non-methylated), of the total of 582 MZ pairs investigated, 13 (2.2%) were clustered to different *nc886* status groups. Of these discordant pairs, one co-twin was intermediately methylated while the other co-twin was either imprinted or non-methylated in all cases, i.e. we identified no twin pairs where one co-twin was imprinted while the other was non-methylated (Supplementary Table 3).

Across all twin pair datasets, the absolute difference in the *nc886* methylation level between MZ co-twins was below 0.05 for 88.2% of the pairs. Only 1.0% of pairs had a methylation beta value difference larger than 0.20. Only one MZ pair across all datasets presented a methylation beta value difference over 0.30 (Table 1). For this pair, the methylation beta values for *nc886* locus were 0.38 and 0.71, suggesting that one twin was imprinted and the other had gained methylation also in the paternal allele of *nc886*. These results are in line with our finding that there were no imprinted/non-methylated MZ twin pairs.

We also investigated the absolute difference in the *nc886* methylation level in MZ twin pairs for whom we had information on chorionicity and amnionicity. In DCDA (separated between days 1-3 after fertilization, 22 pairs) and MCDA (separated after day 3, but prior to

implantation, 9 pairs) twin pairs, we observed that in 64.5% of the twin pairs the within-pair difference in their median methylation beta values was below 0.025. For the remaining twin pairs, the within-pair difference was 0.025-0.05 in 19.4% of the pairs and above 0.05 in 16.1% of the pairs. In contrast, in MCMA twin pairs (18 pairs), which are separated only after implantation, the within-pair difference in median methylation value was below 0.025 for all pairs.

Four of the five twin cohorts also contained DZ twin pairs. Of these, 30.3%, 35.0%, 29.0%, and 32.1% were *nc886* methylation status (imprinted/other) discordant pairs (Table 1). Given the proportions of the *nc886* methylation status groups in these four datasets, with random pairings, the expected proportion of discordant pairs would be 39.8%, 37.7%, 35.7%, and 38.8%, respectively (details on proportions and expected proportions in Supplementary Table 4). For all four datasets, the proportion of identified discordant pairs was lower than expected, and for FTC, the largest of the available cohorts, the difference was statistically significant (29.0% vs 35.7%, χ^2 -test p-value = 0.02).

The methylation level of nc886 differs in the cerebellum and in skeletal muscle

Dataset GSE64491 consists of DNA methylation data for 30 tissues from a 112-year-old female, with tissues analysed in duplicates or quadruplicates. As seen in Figure 2 and Supplementary Figure 7, the methylation level of the *nc886* locus was higher in the cerebellum as compared to other tissues, with the methylation beta value above 0.70 for most probes in this locus. In addition, muscle and diaphragm showed slightly higher methylation beta values as compared to other tissues. For other tissues in dataset GSE64491, variation in the level of methylation at the *nc886* locus between tissues was similar in magnitude to what can be observed between the replicates in the data (Supplementary Table 5). Interestingly, unlike

skeletal muscle and diaphragm, the *nc886* methylation level of heart was not elevated, but was comparable to other tissues in the dataset (Supplementary Figure 7). As a reference, we examined the methylation level of six known imprinted genes and did not observe a marked difference in the methylation level between the cerebellum or muscle and other tissues (Supplementary Figure 8). For these six genes, variation in the methylation level between tissues and between the replicates was smaller as compared to the *nc886* locus (Supplementary Table 5).

To verify the methylation level of the *nc886* locus in the cerebellum, dataset GSE134379, containing methylation data from the cerebellum and middle temporal gyrus (MTG), and GSE72778, containing data from the cerebellum and five other brain regions (frontal lobe, hippocampus, midbrain, occipital lobe and temporal lobe) were investigated. In both data-sets, regions of the cerebrum and midbrain present a bimodal *nc886* methylation pattern, while in the cerebellum, the median methylation level follows a unimodal distribution and is higher, close to 0.70 (Figure 3, Supplementary Figure 9).

The methylation level of the *nc886* locus in the cerebellum was high both in individuals with an imprinted and a non-methylated *nc886* (clustered according to the methylation levels in the cerebrum or MTG) in both datasets studied (Figure 4 and Supplementary Figure 10). Despite the unimodal distribution of the *nc886* methylation level in the cerebellum, we observed a difference in the methylation level of the *nc886* locus in the cerebellum between imprinted and non-methylated individuals (p-value < 0.001 in both datasets). The six known imprinted genes analysed did not present differences in the methylation level between the cerebellum and other brain regions (Supplementary Figure 11).

To further investigate the methylation level of the *nc886* locus in muscle, we investigated six additional datasets, three of which consisted only of muscle samples (GSE142141,

GSE151407, GSE171140), and three for which other tissues were also available (GSE61454 - skeletal muscle, visceral and subcutaneous fat, as well as liver; ERMA - muscle tissue and blood; FTC - muscle tissue, adipose tissue, and blood). In all of the datasets, the methylation level of the *nc886* locus in muscle presented a unimodal distribution, centred slightly above 0.50 (Figure 3, Supplementary Figures 12-14). In GSE61454, ERMA and FTC, other tissues presented the expected bimodal methylation distribution at the *nc886* locus (Supplementary Figures 13 and 14). Despite the unimodal methylation level observed in muscle, there was a difference in the methylation level at the *nc886* locus in muscle between imprinted and non-methylated individuals (as clustered based on the *nc886* methylation levels in other tissues) in FTC and GSE61454 (Mann-Whitney U-test p-value < 0.001) (Figure 5, Supplementary Figure 15). Moreover, in FTC, the median methylation levels correlated well across all three tissues (adipose tissue, muscle, blood; $p < 2 \times 10^{-12}$). Individuals presenting intermediately methylated *nc886* in blood also had intermediate methylation levels in other tissues (Supplementary Figure 16).

We also investigated other available datasets for potential atypical methylation patterns in the *nc886* locus. Skin (GSE90124) and buccal swabs (NELLI, GSE128821) displayed the expected bimodal distribution and the expected level of methylation at the *nc886* locus (Supplementary Figure 17). For the placenta (Figure 3, Supplementary Figures 18 and 19), we observed a slight downward shift in the methylation level at the *nc886* locus in four datasets (GSE167885, GSE75248, GSE71678, GSE115508). While the methylation pattern in the placenta followed a bimodal distribution, a clustering analysis could not establish a clear division between imprinted and non-methylated individuals, as compared to the majority of other tissues studied. In dataset GSE115508, we did not observe a corresponding downward shift in the methylation values in amnion or chorion (Supplementary Figure 19). Similarly, foetal cord tissue

(GSE157896) did not display a downward shift in the *nc886* methylation level (Supplementary Figure 1).

In sperm, we observed a methylation level close to 0 at the *nc886* locus, as is expected for a maternally imprinted gene (Supplementary Figure 20). In addition to methylation data from sperm, dataset GSE149318 also contained methylation data from blood of the same individuals. We did not observe a difference in the sperm *nc886* methylation level between blood-derived imprinted and non-methylated individuals (Mann-Whitney U-test p-value > 0.05).

Discussion

Here, we confirm that the proportion of individuals with an imprinted *nc886* locus is approximately 75% in majority of populations and show in 32 datasets from variable historic and geographic origins, that the variation in this proportion is very limited, especially in populations consisting of white singletons. More varying proportions can be observed in populations consisting of other ethnicities, and in twins. We show that in MZ twin pairs, the methylation level of the *nc886* locus is highly similar, especially in MCMA twin pairs. Finally, we confirm that the *nc886* methylation pattern is stable in the majority of somatic tissues, but also describe two exceptions – the cerebellum and skeletal muscle. These findings allow us to refine the hypotheses on timing and determinants of the polymorphically imprinted *nc886* and how the methylation in this locus varies in tissues and individuals originating from the same zygote.

Variation of the nc886 methylation status group proportions is limited across populations

It has previously been reported that, based on individual cohorts, the proportion of individuals with an imprinted *nc886* locus is approximately 75%, with the remaining 25% presenting a non-methylated *nc886* locus (9–11). Here, utilizing 32 cohorts and over 30 000 individuals, we can confirm that the average proportion of imprinted individuals is 75%. Especially in white singletons, the variation in the proportion of individuals with an imprinted *nc886* locus is limited. The lowest proportion of individuals with an imprinted *nc886* is observed in cohorts of East Asian origin, while the highest proportion of individuals with an imprinted *nc886* is observed in cohorts consisting of African American individuals or twins. Previously, in a cohort of 82 Korean women, the proportion of individuals with methylation levels indicating an imprinted *nc886* locus was reported to be 65.9% (87), comparable to our findings here

(Figure 1, Supplementary Table 2). We, and others, have previously shown that genetic variation is not associated with the establishment of the *nc886* methylation pattern (10,11,18), but, as these results are based on mainly white populations, we can't rule out the possibility that other ethnic groups would present genetic variation that would affect the establishment of the *nc886* imprint. Another explanation for these findings is that the lifestyle or environmental conditions of these populations are affecting the proportions of imprinted individuals. However, dataset GSE157896 consists of individuals born in Singapore (53) and dataset GSE55763 consists of individuals of East Asian origin, some of whom have been born in the UK and others in Asia (29), suggesting that, at least in this case, shared genetics, rather than shared environmental conditions, may affect the proportion of imprinted individuals. Our results again highlight the need to include more diverse populations in genetic association studies (88).

The methylation status of nc886 is not associated with sex

We identified no difference in the proportion of imprinted individuals between males and females, with the exception of one dataset - GSE82273. This dataset consists of individuals born with a facial cleft and unaffected controls matched for the time of birth (33,86). In a previous study, facial clefts were associated with hypomethylation of the *nc886* region (89). However, in this study, the *nc886* methylation was analysed as a continuous variable instead of categorical. The prevalence of facial clefts has also been reported to vary according to sex, whether it is female- or male-biased depending on the type of the cleft (90). Unfortunately, the information on the case/control status was not included in dataset GSE82273, and thus we were unable to test whether the observed difference in proportions is due to the sex bias of facial clefts. As we observed no difference in the proportion of imprinted individuals between males

and females in 26 other datasets (in a total of 27 362 individuals), we assume that the observed difference in this one dataset is due to the bias caused by the case/control setting of the dataset and conclude that the *nc886* methylation status is not associated with sex.

nc886 imprint is stable across majority of somatic tissues

Previous studies have shown that the *nc886* methylation status is stable within one individual in all tissues analysed (10,16), and we have previously shown this in different blood cell subtypes (11). Here, we confirm these findings in over 30 somatic tissues but also report exceptions to this pattern, namely cerebellum and skeletal muscle. In both cerebellum and skeletal muscle, visual inspection revealed a unimodal DNA methylation pattern at the *nc886* locus, with a methylation level of approximately 0.70 in the cerebellum and approximately 0.50 in skeletal muscle. Despite the unimodal methylation pattern in these tissues, there was a slight difference in the methylation level of *nc886* between individuals who were imprinted and non-methylated (according to clustering analysis on other tissues). This suggests that the methylation pattern established in early development in these tissues is not completely reset and established anew, but that the methylation level is built upon the existing methylation state. This is further corroborated by our finding that the methylation level of *nc886* in muscle is strongly correlated with the methylation level in other somatic tissues. It is not known what mechanism is responsible for this increase in methylation at the *nc886* locus in these tissues. In contrast to the cerebellum and skeletal muscle, in the placenta, the methylation level of the *nc886* locus was slightly decreased as compared to other analysed tissues. Placenta has been previously described to have aberrant profiles of imprinted genes and also present a multitude of secondary differentially methylated regions (91). In line with previous studies (9,13), we also show here that the *nc886* locus is non-methylated in sperm and see no difference in the

methylation levels between men who present either imprinted or non-methylated *nc886* locus in other tissues.

We, and others, have previously shown that the *nc886* methylation is tightly associated with the level of *nc886* RNAs (11,12,92), with the imprinted individuals having lower levels of these RNAs as compared to the non-methylated individuals. Therefore, we can speculate that the cerebellum and skeletal muscle have lower levels of these RNAs as compared to other tissues, while the placenta has higher levels of these RNAs. As the function of these RNAs is not known (14,15), different regulation patterns in these specific tissues might offer possibilities to further hypothesize their role.

In addition to being stable across tissues, the methylation level of *nc886* has been shown to be stable through follow-up, from childhood to adolescence (16) and from adolescence to adulthood (11). However, in granulosa cells, the methylation level of the *nc886* locus has been shown to be affected by age, with women over the age of 40 showing higher methylation values as compared to women under the age of 30 (93). Granulosa cells are the somatic cell compartment in the follicle and are crucial for oogenesis (94,95). As it has been previously shown that maternal age is associated with the *nc886* methylation status of the offspring (9,11), it is interesting to speculate whether the altered methylation status of the granulosa cells is associated with this phenomenon.

Establishment of nc886 imprint

Current literature suggests that periconceptual conditions affect the establishment of the *nc886* methylation pattern (9–11,16) and that the imprinting of *nc886* is an early embryonic event happening between days four and six after fertilization (13). This notion is in slight conflict with the finding that the variation in the proportion of imprinted individuals is limited

across populations. If periconceptional conditions would have a substantial role in the establishment of the methylation of the *nc886* locus, one would expect to see more differences between cohorts from different countries or between different birth cohorts and, in contrast, fewer differences between DZ twin pairs. For example, we see only minor differences in the prevalence of imprinted individuals in Finnish cohorts born in the 1960s and 1970s (YFS; 73.5%), the 1950s through 1980s (FTC; 75.7%) or in 2007/2008 (NELLI; 74.1%), even though the nutritional status of Finnish expecting mothers has drastically changed during this time (96,97). Furthermore, if the non-methylated status would be caused by adverse pregnancy conditions, such as lack of energy or certain nutrients, one could expect that the proportion of non-methylated individuals would be significantly decreased in populations with good nutritional status. For example, if the lack of folate would be causal in the establishment of the non-methylated *nc886* methylation pattern (16,98), the number of these individuals should have been more drastically diminished in cohorts born after the folate supplementation recommendations (97,99). For DZ twins, who share the pregnancy but originate from different zygotes, we showed that approximately 1/3 of pairs are discordant regarding the *nc886* methylation status. While this was slightly fewer pairs than would be expected by chance, the difference was subtle and statistically significant only in one dataset studied.

Another plausible hypothesis explaining the shown associations between *nc886* and periconceptional conditions is that the methylation status is determined already in the oocyte, as suggested also by Carpenter et al. (10), and that the slight variation observed in the *nc886* status proportions at the population level is due to either epigenotype offering a survival advantage in specific pregnancy conditions. The establishment of the *nc886* methylation status already in the oocyte is also supported by the fact that in line with results by Carpenter et al. (9), we observed no MZ twin pairs with one co-twin being imprinted and the other non-methylated. This is also true in the subgroup of twins from FTC, who have been reported to be

dichorionic, and thus have been separated 1-3 days after conception, indicating that the process leading to either non-methylated or imprinted *nc886* loci was completed before this time point.

Our results also suggest that the process that leads to individuals presenting intermediate *nc886* methylation pattern is over before implantation. In MZ twins reported to be separated after implantation, the median methylation in *nc886* is nearly identical (difference less than 2.5%). Furthermore, individuals who present intermediate methylation levels in their blood also present very similar levels in their adipose tissue, in line with previous findings in different blood cell populations (11), suggesting stability of the methylation level, including the intermediate state, through tissue differentiation. When taken together with the temporal stability of also the intermediate methylation levels (11), we can suggest that the ratio of cells with an imprinted or a non-methylated maternal allele of *nc886* in individuals presenting intermediately methylated *nc886* status is established before implantation, concurrently to the global de- and re-methylation waves in the embryo (100) and is then reflected on the individual for the rest of their life.

Previous reports on the association of periconceptional conditions and the *nc886* methylation status (9–11,16) could be explained by selective survival in certain pregnancy conditions, instead of these conditions directly affecting the establishment of the methylation status. As shown by our results, an example of pregnancy conditions where certain *nc886* status might be advantageous/disadvantageous is twin pregnancy. In the population cohorts studied, twin cohorts had high proportions of imprinted individuals, and in DZ twins the number of pairs with discordant *nc886* methylation status was slightly lower than would be expected by chance. This suggests that twin pregnancy might be favourable to foetuses with imprinted *nc886* loci or, in the case of DZ twins, to pairs with concordant *nc886* methylation status.

Limitations of the study

Our results are descriptive in nature, and thus need to be interpreted as hypothesis-generating rather than conclusive. Clustering of individuals into *nc886* methylation status groups is, to some extent, affected by different data pre-processing methods, but we have tried to mitigate this by carefully comparing different pre-processing methods. The paucity of datasets consisting of individuals of multiple ethnicities limits our possibilities to draw firm conclusions on the effect of ethnicity on *nc886* methylation status proportions.

Conclusion

Current literature suggests that the polymorphic imprinting of *nc886* is not due to genetic variation in white populations, but as our results show more variation in the proportion of individuals with an imprinted *nc886* in non-white cohorts, the genetic analyses should be repeated in more diverse populations. Based on our results and current literature, we hypothesize that DNA methylation of the *nc886* locus is established in the growing oocyte and that the variation in the proportion of imprinted individuals in a population could be due to survival advantage/disadvantage in certain pregnancy conditions, illustrated in Supplementary Figure 21. After implantation, the methylation level of *nc886* is preserved across studied somatic tissues, with the exception of cerebellum and skeletal muscle. In all individuals, *nc886* locus gains methylation in these tissues, even though the methylation levels still associate with the *nc886* status established earlier in the development.

Summary points

- Variation in the proportion of individuals with an imprinted *nc886* is very modest, with approximately 75% of individuals being imprinted across populations
- The observed variation is mainly limited to non-white ethnic groups and twin cohorts
- Methylation status of *nc886* was not associated with sex or any of the case/control setting investigated
- Methylation level of *nc886* is increased in cerebellum and in skeletal muscle, but is uniform in other somatic tissues
- Placenta presents lower methylation levels than majority of somatic tissues, but binomial methylation pattern can still be detected
- Monozygotic twins show highly similar *nc886* methylation levels, which is even more pronounced on twins separated at a later date in development
- Approximately 30% of dizygotic twin pairs are discordant for *nc886* methylation status
- We suggest that methylation status of *nc886* is established in the oocyte, and that the slight variation observed across populations could be due to selective survival advantage of the foetus in certain conditions

Figure/Table legends

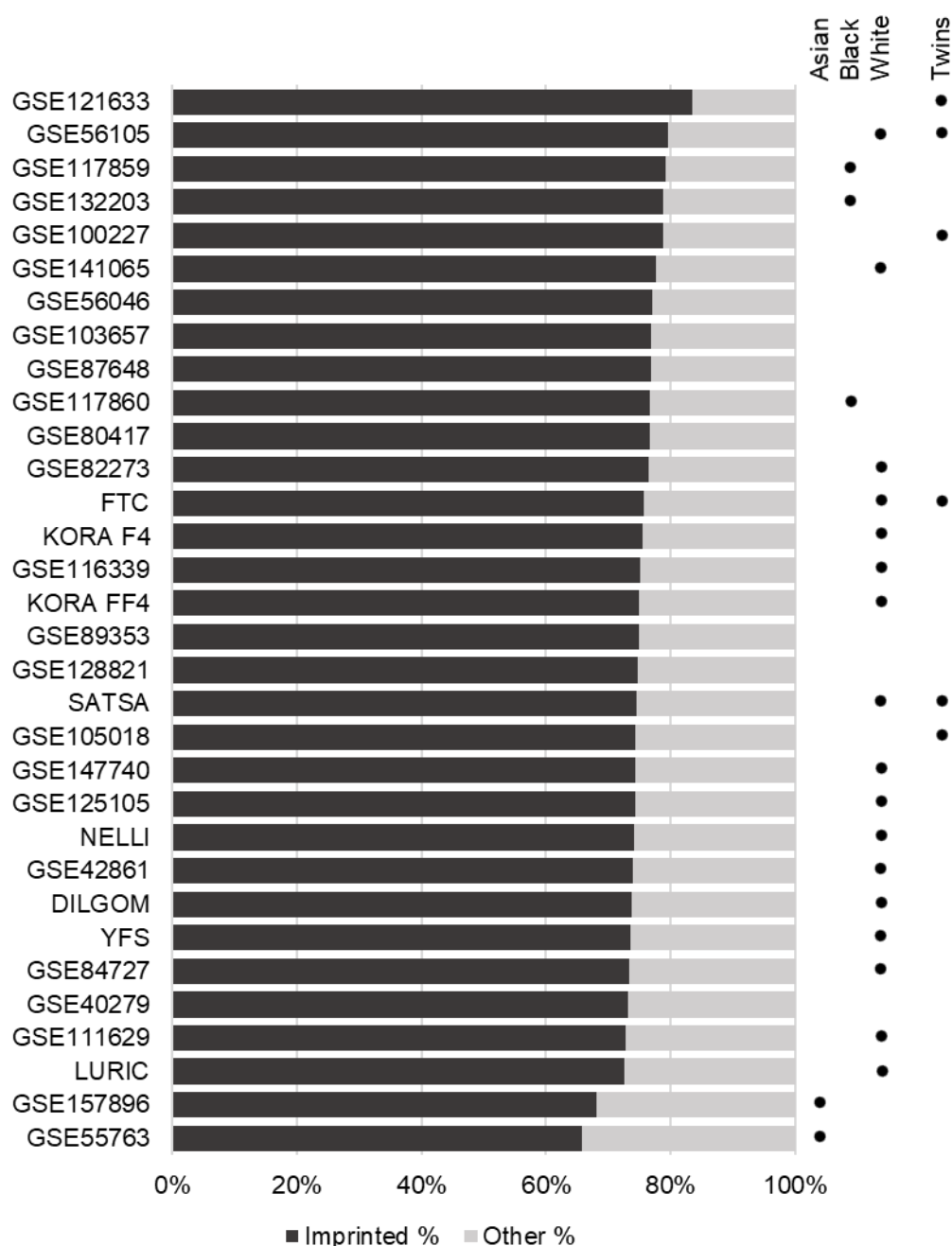


Figure 1. Proportion of imprinted individuals across cohorts utilized in this study. Individuals were clustered as imprinted and other (including non-methylated and intermediately methylated). Ethnicity was not specified for all cohorts utilized. Details of each cohort can be found from Supplementary Tables 1 and 2.

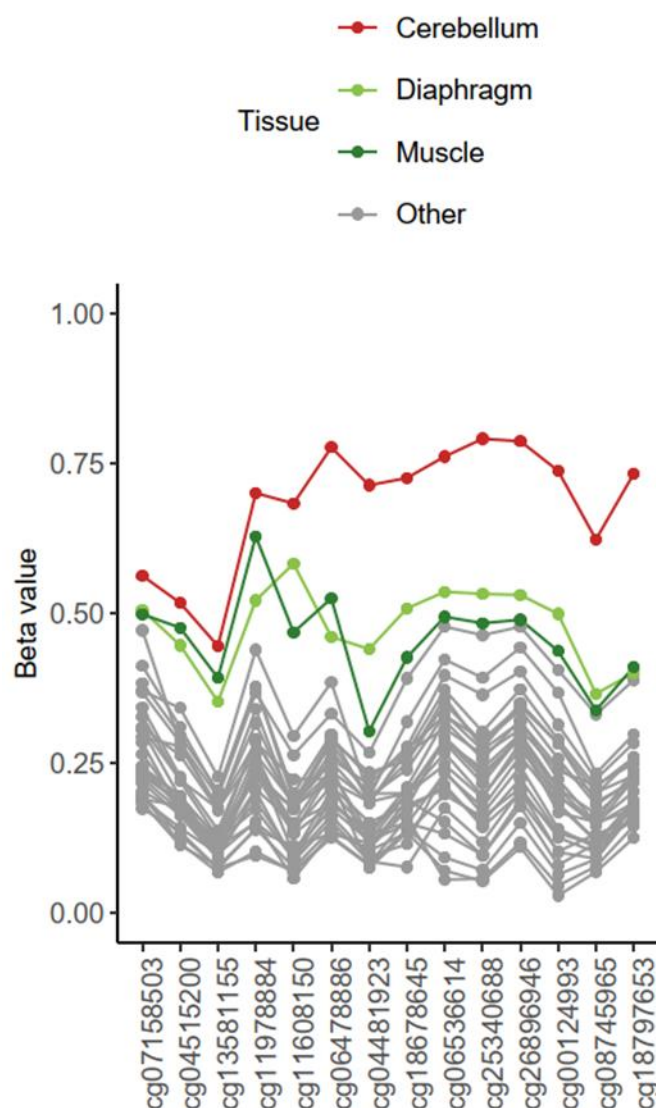


Figure 2. Methylation level of *nc886* locus in 30 tissues of a 112 year-old woman. Cerebellum has considerably higher level of methylation as compared to other tissues. Muscle and diaphragm also show slightly elevated levels of methylation as compared to other tissues. For a figure with all tissues presented in colour, see Supplementary Figure 7.

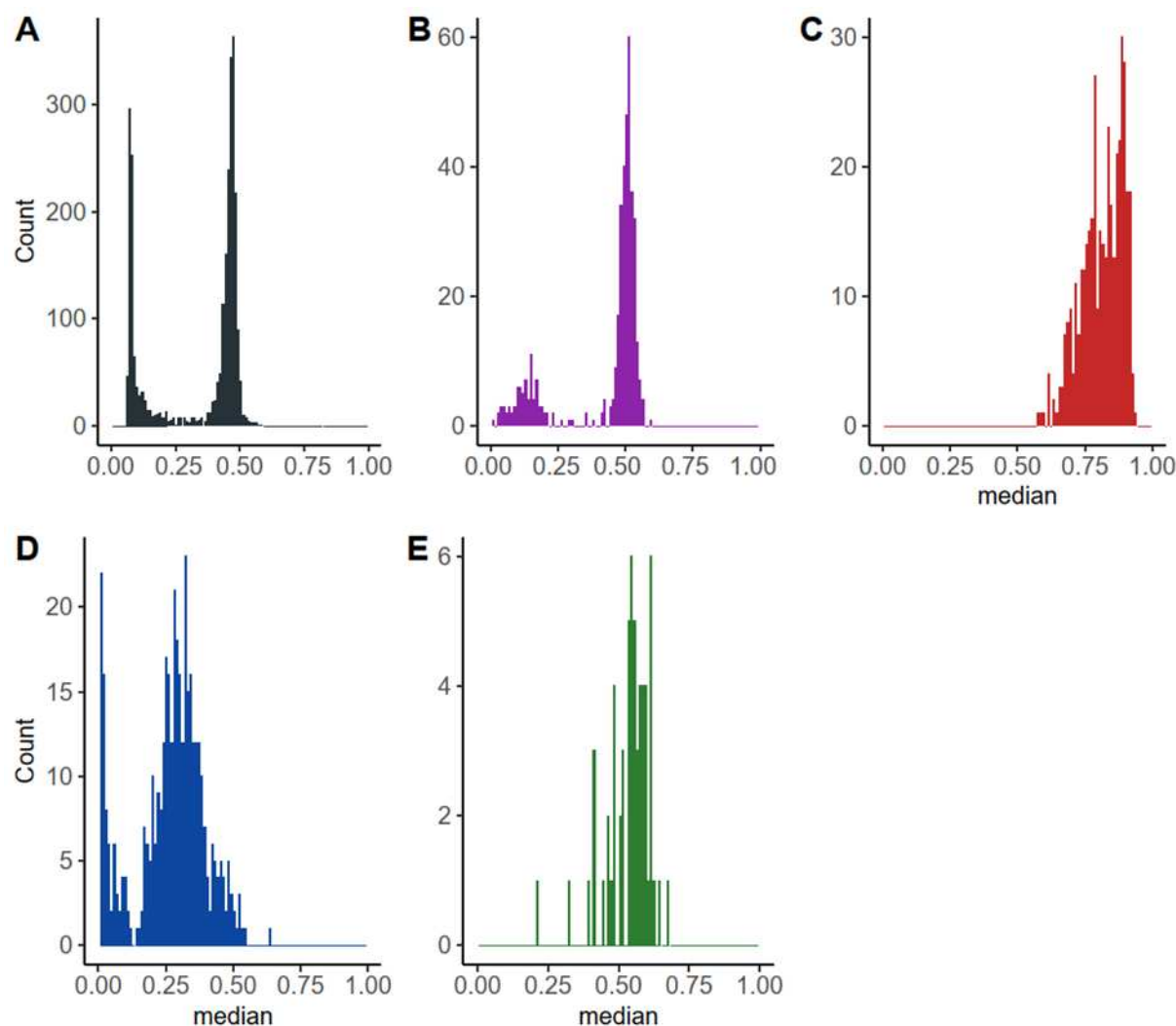


Figure 3. Histograms of *nc886* locus methylation median in different tissues. A) Blood, GSE55763, n = 2664, B) middle temporal gyrus (MTG), GSE134379, n=404, C) cerebellum, GSE134379, n= 404 (same individuals as in B), D) placenta, GSE167885, n = 411, and E) muscle, GSE61454, n = 60. As compared to blood and MTG, cerebellum shows a unimodal distribution with elevated methylation levels in *nc886* locus. Also in muscle (E) *nc886* methylation showed a unimodal distribution. In placenta the methylation level at *nc886* locus presented a bimodal distribution, but the overall methylation level was lower as compared to blood.

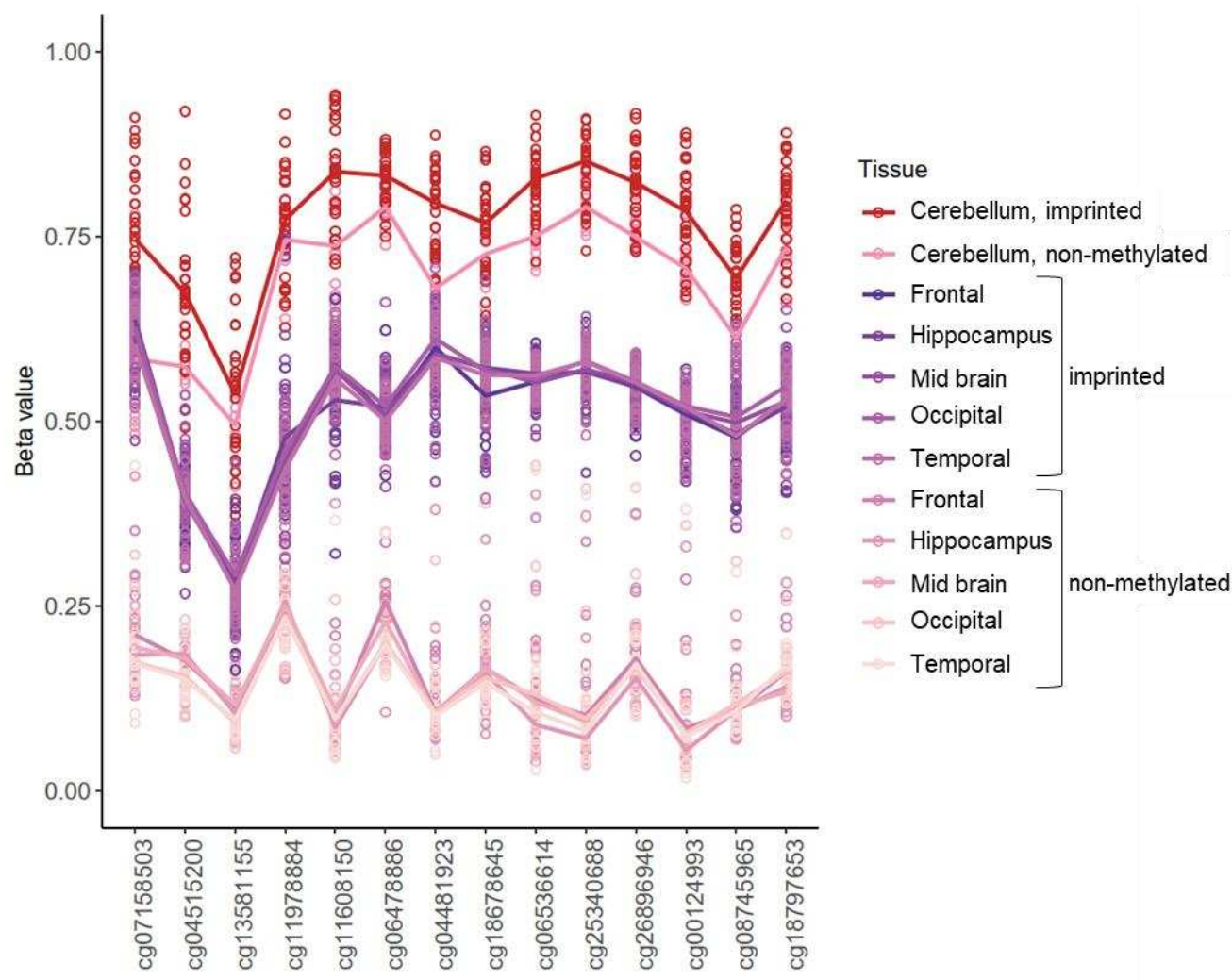


Figure 4. Methylation level of *nc886* in cerebellum and in five other brain regions (GSE72778). Individuals have been grouped to imprinted and non-methylated based on data of the five brain regions where *nc886* displayed a bimodal methylation pattern (frontal lobe, hippocampus, mid brain, occipital lobe and temporal lobe). As compared to other brain regions, cerebellum shows higher levels of methylation in both groups, but there is a statistically significant difference in the methylation between individuals clustered as imprinted and non-methylated. Similar pattern can be observed in dataset GSE134379 (Supplementary Figure 10).

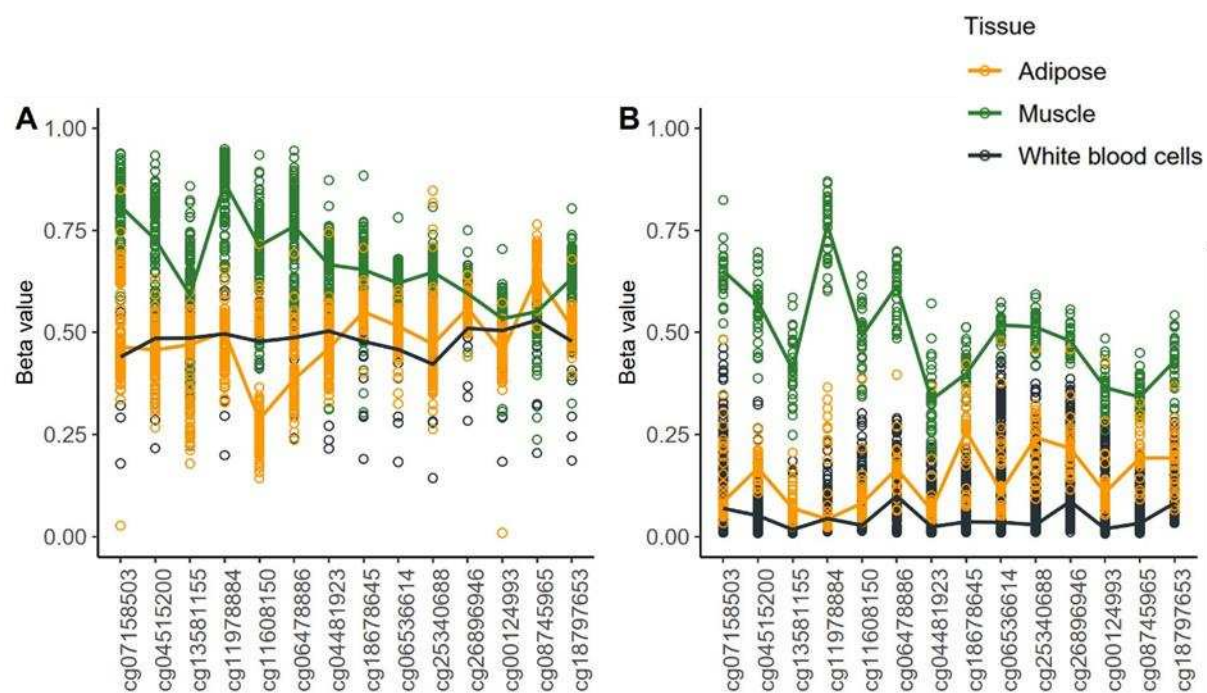


Figure 5. Methylation level of *nc886* locus in FTC in A) imprinted individuals (as clustered based on *nc886* methylation levels in blood) and B) non-methylated individuals. Despite the unimodal methylation level in muscle (Supplementary Figure 13), we observed a difference in the methylation level of *nc886* locus in muscle between imprinted and non-methylated individuals.

Table 1. Number of twin pairs discordant for *nc886* methylation status and the absolute difference in the methylation level across twin pairs.

	FTC				SATSA	GSE105018	GSE100227	GSE61496
	All	DCDA	MCDA	MCMA				
Monozygotic twins								
nc886 status of twin pairs								
Imprinted-Imprinted	387 (74.6%)	18 (81.8%)	5 (55.6%)	16 (88.9%)	61 (73.5%)	313 (73.5%)	55 (83.3%)	127 (81.4%)
Other-Other	126 (24.3%)	4 (18.2%)	4 (44.4%)	2 (11.1%)	18 (21.7%)	109 (25.6%)	11 (16.7%)	26 (16.7%)
Imprinted-Other	6 (1.2%)	0 (0%)	0 (0%)	0 (0%)	4 (4.8%)	4 (1.0%)	0 (0%)	3 (1.9%)
Difference in median methylation between co-twins								
<5%	459 (88.4%)	20 (90.0 %)	6 (66.7%)	18 (100%)	42 (50.6%)	403 (94.6%)	64 (97.0%)	134 (85.9%)
5-10%	42 (8.1%)	1 (4.5%)	2 (22.2%)	0 (0%)	23 (27.7%)	17 (4.0%)	1 (1.5%)	17 (10.9%)
10-20%	11 (2.1%)	1 (4.5%)	1 (11.1%)	0 (0%)	17 (20.5%)	5 (1.2%)	0 (0%)	2 (1.3%)
20-30%	7 (1.3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0.2%)	1 (1.5%)	3 (1.9%)
>30%	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1.2%)	0 (0%)	0(0%)	0 (0%)
Dizygotic twins								
nc886 status of twin pairs								
Imprinted-Imprinted	326 (62.2%)				90 (57.7%)	179 (58.5%)	45 (68.2%)	
Other-Other	46 (8.8%)				16 (10.3%)	27 (8.8%)	1 (1.5%)	
Imprinted-Other	152 (29.0%)				50 (32.1%)	100 (32.7%)	20 (30.3%)	

Reference annotations

**of considerable interest

** Carpenter BL, Zhou W, Madaj Z, DeWitt AK, Ross JP, Grønbæk K, et al. Mother-child transmission of epigenetic information by tunable polymorphic imprinting. *Proc Natl Acad Sci U S A*. 2018 Dec 18;115(51):E11970–7.

Describes how methylation status of nc886 is associated with prenatal environment, and that the methylation status is associated with the offspring's phenotype.

** Carpenter BL, Remba TK, Thomas SL, Madaj Z, Brink L, Tiedemann RL, et al. Oocyte age and preconceptual alcohol use are highly correlated with epigenetic imprinting of a noncoding RNA (nc886). *Proc Natl Acad Sci U S A*. 2021 Mar 23;118(12):e2026580118.

Further confirms the association between periconceptual conditions and offspring's nc886 methylation status and describes that 75% of oocytes also present imprinted nc886 locus.

** Silver MJ, Kessler NJ, Hennig BJ, Dominguez-Salas P, Laritsky E, Baker MS, et al. Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptual environment. *Genome Biology*. 2015 Jun 11;16(1):118.

Identifies nc886 as metastable epiallele, that is responsive to environmental conditions.

** Marttila S, Viiri LE, Mishra PP, Kühnel B, Matias-Garcia PR, Lyytikäinen LP, et al. Methylation status of nc886 epiallele reflects periconceptual conditions and is associated with glucose metabolism through nc886 RNAs. *Clin Epigenetics*. 2021 Jul 22;13(1):143.

Further confirms the association between periconceptual conditions and offspring's nc886 methylation status, describes the association with offspring's phenotype. Describes the association of nc886 methylation and nc886 RNA levels.

*of interest

* Kostiniuk D, Tamminen H, Mishra PP, Marttila S, Raitoharju E. Methylation pattern of polymorphically imprinted nc886 is not conserved across mammalia. PLOS ONE. 2022 Mar 16;17(3):e0261481.

Characterises nc886 methylation in primates, indicating that the locus is not polymorphically imprinted in other species than in humans

* Romanelli V, Nakabayashi K, Vizoso M, Moran S, Iglesias-Platas I, Sugahara N, et al. Variable maternal methylation overlapping the nc886/vtRNA2-1 locus is locked between hypermethylated repeats and is frequently altered in cancer. Epigenetics. 2014 May 6;9(5):783–90.

Describes how nc886 is a metastable epiallele, the establishment of which is an early embryonic event and shows that nc886 is maternally imprinted.

* Treppendahl MB, Qiu X, Søgaaard A, Yang X, Nandrup-Bus C, Hother C, et al. Allelic methylation levels of the noncoding VTRNA2-1 located on chromosome 5q31.1 predict outcome in AML. Blood. 2012 Jan 5;119(1):206–16.

Shows that in population 75% of individuals present imprinted nc886 loci, and that the expression of nc886 RNAs are associated with the methylation level in the locus.

Ethical declarations

YFS: approved by the 1st ethical committee of the Hospital District of Southwest Finland on September 21st, 2010, and by local ethical committees (1st Ethical Committee of the Hospital District of Southwest Finland, Regional Ethics Committee of the Expert Responsibility area of Tampere University Hospital, Helsinki University Hospital Ethical Committee of Medicine, Research Ethics Committee of the Northern Savo Hospital District, and Ethics Committee of the Northern Ostrobothnia Hospital District).

NELLI: This follow-up study and written informed consent procedure were approved by the medical ethics committees of the Pirkanmaa hospital district (R14039)

SATSA: study was approved by the ethics committee at Karolinska Institutet with Dnr 2015/1729-31/5.

LURIC: The study plan was approved by the ethics committee of the State Chamber of Physicians of Rhineland-Palatinate.

KORA: All study methods were approved by the ethics committee of the Bavarian Chamber of Physicians, Munich.

DILGOM: original FINRISK study has been approved by Coordinating Ethics Committee of the HUS Hospital District, decision numbers 229/E0/2006 and 332/13/03/00/2013. FINRISK and DILGOM study materials have been transferred to THL Biobank in accordance with the notification procedure permitted by the Finnish Biobank Act.

FTC: Data collection and analysis were approved by the ethics committee of the Helsinki University Central Hospital (Dnro 249/E5/01, 270/13/03/01/2008, 154/13/03/00/2011).

ERMA: study was approved by the ethics committee of the Central Finland health care district in 2014 (K-SSHP Dnro 8U/2014).

Funding sources:

This research has been supported by Academy of Finland (349708 PPM; 341750, 346509 ES; 275323, 309504, 314181, 335249 EKL; 297908, 328685 MO; 330809, 338395 ER), European Research Council (742927 for MULTIEPIGEN project, OR), Juho Vainio Foundation (LK; ES) Karolinska Institutet Strategic Research Program in Epidemiology (SH), Laboratoriolääketieteen edistämissäätiö sr. (ER), Pirkanmaa Regional Fund of Finnish Cultural Foundation (SM), Päivikki and Sakari Sohlberg foundation (ES), Signe och Ane Gyllenbergs stiftelse (ER), Swedish Research Council (2015-03255; 2019-01272, 2020-06101, SNP&SEQ Technology Platform in Uppsala to SH), the Sigrid Juselius Foundation (MO), the Tampere University Hospital Medical Funds (9AC077, 9X047, 9S054, 9AB059 ER), Yrjö Jahnsson Foundation (20207299 SM; 20217416, 20197181 LK; 20197212 ER).

The Young Finns Study (YFS) has been financially supported by the Academy of Finland: grants 322098, 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; EU Horizon 2020 (grant 755320 for TAXINOMISIS; grant 848146 for To_Aition); and Tampere University Hospital Supporting Foundation.

The DNA methylation measurement in the LURIC Study has been financially supported by the 7th Framework Program RiskyCAD (grant agreement number 305739) of the European Union

and the Competence Cluster of Nutrition and Cardiovascular Health (nutriCARD), which is funded by the German Federal Ministry of Education and Research (grant number 01EA1411A).

The main sources of funding in NELLI follow-up study are Competitive research funding from Pirkanmaa hospital district (9R030, 9S034, 9M053) and Academy of Finland (277079).

The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. This work was supported by a grant (WA 4081/1-1) from the German Research Foundation. The work was further supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the EU Joint Programming Initiative ‘A Healthy Diet for a Healthy Life’ (DIMENSION grant number 01EA1902A).

Acknowledgements: The authors wish to thank all the researchers who have participated in the collection of the utilized datasets and made their data publicly available. The DILGOM data used for the research were obtained from THL Biobank (study number THLBB2021_22). We thank all study participants for their generous participation at THL Biobank and at National FINRISK and DILGOM Studies. We also thank Daria Kostiniuk (BSc) for the in-depth discussions of the establishment of the *nc886* methylation pattern.

References

1. Reik W, Walter J. Genomic imprinting: parental influence on the genome. *Nat Rev Genet.* 2001 Jan;2(1):21–32.
2. Lucifero D, Mann MRW, Bartolomei MS, Trasler JM. Gene-specific timing and epigenetic memory in oocyte imprinting. *Human Molecular Genetics.* 2004 Apr 15;13(8):839–49.
3. Smallwood SA, Kelsey G. De novo DNA methylation: a germ cell perspective. *Trends in Genetics.* 2012 Jan 1;28(1):33–42.
4. Lees-Murdock DJ, Walsh CP. DNA methylation reprogramming in the germ line. *Epigenetics.* 2008 Feb;3(1):5–13.
5. Hudson Q, Kulinski T, Huetter S, Barlow D. Genomic imprinting mechanisms in embryonic and extraembryonic mouse tissues. *Heredity (Edinb).* 2010 Jul;105(1):45–56.
6. Baran Y, Subramaniam M, Biton A, Tukiainen T, Tsang EK, Rivas MA, et al. The landscape of genomic imprinting across diverse adult human tissues. *Genome Res.* 2015 Jul 1;25(7):927–36.
7. Sanchez-Delgado M, Court F, Vidal E, Medrano J, Monteagudo-Sánchez A, Martin-Trujillo A, et al. Human Oocyte-Derived Methylation Differences Persist in the Placenta Revealing Widespread Transient Imprinting. *PLOS Genetics.* 2016 Nov 11;12(11):e1006427.
8. Carli D, Riberi E, Ferrero GB, Mussa A. Syndromic Disorders Caused by Disturbed Human Imprinting. *J Clin Res Pediatr Endocrinol.* 2020 Mar 19;12(1):1–16.
9. Carpenter BL, Zhou W, Madaj Z, DeWitt AK, Ross JP, Grønbæk K, et al. Mother-child transmission of epigenetic information by tunable polymorphic imprinting. *Proc Natl Acad Sci U S A.* 2018 Dec 18;115(51):E11970–7.
10. Carpenter BL, Remba TK, Thomas SL, Madaj Z, Brink L, Tiedemann RL, et al. Oocyte age and preconceptual alcohol use are highly correlated with epigenetic imprinting of a noncoding RNA (nc886). *Proc Natl Acad Sci U S A.* 2021 Mar 23;118(12):e2026580118.
11. Marttila S, Viiri LE, Mishra PP, Kühnel B, Matias-Garcia PR, Lyytikäinen LP, et al. Methylation status of nc886 epiallele reflects periconceptional conditions and is associated with glucose metabolism through nc886 RNAs. *Clin Epigenetics.* 2021 Jul 22;13(1):143.
12. Treppendahl MB, Qiu X, Søgaaard A, Yang X, Nandrup-Bus C, Hother C, et al. Allelic methylation levels of the noncoding VTRNA2-1 located on chromosome 5q31.1 predict outcome in AML. *Blood.* 2012 Jan 5;119(1):206–16.
13. Romanelli V, Nakabayashi K, Vizoso M, Moran S, Iglesias-Platas I, Sugahara N, et al. Variable maternal methylation overlapping the nc886/vtRNA2-1 locus is locked between hypermethylated repeats and is frequently altered in cancer. *Epigenetics.* 2014 May 6;9(5):783–90.

14. Lee YS. Are We Studying Non-Coding RNAs Correctly? Lessons from nc886. *International Journal of Molecular Sciences*. 2022 Jan;23(8):4251.
15. Fort RS, Garat B, Sotelo-Silveira JR, Duhagon MA. vtRNA2-1/nc886 produces a small RNA that contributes to its tumor suppression action through the microRNA pathway in prostate cancer. *Non-coding RNA*. 2020;
16. Silver MJ, Kessler NJ, Hennig BJ, Dominguez-Salas P, Laritsky E, Baker MS, et al. Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. *Genome Biology*. 2015 Jun 11;16(1):118.
17. Lokk K, Modhukur V, Rajashekar B, Märtens K, Mägi R, Kolde R, et al. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biology*. 2014 Apr 1;15(4):3248.
18. Dugué PA, Yu C, McKay T, Wong EM, Joo JE, Tsimiklis H, et al. Vtrna2-1: Genetic variation, heritable methylation and disease association. *International Journal of Molecular Sciences*. 2021;22(5):1–18.
19. van Dijk SJ, Peters TJ, Buckley M, Zhou J, Jones PA, Gibson RA, et al. DNA methylation in blood from neonatal screening cards and the association with BMI and insulin sensitivity in early childhood. *Int J Obes (Lond)*. 2018 Jan;42(1):28–35.
20. Shaoqing Y, Ruxin Z, Guojun L, Zhiqiang Y, Hua H, Shudong Y, et al. Microarray Analysis of Differentially Expressed microRNAs in Allergic Rhinitis. *Am J Rhinol&Allergy*. 2011 Nov 1;25(6):e242–6.
21. Suojalehto H, Lindström I, Majuri ML, Mitts C, Karjalainen J, Wolff H, et al. Altered MicroRNA Expression of Nasal Mucosa in Long-Term Asthma and Allergic Rhinitis. *IAA*. 2014;163(3):168–78.
22. Sharbati J, Lewin A, Kutz-Lohroff B, Kamal E, Einspanier R, Sharbati S. Integrated microRNA-mRNA-analysis of human monocyte derived macrophages upon *Mycobacterium avium* subsp. *hominissuis* infection. *PLoS One*. 2011;6(5):e20258.
23. Asaoka T, Sotolongo B, Island ER, Tryphonopoulos P, Selvaggi G, Moon J, et al. MicroRNA signature of intestinal acute cellular rejection in formalin-fixed paraffin-embedded mucosal biopsies. *Am J Transplant*. 2012 Feb;12(2):458–68.
24. Lin CH, Lee YS, Huang YY, Tsai CN. Methylation status of vault RNA 2-1 promoter is a predictor of glycemic response to glucagon-like peptide-1 analog therapy in type 2 diabetes mellitus. *BMJ Open Diabetes Res Care*. 2021 Mar;9(1):e001416.
25. Barker DJP, Osmond C. INFANT MORTALITY, CHILDHOOD NUTRITION, AND ISCHAEMIC HEART DISEASE IN ENGLAND AND WALES. *The Lancet*. 1986 May 10;327(8489):1077–81.
26. Kostiniuk D, Tamminen H, Mishra PP, Marttila S, Raitoharju E. Methylation pattern of polymorphically imprinted nc886 is not conserved across mammalia. *PLOS ONE*. 2022 Mar 16;17(3):e0261481.

27. Pilvar D, Reiman M, Pilvar A, Laan M. Parent-of-origin-specific allelic expression in the human placenta is limited to established imprinted loci and it is stably maintained across pregnancy. *Clinical Epigenetics*. 2019 Jun 26;11(1):94.
28. Wilkinson LS, Davies W, Isles AR. Genomic imprinting effects on brain development and function. *Nat Rev Neurosci*. 2007 Nov;8(11):832–43.
29. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002 Jan 1;30(1):207–10.
30. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013 Jan 24;49(2):359–67.
31. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biology*. 2015 Feb 15;16(1):37.
32. Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald R, et al. An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biology*. 2016 Aug 30;17(1):176.
33. Markunas CA, Wilcox AJ, Xu Z, Joubert BR, Harlid S, Panduri V, et al. Maternal Age at Delivery Is Associated with an Epigenetic Signature in Both Newborns and Adults. *PLOS ONE*. 2016 Jul 6;11(7):e0156361.
34. Ventham NT, Kennedy NA, Adams AT, Kalla R, Heath S, O’Leary KR, et al. Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nat Commun*. 2016 Nov 25;7(1):13507.
35. Barbosa M, Joshi RS, Garg P, Martin-Trujillo A, Patel N, Jadhav B, et al. Identification of rare de novo epigenetic variations in congenital disorders. *Nat Commun*. 2018 May 25;9(1):2064.
36. Li S, Wong EM, Joo JE, Jung CH, Chung J, Apicella C, et al. Genetic and Environmental Causes of Variation in the Difference Between Biological Age Based on DNA Methylation and Chronological Age for Middle-Aged Women. *Twin Res Hum Genet*. 2015 Dec;18(6):720–6.
37. Chuang YH, Paul KC, Bronstein JM, Bordelon Y, Horvath S, Ritz B. Parkinson’s disease is associated with DNA methylation levels in human blood and saliva. *Genome Medicine*. 2017 Aug 30;9(1):76.
38. Curtis SW, Cobb DO, Kilaru V, Terrell ML, Kennedy EM, Marder ME, et al. Exposure to polybrominated biphenyl (PBB) associates with genome-wide DNA methylation differences in peripheral blood. *Epigenetics*. 2019 Jan 2;14(1):52–66.
39. Zhang X, Hu Y, Aouizerat BE, Peng G, Marconi VC, Corley MJ, et al. Machine learning selected smoking-associated DNA methylation signatures that predict HIV prognosis and mortality. *Clin Epigenetics*. 2018 Dec 13;10(1):155.

40. Kurushima Y, Tsai PC, Castillo-Fernandez J, Couto Alves A, El-Sayed Moustafa JS, Le Roy C, et al. Epigenetic findings in periodontitis in UK twins: a cross-sectional study. *Clinical Epigenetics*. 2019 Feb 13;11(1):27.
41. Arloth J, Eraslan G, Andlauer TFM, Martins J, Iurato S, Kühnel B, et al. DeepWAS: Multivariate genotype-phenotype associations by directly integrating regulatory information using deep learning. *PLOS Computational Biology*. 2020 Feb 3;16(2):e1007616.
42. Kilaru V, Knight AK, Katrinli S, Cobb D, Lori A, Gillespie CF, et al. Critical evaluation of copy number variant calling methods using DNA methylation. *Genetic Epidemiology*. 2020;44(2):148–58.
43. Robinson O, Chadeau Hyam M, Karaman I, Climaco Pinto R, Ala-Korpela M, Handakas E, et al. Determinants of accelerated metabolomic and epigenetic aging in a UK cohort. *Aging Cell*. 2020 Jun;19(6):e13149.
44. Nuotio ML, Pervjakova N, Joensuu A, Karhunen V, Hiekkalinna T, Milani L, et al. An epigenome-wide association study of metabolic syndrome and its components. *Sci Rep*. 2020 Nov 25;10(1):20567.
45. Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. *PLOS ONE*. 2013 May 17;8(5):e63812.
46. Laaksonen J, Mishra PP, Seppälä I, Raitoharju E, Marttila S, Mononen N, et al. Mitochondrial genome-wide analysis of nuclear DNA methylation quantitative trait loci. *Hum Mol Genet*. 2021 Nov 19;ddab339.
47. Wang Y, Karlsson R, Lampa E, Zhang Q, Hedman ÅK, Almgren M, et al. Epigenetic influences on aging: a longitudinal genome-wide methylation study in old Swedish twins. *Epigenetics*. 2018 Sep 2;13(9):975–87.
48. Hannon E, Knox O, Sugden K, Burrage J, Wong CCY, Belsky DW, et al. Characterizing genetic and environmental influences on variable DNA methylation using monozygotic and dizygotic twins. *PLOS Genetics*. 2018 Aug 9;14(8):e1007544.
49. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol*. 2013 Feb;31(2):142–7.
50. Reynolds LM, Taylor JR, Ding J, Lohman K, Johnson C, Siscovick D, et al. Age-related variations in the methylome associated with gene expression in human monocytes and T cells. *Nat Commun*. 2014 Nov 18;5(1):5366.
51. McRae AF, Powell JE, Henders AK, Bowdler L, Hemani G, Shah S, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biology*. 2014 May 29;15(5):R73.
52. Everson TM, Marsit CJ, Michael O'Shea T, Burt A, Hermetz K, Carter BS, et al. Epigenome-wide Analysis Identifies Genes and Pathways Linked to Neurobehavioral Variation in Preterm Infants. *Sci Rep*. 2019 Apr 19;9(1):6322.

53. Huang JY, Cai S, Huang Z, Tint MT, Yuan WL, Aris IM, et al. Analyses of child cardiometabolic phenotype following assisted reproductive technologies using a pragmatic trial emulation approach. *Nat Commun.* 2021 Sep 23;12(1):5613.
54. van Dongen J, Gordon SD, McRae AF, Odintsova VV, Mbarek H, Breeze CE, et al. Identical twins carry a persistent epigenetic signature of early genome programming. *Nat Commun.* 2021 Sep 28;12(1):5618.
55. Tan Q, Frost M, Heijmans BT, von Bornemann Hjelmberg J, Tobi EW, Christensen K, et al. Epigenetic signature of birth weight discordance in adult twins. *BMC Genomics.* 2014 Dec 4;15(1):1062.
56. Horvath S, Mah V, Lu AT, Woo JS, Choi OW, Jasinska AJ, et al. The cerebellum ages slowly according to the epigenetic clock. *Aging.* 2015 May 11;7(5):294–306.
57. Horvath S, Langfelder P, Kwak S, Aaronson J, Rosinski J, Vogt TF, et al. Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels. *Aging.* 2016;8(7).
58. Bonder MJ, Kasela S, Kals M, Tamm R, Lokk K, Barragan I, et al. Genetic and epigenetic regulation of gene expression in fetal and adult human livers. *BMC Genomics.* 2014 Oct 4;15(1):860.
59. Voisin S, Harvey NR, Haupt LM, Griffiths LR, Ashton KJ, Coffey VG, et al. An epigenetic clock for human skeletal muscle. *J Cachexia Sarcopenia Muscle.* 2020 Aug;11(4):887–98.
60. Sillanpää E, Heikkinen A, Kankaanpää A, Paavilainen A, Kujala UM, Tammelin TH, et al. Blood and skeletal muscle ageing determined by epigenetic clocks and their associations with physical activity and functioning. *Clinical Epigenetics.* 2021 May 17;13(1):110.
61. Tuominen PPA, Husu P, Raitanen J, Luoto RM. Rationale and methods for a randomized controlled trial of a movement-to-music video program for decreasing sedentary time among mother-child pairs. *BMC Public Health.* 2015 Oct 5;15:1016.
62. Roos L, Sandling JK, Bell CG, Glass D, Mangino M, Spector TD, et al. Higher Nevus Count Exhibits a Distinct DNA Methylation Signature in Healthy Human Skin: Implications for Melanoma. *J Invest Dermatol.* 2017 Apr;137(4):910–20.
63. Åsenius F, Gorrie-Stone TJ, Brew A, Panchbhaya Y, Williamson E, Schalkwyk LC, et al. The DNA methylome of human sperm is distinct from blood with little evidence for tissue-consistent obesity associations. *PLoS Genet.* 2020 Oct;16(10):e1009035.
64. Jenkins TG, James ER, Alonso DF, Hoidal JR, Murphy PJ, Hotelling JM, et al. Cigarette smoking significantly alters sperm DNA methylation patterns. *Andrology.* 2017;5(6):1089–99.
65. Green BB, Karagas MR, Punshon T, Jackson BP, Robbins DJ, Houseman EA, et al. Epigenome-Wide Assessment of DNA Methylation in the Placenta and Arsenic Exposure in the New Hampshire Birth Cohort Study (USA). *Environmental Health Perspectives.* 2016 Aug;124(8):1253–60.

66. Paquette AG, Houseman EA, Green BB, Lesueur C, Armstrong DA, Lester B, et al. Regions of variable DNA methylation in human placenta associated with newborn neurobehavior. *Epigenetics*. 2016 Aug 2;11(8):603–13.
67. Konwar C, Price EM, Wang LQ, Wilson SL, Terry J, Robinson WP. DNA methylation profiling of acute chorioamnionitis-associated placentas and fetal membranes: insights into epigenetic variation in spontaneous preterm births. *Epigenetics & Chromatin*. 2018 Oct 29;11(1):63.
68. Bhattacharya A, Freedman AN, Avula V, Harris R, Liu W, Pan C, et al. Placental genomics mediates genetic associations with complex health traits and disease. *Nat Commun*. 2022 Feb 4;13(1):706.
69. Borodulin K, Tolonen H, Jousilahti P, Jula A, Juolevi A, Koskinen S, et al. Cohort Profile: The National FINRISK Study. *Int J Epidemiol*. 2018 Jun 1;47(3):696–696i.
70. Kaprio J. The Finnish Twin Cohort Study: An Update. *Twin Research and Human Genetics*. 2013 Feb;16(1):157–62.
71. Kaprio J, Bollepalli S, Buchwald J, Iso-Markku P, Korhonen T, Kovanen V, et al. The Older Finnish Twin Cohort — 45 Years of Follow-up. *Twin Research and Human Genetics*. 2019 Aug;22(4):240–54.
72. Kaidesoja M, Aaltonen S, Bogl LH, Heikkilä K, Kaartinen S, Kujala UM, et al. FinnTwin16: A Longitudinal Study from Age 16 of a Population-Based Finnish Twin Cohort. *Twin Research and Human Genetics*. 2019 Dec;22(6):530–9.
73. Rose RJ, Salvatore JE, Aaltonen S, Barr PB, Bogl LH, Byers HA, et al. FinnTwin12 Cohort: An Updated Review. *Twin Research and Human Genetics*. 2019 Oct;22(5):302–11.
74. Kovanen V, Aukee P, Kokko K, Finni T, Tarkka IM, Tammelin T, et al. Design and protocol of Estrogenic Regulation of Muscle Apoptosis (ERMA) study with 47 to 55-year-old women's cohort: novel results show menopause-related differences in blood count. *Menopause*. 2018 Sep;25(9):1020–32.
75. Holle R, Happich M, Löwel H, Wichmann HE, Group null for the MS. KORA - A Research Platform for Population Based Health Research. *Gesundheitswesen*. 2005 Aug;67(S 1):19–25.
76. Winkelmann BR, März W, Boehm BO, Zotz R, Hager J, Hellstern P, et al. Rationale and design of the LURIC study - a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics*. 2001 Feb;2(1s1):S1–73.
77. Luoto RM, Kinnunen TI, Aittasalo M, Ojala K, Mansikkamäki K, Toropainen E, et al. Prevention of gestational diabetes: design of a cluster-randomized controlled trial and one-year follow-up. *BMC Pregnancy Childbirth*. 2010 Aug 3;10:39.
78. Pedersen NL. Swedish Adoption/Twin Study on Aging (SATSA), 1984, 1987, 1990, 1993, 2004, 2007, and 2010: Version 2 [Internet]. ICPSR - Interuniversity Consortium

for Political and Social Research; 2015 [cited 2022 Mar 18]. Available from: <http://www.icpsr.umich.edu/icpsrweb/NACDA/studies/3843/version/2>

79. Raitakari OT, Juonala M, Rönnemaa T, Keltikangas-Järvinen L, Räsänen L, Pietikäinen M, et al. Cohort Profile: The Cardiovascular Risk in Young Finns Study. *International Journal of Epidemiology*. 2008 Dec 1;37(6):1220–6.
80. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014 May 15;30(10):1363–9.
81. Touleimat N, Tost J. Complete pipeline for Infinium® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*. 2012 Jun;4(3):325–41.
82. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genom Data*. 2016 May 26;9:22–4.
83. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 2013 May 1;14(1):293.
84. Hernandez Mora JR, Tayama C, Sánchez-Delgado M, Monteagudo-Sánchez A, Hata K, Ogata T, et al. Characterization of parent-of-origin methylation using the Illumina Infinium MethylationEPIC array platform. *Epigenomics*. 2018;10(7).
85. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet Diabetes Endocrinol*. 2015 Jul;3(7):526–34.
86. Wilcox AJ, Lie RT, Solvoll K, Taylor J, McConaughy DR, Abyholm F, et al. Folic acid supplements and risk of facial clefts: national population based case-control study. *BMJ*. 2007 Mar 3;334(7591):464.
87. You YA, Kwon EJ, Hwang HS, Choi SJ, Choi SK, Kim YJ. Elevated methylation of the vault RNA2-1 promoter in maternal blood is associated with preterm birth. *BMC Genomics*. 2021 Jul 10;22(1):528.
88. Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. *Cell*. 2019 Mar 21;177(1):26–31.
89. Gonseth S, Shaw GM, Roy R, Segal MR, Asrani K, Rine J, et al. Epigenomic profiling of newborns with isolated orofacial clefts reveals widespread DNA methylation changes and implicates metastable epiallele regions in disease risk. *Epigenetics*. 2019;
90. Ahmed MK, Bui AH, Taioli E. Epidemiology of Cleft Lip and Palate [Internet]. *Designing Strategies for Cleft Lip and Palate Care*. IntechOpen; 2017 [cited 2022 May 13]. Available from: <https://www.intechopen.com/chapters/53918>

91. Hanna CW, Peñaherrera MS, Saadeh H, Andrews S, McFadden DE, Kelsey G, et al. Pervasive polymorphic imprinted methylation in the human placenta. *Genome Res.* 2016 Jun;26(6):756–67.
92. Cao J, Song Y, Bi N, Shen J, Liu W, Fan J, et al. DNA Methylation-Mediated Repression of miR-886-3p Predicts Poor Outcome of Human Small Cell Lung Cancer. *Cancer Research.* 2013 May 30;73(11):3326–35.
93. Olsen KW, Castillo-Fernandez J, Zedeler A, Freiesleben NC, Bungum M, Chan AC, et al. A distinctive epigenetic ageing profile in human granulosa cells. *Human reproduction* (Oxford, England). 2020;
94. Eppig JJ. Oocyte control of ovarian follicular development and function in mammals. *Reproduction.* 2001 Dec 1;122(6):829–38.
95. Gilchrist RB, Lane M, Thompson JG. Oocyte-secreted factors: regulators of cumulus cell function and oocyte quality. *Human Reproduction Update.* 2008 Apr 1;14(2):159–77.
96. Prättälä R. Dietary changes in Finland--success stories and future challenges. *Appetite.* 2003 Dec;41(3):245–9.
97. Erkkola M, Karppinen M, Järvinen A, Knip M, Virtanen SM. Folate, vitamin D, and iron intakes are low among pregnant Finnish women. *Eur J Clin Nutr.* 1998 Oct;52(10):742–8.
98. Steegers-Theunissen RPM, Twigt J, Pestinger V, Sinclair KD. The periconceptional period, reproduction and long-term health of offspring: The importance of one-carbon metabolism. *Human Reproduction Update.* 2013;
99. Becker W, Lyhne N, Pedersen AN, Aro A, Fogelholm M, Phórsdóttir P, et al. Nordic Nutrition Recommendations2004 - integrating nutrition and physical activity. *Food & Nutrition Research.* 2004;178–87.
100. Zeng Y, Chen T. DNA methylation reprogramming during mammalian development. Vol. 10, *Genes.* 2019.