Ribosomal DNA copy number is associated with body mass in humans and other mammals

Authors: Pui Pik Law[1,2*], Francisco Rodriguez-Algarra[2*], Fredrika Asenius[3], Maria Gregori[3] Robert AE Seaborne[2†], Selin Yildizoglu[2], James RC Miller[1], Robin Mesnage[1], Michael N Antoniou[1], Weilong Li[4], Qihua Tan[5], Sara L Hillman[3], Vardhman K Rakyan[2], David J Williams[3], Michelle L Holland[1].

Key words: Body mass, genomics, genetics, ribosomal DNA, repetitive elements

Affiliations:

[1]Department of Medical and Molecular Genetics, School of Basic and Medical Biosciences, King's College London, UK.

[2]The Blizard Institute, School of Medicine and Dentistry, Queen Mary University of London, UK.

[3]UCL EGA Institute for Women's Health, University College London, London, UK.

[4]Population Research Unit, University of Helsinki, Helsinki, Finland.

[5]Epidemiology, Biostatistics and Biodemography, Department of Public Health, University of Southern Denmark, Denmark.

†Present address: Centre for Human and Applied Physiological Studies, King's College London, UK.

Corresponding author. Email: michelle.holland@kcl.ac.uk

*Equal contribution

SUMMARY

Body mass results from a complex interplay between genetics and environment. The contribution of genetic variation to body mass has been extensively studied, but due to the technical limitations of platforms used for population scale studies, repetitive parts of the genome have not previously been considered. Here we applied genome-wide approaches to identify an association between adult body mass and the copy number (CN) of 45S-ribosomal DNA (rDNA). rDNA codes for the rRNA components of the ribosome and exists in hundreds of copies/cell in mammals. Inter-individual variation in rDNA CN has not previously been documented but has not been associated with a mammalian phenotype. Here, we show that germ-line inherited rDNA CN is associated with post-pubertal growth rate in rats and BMI in adult humans. rDNA CN variation is not associated with rRNA transcription rates in adult tissues, suggesting the mechanistic link occurs earlier in development. This is supported by our observation that it is growth rate and body mass, rather than obesity that are associated with copy number and that these phenotypes emerge prior to adulthood. We present hypotheses for future investigation into the mechanistic basis of this association.

INTRODUCTION

Lifestyle changes have driven a relentless increase in the incidence of obesity[1]. Current interventions have proven insufficient to curb this trend, therefore, understanding the basis of an individual's response to an obesogenic environment is of great interest. Genome-wide association studies (GWAS) have contributed to our understanding of the genetic influence over body mass, yet so far only part of the heritability estimated by family and twin studies can be explained[2,3]. Epigenetic mechanisms have also been explored in depth but largely using array technologies that only partially capture the DNA methylome[4]. Due to technical limitations, repetitive parts of the genome, such as 45S-ribosomal DNA (rDNA) have been excluded from such studies[5]. Here we have utilized whole-genome approaches, inclusive of repetitive regions and have identified an association between rDNA copy number (CN) and body mass in humans and rats.

RESULTS

Obese individuals have lower rDNA copy number

Previously we identified hypermethylated rDNA in inbred adult mice that had been exposed to protein restriction (PR) via maternal diet from the period of conception to weaning[6,7]. This exposure resulted in growth restriction. Intriguingly, the hypermethylation of rDNA was inversely correlated with the weight of the PR-exposed animals at the end of the exposure period. This raised the question of whether the functional genomics of rDNA may be linked to body weight regulation.

To address if altered DNA methylation of repetitive genomic elements, inclusive of the rDNA is associated with body mass index (BMI) in humans, we analysed whole blood from lean (BMI<25 kg/m^2) or obese (BMI>30 kg/m^2) human males (**Fig. 1A,** $p<0.0001$) using whole genome bisulfite sequencing (WGBS). There was no difference in the mean age of the lean and obese groups (~36 years old, Table S1) and no history of diagnosed comorbidities or medication use across the entire cohort. However, there were significant differences between the lean and obese groups relating to anthropomorphic measurements, blood pressure and serum markers (Table S1), consistent with the obese group having a metabolic syndrome associated with higher risk of cardiovascular disease and Type 2 Diabetes[8].

The WGBS data was mapped to a reference genome modified to include a representative copy of the rDNA since the rDNA clusters present on chromosomes 13, 14, 15, 21 and 22 in humans are not included in the genome assembly[5,9]. The sequencing depth of these data (~14x), is sufficient to produce high resolution quantitation of DNA methylation when considering groups of genomic features, or mapping to a consensus for a multi-copy genomic element, such as the rDNA (Table S2). When all reads mapping to genomic features such as exons, introns or a range of repetitive genomic elements were collectively considered, there was no difference in DNA methylation between the lean and obese groups (Fig. S1), with the exception of rDNA, for which methylation was ~8% less across the entire promoter and transcriptional unit (~ 14 kB) in the obese compared to lean cohort (**Fig. 1B**, p= 0.0089 & **Fig. 1C**). This suggests that rDNA is the only genomic feature, when collectively considered, that demonstrates altered DNA methylation in association with obesity.

As we and others have previously shown that DNA methylation levels at rDNA are associated with inter-individual variation in rDNA CN[10-12], we next queried if this epigenetic-genetic interaction is also observed in this cohort. rDNA CN was estimated from the WGBS data using an established approach and after applying strict criteria for data quality control[13,14] (Table S3). Consistent with previous reports, we observed that rDNA CN and DNA methylation are highly correlated in blood, such that individuals with higher rDNA CN also have more rDNA methylation (**Fig. 1D**, r=0.75, p<0.0001). This relationship was present in both the lean (Spearman r=0.5738, p=0.0007, n=31) and obese (Spearman r=0.7870, p<0.0001, n=32) groups when considered separately as well as overall. Reconcilable with the observation of hypomethylation of rDNA in the obese group, rDNA CN was also significantly lower in the obese compared to lean group (**Fig. 1E**, p=0.0024). This was not accounted for by differences in the ethnic composition of the lean and obese groups (Table S4), or sequencing parameters (Fig. S2). Furthermore, the trend for lower CN in obese individuals was observed in multiple ethnicities (Fig. S3). These observations demonstrate that total rDNA methylation and CN are positively correlated in humans and that rDNA CN variation in blood is associated with variation in a complex trait in humans, BMI.

The rDNA CN association with BMI is not due to cell type differences but interventions targeting metabolic phenotypes affect the genotype-phenotype association

To validate the observed association between rDNA CN and BMI, we re-analysed published reduced representation bisulfite sequencing data (RRBS) derived from the adipose tissue of Finnish males 45-67 years of age from the METSIM study[15,16]. We devised a methodology for estimating rDNA CN from RRBS for determining relative (rather than absolute) CN across individuals. The method was cross-validated with the previously published method used for the WGBS data above, which in turn we have previously validated using an independent methodology, digital droplet PCR[10] (and Fig. S4). In this cohort, we also observed a strong positive correlation between rDNA CN and methylation, confirming its presence in at least two different human tissues (Fig. S5).

The individuals in the METSIM cohort were not selected for clinical obesity (BMI>30 kg/m^2). Therefore, a continuous spectrum of BMI is represented with the majority of individuals included in this analysis classified as clinically "overweight" (25<BMI<30 kg/m^2). As such, we performed a correlation analysis between the relative rDNA CN and an individual's BMI, again revealing a negative correlation that passed a nominal significance threshold (**Fig. 2A**. r=-0.18, p=0.02). Consistent with this observation, rDNA methylation also demonstrated a negative correlation with BMI, although the effect did not pass the nominal threshold of p<0.05 (**Fig. 2B,** r=-0.13, p=0.09). As this cohort is older and has a larger age range than the discovery cohort, we calculated age-adjusted BMI and found that this did not have a significant effect on the correlation between rDNA CN or methylation with BMI (Fig. S6).

Ageing[17] and body mass[18] have been associated with changes in blood cell proportions and adipose infiltration[19]. However, total rDNA CN has previously been shown to be consistent across multiple tissues in mice[20]. To the best of our knowledge, cell-type specific variation in rDNA CN has not been characterised in humans. We therefore enquired whether there was any cell-type specific variation in rDNA CN using whole genome bisulphite sequencing (WGBS) data derived from multiple purified blood (Fig. S7) and other cell types from donors[21]. This confirmed that although there is extensive inter-individual rDNA CN variation, there is no systematic rDNA CN difference across cell types or evidence of substantial somatic drift (Table S5). Therefore, the rDNA CN association with BMI is not a downstream artifact of altered blood cell composition across samples.

The METSIM validation cohort has a significant proportion (69/169) of individuals regularly taking one or more medications, most commonly statins being. Therefore, we examined the

association of medication status with anthropomorphic and metabolic measurements. Although there was no difference in age or BMI in the medicated compared to the non-medicated groups, the medicated group had elevated measurements for waist circumference, waist to hip ratio, diastolic blood pressure, fasting insulin levels and the Homeostatic Model Assessment for Insulin Resistance (HOMA-IR) (Table S6). Collectively, these findings indicate a higher prevalence of metabolic syndrome in these individuals[8]. However, unlike the younger discovery cohort, this was not reflected in higher low-density lipoprotein levels, consistent with mitigation by statin use[22].

To investigate whether medication or other interventions that modify anthropomorphic and metabolic phenotypes might influence the strength of genotype-phenotype associations, we next considered the medicated and non-medicated groups separately. Intriguingly, the correlation between BMI and both rDNA CN and methylation strengthened when the non-medicated group was considered separately (**Fig. 2C**, r=-0.30, p=0.0025 & **2D,** r=-0.25, p=0.01), but disappeared when the medicated group were analysed alone (**Fig. 2E**, r=0.05, p=0.71 & **2F**, r=0.07, p=0.59). This could not be explained by differences in sequencing quality (Fig. S8) after sample exclusion based on strict QC of the sequencing data (Table S7). Similar findings were observed after adjusting BMI for age (Fig. S9). Taken together, these results support a quantitative relationship between rDNA CN and BMI in adults. This has been observed in two independent, ethnically different populations and holds across two different tissues. However, we find that lifestyle interventions that influence phenotype can reduce the genotype-phenotype correlation.

Comparison of the relationship between rDNA CN and the anthropomorphic and metabolic traits captured in both cohorts suggests that body mass is the primary measured trait associated with rDNA CN variation, as BMI and waist circumference were the only variables significantly associated across both the extreme discovery and METSIM validation cohorts (Table S8). Other variables did show cohort-specific negative correlations with rDNA CN, such as C-reactive protein, fasting insulin levels and the HOMA-IR exclusively in the discovery cohort and diastolic and systolic blood pressure in the unmedicated subgroup of the replication cohort. As these are all established sequelae of being overweight, it is likely this is explained by differences in the age and phenotypic extremes between the cohorts. rDNA methylation was less strongly correlated with BMI and other traits than rDNA CN, supporting

the conclusion that it is the genetic rDNA CN association with BMI that is driving the altered DNA methylation profiles we initially observed.

rDNA CN is not discordant in twins with divergent BMI

The observation that rDNA CN negatively correlates with adult BMI raises questions about the origin of the association. Two scenarios are plausible, i) germline inherited rDNA CN, through an unknown mechanism, can influence BMI variation, or ii) environmental exposures and/or metabolic changes occurring concomitantly with increasing BMI may lead to rDNA (epi)genetic instability and rDNA CN loss. To our knowledge, there is no precedent for a human trait associated with germline inherited rDNA CN, but rDNA CN changes have been observed in human cancers[23,24].

To address the issue of causation, we obtained RRBS data derived from whole blood of monozygotic twins discordant for BMI and of a single ethnicity. The original study excluded individuals with diagnosed comorbidities[25]. The cohort characteristics for the 10 male and 14 female monozygotic twin pairs included in this analysis after data quality control (Table S9, Fig. S10) are shown in Table S10. The twins were confirmed to have highly discordant BMI (**Fig. 3A,** $p<0.0001$), with both the leaner and heavier twins spanning all clinical BMI group classifications. However, leaner co-twins were enriched in the clinically "lean" range (BMI<25 kg/m^2), and the heavier co-twins enriched in the clinically "overweight" range (25 kg/m^2<BMI<30 kg/m^2). Despite discordance in BMI, there was no difference observed for the rDNA CN within twin pairs (**Fig.3B**, $p=0.81$). Consistent with this, there was not any association between within-twin rDNA CN variation and twin age, or BMI discordance (Fig. S11). As with the previous cohorts, there was a positive correlation between rDNA CN and methylation (Fig. S12), and in line with this, there was no rDNA methylation differences between discordant co-twins (**Fig. 3C**, $p=0.70$). These results, together with the cell type specific analysis above, support the idea that rDNA CN is somatically stable and favours the theory that germline inherited rDNA CN is associated with adult BMI, rather than environmentally driven effects on adult BMI acting to cause somatic drift in rDNA CN.

rDNA CN is not associated with genetic variation in other parts of the genome that are implicated in body mass regulation

As BMI has a strong genetic as well as environmental component, we next sought to address whether there is a direct association between rDNA CN and previously identified single nucleotide variants (SNVs) associated with BMI from GWAS. To this end, we retrieved summary statistics from a meta-analysis of BMI including ~ 700 000 individuals[26] and utilised these as base data for calculating BMI polygenic risk score (PRS) for individuals in the 1000 Genome Project[27]. Only individuals which have also independently been determined to have WGS data of sufficient quality for rDNA CN estimation were included[14]. PRS scores calculated using only the highly significant, near-independent SNVs previously identified[26], or a more relaxed significance threshold produced highly correlated PRS scores, as expected (Fig. S13). We then asked if the PRS calculated from BMI-associated SNVs explains variance in rDNA CN. We did not find any models that could explain variance in rDNA CN (top model ($p<0.005$ threshold) goodness of fit $R^2=0.00063$, $p=0.19$). This is reflected when the PRS scores are directly correlated with the rDNA CN estimates (**Fig. 3D**, $r=-0.03$, $p=0.19$). In an alternative approach, we also performed a GWAS for rDNA CN in the same cohort. Only one SNV was identified that passed the nominal $p<5 \times 10^{-8}$ threshold for genome-wide correction (Chr5: 159614102, $p=3.7\times10^{-8}$). This position is only variable in some East Asian populations[28], suggesting that it is unlikely to be confounding results in the predominantly European discovery and validation cohorts. Furthermore, this SNV has not previously been associated with body mass or related traits. Taken together, these findings suggest that rDNA CN is not strongly influenced by inter-individual genetic differences in other parts of the genome that contribute to BMI variation.

rDNA CN is associated with post-pubertal growth rate in rats

We next addressed whether an association between rDNA CN and body mass also exists in non-human mammals. To address this, we leveraged published RRBS data generated from the liver of female, outbred (Sprague-Dawley) rats[29,30]. The authors made available weight data that was longitudinally collected throughout the experimental period, weeks 8-19 of age. This period encompasses the onset of sexual maturity in this strain (9-10 weeks of age for females) and is still within a period of growth, which has been reported to extend to 24 weeks of age[31].

After excluding some individuals based on strict quality control of the RRBS data (Fig. S14 and Table S11), we further verified that there was no effect of the study treatments on either

weight, rDNA CN or methylation (Fig. S15). Plotting the longitudinal weight data confirmed that this was still a period of active growth (**Fig. 4A**). The availability of longitudinally collected data allowed us to query whether growth rate correlates with rDNA CN. Indeed, we observed a negative correlation between the weight gained over the study period and rDNA CN (**Fig. 4B**, r=-0.42, p=0.0047), but not methylation (Fig. S16), despite a positive correlation being observed between rDNA CN and methylation, as in all other analysed datasets (Fig. S17). Despite overall weight gain showing a negative correlation with rDNA CN, cross-sectional analysis of each time point demonstrated that this correlation only emerged towards the end of the experimental period, reaching significance at weeks 18 and 19 (**Fig. 4A** and Table S12). Indeed, at these timepoints, we observed a negative correlation between rDNA CN and the body mass of the animals (**Fig. 4C**, r=-0.35, p=0.0206). Collectively, these findings suggest that the correlation between rDNA CN and body mass is not unique to humans and furthermore, becomes manifest between the ages of puberty and early adulthood.

rDNA CN is not associated with altered nascent transcription in adult tissues

Having demonstrated an association with rDNA CN and body mass regulation across two mammalian species, we next asked whether rDNA CN variation is associated with nascent rRNA transcript levels in adult somatic tissues. Previously, we have shown that expression of specific, well-defined rRNA haplotypes in mice reflects their relative contribution to total rDNA CN after adjusting for the silencing of the methylated copies[10]. Similar haplotypes in humans are yet to be defined. Comparing the total rDNA copy number before and after adjusting for methylation in three independent data sets derived from inbred mice, outbred mice and human lymphoblastoid cell lines (LCLs) did not reveal an association with nascent rDNA transcription rates (Fig. S18). This supports the conclusion that the association between rDNA CN and body mass is not due to altered transcription rate in mature tissues and if the mechanism is associated directly with rDNA transcription, then this occurs earlier in development, prior to compensation of rDNA CN variation through the initiation of epigenetic silencing upon the initiation of differentiation[32]. This finding agrees with the timing at which the emergence of the genotype-phenotype association was observed in the rats.

DISCUSSION

rDNA CN is highly variable in humans[9,10,14]. This variation is thought to derive from a very high frequency of meiotic rearrangement at the rDNA repeats[33]. However, germline genetic variation at rDNA has not, to the best of our knowledge, been studied in relation to human trait variation as it requires sequencing data rather than genotyping arrays. Here we provide evidence that germ-line variation in rDNA CN is associated with body mass in mammals using multiple cohorts and tissues, and both extreme and continuous phenotypes.

Somatic rDNA CN instability has been reported previously in human diseases, such as cancer and neurodegenerative disorders[20,23,24,34]. However, in these cases there is clear evidence that rDNA CN variation is consequential of the disease process and shows no directional association, or predictive value. There are some rare examples of cancers with specific genetic lesions that do lead to tumour-specific rDNA CN loss[20,35]. These are restricted to epigenetic regulators important for the maintenance of heterochromatin at repetitive genomic regions, including the rDNA[35] or tumour suppressors involved in chromatin stability[36]. However, the evidence from the monozygotic twins and multiple cell types from the same donors analysed here suggests that rDNA CN is largely somatically stable.

A caveat of our findings is that at this stage we are limited to identifying an association between rDNA CN and body mass. Changes in inflammatory cell types are known to occur in obesity[37]. However, our analyses demonstrate that rDNA CN does not systematically vary with cell type and furthermore, the association of rDNA CN with growth rate in rats in the absence of an obesogenic environment suggests that the origin of the association is not with obesity *per se*, but rather through a mechanism that more fundamentally regulates organismal growth. Interestingly, a human longitudinal study has recently linked the rate of BMI change during the post-pubertal to young adult ages to higher risk for obesity in mid-life, independently of the actual BMI during this period[38]. This implies that factors influencing growth rate during this period are associated with later disease risk.

Our data suggests that epigenetic silencing, associated with DNA methylation compensates for higher rDNA CN, essentially serving to “normalise” the rate of rRNA transcription in adult tissues. This, together with the phenotype-genotype association emerging by early adulthood, suggests that the physiological basis for the association between rDNA CN and

BMI may occur earlier in development. The dynamic regulation of rDNA transcription is essential for embryonic development in mammals[32], with pluripotency requiring a lack of rDNA silencing to produce a translational state essential for maintaining stemness and inhibiting genes required for differentiation[36,39]. Silencing of the rDNA units begins with cellular differentiation but stabilises slowly. Could then rDNA CN influence cell fate decisions at these earlier timepoints? Interestingly, disease caused by mutations in ribosomal proteins producing a reduction in total ribosome levels, without a change in composition have been shown to influence cell fate decisions through selectively influencing translation of lineage specifiers[40]. This raises the intriguing possibility that variation in rDNA CN may produce subtle effects on cell commitment in development which ultimately impact growth trajectories later on. An alternative hypothesis is based on observations from Drosophila, where variations in 35S-rDNA CN have been shown to influence the expression of other genes by altering genome-wide chromatin structure[41,42]. Exploring these hypotheses is beyond the scope of the work presented here and will be challenging, as rDNA CN in mammals is not amenable to manipulation by reverse genetic approaches, even in the era of CRISPR. This is due to the large tandem arrays of rDNA with very high sequence homology, combined with a lack of knowledge of their organisation beyond the recently released telomere-to-telomere genome assembly which is derived from a single cell line from a hydatiform mole that is homozygous[5]. Nonetheless, the first demonstration that rDNA CN may be associated with a biomedically relevant human phenotype provides the impetus for such investigations in the future.

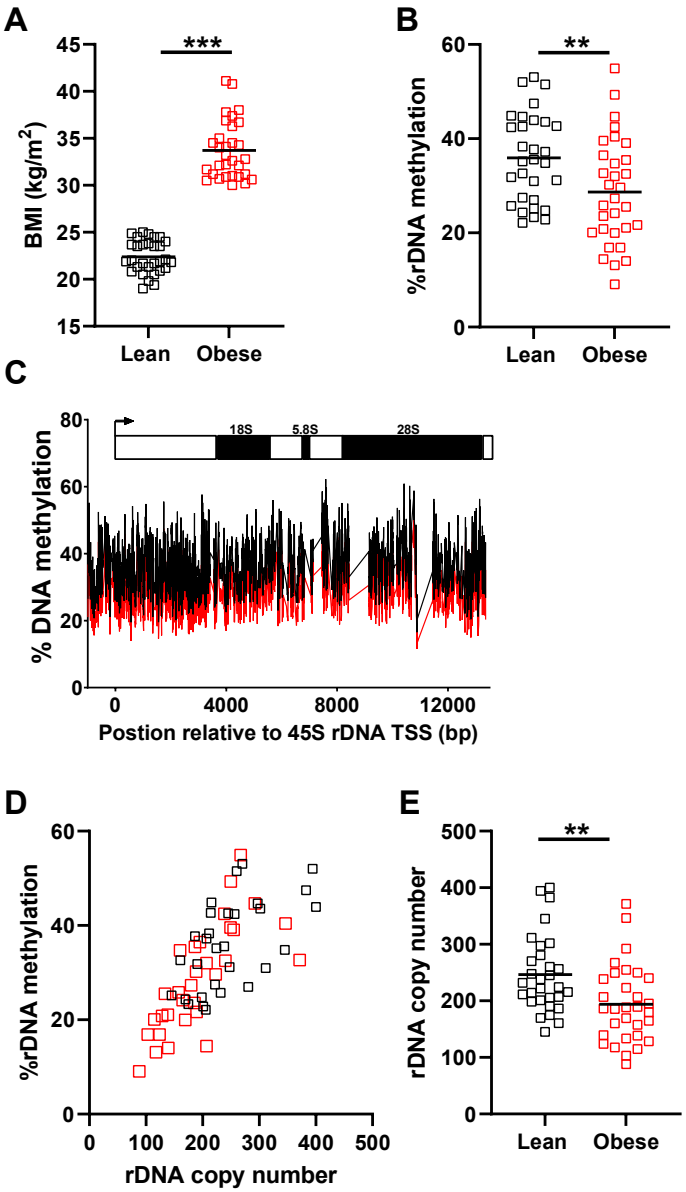

Figure 1. Methylation of rDNA is positively correlated with copy number in blood and associated with obesity. Lean (BMI<25 kg/m^2, n =30, black), Obese (BMI>30 kg/m^2, n=31, red), mean indicated **(A)** Body mass index (BMI) distributions for males in an age-matched cohort are different (Mann-Whitney test, p<0.0001). **(B)** 45S rDNA is hypomethylated in obese males. Average DNA methylation from whole-genome bisulfite sequencing data (Mann-Whitney test, p = 0.0089). **(C)** Mean methylation is lower in obese males across the entire promoter and coding region for the rDNA unit. **(D)** Average rDNA methylation estimated across the promoter (-1000 bp upstream) and the entire transcribed region is positively correlated with rDNA copy number (Total cohort: Spearman r=0.7613, p<0.0001, n=61; Lean only: Spearman r=0.5942, p=0.0005; Obese only: Spearman r=0.8081, p<0.0001). **(E)** Obesity is associated with lower total rDNA copy number within the multi-ethnic cohort (Mann-Whitney test, p=0.0024).

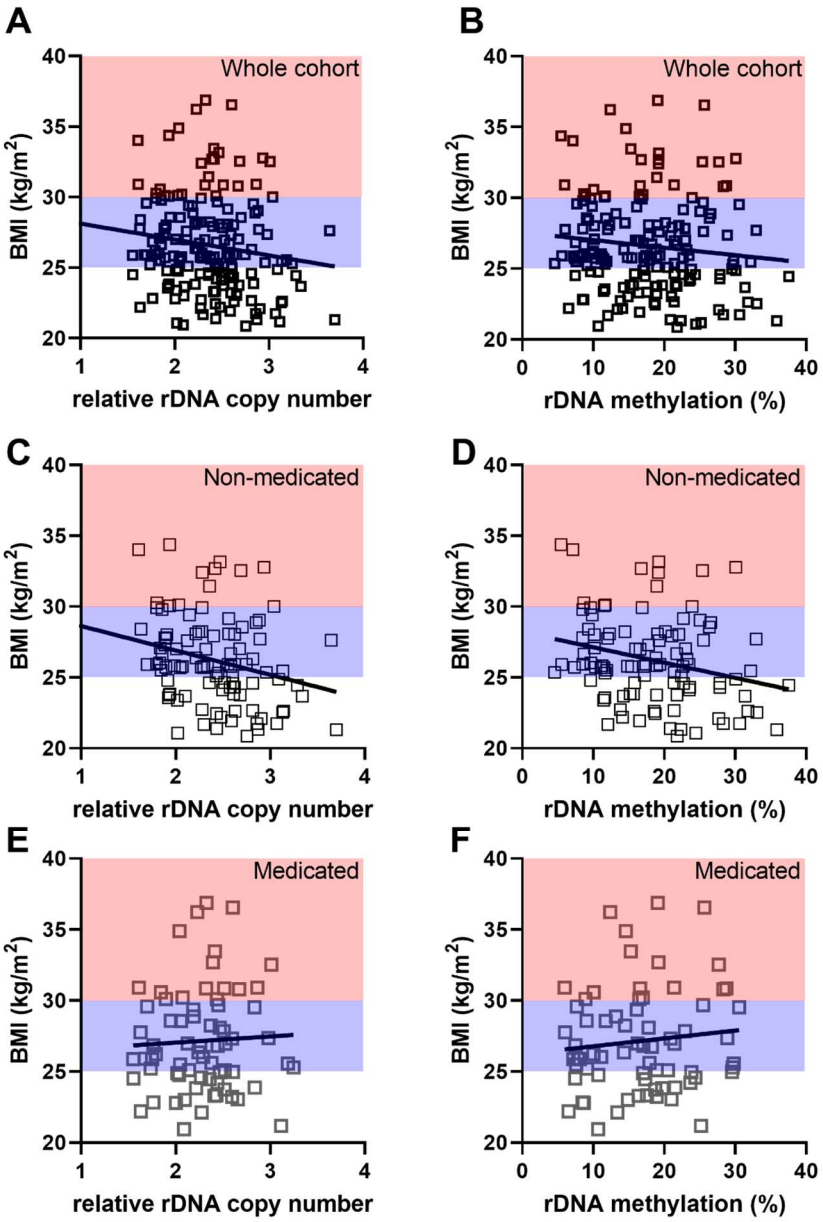

Figure 2. rDNA copy number negatively correlates with BMI but the association is diminished by lifestyle interventions. BMI clinical classifications are indicated (Lean (BMI<25 kg/m^2) black, Overweight (25 kg/m^2<BMI<30 kg/m^2) blue, Obese (BMI>30 kg/m^2) red). **(A)** Relative rDNA copy number is negatively correlated BMI when the entire cohort is considered (Spearman r=-0.1785, p=0.0202, n=169). **(B)** rDNA methylation is not correlated with BMI when the entire cohort is considered (Spearman r=-0.1320, p=0.0872, n=169). **(C)** Relative rDNA copy number is negatively correlated with BMI when only non-medicated individuals are considered (Spearman r=-0.2992, p=0.0025, n=100). **(D)** rDNA methylation is negatively correlated with BMI when only non-medicated individuals are considered (Spearman r=-0.2487, p=0.0126, n=100). **(E)** Relative rDNA copy number is not correlated with BMI when only non-medicated individuals are considered (Spearman r=0.04607, p=0.7070, n=69). **(F)** rDNA methylation is not correlated with age-adjusted-BMI when only medicated individuals are considered (Spearman r=0.06675, p=0.5858, n=69).

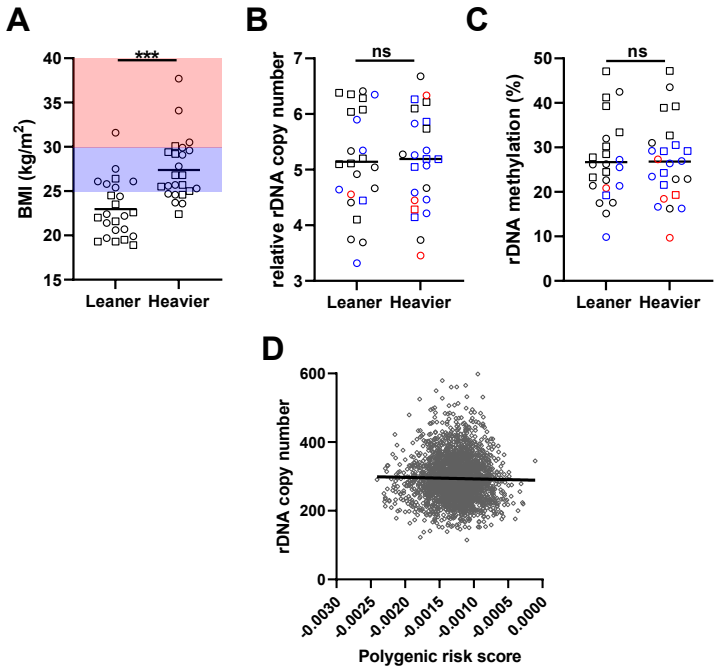

Figure 3. BMI does not induce rDNA copy number variation and is not influenced by BMI-associated genetic variation in the rest of the genome. BMI clinical classifications are indicated (lean (BMI<25 kg/m^2) black, overweight (25 kg/m^2<BMI<30 kg/m^2) blue, obese (BMI>30 kg/m^2) red. Sex is indicated by symbol shape, circle = female (14 MZ twin pairs), square = male (10 MZ twin pairs). **(A)** BMI is discordant between twins (Wilcoxon matched-pairs signed rank test, p<0.0001) and significantly paired (Spearman r=0.8723, p<0.0001). **(B)** Relative rDNA copy number is not different between twins (Wilcoxon matched-pairs signed rank test, p = 0.8115) and is significantly paired (Spearman r = 0.9670, p<0.0001). **(C)** rDNA methylation is not different between twins (Wilcoxon matched-pairs signed rank test, p = 0.7048) and is significantly paired (Spearman r = 0.9783, p<0.0001). **(D)** BMI polygenic risk scores are not correlated with rDNA copy number (Spearman r= -0.02701, p=0.1868, n=2390).

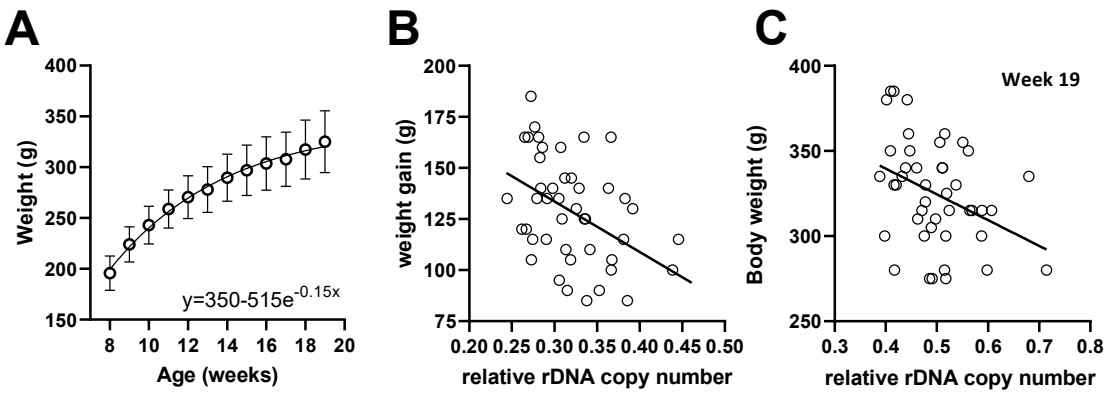

Figure 4. Weight gain from puberty to early adulthood negatively correlates with rDNA copy number in Sprague-Dawley rats. RRBS data derived from liver of female Sprague-Dawley rats at the time of sacrifice. **(A)** Weekly weight data fits an exponential plateau curve (R^2=0.7232). **(B)** Weight gain (week 8-19) is negatively correlated with rDNA copy number (Spearman r= -0.4116, p=0.0055, n=44). **(C)** Absolute weight is negatively correlated with rDNA copy number at week 19 (Spearman r= -0.3481, p=0.0206, n=44).

MATERIALS AND METHODS

Sample information

Mixed ethnicity obese discovery cohort: Whole blood samples were collected from participants recruited from University College London Hospital (UCLH) May 2016—March 2019. Participants were phenotyped with regards to BMI, waist circumference, systolic and diastolic blood pressure, blood lipids, fasting insulin and glucose levels and C-reactive protein (CRP). Two groups of participants were included; lean (BMI <25kg/m^2) and obese (BMI >30kg/m^2). Summary phenotypic data for each group is detailed in Table S1. BMI was determined with light clothing, by a trained researcher in the same visit at which blood samples were obtained. Peripheral blood samples were centrifuged at 3000g for 15 minutes within one hour of venepuncture and the buffy coat stored at -80^{0}C.

Ethics approval and consent to participate: Ethical approval for the study was granted from the South East Coast—Surrey Research Ethics Committee on 28 September 2015 (REC reference number 15/LO/1437, IRAS project ID 164459). The study was also registered with the University College London Hospital Joint Research Office (Project ID 15/0548). All participants provided written, informed consent.

Data Generation

DNA extraction: DNA was extracted from 200 ml of buffy coat using the Qiagen QIAamp DNA Blood Mini Kit (Qiagen, Cat No. 51106) according to the manufacturer's instructions and including RNA digestion. Purity of the extracted DNA was confirmed on a Nanodrop (ThermoFisher, cat. No ND-ONEC-W) and the concentration determined using the QuBit dsDNA HS Assay Kit (ThermoFisher, Cat No. Q32854).

Whole genome bisulfite sequencing library construction and sequencing: Genomic DNA was diluted to 10 ng/μl, and 100 μl sonicated using a Bioruptor® Pico (Diagenode, Cat No. B01060010) to achieve a 500-600 bp size range, which was confirmed using a TapeStation High Sensitivity D1000 System (Agilent, Cat No. 5067-5584 & 5067-5585). Once the desired size range was achieved, 200 ng of sonicated DNA was subjected to bisulfite conversion using the EZ DNA Methylation-Gold™ Kit (Zymo, Cat No. D5006). Libraries were then made using

the Accel-NGS Methyl-Seq DNA Library Kit with unique dual indices following the size-selection guidance provided with the kit and 10 cycles of amplification to minimise clonality. The removal of all adaptors and quantification were confirmed with TapeStation and QuBit before libraries were pooled into equimolar 12-plex pools and subjected to 150 bp paired-end sequencing on a NovaSeq6000 (GeneWiz).

Mouse digital droplet PCR for rDNA copy number: This assay was performed as described previously[10].

Mouse nascent 45S-rRNA qRTPCR: This assay was performed on kidney samples from mice of different strains as described previously[7].

Human nascent 45S-rRNA qRTPCR: RNA was isolated using TRIzol as per manufacturer's instructions, quality assured via RNA nano bioanalyser (Agilent) and 500 ng of total RNA was reverse transcribed using random hexamers (NEB ProtoScript®II). Real-time qPCR was performed using QuantiTect SYBR Green qPCR mix (Qiagen). Primers to amplify the precursor of the human rRNA and the housekeeping control ACTB were taken from previously published work and GAPDH primers designed as F- 5'-CCATCACCATGTTCCAGGAG-3', and R- 5'-CCTGCTTCACCACCTTCTTG-3' [43,44].

External data sources

Justification for cohort selection

rDNA is only captured using long or short-read sequencing approaches. Therefore, for methylation analysis, we were restricted to selecting cohorts analysed using bisulfite-sequencing based approaches. Furthermore, the strong positive correlation between rDNA CN and methylation provided a useful additional quality control. This additional quality control is absent in WGS datasets. Therefore, we limited our analysis of WGS data sources for which the rDNA CN had been previously established and rigorously quality assessed[14].

Validation cohort (METSIM Adipose tissue): Raw RRBS data for this cohort was downloaded from GEO (GSE87893). Limited phenotype data for the samples included were made available through collaboration[15]. Summary phenotypic data from this cohort for the samples included in these analyses can be found in Table S6.

Monozygotic twin cohort: Raw reduced representation bisulfite sequencing data from the monozygotic twin cohort was made available through collaboration and similarly, can be

made available by request to Q. Tan[25]. Data is derived from whole blood from twins that have no diagnosed illness or medications. Summary phenotypic data from this cohort for the twin pairs included in this analysis are provided in Table S10.

1000 Genome Project: rDNA CN estimates and single nucleotide variant calls (SNV) were obtained from published sources[14,45].

Rat data with longitudinal weights: Raw RRBS data for this cohort was downloaded from GEO (GSE157551). Longitudinal weight measurements were made available through collaboration[29,30].

Methylation atlas: WGBS data generated from sorted and pure human cell populations was retrieved from the European Genome Archive (Dataset ID EGAD00001009789). Only individuals with multiple cell types were included in the analyses. This left 67 samples from 18 donors and included at least 26 cell types represented by 2 or more donors.

Mouse data for rDNA copy number and/or methylation quantitation: This data if not described above has been generated as part of previously published work[7,10].

Data Analysis

Reference sequences: Genomic reference sequences used throughout this study were generated as follows to alleviate potential coverage loss and spurious alignments. The repetitive element appearing in the rDNA IGS closest to the 3' end of the unit was identified from the publicly available annotations for the human rDNA unit reference (Genbank accession KY962518.1). The midpoint of this repetitive element, which is located 2120 base pairs upstream of the TSS, was employed as breakpoint for creating a "looped" rDNA unit. In particular, the bases downstream of the breakpoint up to the end of the rDNA unit reference were prepended to the bases upstream of the breakpoint, which improves read coverage around the TSS by avoiding reads being discarded due to split alignments.

To minimise the risk of sequencing reads from locations outside the rDNA being spuriously mapped to the rDNA, identified rDNA pseudocopies in the Hg38 assembly were masked, and the "looped" rDNA reference mentioned above was appended. Masked regions and their

genomic coordinates, including an entire rDNA unit located in an unplaced contig (Genbank accession GL000220.1) are shown in Table S13.

A human exome reference was obtained and adapted as described previously [13], with exon sequences and their annotations being downloaded from the EMBL/EBI repository. In particular, exons from the sex chromosomes and smaller than 300 bases were removed, as were sequences with significant similarity as reported by blastn version 2.7.1+ in --ungapped mode. This left a total of 12,898 exon sequences in the adapted reference.

In the case of rat data, the most complete rDNA consensus sequence available presently spans solely from the 5' end of the 18S to the 3' end of the 28S (Genbank accession V01270.1). A blastn comparison between V01270.1 and the rat whole genome assembly Rn7 revealed several apparent partial rDNA pseudocopies scattered across the assembly, plus three end to end matches. These three complete pseudocopies, whose coordinates are indicated in Table S13, were thus masked. To minimise the potential detrimental effects of the partial pseudocopies, only the sequence corresponding to the 18S (positions 1 to 1874) was appended to the masked Rn7 assembly as an additional contig.

Sequencing data processing: Initial processing of data was performed using fastqc version 0.11.9 to identify failed libraries for exclusion. Data was then trimmed for base quality and adaptor removal using trimgalore version 0.6.5, with the –-paired and --rrbs options enabled when appropriate. Remaining parameters were set to the default values.

Alignments to the reference sequence were performed using bismark version 0.23.0 with underlying bowtie2 for the WGBS and RRBS data sets. Bisulfite conversion of the reference sequence for WGBS/RRBS alignments was performed using bismark_genome_preparation. Alignment output files were then sorted, indexed and filtered to retain only reads uniquely mapping to the rDNA reference using samtools version 1.10.

Methylation data from the WGBS/RRBS data was extracted from the bismark alignments using bismark_methylation_extractor.

Data QC and sample exclusion: Samples were excluded on the basis of bismark report data (Tables S3, S6, S8, S10). Namely, if they had extremely high or low uniquely mapped reads, poor mapping efficiency, poor bisulfite conversion (as evidenced by high non-CpG methylation values). In the case of the RRBS datasets, samples that were extreme outliers

with regards to total CpG methylation were also excluded as this indicates a potential problem with enzyme digestion or size selection altering the genomic regions captured compared to other samples in the same set.

rDNA CN estimation: rDNA CN was estimated from WGBS using a previously described method[10,13]. This involved aligning reads to the exome reference described above to obtain the average read depth for each sample using samtools depth. The average read depth mapped to the 18S subunit from the whole genome + rDNA alignments were then normalised to the exome value and CN calculated as 2 x (18S /exome average read depth) for each sample. We previously validated this method of estimating the relative total CN of rDNA across samples by comparison to digital droplet PCR[10].

To estimate the relative rDNA CN across samples from the RRBS data, we developed an alternative approach due to the patchy coverage of exomes in these libraries. In this approach the number of reads aligned to the rDNA reference as reported by samtools idxstats was divided by the total number of paired-end alignments listed in the corresponding bismark report. This method was cross-validated with the methodology using whole-genome data and showed a high correlation (Fig. S4).

Genomic feature methylation and coverage estimates: Genomic feature methylation was extracted using R package methylKit version 1.16.0[46]. Genomic features and repetitive DNA elements were defined based on hg38 genomic annotation Reference Sequence (RefSeq) and RepeatMasker database acquired from UCSC table browser respectively[47]. Regions 1kb upstream and downstream of the transcription start site of the reference genome were considered as promoters. For rDNA analysis, only CpGs covered by at least 50 unique reads were used whereas all CpGs were analysed for rest of the genome. Coverages per genomic features were estimated by calculating average number of reads that cover all CpG sites within each annotated genomic feature for each sample revealed by methylKit methRead() function.

Genotype association between BMI PRSs and rDNA CN: The base data were obtained from the meta-analysis in[26], with files hosted within the GIANT consortium data site of the Broad Institute. In particular, the two summary files for BMI analysis (updated after June 25, 2018) were retrieved, one containing all considered loci (hereinafter referred to as the “COMPLETE”

file, with 2,336,269 variants), and another including Conditional and Joint (COJO) -transformed p-values of only significant hits (hereinafter, "COJO" file, with 941 variants).

Base data SNPs were filtered for quality using the munge_sumstats.py script from ldsc version 1.0.1[48], with the options --snp SNP, --N-col N, --a1 Tested_Allele, --a2 Other_Allele, --frq Freq_Tested_Allele_in_HRS and --n-min 100000 for both COMPLETE and COJO input files. For the COJO input, the options --p P_COJO, --signed-sumstats BETA_COJO,0 and --ignore P,SE,BETA were also included to ensure the intended estimates were employed. This filtering left 1,977,697/2,336,269 SNPs from the COMPLETE input and 802/941 from the COJO input.

Although not explicitly specified, the genomic coordinates indicated in the base data files appear to refer to the GRCh37 assembly (e.g., rs1000096 is listed at chr4:38,692,835, instead of chr4:38,691,214 as it would correspond in GRCh38). For consistency with the target data, remaining locations in the base data were thus converted to GRCh38 coordinates using the liftOver function of the rtracklayer package version 1.54.0 in R with the hg19toHg38.over.chain file obtained from UCSC.

Target data vcf files were retrieved from the 1000 Genomes Project FTP servers. In particular, information for 3202 individuals was obtained for the autosomal chromosomes from the 20201028_3202_raw_GT_with_annot folder within the 1000G_2504_high_coverage collection. These were initially filtered at chromosome level with plink2 (version 2.0-20200328) --make-bed, keeping solely the SNPs remaining on the base data using the --extract bed1 option, and QC parameters --mind 0.01, --geno 0.01, --maf 0.01, --hw2 1e-6, and --max-alleles 2. The --set-missing-var-ids @:# option was also included to avoid apparent duplicate names, and a list of successful loci was requested with the --write-snplist option. The per-chromosome filtered outputs were then merged using plink2 --pmerge-list. 1,196,533 SNPs remained on the COMPLETE case after this step, and 467 for the COJO input. These were further pruned to remove highly-correlated SNPs using plink2's --indep-pairwise 200 50 0.25 option, where the values represent the window size, step size and maximum LD r^2 threshold allowed. A total of 1,002,886 and 8 SNPs were removed in this step on the COMPLETE and COJO cases, respectively.

The remaining SNPs were then employed to calculate the F coefficient for heterozygosity of each sample using plink2 --het. All individuals with F coefficients more than 3 standard deviations away from the overall mean were removed, leaving 3198 and 3199 samples in the

COMPLETE and COJO cases, respectively. These were further pruned with plink2 --king-cuttof 0.125 to avoid closely-related samples biasing the results, with 0.125 representing the relatedness level of second-degree relatives. To ensure reproducibility, a fixed seed value (1986) was also included. This pruning step left 2575 and 2502 individuals for the COMPLETE and COJO analyses, respectively. The final target data files were then generated with plink2 --make-bed specifying the remaining variants and individuals

Clumping of the remaining variants – retaining only weakly-correlated SNPs most associated with the phenotype of interest – is not yet available on plink2, so plink --clump from version 1.9-170906 was employed instead, with parameters --clump-p1 1, --clump-r2 0.1, and --clump-kb 250. Whereas no significant clumps were identified for the COJO analysis, 1414 clumps from the top 2213 variants were generated for the COMPLETE input in this manner.

Poligenic risk scores (PRSs) were finally computed using plink2 --score at different p-value thresholds indicated with the --q-score-range option. The generated values at each threshold were then employed to construct linear models with rDNA CN estimates as independent variable, as well as the 6 first principal components calculated on the pre-clumping variants with plink2 --pca, sex and 1000 genomes population as covariates. The goodness of fit of each model was then estimated as the difference between the R^2 of the model itself minus the R^2 of a null model without the PRSs.

Genome-wide association study for rDNA CN: Association between SNVs and rDNA CN was performed using the plink2 --glm command from plink release 2.0-20200328, with parameters --mind 0.05, --geno 0.05, --maf 0.05, and --hwe 0.001 for pre-filtering.

General statistical analysis: Final figures and statistical analysis were performed using GraphPad Prism (v9.2.0) and R (v4.1.1). Specific statistical analyses are indicated in the respective figure and table legends.

REFERENCES

1. Collaboration, N.C.D.R.F. (2017). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. Lancet *390*, 2627-2642. 10.1016/S0140-6736(17)32129-3.
2. Khera, A.V., Chaffin, M., Wade, K.H., Zahid, S., Brancale, J., Xia, R., Distefano, M., Senol-Cosar, O., Haas, M.E., Bick, A., et al. (2019). Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. Cell *177*, 587-596 e589. 10.1016/j.cell.2019.03.028.
3. Loos, R.J.F., and Yeo, G.S.H. (2022). The genetics of obesity: from discovery to biology. Nat Rev Genet *23*, 120-133. 10.1038/s41576-021-00414-z.
4. Wahl, S., Drong, A., Lehne, B., Loh, M., Scott, W.R., Kunze, S., Tsai, P.C., Ried, J.S., Zhang, W., Yang, Y., et al. (2017). Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature *541*, 81-86. 10.1038/nature20784.
5. Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., et al. (2022). The complete sequence of a human genome. Science *376*, 44-53. 10.1126/science.abj6987.
6. Danson, A.F., Marzi, S.J., Lowe, R., Holland, M.L., and Rakyan, V.K. (2018). Early life diet conditions the molecular response to post-weaning protein restriction in the mouse. BMC Biol *16*, 51. 10.1186/s12915-018-0516-5.
7. Holland, M.L., Lowe, R., Caton, P.W., Gemma, C., Carbajosa, G., Danson, A.F., Carpenter, A.A., Loche, E., Ozanne, S.E., and Rakyan, V.K. (2016). Early-life nutrition modulates the epigenetic state of specific rDNA genetic variants in mice. Science *353*, 495-498. 10.1126/science.aaf7040.
8. Huang, P.L. (2009). A comprehensive definition for metabolic syndrome. Dis Model Mech *2*, 231-237. 10.1242/dmm.001180.
9. Parks, M.M., Kurylo, C.M., Dass, R.A., Bojmar, L., Lyden, D., Vincent, C.T., and Blanchard, S.C. (2018). Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression. Sci Adv *4*, eaao0665. 10.1126/sciadv.aao0665.
10. Rodriguez-Algarra, F., Seaborne, R.A.E., Danson, A.F., Yildizoglu, S., Yoshikawa, H., Law, P.P., Ahmad, Z., Maudsley, V.A., Brew, A., Holmes, N., et al. (2022). Genetic variation at mouse and human ribosomal DNA influences associated epigenetic states. Genome Biol *23*, 54. 10.1186/s13059-022-02617-x.
11. Shea, J.M., Serra, R.W., Carone, B.R., Shulha, H.P., Kucukural, A., Ziller, M.J., Vallaster, M.P., Gu, H., Tapper, A.R., Gardner, P.D., et al. (2015). Genetic and Epigenetic Variation, but Not Diet, Shape the Sperm Methylome. Dev Cell *35*, 750-758. 10.1016/j.devcel.2015.11.024.
12. Hori, Y., Shimamoto, A., and Kobayashi, T. (2021). The human ribosomal DNA array is composed of highly homogenized tandem clusters. Genome Res *31*, 1971-1982. 10.1101/gr.275838.121.
13. Gibbons, J.G., Branco, A.T., Godinho, S.A., Yu, S., and Lemos, B. (2015). Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. Proc Natl Acad Sci U S A *112*, 2485-2490. 10.1073/pnas.1416878112.
14. Hall, A.N., Turner, T.N., and Queitsch, C. (2021). Thousands of high-quality sequencing samples fail to show meaningful correlation between 5S and 45S ribosomal DNA arrays in humans. Sci Rep *11*, 449. 10.1038/s41598-020-80049-y.
15. Orozco, L.D., Farrell, C., Hale, C., Rubbi, L., Rinaldi, A., Civelek, M., Pan, C., Lam, L., Montoya, D., Edillor, C., et al. (2018). Epigenome-wide association in adipose tissue from the METSIM cohort. Hum Mol Genet *27*, 1830-1846. 10.1093/hmg/ddy093.
16. Laakso, M., Kuusisto, J., Stancakova, A., Kuulasmaa, T., Pajukanta, P., Lusis, A.J., Collins, F.S., Mohlke, K.L., and Boehnke, M. (2017). The Metabolic Syndrome in Men study: a resource for

studies of metabolic and cardiovascular diseases. J Lipid Res *58*, 481-493. 10.1194/jlr.O072629.

17. Mitchell, E., Spencer Chapman, M., Williams, N., Dawson, K.J., Mende, N., Calderbank, E.F., Jung, H., Mitchell, T., Coorens, T.H.H., Spencer, D.H., et al. (2022). Clonal dynamics of haematopoiesis across the human lifespan. Nature *606*, 343-350. 10.1038/s41586-022-04786-y.
18. Desai, M.Y., Dalal, D., Santos, R.D., Carvalho, J.A., Nasir, K., and Blumenthal, R.S. (2006). Association of body mass index, metabolic syndrome, and leukocyte count. Am J Cardiol *97*, 835-838. 10.1016/j.amjcard.2005.10.021.
19. Nishimura, S., Manabe, I., Nagasaki, M., Eto, K., Yamashita, H., Ohsugi, M., Otsu, M., Hara, K., Ueki, K., Sugiura, S., et al. (2009). CD8+ effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. Nat Med *15*, 914-920. 10.1038/nm.1964.
20. Xu, B., Li, H., Perry, J.M., Singh, V.P., Unruh, J., Yu, Z., Zakari, M., McDowell, W., Li, L., and Gerton, J.L. (2017). Ribosomal DNA copy number loss and sequence variation in cancer. PLoS Genet *13*, e1006771. 10.1371/journal.pgen.1006771.
21. Loyfer, N., Magenheim, J., Peretz, A., Cann, G., Bredno, J., Klochendler, A., Fox-Fisher, I., Shabi-Porat, S., Hecht, M., Pelet, T., et al. (2023). A DNA methylation atlas of normal human cell types. Nature *613*, 355-364. 10.1038/s41586-022-05580-6.
22. Wurtz, P., Wang, Q., Soininen, P., Kangas, A.J., Fatemifar, G., Tynkkynen, T., Tiainen, M., Perola, M., Tillin, T., Hughes, A.D., et al. (2016). Metabolomic Profiling of Statin Use and Genetic Inhibition of HMG-CoA Reductase. J Am Coll Cardiol *67*, 1200-1210. 10.1016/j.jacc.2015.12.060.
23. Stults, D.M., Killen, M.W., Williamson, E.P., Hourigan, J.S., Vargas, H.D., Arnold, S.M., Moscow, J.A., and Pierce, A.J. (2009). Human rRNA gene clusters are recombinational hotspots in cancer. Cancer Res *69*, 9096-9104. 10.1158/0008-5472.CAN-09-2680.
24. Valori, V., Tus, K., Laukaitis, C., Harris, D.T., LeBeau, L., and Maggert, K.A. (2020). Human rDNA copy number is unstable in metastatic breast cancers. Epigenetics *15*, 85-106. 10.1080/15592294.2019.1649930.
25. Li, W., Zhang, D., Wang, W., Wu, Y., Mohammadnejad, A., Lund, J., Baumbach, J., Christiansen, L., and Tan, Q. (2019). DNA methylome profiling in identical twin pairs discordant for body mass index. Int J Obes (Lond) *43*, 2491-2499. 10.1038/s41366-019-0382-4.
26. Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., Visscher, P.M., and Consortium, G. (2018). Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. Hum Mol Genet *27*, 3641-3649. 10.1093/hmg/ddy271.
27. Choi, S.W., Mak, T.S., and O'Reilly, P.F. (2020). Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc *15*, 2759-2772. 10.1038/s41596-020-0353-1.
28. Phan L., J.Y., Zhang H., Qiang W., Shekhtman E., Shao D., Revoe D., Villamarin R., Ivanchenko E., Kimura M., Wang Z.Y., Hao L., Sharopova N., Bihan M., Sturcke A., Lee M., Popova N., Wu W., Bastiani C., Ward M., Holmes J.B., Lyoshin V., Kaur K., Moyer E., Feolo M., Kattman B.L. (2021). ALFA: allele frequency aggregator. National Center for Biotechnology Information, U.S. National Library of Medicine.
29. Mesnage, R., Ibragim, M., Mandrioli, D., Falcioni, L., Tibaldi, E., Belpoggi, F., Brandsma, I., Bourne, E., Savage, E., Mein, C.A., and Antoniou, M.N. (2022). Comparative Toxicogenomics of Glyphosate and Roundup Herbicides by Mammalian Stem Cell-Based Genotoxicity Assays and Molecular Profiling in Sprague-Dawley Rats. Toxicol Sci *186*, 83-101. 10.1093/toxsci/kfab143.
30. Mesnage, R., Teixeira, M., Mandrioli, D., Falcioni, L., Ibragim, M., Ducarmon, Q.R., Zwittink, R.D., Amiel, C., Panoff, J.M., Bourne, E., et al. (2021). Multi-omics phenotyping of the gut-liver axis reveals metabolic perturbations from a low-dose pesticide mixture in rats. Commun Biol *4*, 471. 10.1038/s42003-021-01990-w.

31. Ghasemi, A., Jeddi, S., and Kashfi, K. (2021). The laboratory rat: Age and body weight matter. EXCLI J *20*, 1431-1445. 10.17179/excli2021-4072.
32. Xie, S.Q., Leeke, B.J., Whilding, C., Wagner, R.T., Garcia-Llagostera, F., Low, Y., Chammas, P., Cheung, N.T., Dormann, D., McManus, M.T., and Percharde, M. (2022). Nucleolar-based Dux repression is essential for embryonic two-cell stage exit. Genes Dev *36*, 331-347. 10.1101/gad.349172.121.
33. Stults, D.M., Killen, M.W., Pierce, H.H., and Pierce, A.J. (2008). Genomic architecture and inheritance of human ribosomal RNA gene clusters. Genome Res *18*, 13-18. 10.1101/gr.6858507.
34. Hallgren, J., Pietrzak, M., Rempala, G., Nelson, P.T., and Hetman, M. (2014). Neurodegeneration-associated instability of ribosomal DNA. Biochim Biophys Acta *1842*, 860-868. 10.1016/j.bbadis.2013.12.012.
35. Udugama, M., Sanij, E., Voon, H.P.J., Son, J., Hii, L., Henson, J.D., Chan, F.L., Chang, F.T.M., Liu, Y., Pearson, R.B., et al. (2018). Ribosomal DNA copy loss and repeat instability in ATRX-mutated cancers. Proc Natl Acad Sci U S A *115*, 4737-4742. 10.1073/pnas.1720391115.
36. Durand, S., Bruelle, M., Bourdelais, F., Bennychen, B., Blin-Gonthier, J., Isaac, C., Huyghe, A., Martel, S., Seyve, A., Vanbelle, C., et al. (2023). RSL24D1 sustains steady-state ribosome biogenesis and pluripotency translational programs in embryonic stem cells. Nat Commun *14*, 356. 10.1038/s41467-023-36037-7.
37. Gregor, M.F., and Hotamisligil, G.S. (2011). Inflammatory mechanisms in obesity. Annu Rev Immunol *29*, 415-445. 10.1146/annurev-immunol-031210-101322.
38. Zhang, T., Whelton, P.K., Xi, B., Krousel-Wood, M., Bazzano, L., He, J., Chen, W., and Li, S. (2019). Rate of change in body mass index at different ages during childhood and adult obesity risk. Pediatr Obes *14*, e12513. 10.1111/ijpo.12513.
39. Bulut-Karslioglu, A., Macrae, T.A., Oses-Prieto, J.A., Covarrubias, S., Percharde, M., Ku, G., Diaz, A., McManus, M.T., Burlingame, A.L., and Ramalho-Santos, M. (2018). The Transcriptionally Permissive Chromatin State of Embryonic Stem Cells Is Acutely Tuned to Translational Output. Cell Stem Cell *22*, 369-383 e368. 10.1016/j.stem.2018.02.004.
40. Khajuria, R.K., Munschauer, M., Ulirsch, J.C., Fiorini, C., Ludwig, L.S., McFarland, S.K., Abdulhay, N.J., Specht, H., Keshishian, H., Mani, D.R., et al. (2018). Ribosome Levels Selectively Regulate Translation and Lineage Commitment in Human Hematopoiesis. Cell *173*, 90-103 e119. 10.1016/j.cell.2018.02.036.
41. Paredes, S., and Maggert, K.A. (2009). Ribosomal DNA contributes to global chromatin regulation. Proc Natl Acad Sci U S A *106*, 17829-17834. 10.1073/pnas.0906811106.
42. Paredes, S., Branco, A.T., Hartl, D.L., Maggert, K.A., and Lemos, B. (2011). Ribosomal DNA deletions modulate genome-wide gene expression: "rDNA-sensitive" genes and natural variation. PLoS Genet *7*, e1001376. 10.1371/journal.pgen.1001376.
43. Tanaka, Y., Okamoto, K., Teye, K., Umata, T., Yamagiwa, N., Suto, Y., Zhang, Y., and Tsuneoka, M. (2010). JmjC enzyme KDM2A is a regulator of rRNA transcription in response to starvation. EMBO J *29*, 1510-1522. 10.1038/emboj.2010.56.
44. Murayama, A., Ohmori, K., Fujimura, A., Minami, H., Yasuzawa-Tanaka, K., Kuroda, T., Oie, S., Daitoku, H., Okuwaki, M., Nagata, K., et al. (2008). Epigenetic control of rDNA loci in response to intracellular energy status. Cell *133*, 627-639. 10.1016/j.cell.2008.03.030.
45. Clarke, L., Fairley, S., Zheng-Bradley, X., Streeter, I., Perry, E., Lowy, E., Tasse, A.M., and Flicek, P. (2017). The international Genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data. Nucleic Acids Res *45*, D854-D859. 10.1093/nar/gkw829.
46. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol *13*, R87. 10.1186/gb-2012-13-10-r87.

47. Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res *32*, D493-496. 10.1093/nar/gkh103.

48. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics, C., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet *47*, 291-295. 10.1038/ng.3211.

Acknowledgments: We thank Christopher Bell (QMUL) and Sue Ozanne (IMS, Cambridge) for their helpful comments on the manuscript.

Funding: Academy of Medical Sciences Springboard Award (SBF003\1026) (MLH), Medical Research Council (MR/P011799/1) (DJW, VKR, MLH, SH), Barts Charity Grant (MGU0604) (VKR), Rosetrees grant (Seedcorn 2021\100182) (VKR), Rosetrees grant (CF-2021-2\109) (VKR), BBSRC grant (BB/R00675X/1) (VKR & FRA). DJW's salary is partly supported by the National Institute for Health and Care Research University College London Hospitals Biomedical Research Centre.

Author contributions:

Conceptualization: MLH, VKR, DJW

Methodology: MLH, VKR, DJW, PPL, FR-A

Investigation: PPL, FR-A, FA, MG, MLH, REAS, RM, MNA, WL, QL, SY

Visualization: MLH, PPL, FR-A

Funding acquisition: MLH, VKR, DJW, SLH

Project administration: MLH, VKR, DJW

Supervision: MLH, VKR, DJW

Writing – original draft: MLH, PPL, FR-A

Writing – review & editing: MLH, VKR, FR-A, PPL, REAS, DJW

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Data generated for this study will be available from Bioproject ID: PRJNA817350 upon manuscript acceptance.

External data sources were available in public repositories: GEO (GSE87893), (GSE157551). According to the Danish and EU legislations, transfer and sharing of individual-level data require prior approval from the Danish Data Protection Agency and require that data sharing requests are dealt with on a case-by-case basis. For this reason, the raw data on the

monozygotic twin cohort cannot be deposited in a public database but can be made available through collaboration or upon request.