Reviewed Preprint
v1 • November 20, 2024
Not revised

Reviewed Preprint
v2 • July 23, 2026
Revised by authors

Estimating bone marrow adiposity from head MRI and identifying its genetic architecture

Tobias Kaufmann[1,2] ✉, **Pål Marius Bjørnstad**[3], **Martin Falck**[3], **Stener Nerland**[1,4], **Kevin O'Connell**[1], **Oleksandr Frei**[1,5], **Ole A Andreassen**[1,6], **Lars T Westlye**[1,6,7], **Srdjan Djurovic**[3,8], **Timothy Hughes**[1,3] ✉

[1]Norwegian Centre for Mental Disorders Research, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway • [2]Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health, University of Tübingen, Tübingen, Germany • [3]Department of Medical Genetics, Oslo University Hospital, Oslo, Norway • [4]Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, Norway • [5]Centre for Bioinformatics, Department of Informatics, University of Oslo, Oslo, Norway • [6]KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway • [7]Department of Psychology, University of Oslo, Oslo, Norway • [8]NORMENT, Department of Clinical Science, University of Bergen, Bergen, Norway

✉ **For correspondence:**
tobias.kaufmann@med.uni-tuebingen.de
timothy.hughes@medisin.uio.no

Competing interests: OAA has received speaker's honorarium from Lundbeck and Sunovion, and is a consultant to HealthLytix. Other authors report no competing interests.

Funding: See page 19

Reviewing editor: Clifford J Rosen, Maine Medical Center Research Institute, United States

© 2024, Kaufmann et al. This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

eLife Assessment

This study introduces an artificial-intelligence tool that estimates fat in skull bone marrow from routine brain scans, enabling large studies that were previously impractical. The evidence for the method's repeatability and for identifying genetic links is **convincing** overall. The authors identify genes, diseases, and other biological characteristics that are linked to skull marrow fat, which represents an **important** advance. The work will be of most interest to researchers using large imaging biobanks and those studying ageing-related changes across bone, blood, and brain.

https://doi.org/10.7554/eLife.101499.2.sa4

Abstract

Bone marrow adiposity changes radically through the lifespan, but this phenomenon is poorly characterised and understood in humans. Large datasets of magnetic resonance imaging (MRI) scans of the head have been collected to study the human brain, but also contain unexploited information about other organs. We developed an artificial neural network that localises calvarial bone marrow in T1-weighted MRI head scans, enabling us to study its composition in several large MRI datasets, and to model sex-dimorphic age trajectories, including the effect of menopause. We revealed high heritability in single-nucleotide polymorphism and twin data, and identified 41 genetic loci significantly associated with the trait, including six sex-specific loci. Integrating mapped genes with existing bone marrow single-cell RNA-sequencing data revealed patterns of adipogenic lineage differentiation and lipid loading. Finally, we identified significant genetic correlations with several human traits, including cognitive ability and Parkinson's disease, which is intriguing in light of the recently discovered channels that link calvarial bone marrow to the meninges.

Introduction

Bone marrow (BM) is a tissue located within trabecular bone which produces the cellular components of blood. Beginning in childhood, it undergoes a radical conversion from red to yellow adipose marrow (1) which begins in the long bones of the appendicular skeleton, but progresses more slowly in the flat bones of the axial skeleton (2). By middle-age, BM adipose tissue accounts for 50-70% of total BM volume in healthy humans (3). Men have higher bone marrow

adiposity (BMA) than women, but post-menopausal women experience a rapid rise in BMA (4). The biological causes and consequences of this phenomenon are poorly understood. It is established that BM fat has a distinct function to white and brown adipose tissue depots (5, 6), but a precise definition of that function and the evolutionary basis of the sex-dimorphism remain elusive. What is known is that BM adipocytes and bone-forming osteoblasts both result from the bifurcating differentiation of mesenchymal stem cells in the BM niche and this lies at the root of the well-established connection between BMA and osteoporosis (7–9). Crucially, BM adipocytes have also been shown to affect the hematopeoitic microenvironment (10).

Magnetic resonance imaging (MRI) scans of the head are primarily thought of as a powerful non-invasive tool for investigating the structure and function of the human brain, but these scans also contain a wealth of imaging information about other organs, such as BM. The calvarial bones contain a BM layer of approximately 600 cm^2 which is several millimetres thick (11, 12) and has recently been shown to be connected to the meninges by direct vascular channels through the inner table of the calvarium (13). These channels transport cerebrospinal fluid from the subarachnoid space to the cranial BM via a perivascular route (14, 15) and myeloid blood cells travel in the opposite direction (13, 16, 17). Thus, there is a clear potential for the state of calvarial BM to directly affect the brain. Another specific feature of calvarial BM is that it converts from red to yellow much later in life than the BM of appendicular bones, thus its composition may be a useful normative marker of ageing.

The gold standard for assessing BM composition is magnetic resonance spectroscopy (MRS), but it is limited by its low spatial resolution, long acquisition time, and need for a predefined volume of interest, making it less well-suited for calvarial BM. Chemical shift imaging techniques offer higher resolution and shorter acquisition time, but such data are not routinely acquired. This has led to a growing interest in using conventional T1-weighted MRI to assess BM composition, with studies reporting robust correlations between BM fat volume and fraction measured with MRS, chemical shift imaging, and T1-weighted pulse sequences (18–20). Importantly, T1-weighted signal intensity is mainly determined by fat and water content (20–23) rendering it well-suited to assess BM composition.Large-scale neuroimaging studies routinely acquire T1-weighted MRI head scans, but chiefly focus on the brain. Thus, these datasets present a hitherto unexplored opportunity to address three unresolved aspects of BM biology: First, quantifying *in vivo* BM composition characteristics at population-scale, including age trajectories and sex-dimorphism, and its relationship to other traits such as bone mineral density (BMD). Second, estimating the heritability and genetic architecture of calvarial BM composition. Third, determining whether the composition of calvarial BM, which envelops the neocortex and has direct channels to it, has detectable effects on the brain.

We developed an artificial neural network-based method for locating calvarial BM in T1-weighted MRI head scans. Since T1-weighted MRI signal intensity is largely determined by fat content, we interpreted the averaged signal intensity as a proxy for calvarial BMA. We conducted extensive validation of the proposed measure using several publicly available MRI resources. We then applied this method to MRI head scans from the *UK Biobank* (24) to extract information on BMA, its spatial distribution in the calvarium and to model its main demographic determinants, specifically its relationship with age and sex, including female-specific factors. The size of the sample (n=33,042) enabled us to conduct a well-powered genome-wide analysis (GWAS) to identify gene associations with the proposed measure of BMA and, using a single-cell RNA sequencing (scRNAseq) dataset of BM mesenchymal cells, we gained insight into which genes drive differentiation into the adipogenic lineage and which drive lipid loading. Finally, we investigated the genetic overlap between BMA and twelve brain and body traits.

Results

Locating calvarial BM in head MRI scans

Our procedure for locating the BM and measuring its T1-weighted signal intensity involved 1. identifying the skin surface of the head, 2. identifying the calvarial part of the head using the top of the cerebellum and the accumbens as reference points, 3. for each point of the calvarial surface extracting an intensity array 25 mm into the skull (each layer is 0.5 mm thick), and 4. applying an artificial neural network to identify which layers of the intensity array correspond to BM (Figure 1). This network has 50 input nodes, 3 internal layers, and 6 output nodes (upper and lower bounds of outer table, BM, inner table). We trained the network on simulated data incorporating the natural intra- and inter-individual variation in thickness and intensity of the relevant anatomical structures (Figure 1A, S1). To obtain the BM signal intensity for an individual datapoint of the calvarium, we used the network model to estimate the location of the BM within the datapoint's intensity array and averaged these BM intensities to get the BM intensity for that datapoint. Then, we averaged these datapoint intensities across the calvarium to produce the global BMA measure for the scan.

Procedure validation on simulated data

We evaluated the performance of the neural network on simulated data using two metrics (Figure 1B): 1. the overlap between the predicted and the true BM location, and 2. the ratio between the predicted intensity of the BM and the true intensity of the BM. The median overlap between the predicted BM and the true BM in the simulated dataset was 0.76 with the lowest value (0.4) obtained in calvaria with a thin bone and the highest value (0.88) obtained in calvaria with thick bone. Poor overlap (below 0.7) was almost only observed in the thinnest bone and is explained by the fact that when the BM part of the bone is only a few layers thick (1 layer = 0.5 mm), an error by one layer will inevitably lead to a substantial fall in overlap. However, this did not result in a corresponding fall in the ratio of the predicted intensity of BM to its true intensity, because the typical BM intensity was only marginally higher than the neighbouring bone intensity. This property of the typical relative intensities of these anatomic structures also explained why the intensity ratio at high overlap is not centred on 1: any misidentification of cortical bone as BM, will typically result in an underestimate of true BM intensity (Figure S2). The neural network performed well and intensity ratios were in the range 0.9-1.1 for the vast majority of head types (Figure S3). It was only for heads with thin bone and either very high or very low BM intensity that we observed intensity ratios below or above this range (Figure S3B). Variation in subcutaneous fat thickness had no systematic impact on the performance (Figure S3C).

Procedure validation on real data and heritability estimate

The procedure also performed well on real data. We first tested a dataset from the *Consortium for Reliability and Reproducibility* (25) consisting of 30 individuals that were scanned 10 times each at 3-day intervals. The scans passing quality control displayed mean intensity of the outer table and BM at the levels that we would expect (Figure 1D): very low variation between scans of the same individual, outer table intensities centred on similar values with low variation across individuals, BM intensity always higher than outer table intensity, and high variation of BM intensity between individuals, but little within individuals. These scans were also inspected visually to verify that structure boundaries were correctly identified. Figure S4 provides examples of network performance in data of individuals with different levels of subcutaneous fat thickness, BMA, and bone thickness. Furthermore, we performed several assessments of the impact of different intensity normalization approaches on the obtained BMA estimates and compared estimates based on T1-weighted images to estimates obtained from T1 relaxation maps derived from a quantitative MRI pulse sequence (MP2RAGE). The BMA measure derived from T1-weighted images significantly correlated with BMA derived from the quantitative T1 relaxation maps and was not prone to differences in standard image normalization procedures (Figure S9-S12), supporting the validity of the proposed approach.

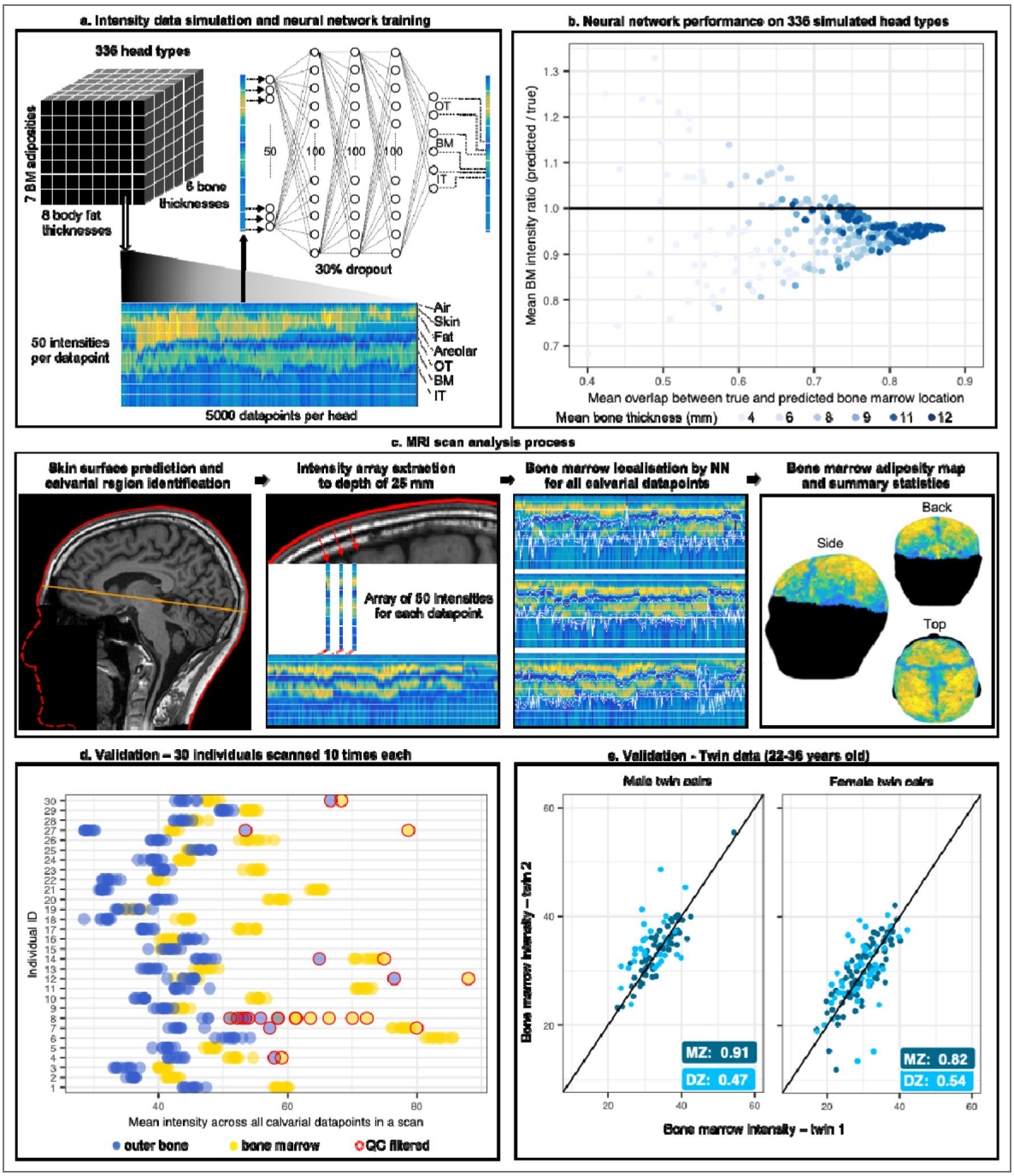


Figure 1. Overview of the artificial neural network training, application, and validation.

a. The data used for training the network were generated by simulating 336 different head types (with an example simulation for one head type provided here), where a head type is defined by mean bone thickness, mean BM intensity and mean body fat thickness. There are 5000 datapoints per head each with 50 intensities per datapoint. OT: outer table, IT: inner table, BM: bone marrow. **b.** We evaluated the performance of the neural network using two metrics: 1. the overlap between the predicted and the true BM location (defined by the simulation), and 2. the ratio between the predicted signal intensity of the BM and the true intensity of the BM. **c.** Illustration of the process of extracting the intensity arrays from the MRI scan, locating the BM, and computing its intensity. **d.** The algorithm was validated on a real dataset consisting of 30 individuals that were scanned 10 times each at 3-day intervals. For each scan, we computed the intensity of the outer bone (blue) and of the BM (yellow) which displayed high concordance across scans of the same individuals. The quality control measures correctly identified scans not producing reliable measures (13 of 300, 11 of which fail due to too few datapoints). **e.** Application of the algorithm to a cohort of monozygotic and same-sex dizygotic twins. Correlation in BMA between twin pairs in coloured boxes.

The monozygotic (MZ) twin data from the *Human Connectome Project* cohort (26) (Figure 1E) provided a second validation of the procedure through the high concordance in mean BM signal intensity between MZ twins, 0.91 in male pairs and 0.82 in female pairs. Further, using the data on the same-sex MZ and DZ twin pairs, we showed that BMA has an estimated broad-sense heritability of 91% in males and 60% in females (Table S1). We estimated a strong influence of environmental factors in females, one of which is highly likely to be estrogen levels as these vary widely between women of child-bearing age, with particularly large changes during pregnancy when estrogen levels are orders of magnitude higher, and the common use of estrogen-based contraceptives.

The calvarial BMA phenotype

In the 16,140 male and 16,902 female white British individuals from the *UK Biobank* for which head imaging data is available, we observed that calvarial BMA had strongly sex-dimorphic features. Males had a higher mean BMA than females (Figure 2A) and showed no increase at the group level in BMA levels in the age range covered by the biobank sample (45 to 82 years) implying that the rise in BMA must have occurred earlier in life (Figure 2B). In females the BMA level was highly dynamic in this age range increasing by 50% in the same age interval and thereby reaching male levels at age 70. Hormone replacement therapy (HRT) was used at some point in life by 38% of the women in the study and this therapeutic intervention slowed the rise in BMA, resulting in average BMA levels for treated women that did not exceed the level of males at age 80.

BM adipocytes and osteoblasts differentiate from the same mesenchymal progenitor cells and this would appear to drive the known negative correlation between BMA and bone mineral density (BMD). We replicated this effect in our data (Figure 2C – 14,262 males and 14,831 females). This relationship is also dimorphic with females displaying a lower level of calvarial BMD for a given level of BMA. Further, the size of the female BMD deficit increases as BMA levels rise (a similar pattern is observed with BMI in Figure 2D). We quantify the determinants of calvarial BMD by linear regression (Table S2). These regressions confirm the known effects of age and sex, and that the effect of BMA on BMD is about twice as large in females relative to males. Our regression results suggest that most of the effect of HRT on BMD is mediated by its effect on BMA (Table S2) since: 1. the statistical significance of HRT as an explanatory variable for BMD disappears when BMA is included as an explanatory variable in the regression and 2. the R^2 of the regression increases by a factor of 4 to 30%.

Genetic architecture

The strong twin-based heritability of BMA emphasised the importance of a genome-wide association study (GWAS) of the trait. We used the *UK Biobank* cohort (Table S3) and performed sex-specific covariate regression prior to separate discovery GWASs on males (n=16,140) and females (n=16,902). We found the male and female GWASs to have low genomic inflation (Figure 3A and Table S4) and to be significantly genetically correlated (Rg=.94, P=6e-27, Figure 3B), allowing us to combine these samples to perform a joint discovery GWAS across males and females (n=33,042). Further, we performed an independent replication GWAS (n=4,958) on non-white British males and females.

SNP-based heritability (h^2_{SNP}) estimated using linkage disequilibrium score regression confirmed significant heritability of the joint discovery GWAS, (h^2_{SNP} = 31.5%) (Table S4). We identified 168 genome-wide significant independent SNPs (P<5e-8), corresponding to 41 independent genomic loci (Table S5). The replication GWAS also displayed significant heritability (h^2_{SNP} = 28.6%). Out of the 168 significant discovery SNPs, 62% replicated at P<.05, and 39% of the 41 lead SNPs replicated at P<.05 (Table S6). One locus replicated at genome-wide significance (P<5e-8). Furthermore, 92.7% of the lead SNPs of the discovery sample showed same effect direction in the replication sample (Table S5).

Figure 3C depicts genes mapped from the 41 loci associated with the calvarial BMA. The most significant top lead SNP was located in the WNT16 gene (P=2.6e-143), which is also the top hit in BMD GWAS (27, 28) (Figure S12). The ESR1 locus displayed a notable sex-dimorphic pattern (Figure

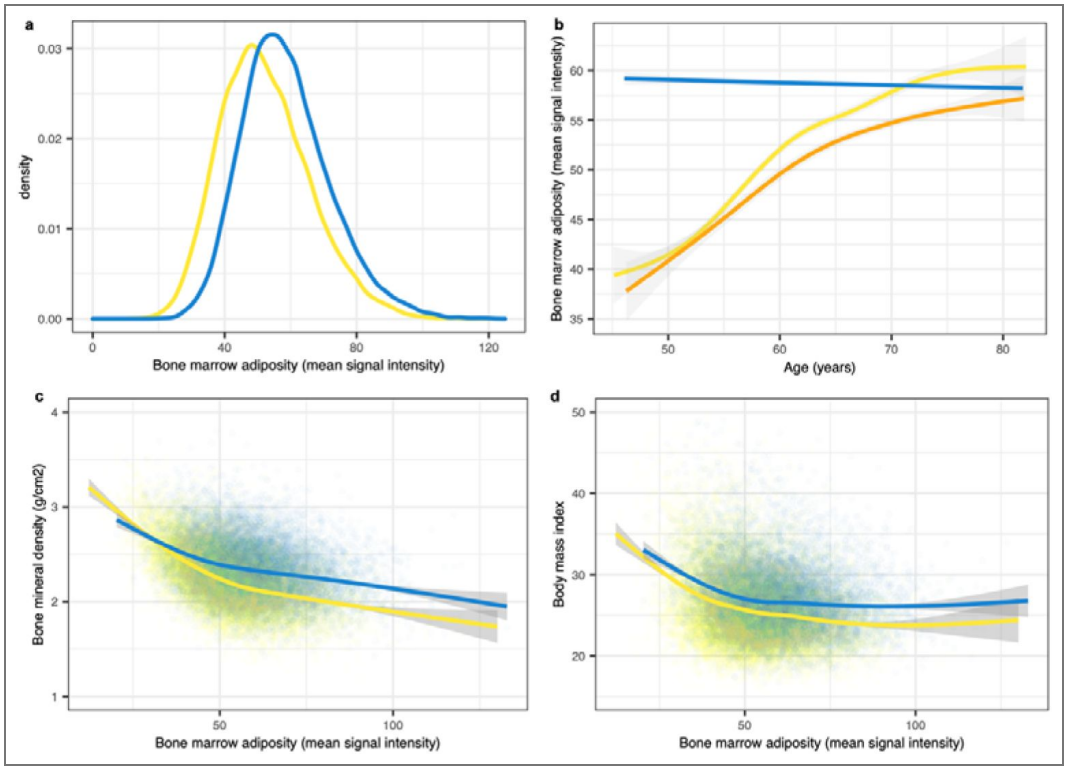


Figure 2. The calvarial BMA phenotype

a. BM intensity distribution in males (blue) and females (yellow) **b.** BMA age trajectories for males and females. Males in blue and females in yellow (never HRT) and orange (ever HRT). The effects of age and HRT on female BMA levels were both statistically significant (p-value < 2.0E-16) **c.** Sex-dimorphic relationship between BMA and calvarial bone mineral density. **d.** Relationship between BMA and BMI. Fitted lines in B, C, and D were computed using a general additive model and shading represents 95% confidence intervals.

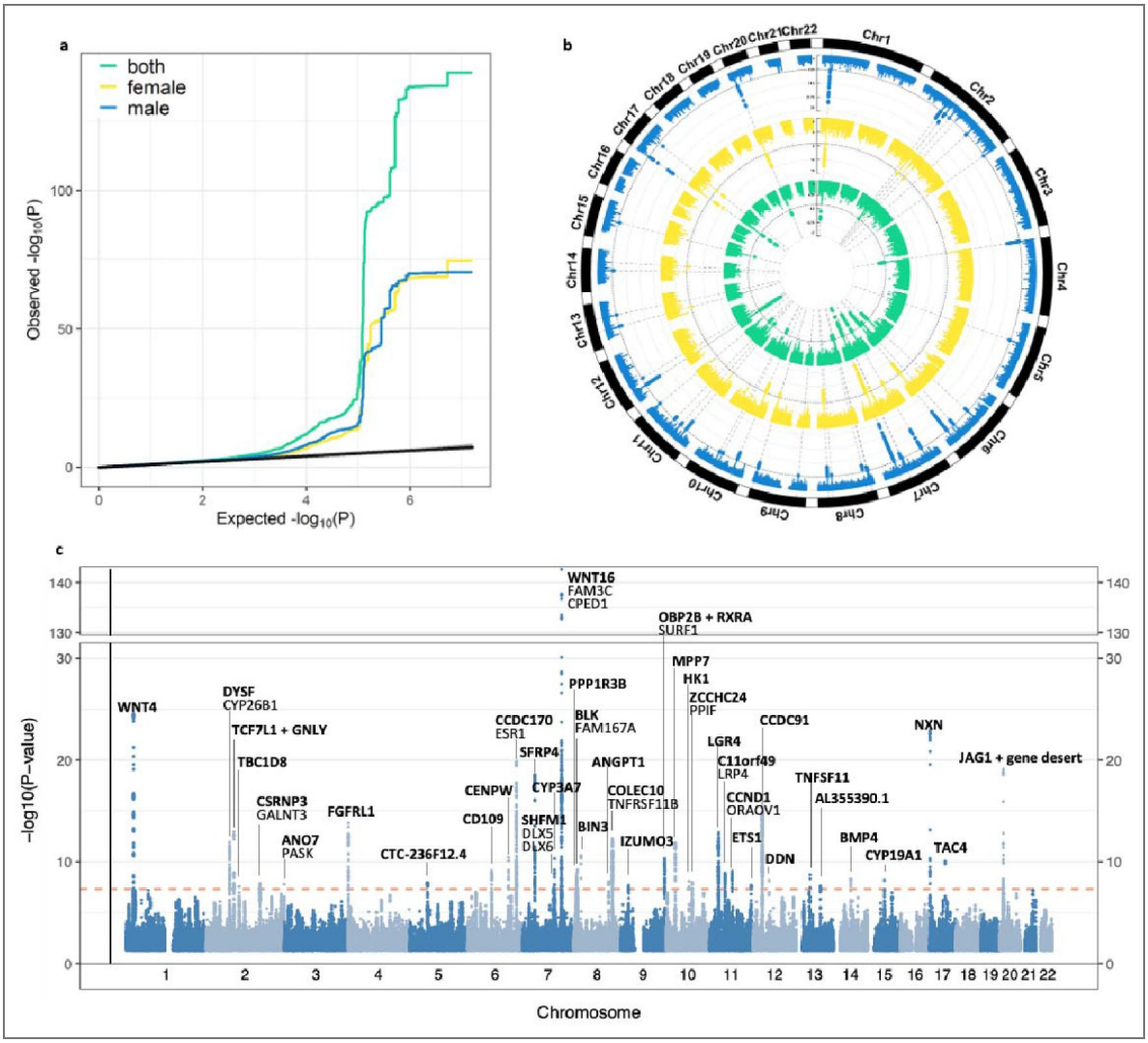


Figure 3. **BMA genome-wide significant loci**

a. QQ-plot of p-values for the male, female and joint discovery GWAS **b.** Comparison of male (outer), female (middle), and joint (inner) discovery GWAS significant loci. **c**. Manhattan plot of the 41 genome-wide significant loci in the joint discovery GWAS named after gene closest to the top lead SNP (bold). Other genes in each locus for which genome-wide significant SNPs were eQTLs, are listed below the closest gene. Three loci were closely adjacent to another more significant locus and are marked "+". SNPs with $p < 0.05$ not plotted.

S13). In females, the top lead SNP overlapped the ESR1 gene, but another lead SNP overlapped the neighbouring CCDC170 gene. In males, the signal in ESR1 was absent, but a genome-wide significant signal was located in the CCDC170 gene, suggesting that the ESR1 gene might not be the only source of an association signal in this locus.

A comparison of the male and female effect sizes of the top lead SNPs of each locus revealed 6 loci in which there is a more than two-fold difference between the sexes (loci 10, 18, 26, 30, 32, 37 in Table S5). In all cases, it is the male effect size that is larger. Further, each locus is genome-wide significant in males, but not in females, thus suggesting that these loci may be genetic drivers of the sex-dimorphism.

Gene expression in BM mesenchymal stem cell differentiation

Mesenchymal stem cells of the BM niche commit to either the adipogenic or the osteogenic lineage (Figure 4A) and both the number committing to the adipogenic lineage and their level of lipid-loading influences the total level of BMA. This aspect of BM biology is shared between humans and mice (29), so we made use of an existing mouse scRNAseq dataset of BM mesenchymal lineage cells (30) to study variation in the expression of BMA-associated genes as cells differentiate (Figure 4B). For any given gene, the percent of cells expressing a gene was relatively stable across different cell types. However, the mean scaled expression varied considerably between differentiation stages and the clustering of expression profiles revealed that most genes fall into one of the two largest groups of profiles: either genes with their peak of expression at the lineage commitment stage (LMP and LCP) or within the adipogenic lineage (MALP). It was also noteworthy that the genes peaking in MALPs mostly showed low expression prior to that stage. This differed from the ramping up observed in mesenchymal progenitors and LCPs. Only one of the closest genes saw their peak of expression in osteoblasts, but it was unexpected to observe that 6 of the closest genes see a strong peak in osteocytes. It remains to be investigated whether the closest gene might not be the source of the association or whether gene expression in osteocytes may have an influence on BMA.

Genetic correlation and overlap

We investigated genetic correlation and overlap with 13 traits. We selected body traits (BMD, BMI, waist-to-hip ratio, systolic and diastolic blood pressure, type-2 diabetes, coronary artery disease) that have a logical connection to BMA given the mesenchymal stem cells origin of BM adipocytes and their role in bone, fat, and vasculature (30). For the brain, we selected traits reflecting cognition (general cognitive ability and educational attainment) and disorders that are prevalent in adulthood (insomnia, multiple sclerosis, Parkinson's disease, Alzheimer's disease) since it is primarily in adulthood that the adiposity of calvarial BM experiences a substantial change (Figure 2B).

We identified strong negative genetic correlations between BMA and both BMD and BMI, and more subtle global genetic correlations between BMA and Parkinson's disease, general cognitive ability, and educational attainment (Table 1). However, global genetic correlations lack the ability to assess genetic overlap in the presence of mixed effect directions. So, to better understand the genetic overlap between BMA and other traits of interest, we also applied MiXeR (31). MiXeR uses trait polygenicity, defined as the number of trait-influencing variants required to explain 90% of SNP-based heritability, to model the number of shared trait-influencing-variants between two traits as well as the correlation within this shared component. MiXeR estimates that 99% of BMA trait-influencing variants are shared with BMD (497 of 500 variants), and that the correlation of this shared component is -0.95, indicating high genetic overlap with discordant effect. The proportion of BMA trait-influencing variants also estimated to influence general cognitive ability (0.57) and educational attainment (0.67) was lower than observed for BMD, but the correlations within their shared components were similarly large (COG = 0.93, EDU = 0.90). These patterns of high overlap and high effect correlation within this overlap are suggestive of vertical pleiotropy

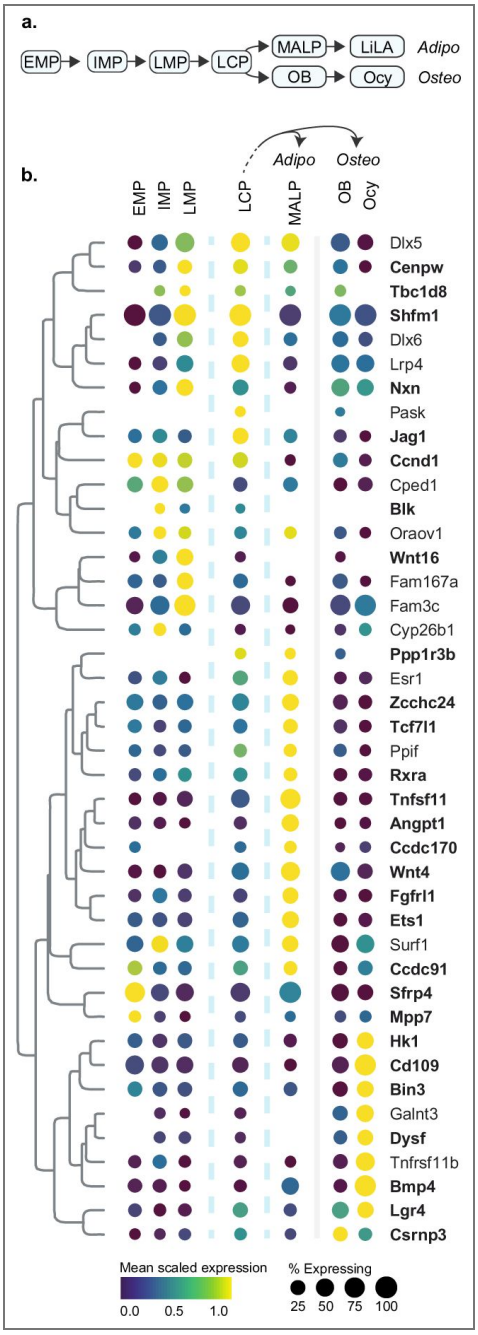


Figure 4. Gene expression in mesenchymal lineage cells of mouse BM.

a. The differentiation of BM mesenchymal cells. Early (EMP), intermediate (IMP), and late mesenchymal progenitors (LMP). Lineage committed progenitors (LCP). Marrow adipogenic lineage precursor (MALP). Lipid-laden Adipocyte (LiLA). Osteoblast (OB). Osteocyte (Ocy). **b**. A clustering of the gene expression profiles across cell types. Gene names are for mouse genes (these broadly match human gene names). Of the 40 genes which are closest to lead SNP for each locus, 28 had expression in the scRNAseq data (bold). LiLA expression data was not available in this dataset.

i.e. a molecular mechanism influencing one trait, that in turn influences a second trait, such that most of the variants driving the first trait either have the same or the opposite direction of effect on the second trait.

In Parkinson's Disease, where we also estimated statistically significant genetic correlation with BMA, MiXeR estimates a strong negative correlation of effect sizes, but this is within a much lower percentage of BMA trait-associated variants (14%). Similarly, BMA displayed a high correlation of effect sizes (-0.69) with Type-2 diabetes but within a moderate overlap, and a low correlation of effect sizes with both systolic and diastolic blood pressure, but with an overlap of 95%. These patterns are more indicative of horizontal pleiotropy (i.e. molecular mechanisms influencing multiple traits) than vertical pleiotropy.

Discussion

The development of a method for locating BM in MRI head scans enabled us to characterize calvarial BMA in several datasets and to quantify, *in vivo* and at population-scale, BMA's age trajectory, sex-dimorphism, and relationship to BMD. We focused on BM adiposity rather than volume, as this trait was considered more likely to be influencing other traits of interest, such as the blood (hematopoiesis), bone (osteoporosis), and brain health (through transcortical channels). Indeed, through a well-powered GWAS, we identified the main genetic drivers of this phenotype and we observed significant genetic correlation and overlap with several traits, including Parkinson's disease and general cognitive ability.

Estimating BMA in MRI head scans

BMA can be quantified in whole-body scans (32), but calvarial BM has a number of features which make it particularly well-suited to the study of BMA. First, conversion to yellow BM is slower in the flat bones (33), such that sex-dimorphic trajectories are observable in middle-aged cohorts. Second, there are many large collections of T1-weighted MRI scans from which calvarial BMA can potentially be extracted and, although T1-weighted pulse sequences are not specifically designed for BMA quantification (34), this enables analyses that have not previously been possible, such as GWAS and accurate estimation of lifespan sex differences. Finally, it has a simple "sandwiched" structure that we were able to model and then generate simulated calvaria which we use as training data for an artificial neural network. Our validation procedures demonstrate that this automated BM location procedure delivers good performance on real data, even though it was trained on simulated data and analyses each datapoint independently of all others. A more sophisticated model (35, 36), trained directly on annotated real data, may raise performance further, for example by improving the prediction of the lower bound of the BM. However, this would require a substantial manual annotation effort and would face the difficult challenge of distinguishing low-adiposity BM from calvarial bone. Both these issues are circumvented by our approach.

The calvarial BMA and BMD phenotypes

The sex-dimorphic nature of BM adiposity is qualitatively well-established (4), but has not previously been quantified in a large human cohort. Males have a high and stable average level of BMA over the covered age range (45 to 82 years of age), indicating that the dynamic male phase occurs earlier in life. The youngest females in our sample have a low level which ramps at a constant rate from the beginning to the end of the range with HRT significantly reducing the rate of increase. Interestingly, however, by the age of 70 (for women not using HRT), female BMA has risen to the same level as males. It appears highly likely that the menopause and the accompanying drop in natural estrogen levels is part of what drives the rise in BMA levels and that estrogen treatment can delay the rise. We emphasise that these trajectories are for calvarial BM and are most probably different in profile and levels to other BM, since it is known that axial flat bones convert later and perhaps less completely than the appendicular long bones. However, the nature of the sex-dimorphism and the effect of estrogen is likely shared with other BM.

Trait	Genetic correlation (LDSR)			Genetic overlap (MiXeR)				
	All subjects			Partitioning of causal variants Mean nb per partition			Overlap as fraction of all BMA causal variants	Corr. of effect sizes within overlap (s.e)
	r_g	*P*	*Q*	Unique to BMA	Overlap	Unique to trait		
Bone mineral density	-0.41	1.5E-22	1.8E-21	3	497	1471	0.99	-0.95 (0.05)
Body mass index	-0.20	9.3E-15	5.6E-14	NA*	NA*	NA*	NA*	NA*
Waist-to-hip ratio	-0.10	1.2E-02	2.60E-02	63	436	3300	0.87	-0.48 (0.16)
Blood pressure systolic	-0.02	6.5E-01	6.8E-01	25	475	3921	0.95	-0.25 (0.05)
Blood pressure diastolic	-0.03	3.4E-01	4.1E-01	27	473	3541	0.95	-0.21 (0.04)
Coronary artery disease	-0.10	3.6E-02	7.2E-02	275	225	1140	0.45	-0.40 (0.11)
Type-2 diabetes	-0.07	6.5E-02	1.1E-01	361	139	752	0.28	-0.69 (0.16)
Insomnia	-0.05	2.1E-01	2.8E-01	248	252	8868	0.50	-0.45 (0.32)
Multiple Sclerosis	-0.02	6.8E-01	6.8E-01	474	26	86	0.05	0.45 (0.14)
Parkinson's disease	-0.15	1.7E-03	4.1E-03	432	68	844	0.14	-0.75 (0.14)
Alzheimer's disease	-0.13	9.2E-02	1.4E-01	485	15	52	0.03	-0.61 (0.24)
General cognitive ability	0.09	1.1E-03	3.3E-03	217	283	10802	0.57	0.93 (0.12)
Educational attainment	0.10	9.9E-05	4.0E-04	166	334	12855	0.67	0.90 (0.11)

Table 1. Genetic correlation and overlap of BMA with other traits

r_g: genetic correlation. *Q*-values are FDR corrected *P*-values. Genetic overlap: number of causal variants that overlap (i.e. are shared) between BMA and the second trait. Total number of estimated causal variants for a trait (i.e. explaining 90% of SNP heritability) is the sum of those unique to trait and overlap (large total is indicative of high polygenicity). NA*: The publicly available summary statistics of the BMI GWAS contained only 3.3M SNPs, it was thus not possible to apply the MiXeR method to this dataset. Sex-specific genetic correlations, standard errors for causal variant partitioning, and Akaike Information criterion values (Table S9). Blue highlighting for: genetic correlation p < 0.01, overlap fraction > 0.5, correlation of effect sizes within overlapping > 0.5 or < -0.5.

Women at risk of osteoporosis are often prescribed HRT in order to improve BMD. Our regression results (Table S2) indicate that the effect of HRT on BMD is mostly mediated by its effect on BMA. Further, we suspect that the effect of BMA on BMD may be underestimated given that the prescription of HRT is biased towards women with low BMD and high BMA. Future studies, including information on why HRT was prescribed and for how long, may use our BMA quantification method to further detail this important relationship.

BMA heritability estimates

Our findings showed that male SNP heritability (43%) is half the broad-sense heritability (91%). In females, SNP heritability is only 23% which is low relative to men and relative to the broad-sense heritability of 60%. Since the vast majority of women in the *UK Biobank* sample have probably not reached their ultimate BMA levels, we suspect that the genetic effects on the phenotype may be underestimated in this female GWAS sample. Despite these features of the sample and its relatively modest size, we do observe good genetic correlation between the sexes of the discovery sample (Figure 3b) and cross-ethnic replication in a limited sample of 4,958 individuals. Based on these GWAS results, MiXeR estimates that BMA has approximately 500 causal variants, indicating low polygenicity relative to a typical polygenic trait such as BMD, which we estimated to have approximately two thousand causal variants.

The high broad-sense heritability estimates indicate that environmental effects on BMA are limited relative to genetics. However, a limitation of our broad-sense heritability estimates is that the age range of the twin dataset is much younger than the *UK Biobank* dataset from which we compute SNP heritability. This is particularly relevant for females, first because fluctuations in estrogen levels are substantial during child-bearing age due to pregnancy and contraception, and second because females only reach their ultimate BMA level between the ages of 70 and 80 (the upper end of the age range). Both the broad-sense and SNP heritability estimates should therefore be interpreted with caution for females.

Genetic correlation and overlap with other traits

We selected 7 body and 6 brain traits for which we computed genetic correlation, and obtained significant results for BMD, BMI, Parkinson's disease, general cognitive ability and educational attainment. Genetic correlation has two limitations. First, it tends to underestimate the true correlation in the case of mixed effects and, second, it provides no indication of whether the relationship between the traits is driven by horizontal or by vertical pleiotropy. MiXeR estimates the causal variants for each trait and thereby can estimate the overlapping fraction as well as the correlation of effect within this set. A high overlap of BMA causal variants with those of the other trait and a high correlation of effects within this set is suggestive of causal effect (vertical pleiotropy). Only 3 of the 12 traits displayed a clear combination of such features. One of these is BMD, which can be considered as a positive control since the BMA and BMD traits are known to be negatively correlated and the bifurcating differentiation of mesenchymal stem cells provides a causal cellular mechanism for this relationship between traits. The other two traits are educational attainment and general cognitive ability. Since these are two highly correlated traits (37), a high overlap and correlation of genetic effects for both traits with BMA may be consistent with the hypothesis that BMA could have a causal effect on cognition (Table 1).

GWAS results and differentiation of mesenchymal progenitors

Previous studies of mesenchymal cell differentiation in animal and cell models have pointed towards the Wnt/β-catenin signalling pathway being an important determinant of BMA (29). Our large-scale phenotyping of BMA in a human cohort enabled us to perform a hypothesis-free quantification of its genetic architecture which confirmed the strong effect of genes related to this pathway, including WNT16 (38), WNT4 (39), NXN (40). Further, we found associations in genes (BMP4 (41, 42), DLX5 (9, 43), LGR4 (44)) or close paralogs of genes (LRP4/LRP5(45), SFRP4/SFRP1 (46)) that were known to specifically influence adiposity. We also discovered and quantified the effect of many interesting novel loci, such as FGFRL1 which has a pleiotropic role in hypertension

resistance and bone growth in giraffes (47). The strong effects of the CCDC170/ESR1 and WNT16 loci (Table S5) were not unexpected given previous results from the BMD GWAS (28), but the pattern of the genetic signal illustrates the importance of fine mapping for these and other loci (Figures S9-10).

Recent scRNAseq studies of the BM niche have provided a comprehensive insight into BM cell types and the changes in gene expression during differentiation (48–50). Zhong et al. focused specifically on mesenchymal lineage cells and produced a detailed transcriptomic profile of progenitor maturation and bifurcating differentiation into adipogenic and osteogenic lineages (30). Interestingly, they identify a new non-proliferative cell type in the adipogenic lineage that does not contain lipid droplets (MALP: marrow adipogenic lineage precursors). In our clustering of GWAS genes according to their scRNAseq transcriptomic profiles, we find 17 genes which ramp expression in mesenchymal progenitors with the peak of expression occurring in LMP or LCP, and 16 genes with very low expression in mesenchymal progenitors and a distinct peak at the MALP stage. Since BMA is primarily determined by two potentially asynchronous processes, the number of adipocytes and the level of lipid-loading, this suggests that the first set (including WNT16) controls the bifurcation between the two main lineages, whilst the second set (including ESR1) controls MALPs further differentiation into lipid-laden adipocytes.

MALPs outnumber lipid-laden adipocytes by orders of magnitude in young mice and exist as pericytes and stromal cells that form a ubiquitous 3D network inside the marrow cavity (30). Evidence points to this network playing pivotal roles in maintaining marrow vasculature (30), suppressing osteogenic differentiation (51), and probably also influencing hematopoiesis (10). We tentatively speculate that calvarial MALPs may be involved in sensing perivascular flows of CSF from the meninges to the BM and in influencing the BM's hematopoietic response, and that this might be the basis of the observed genetic overlap between BMA and some cerebral traits. However, more conventional anatomical pathways may also be relevant, as the diploë and inner table communicate with meningeal and dural venous systems through diploic veins.

In addition to these recently described channels, classical cranial venous anatomy provides other potential interfaces between calvarial marrow and the intracranial compartment, including diploic and emissary venous connections with meningeal veins and dural venous sinuses.

Limitations and future validations

Several previous studies have used T1-weighted MRI to study BM composition (52–54). Validation studies have reported correlations of 0.99 with BM fat volume and of 0.71 with BM fat fraction measured with the IDEAL technique in the tibia of children (18). Similarly, BM fat volume estimated with T1-weighted images have shown strong correlations with MRS (r=0.88) and the Dixon method (r=0.79) for vertebrae and with the Dixon method for femoral neck (r=0.86) in menopausal women (19). A score based on T1-weighted signal intensities in the vertebrae was found to be a promising marker for osteoporosis and correlated with BMD measured with DEXA (r=-0.68) (55). While most prior studies have focused on non-cranial BM, one study found that T1-weighted signal intensity in the calvarium was a sensitive (93%) and specific (86%) marker for infiltrative processes including lymphoma, leukaemia and chronic anaemia known to affect BM composition (20).

Whereas prior studies have relied on manual grading and selection of segmentation thresholds and regions of interest, our approach automatically locates calvarial BM. This improves repeatability and allows for deployment in large datasets such as the UK Biobank. Although the main determinant of the T1-weighted signal intensity is fat, making it well-suited to assess BMA, weighted MRI pulse sequences present notable challenges. Importantly, T1-weighted signal intensity is influenced by scanner-related variation (56).

To ensure between-subject comparability, we used the intensity normalised nu.mgz volume output by FreeSurfer. We validated this approach through comparison with well-established intensity normalization methods: Kernel Density Estimation (KDE), WhiteStripe (WS), Gaussian Mixture Model (GMM), Fuzzy C-Means (FCM), and Z-score normalization (ZS). We found the highest test-

retest reliability with our approach (Figure S9), and, together with KDE (based on reference signal intensity in WM), the highest correlation with quantitative T1 relaxation maps (Figure S10). Together with the significant correlations with BMD (Figure S11), as measured with DEXA, and significant effects of osteoporosis status (Figure S12), these results demonstrate the biological relevance of the proposed measure of BM composition.

Nevertheless, we recognise the need for further validation to disentangle the relative contributions of adiposity and other features of BM composition. As such, caution is still warranted on the biological interpretation of measures based on T1-weighted images, and future sensitivity and robustness analyses are needed. Given the plausible biological findings in our study (sex-dimorphic age trajectories mapping known BMA biology, discovery of genes previously identified through cellular experiments as critical to adipocyte biology), we are confident that the benefits of exploiting the existing large samples of T1-weighted images outweighs the limitations. Indeed, future studies can build on our developments to further validate the proposed measure of bone marrow composition and to study its effect on bone, blood, and brain.

Materials and Methods

Experimental design

This study had several sequential subgoals: 1. obtain data on bone marrow composition from head MRI scans, 2. determine the sex-dimorphic age trajectories, 3. identify the genes associated with this trait, 4. determine which of these genes control the bifurcating (osteogenic/adipogenic) differentiation and which control lipid loading and 5. investigate whether the bone marrow composition trait has genetic overlap with other related body or brain traits. Each of these study components requires specific methodology and several data sources are used. These are described separately in the following subsections.

Cohorts

Hangzhou Normal University (HNU) data as part of the Consortium for Reliability and Reproducibility (25)

We accessed the openly available test-retest reliability dataset from *http://fcon_1000.projects.nitrc.org*. In this study, T1 MRI was acquired on a 3T GE Discovery MR750 scanner using an 8-channel head coil with a repetition time of 8.06ms, 8° flip angle and 250 mm field of view at a resolution of 1x1x1 mm. We included all available data from this sample, specifically data from 30 healthy adults (50% female) aged 20-30 years (mean: 24.4, sd: 2.4 years) that were scanned ten times across one month, one scan every three days.

MICA-MICs data (57)

We accessed the openly available multimodal dataset from *https://portal.conp.ca/dataset?id=projects/mica-mics*. In this study, MRI data was acquired on a 3 T Siemens Magnetom Prisma-Fit equipped with a 64-channel head coil. T1 MRI was acquired with a 3D magnetization-prepared rapid gradient-echo sequence (MPRAGE; 0.8 mm isotropic voxels, matrix = 320 × 320, 224 sagittal slices, TR = 2300 ms, TE = 3.14 ms, TI = 900 ms, flip angle = 9°, iPAT = 2, partial Fourier = 6/8). Quantitative T1 MRI was acquired with a 3D-MP2RAGE sequence (0.8 mm isotropic voxels, 240 sagittal slices, TR = 5000 ms, TE = 2.9 ms, TI 1 = ld0 ms, T1 2 = 2830 ms, flip angle 1 = 4°, flip angle 2 = 5°, iPAT = 3, bandwidth = 270 Hz/px, echo spacing = 7.2 ms, partial Fourier = 6/8).

Human Connectome Project data (58)

We accessed openly available data from *https://db.humanconnectome.org*. In this study, T1 MRI was acquired on a customized 3T Siemens Skyra scanner with a repetition time of 2400ms, 8° flip angle, 224mm field of view and a voxel size of 0.7mm. We included data from 430 twins from this sample, including 174 males (60 dizygotic and 114 monozygotic, age range 22 to 34) and 256 females (96 dizygotic and 160 monozygotic, age range 26 to 36).

UK Biobank (24)

We accessed data from the UK Biobank study resource (access code 27412; *https://www.ukbiobank.ac.uk*). In this study, T1 MRI data was acquired on three identical 3T Siemens Skyra scanners with a repetition time of 2000ms, 8° flip angle, 256mm field of view and a resolution of 1x1x1mm. We split the available sample into a discovery set comprising 33,042 white British individuals (51% females) aged 45-82 years (mean 64.9, sd 7.5 years), and an independent replication set comprising all other individuals (N=4958, 52% female, age mean 62.9, sd 7.7, range 45-81 years). We also obtained head bone mineral density measurements produced by a DXA system (code 23226-2.0), body mass index (code 21001-2.0), and data on whether women ever used hormone-replacement therapy (code 2814-2.0). BMD data was only available for 29,093 of the samples for which MRI head scans were available.

Ethical approvals

The UK Biobank was approved by the National Health Service National Research Ethics Service (ref. 11/NW/0382). The Human Connectome Project was approved by the Washington University institutional review board. Hangzhou Normal University study protocol and data sharing was approved by its local ethics committee.

FreeSurfer preprocessing and extraction of datapoint intensities

We processed all T1 MRI data using the standard recon-all pipeline in FreeSurfer 5.3.0. T1 intensities were extracted from the nu image, which is corrected for intensity non-uniformities and intensity normalized via scaling. Next, we fed the intensity-normalized volumes derived via recon-all into the FreeSurfer mri_watershed program to delineate the surfaces of the brain, skull and skin. We performed a stepwise extraction of intensities from layers following the shape of the outer skin, walking inward into the skull in steps of 0.5mm up to 25mm inward which produces an intensity array of length 50 for each datapoint. Since watershed delineates the skin around the entire head, we disregarded parts that are not directly on top of the cortex. For this, we identified the coordinates of the upper part of the cerebellum and the accumbens from the individual level FreeSurfer aseg parcellation and defined a plane in 3D space that intersected these two points, disregarding all data points in coordinates below the plane. The resulting stream of layerwise intensities therefore comprised data on top of the cortex and was fed one-by-one into an artificial neural network model to identify and define the borders of the bone marrow (BM) for that datapoint. See supplementary materials for more detail on this workflow, including software versions.

Intensity array data simulation

In order to train the neural network model, we generated a large synthetic dataset of intensity arrays, with known boundaries between anatomical structures, by simulating the thickness and intensity of the different structures located between the outer skin and the sub-arachnoid space. The simulation incorporated the following real-world complexities:

1. Different anatomical architectures (skin, subcutaneous fat, aponeurosis, outer table, BM, inner table, dura mater, arachnoid space), including when a structure is not present throughout the calvarium.
2. Variation in thickness and intensity between vertices (on the same calvarium)
3. A wide variety of different calvarium types with different combinations of levels of BM adiposity, bone thickness, and subcutaneous adiposity.
4. "Blurring" between anatomic structures was present in real data and was modelled in our simulations:
    a. blurring that occurred because of the voxel resolution of the MRI scanner relative to the underlying structures
    b. blurring that occurred because the layers for a data point usually do not match with voxel boundaries.

The parameters used in the simulation (primarily mean and variance of both thickness and intensity of each of the different anatomical structures within a calvarium) were estimated by reviewing the MRI scans of a wide variety of real subjects from the MRI scan collections included in this study. We simulated 336 different calvaria with the following mean characteristics: 7 BM intensities (40, 50, 70, 90, 110, 140, 170) x 6 bone thicknesses (4, 6, 8, 9, 11, 12 mm) x 8 body fat thicknesses (0, 1, 2, 3, 4, 6, 9, 12 mm) and for each calvarium we generated 5000 intensity arrays. In the iterative process of simulating data, training the network, and testing its performance on real data, we observed that incorporating the characteristics of extreme head types in the simulation parameters was critical to obtaining good performance on real data. Fine-tuning of model hyperparameters played a much lesser role. The R code and all parameters are openly available [made public upon acceptance] (bmSimulation_50_versionA.r).

Artificial neural network architecture and training

The network is a feed-forward fully connected network. There are 50 input nodes (for the 50 intensities of a datapoint), three internal layers with 100 nodes each (ReLU activation function), and six output nodes (upper and lower bound for each of outer table, BM, and inner table). The model was implemented in the Keras framework on top of the Tensorflow backend, and the training of the model was performed in Google Colab (59). All vertices from all calvaria were randomly shuffled, and then two thirds of the data were used for training the model, and one third was used as the validation set, to measure the performance. The loss function, defined as the sum of the mean squared errors of the six outputs, was minimised using the Adam algorithm (60), with a learning rate of 0.1 and a batch size of 8192. We used dropout to regularise the model (61), with a 30 % dropout rate in all the hidden layers. The model is trained over 300 epochs. The performance of the network plateaus before 300 epochs (Figure S1), suggesting that the optimisation has converged on the best parameters for this model architecture.

To compute the overlap of the true and the predicted intervals, the predictions were first rounded to integers, so they refer to specific layers in the intensity vector. Then the overlap is defined as the number of layers that are included in both the predicted and the true interval, divided by the number of layers in the longest of the true and the predicted interval. For each datapoint of a scan, we computed the mean BM intensity from the part of the intensity array predicted to be BM and then computed the mean and standard deviation across all datapoints in the scan. We did the same for the outer and inner table. The Python code is openly available *[made public upon acceptance]* (bmDetectNN50_randomAux_training_A.ipynb)

Additional validation of the BMA estimation approach

In the BMA estimation pipeline, we used intensity normalized T1 images generated in the standard recon-all processing stream of FreeSurfer. To assess the effect of intensity normalization, we estimated BMA with five additional frequently used intensity normalization methods (62). We calculated test-retest reliability (Figure S9) and, in a dataset where quantitative T1 data was available (57), compared BMA estimates derived from quantitative T1 with our semi-quantitative approach, as well as each of the five additional intensity normalization methods (Figure S10). Finally, we further assessed the semi-quantitative nature of our approach by comparing outer-bone intensity with bone mineral density as measured by DEXA scans (Figure S11).

ANN Quality control measures

We performed the BM location procedure on the intensity data from 42,068 heads in the UK Biobank. 420 of these produced no output and 2,234 produced a calvarium with less than 5,000 vertices most likely related to poor image quality (Figure S5), leaving 39,414 samples.

We used two additional QC metrics to filter out calvaria where BM location was likely to have failed. First, we set an upper limit of 30 on the standard deviation of the intensity of the outer table as scans with higher values were clear outliers and were probably cases where the location of both outer table and BM has failed (Figure S6). Second, for each calvarium, we computed the

Mahalonobis distance for all vertices in the two dimensions “first layer of the BM” and “BM intensity” (Figure S7). By manual inspection we found that data points with MD > 25 often had errors in BM layer identification, typically where the network had erroneously predicted a higher and more intense layer to be the BM. We considered a calvarium as failing this QC criterium if more than 0.5% of vertices have MD > 25. This criterium is very strict as errors on only 0.5% of data points in a calvarium would not significantly affect the average BM intensity for a calvarium. However, we wanted the data to be usable to generate individual calvarial BMA maps and were only willing to tolerate a very low percentage of data points with errors in each calvarium. These two additional QC filters were complementary (Figure S8) and resulted in a final number of 39,207 scans passing QC.

Twin heritability estimates

We used the monozygotic and same sex dizygotic twins from the *Human Connectome Project* dataset to fit the NACE model using the twinlm function of the mets package (63) in R (Table S1).

GWAS methods

We followed standard quality control procedures for preparing UK Biobank genetic data. Specifically, we removed individuals missing more than 1% of genotyping data, and single nucleotide polymorphisms (SNP) missing in more than 5% of individuals or with a minor allele frequency below 0.01, yielding approximately 7.9M SNPs. We split the available sample into a white British discovery set and a non-white replication set, as detailed under *cohorts*. Within each of these samples, we performed sex-specific covariate regression prior to genome-wide-association analysis (GWAS). For males, we residualized our estimates of BM adiposity from age (using orthogonal polynomials of degree 2), scanning site, and the first 20 genetic principal components for population level stratification. For females, we applied the same model yet also controlling for hormonal replacement therapy as this is known to affect BMA (64) (binary: ever used HRT therapy yes/no).

The residuals were fed into GWAS using plink (version 2) (65). Using the discovery sample, we first performed a GWAS in females only (N=16,902), and a GWAS in males only (N=16,140). After validating their significant genetic correlation (Rg=.94, P=6e-27), we performed a GWAS in the full discovery sample, merging the residuals from the female model with the residuals from the male model. Likewise, we performed a GWAS in the replication sample merging the residuals from a male and a female model in a joint GWAS (no sex-specific GWAS performed due to power in this sample). The resulting summary statistics were standardized using the python_convert toolbox (*https://github.com/precimed/python_convert*), removing the MHC region in chromosome 6 between base pairs 26M and 34M before feeding them into genetic correlation analysis.

GWAS summary statistics are available on the FUMA website (ID XXXX) *[made public upon acceptance]*

GWAS results were post-processed using the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) tool (66) to define genomic loci, significant independent SNPs, lead SNPs, and top lead SNPs. We used default settings which were: N = NA, Ncol = N, exMHC = 1, MHCopt = annot, extMHC = NA, ensembl = v92, genetype = protein_coding, leadP = 5e-8, gwasP = 0.05, r2 = 0.6, r2_2 = 0.1, refpanel = 1KG/Phase3, pop = EUR, MAF = 0, refSNPs = 1, mergeDist = 250. FUMA was also used to perform an eQTL analysis of the GWAS results using the GTEX v8 expression dataset. Genes, which were not the closest gene to the top lead SNP of a locus, but for which genome-wide significant SNPs were strong eQTLs, were annotated in the Manhattan plot (Figure 3).

Single-cell data and analysis

Single cell RNAseq data of BM mesenchymal lineage cells (30) was downloaded from the Single Cell Portal (*https://singlecell.broadinstitute.org/single_cell/study/SCP1017*) and used to create a Seurat 4.0 object in R-studio (count, embedding and meta-data). This was an integrated dataset of 7585 cells that contains endosteal Td+ BM cells from 1-month-old (1M) and 1.5-month-old (1.5M)

male Col2:Td mice. We did not modify the cell clustering or nomenclature, other than excluding two chondrocyte clusters as chondrocytes originate from the resting zone progenitors, but not BM MSCs (30). The dataset does not include lipid-laden adipocytes presumably due to the difficulty in analyzing these large and fatty cells using this method.

We searched the data from Zhong et al for the 54 genes displayed in the Manhattan plot (Figure 3). Human gene names were translated to mouse gene names and, if no similar name was identified, we searched for orthologs. For 3 of the 54 human genes, we could not identify a mouse gene and, for an additional 9 genes, the mouse gene had no expression in the scRNAseq data (Tables S7 and S8). Cluster gene expression was plotted as a dotplot (ggplot2) and patterns in gene expression were grouped by alignment to a dendrogram (ggtree) based on Euclidean distance. The expression of each gene was scaled by the gene's mean expression across clusters (i.e. cell types). Zero expression dots were removed.

GWAS downstream analysis and genetic architecture

We estimated SNP heritability and genomic inflation of the resulting GWAS summary statistics using LD-score regression (67). The same tool was used for estimating genetic correlations with bone mineral density (28), body mass index (68), waist-to-hip ratio (69), blood pressure (70) (systolic and diastolic), coronary artery disease (71), type-2 diabetes (72), multiple sclerosis (73), Parkinson's disease (74), Alzheimer's disease(75), general cognitive ability (37), educational attainment (76), and insomnia (77).

To estimate genetic overlap between pairs of traits we used bivariate MiXeR v1.3 (*https://github.com/precimed/mixer*). MiXeR(31) uses a Gaussian causal mixture model to estimate the total number of trait-influencing variants that explains 90% of SNP heritability (i.e., polygenicity). In this case, a trait-influencing variant is a common variant with a direct effect on the trait of interest excluding effects due to LD. Parameters are estimated with 20 iterations of the mixture models and the mean of each estimate is quantified along with the standard deviation over the 20 iterations. Given two traits of interest, MiXeR models the trait-influencing variants unique to each trait (non-overlapping) as well the number of shared trait-influencing variants (overlapping). To assess model fit, MiXeR employs the difference in Akaike information criterion between the MiXeR model and an infinitesimal model with a positive value indicative of good model fit.

Statistical analysis

All statistical analyses are described in the relevant subsections above.

Data availability

All data are available in the main text or the supplementary materials. This research has been conducted using the *UK Biobank Resource* (78) (access code 27412), the *Human Connectome Project* (58), data from Hangzhou Normal University as part of the data provided by the *Consortium for Reliability and Reproducibility* (25)*, and the MICA-MICs dataset which is available on the Canadian Open Neuroscience Platform's data portal. Human Connectome Project* data was provided by the WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. We obtained GWAS summary statistics from the following consortia: Genetic Factors for Osteoporosis Consortium, Genetic Investigation of ANthropometric Traits, ICBP International Consortium for Blood Pressure, Diabetes Genetics Replication And Meta-analysis Consortium, Coronary Artery Disease Genome-wide Replication and Meta-analysis Consortium, Complex Traits Genetics Consortium, Social Science Genetic Association Consortium, International Multiple Sclerosis Genetics Consortium, International Parkinson Disease Genomics Consortium, Psychiatric Genomics Consortium. BMA GWAS summary statistics are available on the FUMA website. All code

and software needed to generate the results is available as part of public resources: LD score regression (*https://github.com/bulik/ldsc*), FUMA (*https://fuma.ctglab.nl/*), MiXeR (*https://github.com/precimed/mixer*), BM location (simulation, training, and model).

Additional information

Funding

We acknowledge the following funding sources. TK: RCN (276082, 323961). TH: RCN (295679). OAA: RCN (223273) and EU's H2020 RIA (grant 847776). LTW: EU H2020 ERC StG (Grant 802998); RCN (223273, 249795, 248238, 286838); the South-Eastern Norway Regional Health Authority (2014097, 2015044, 2015073, 2016083, 2018037, 2018076. 2019101). We would like to thank everyone involved with the collection and sharing of imaging and genetics data used in this study. Analysis was performed on the TSD computer facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT) and on resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway.

Author contributions

Conceptualisation: TK SD TH

Methodology: TK SN TH

Investigation: TK PMB MF SN KOC TH

Visualization: TK MF TH

Writing – original draft: TK PMB MF SN KOC TH

Writing – review and editing: TK PMB MF SN KOC AF OAA LTW SD TH

Funding

Funder	Grant reference number	Author
Norges Forskningsråd (Forskningsrådet)	276082	Tobias Kaufmann
Norges Forskningsråd (Forskningsrådet)	323961	Tobias Kaufmann
Norges Forskningsråd (Forskningsrådet)	295679	Timothy Hughes

Author ORCID iDs

Srdjan Djurovic: https://orcid.org/0000-0002-8140-8061
Timothy Hughes: https://orcid.org/0000-0002-8022-6012

Additional files

Supplementary materials.

References

1. **Moore SG,** Dawson KL (1990) Red and yellow marrow in the femur: age-related changes in appearance at MR imaging. *Radiology* **175**:219-223 https://doi.org/10.1148/radiology.175.1.2315484 | PubMed
2. **Veldhuis-Vlug AG,** Rosen CJ (2018) Clinical implications of bone marrow adiposity. *J. Intern. Med* **283**:121-139 https://doi.org/10.1111/joim.12718 | PubMed
3. **Cawthorn WP,** Scheller EL (2017) Editorial: Bone marrow adipose tissue: Formation, function, and impact on health and disease. *Front. Endocrinol* **8**:112 https://doi.org/10.3389/fendo.2017.00112 | PubMed

4. **Griffith JF,** Yeung DKW, Ma HT, Leung JCS, Kwok TCY, Leung PC (2012) Bone marrow fat content in the elderly: A reversal of sex difference seen in younger subjects. *Journal of Magnetic Resonance Imaging* **36**:225-230 https://doi.org/10.1002/jmri.23619 | PubMed
5. **Hardouin P,** Rharass T, Lucas S (2016) Bone Marrow Adipose Tissue: To Be or Not To Be a Typical Adipose Tissue?. *Front. Endocrinol* **7** https://doi.org/10.3389/fendo.2016.00085 | PubMed
6. **Suchacki KJ,** Tavares AAS, Mattiucci D, Scheller EL, Papanastasiou G, Gray C, Sinton MC, Ramage LE, McDougald WA, Lovdel A, *et al.* (2020) Bone marrow adipose tissue is a unique adipose subtype with distinct roles in glucose homeostasis. *Nat. Commun* **11**:3097 https://doi.org/10.1038/s41467-020-16878-2 | PubMed
7. **Meunier P,** Aaron J, Edouard C, Vignon G (1971) Osteoporosis and the replacement of cell populations of the marrow by adipose tissue. A quantitative study of 84 iliac bone biopsies. *Clin. Orthop. Relat. Res.* **80**:147-154 https://doi.org/10.1097/00003086-197110000-00021 | PubMed
8. **Justesen J,** Stenderup K, Ebbesen EN, Mosekilde L, Steiniche T, Kassem M (2001) Adipocyte tissue volume in bone marrow is increased with aging and in patients with osteoporosis. *Biogerontology* **2**:165-171 https://doi.org/10.1023/a:1011513223894 | PubMed
9. **Rauch A,** Haakonsson AK, Madsen JGS, Larsen M, Forss I, Madsen MR, Van Hauwaert EL, Wiwie C, Jespersen NZ, Tencerova M, *et al.* (2019) Osteogenesis depends on commissioning of a network of stem cell transcription factors that act as repressors of adipogenesis. *Nature Genetics* **51**:716-727 https://doi.org/10.1038/s41588-019-0359-1 | PubMed
10. **Naveiras O,** Nardi V, Wenzel PL, Hauschka PV, Fahey F, Daley GQ (2009) Bone-marrow adipocytes as negative regulators of the haematopoietic microenvironment. *Nature* **460**:259-263 https://doi.org/10.1038/nature08099 | PubMed
11. **Lillie EM,** Urban JE, Lynch SK, Weaver AA, Stitzel JD (2016) Evaluation of Skull Cortical Thickness Changes With Age and Sex From Computed Tomography Scans. *J. Bone Miner. Res* **31**:299-307 https://doi.org/10.1002/jbmr.2613 | PubMed
12. **Law SK** (1993) Thickness and resistivity variations over the upper surface of the human skull. *Brain Topogr* **6**:99-109 https://doi.org/10.1007/bf01191074 | PubMed
13. **Herisson F,** Frodermann V, Courties G, Rohde D, Sun Y, Vandoorne K, Wojtkiewicz GR, Masson GS, Vinegoni C, Kim J, *et al.* (2018) Direct vascular channels connect skull bone marrow and the brain surface enabling myeloid cell migration. *Nat. Neurosci* **21**:1209-1217 https://doi.org/10.1038/s41593-018-0213-2 | PubMed
14. **Pulous FE,** Cruz-Hernández JC, Yang C, Paccalet A, Wojtkiewicz G, Capen D, Brown D, Wu JW, Schloss MJ, Vinegoni C, *et al.* (2022) Cerebrospinal fluid can exit into the skull bone marrow and instruct cranial hematopoiesis in mice with bacterial meningitis. *Nature Neuroscience* **25**:567-576 https://doi.org/10.1038/s41593-022-01060-2 | PubMed
15. **Mazzitelli JA,** Smyth LCD, Cross KA, Dykstra T, Sun J, Du S, Mamuladze T, Smirnov I, Rustenhoven J, Kipnis J (2022) Cerebrospinal fluid regulates skull bone marrow niches via direct access through dural channels. *Nat. Neurosci* 1-6 https://doi.org/10.1038/s41593-022-01029-1 | PubMed
16. **Cai R,** Pan C, Ghasemigharagoz A, Todorov MI, Förstera B, Zhao S, Bhatia HS, Parra-Damas A, Mrowka L, Theodorou D, *et al.* (2019) Panoptic imaging of transparent mice reveals whole-body neuronal projections and skull–meninges connections. *Nat. Neurosci* **22**:317-327 https://doi.org/10.1038/s41593-018-0301-3 | PubMed
17. **Cugurra A,** Mamuladze T, Rustenhoven J, Dykstra T, Beroshvili G, Greenberg ZJ, Baker W, Papadopoulos Z, Drieu A, Blackburn S, *et al.* (2021) Skull and vertebral bone marrow are myeloid cell reservoirs for the meninges and CNS parenchyma. *Science* **373** https://doi.org/10.1126/science.abf7844 | PubMed
18. **Zhang C,** Slade JM, Miller F, Modlesky CM (2020) Quantifying bone marrow fat using standard T1-weighted magnetic resonance images in children with typical development and in children with cerebral palsy. *Sci. Rep* **10** https://doi.org/10.1038/s41598-019-57030-5 | PubMed

19. **Shen W,** Gong X, Weiss J, Jin Y (2013) Comparison among T1-weighted magnetic resonance imaging, modified dixon method, and magnetic resonance spectroscopy in measuring bone marrow fat. *J. Obes* **2013** https://doi.org/10.1155/2013/298675 | PubMed

20. **Loevner LA,** Tobey JD, Yousem DM, Sonners AI, Hsu WC (2002) MR Imaging Characteristics of Cranial Bone Marrow in Adult Patients with Underlying Systemic Disorders Compared with Healthy Control Subjects. *AJNR Am. J. Neuroradiol* **23** PubMed

21. **Shah LM,** Hanrahan CJ (2011) MRI of spinal bone marrow: Part 1, techniques and normal age-related appearances. *American Journal of Roentgenology* **197**:1298-1308 https://doi.org/10.2214/ajr.11.7005 | PubMed

22. **Simonson TM,** Kao SCS (1992) Normal childhood developmental patterns in skull bone marrow by MR imaging. *Pediatric Radiology* **22**:556-559 https://doi.org/10.1007/bf02015347 | PubMed

23. **Cordes C,** Baum T, Dieckmeyer M, Ruschke S, Diefenbach MN, Hauner H, Kirschke JS, Karampinos DC (2016) MR-Based Assessment of Bone Marrow Fat in Osteoporosis, Diabetes, and Obesity. *Front. Endocrinol* **7**:74 https://doi.org/10.3389/fendo.2016.00074 | PubMed

24. **Sudlow C,** Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, *et al.* (2015) UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med* **12**:e1001779 https://doi.org/10.1371/journal.pmed.1001779 | PubMed

25. **Zuo XN,** Anderson JS, Bellec P, Birn RM, Biswal BB, Blautzik J, Breitner JCS, Buckner RL, Calhoun VD, Castellanos FX, *et al.* (2014) An open science resource for establishing reliability and reproducibility in functional connectomics. *Scientific Data* **1**:1-13 https://doi.org/10.1038/sdata.2014.49 | PubMed

26. **Glasser MF,** Smith SM, Marcus DS, Andersson JLR, Auerbach EJ, Behrens TEJ, Coalson TS, Harms MP, Jenkinson M, Moeller S, *et al.* (2016) The Human Connectome Project's neuroimaging approach. *Nature Neuroscience* **19**:1175-1187 https://doi.org/10.1038/nn.4361 | PubMed

27. **Koller DL,** Zheng HF, Karasik D, Yerges-Armstrong L, Liu CT, McGuigan F, Kemp JP, Giroux S, Lai D, Edenberg HJ, *et al.* (2013) Meta-analysis of genome-wide studies identifies WNT16 and ESR1 SNPs associated with bone mineral density in premenopausal women. *J. Bone Miner. Res* **28**:547-558 https://doi.org/10.1002/jbmr.1796 | PubMed

28. **Morris JA,** Kemp JP, Youlten SE, Laurent L, Logan JG, Chai RC, Vulpescu NA, Forgetta V, Kleinman A, Mohanty ST, *et al.* (2019) An atlas of genetic influences on osteoporosis in humans and mice. *Nat. Genet* **51**:258-266 https://doi.org/10.1038/s41588-018-0302-x | PubMed

29. **Tencerova M,** Kassem M (2016) The bone marrow-derived stromal cells: Commitment and regulation of adipogenesis. *Front. Endocrinol* **7**:127 https://doi.org/10.3389/fendo.2016.00127 | PubMed

30. **Zhong L,** Yao L, Tower RJ, Wei Y, Miao Z, Park J, Shrestha R, Wang L, Yu W, Holdreith N, *et al.* (2020) Single cell transcriptomics identifies a unique adipose lineage cell population that regulates bone marrow environment. *eLife* **9** https://doi.org/10.7554/elife.54695 | PubMed

31. **Frei O,** Holland D, Smeland OB, Shadrin AA, Fan CC, Maeland S, O'Connell KS, Wang Y, Djurovic S, Thompson WK, *et al.* (2019) Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nature Communications* **10**:1-11 https://doi.org/10.1038/s41467-019-10310-0 | PubMed

32. **Morris DM,** Wang C, Papanastasiou G, Gray CD, Badr S, Paccou J, Semple SIK, MacGillivray T, Cawthorn WP (2022) A novel deep learning method for large-scale analysis of bone marrow adiposity using UK Biobank Dixon MRI data. *medRxiv* https://doi.org/10.1101/2022.12.06.22283151

33. **Chan BY,** Gill KG, Rebsamen SL, Nguyen JC (2016) MR Imaging of Pediatric Bone Marrow. *Radiographics : a review publication of the Radiological Society of North America, Inc* **36**:1911-1930 https://doi.org/10.1148/rg.2016160056 | PubMed

34. **Karampinos DC,** Ruschke S, Dieckmeyer M, Diefenbach M, Franz D, Gersing AS, Krug R, Baum T (2018) Quantitative MRI and spectroscopy of bone marrow. *J. Magn. Reson. Imaging* **47**:332-353 https://doi.org/10.1002/jmri.25769 | PubMed

35. **Zopes J,** Platscher M, Paganucci S, Federau C (2021) Multi-Modal Segmentation of 3D Brain Scans Using Neural Networks. *Front. Neurol* **12** https://doi.org/10.3389/fneur.2021.653375 | PubMed

36. **Çiçek Ö,** Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O (2016) 3D U-net: Learning dense volumetric segmentation from sparse annotation. *Lect. Notes Comput. Sci* **9901 LNCS**:424-432 https://doi.org/10.1007/978-3-319-46723-8_49

37. **Savage JE,** Jansen PR, Stringer S, Watanabe K, Bryois J, De Leeuw CA, Nagel M, Awasthi S, Barr PB, Coleman JRI, *et al.* (2018) Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics* **50**:912-919 https://doi.org/10.1038/s41588-018-0152-6 | PubMed

38. **Shen J,** Chen X, Jia H, Meyers CA, Shrestha S, Asatrian G, Ding C, Tsuei R, Zhang X, Peault B, *et al.* (2018) Effects of WNT3A and WNT16 on the Osteogenic and Adipogenic Differentiation of Perivascular Stem/Stromal Cells. *Tissue Eng. Part A* **24**:68 https://doi.org/10.1089/ten.tea.2016.0387 | PubMed

39. **Zhang Q,** Pan Y, Ji J, Xu Y, Zhang Q, Qin L (2021) Roles and action mechanisms of WNT4 in cell differentiation and human diseases: a review. *Cell Death Discovery* **7**:1-10 https://doi.org/10.1038/s41420-021-00668-w | PubMed

40. **Funato Y,** Miki H (2007) Nucleoredoxin, a novel thioredoxin family member involved in cell growth and differentiation. *Antioxid. Redox Signal* **9**:1035-1057 https://doi.org/10.1089/ars.2007.1550 | PubMed

41. **Bowers RR,** Kim JW, Otto TC, Lane MD (2006) Stable stem cell commitment to the adipocyte lineage by inhibition of DNA methylation: role of the BMP-4 gene. *Proc. Natl. Acad. Sci. U. S. A* **103**:13022-13027 https://doi.org/10.1073/pnas.0605789103 | PubMed

42. **Bowers RR,** Lane MD (2007) A role for bone morphogenetic protein-4 in adipocyte development. *Cell Cycle* **6**:385-389 https://doi.org/10.4161/cc.6.4.3804 | PubMed

43. **Kim YJ,** Lee MH, Wozney JM, Cho JY, Ryoo HM (2004) Bone morphogenetic protein-2-induced alkaline phosphatase expression is stimulated by Dlx5 and repressed by Msx2. *J. Biol. Chem* **279**:50773-50780 https://doi.org/10.1074/jbc.m404145200 | PubMed

44. **Sun P,** Jia K, Zheng C, Zhu X, Li J, He L, Siwko S, Xue F, Liu M, Luo J (2019) Loss of Lgr4 inhibits differentiation, migration and apoptosis, and promotes proliferation in bone mesenchymal stem cells. *J. Cell. Physiol* **234**:10855-10867 https://doi.org/10.1002/jcp.27927 | PubMed

45. **Qiu W,** Andersen TE, Bollerslev J, Mandrup S, Abdallah BM, Kassem M (2007) Patients with high bone mass phenotype exhibit enhanced osteoblast differentiation and inhibition of adipogenesis of human mesenchymal stem cells. *J. Bone Miner. Res* **22**:1720-1731 https://doi.org/10.1359/jbmr.070721 | PubMed

46. **Taipaleenmäki H,** Abdallah BM, AlDahmash A, Säämänen AM, Kassem M (2011) Wnt signalling mediates the cross-talk between bone marrow derived pre-adipocytic and pre-osteoblastic cell populations. *Exp. Cell Res* **317**:745-756 https://doi.org/10.1016/j.yexcr.2010.12.015 | PubMed

47. **Liu C,** Gao J, Cui X, Li Z, Chen L, Yuan Y, Zhang Y, Mei L, Zhao L, Cai D, *et al.* (2021) A towering genome: Experimentally validated adaptations to high blood pressure and extreme stature in the giraffe. *Science Advances* **7**:9459-9476 https://doi.org/10.1126/sciadv.abe9459 | PubMed

48. **Baryawno N,** Przybylski D, Kowalczyk MS, Kfoury Y, Severe N, Gustafsson K, Kokkaliaris KD, Mercier F, Tabaka M, Hofree M, *et al.* (2019) A Cellular Taxonomy of the Bone Marrow Stroma in Homeostasis and Leukemia. *Cell* **177**:1915-1932.e16 https://doi.org/10.1016/j.cell.2019.04.040 | PubMed

49. **Tikhonova AN,** Dolgalev I, Hu H, Sivaraj KK, Hoxha E, Cuesta-Domínguez Á, Pinho S, Akhmetzyanova I, Gao J, Witkowski M, *et al.* (2019) The bone marrow microenvironment at single-cell resolution. *Nature* **569**:222-228 https://doi.org/10.1038/s41586-019-1104-8 | PubMed

50. **Dolgalev I,** Tikhonova AN (2021) Connecting the Dots: Resolving the Bone Marrow Niche Heterogeneity. *Frontiers in Cell and Developmental Biology* **9**:478 https://doi.org/10.3389/fcell.2021.622519 | PubMed

51. **Pino AM,** Miranda M, Figueroa C, Rodríguez JP, Rosen CJ (2016) Qualitative aspects of bone marrow adiposity in osteoporosis. *Front. Endocrinol* **7**:139 https://doi.org/10.3389/fendo.2016.00139 | PubMed

52. **Okada Y,** Aoki S, Barkovich AJ, Nishimura K, Norman D, Kjos BO, Brasch RC (1989) Cranial bone marrow in children: assessment of normal development with MR imaging. *Radiology* **171**:161-164 https://doi.org/10.1148/radiology.171.1.2928520 | PubMed

53. **Kimura F,** Kim KS, Friedman H, Russell EJ, Breit R (1990) MR imaging of the normal and abnormal clivus. *AJR Am. J. Roentgenol* **155**:1285-1291 https://doi.org/10.2214/ajr.155.6.2122682 | PubMed

54. **Shen W,** Chen J, Punyanitya M, Shapses S, Heshka S, Heymsfield SB (2007) MRI-measured bone marrow adipose tissue is inversely related to DXA-measured bone mineral in Caucasian women. *Osteoporos. Int* **18**:641-647 https://doi.org/10.1007/s00198-006-0285-9 | PubMed

55. **Bandirali M,** Di Leo G, Papini GDE, Messina C, Sconfienza LM, Ulivieri FM, Sardanelli F (2015) A new diagnostic score to detect osteoporosis in patients undergoing lumbar spine MRI. *Eur. Radiol* **25**:2951-2959 https://doi.org/10.1007/s00330-015-3699-y | PubMed

56. **Wiltgen T,** Voon C, Van Leemput K, Wiestler B, Mühlau M (2024) Intensity scaling of conventional brain magnetic resonance images avoiding cerebral reference regions: A systematic review. *PLoS One* **19**:e0298642 https://doi.org/10.1371/journal.pone.0298642 | PubMed

57. **Royer J,** Rodríguez-Cruces R, Tavakol S, Larivière S, Li Q, Vos De Wael R, Paquola C, Benkarim O, Park B.-Y, Lowe AJ, *et al.* (2021) An Open MRI Dataset for Multiscale Neuroscience. *bioRxiv* https://doi.org/10.1101/2021.08.04.454795

58. **Van Essen DC,** Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K (2013) The WU-Minn Human Connectome Project: An overview. *Neuroimage* **80**:62-79 https://doi.org/10.1016/j.neuroimage.2013.05.041 | PubMed

59. (no date) Google Colaboratory. https://colab.research.google.com/

60. **Kingma DP,** Ba J (2014) Adam: A Method for Stochastic Optimization. *arXiv* https://doi.org/10.48550/arxiv.1412.6980

61. **Srivastava N,** Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res* **15**:1929-1958

62. **Reinhold JC,** Dewey BE, Carass A, Prince JL (2019) Evaluating the Impact of Intensity Normalization on MR Image Synthesis. *Proceedings of SPIE--the International Society for Optical Engineering* **10949**:126 https://doi.org/10.1117/12.2513089 | PubMed

63. **Holst KK,** Scheike TH, Hjelmborg JB (2016) The Liability Threshold Model for Censored Twin Data. *Comput. Stat. Data Anal* **93**:324-335 https://doi.org/10.1016/j.csda.2015.01.014

64. **Syed FA,** Oursler MJ, Hefferanm TE, Peterson JM, Riggs BL, Khosla S (2008) Effects of Estrogen Therapy on Bone Marrow Adipocytes in Postmenopausal Osteoporotic Women. *Osteoporos. Int* **19**:1323 https://doi.org/10.1007/s00198-008-0574-6 | PubMed

65. **Purcell S,** Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ, *et al.* (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet* **81**:559 https://doi.org/10.1086/519795 | PubMed

66. **Watanabe K,** Taskesen E, Van Bochoven A, Posthuma D (2017) Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**:1-11 https://doi.org/10.1038/s41467-017-01261-5 | PubMed

67. **Bulik-Sullivan B,** Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM, Corvin A, *et al.* (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**:291-295 https://doi.org/10.1038/ng.3211 | PubMed

68. **Yengo L,** Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM (2018) Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet* **27**:3641-3649 https://doi.org/10.1093/hmg/ddy271 | PubMed

69. **Shungin D,** Winkler T, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, Strawbridge RJ, Pers TH, Fischer K, Justice AE, *et al.* (2015) New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**:187-196 https://doi.org/10.1038/nature14132 | PubMed

70. **Evangelou E,** Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, Ntritsos G, Dimou N, Cabrera CP, Karaman I, *et al.* (2018) Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet* **50**:1412-1425 https://doi.org/10.1038/s41588-018-0205-x | PubMed

71. **Nikpay M,** Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, CHopewell J, *et al.* (2015) A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet* **47**:1121-1130 https://doi.org/10.1038/ng.3396 | PubMed

72. **Mahajan A,** Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N, *et al.* (2018) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet* **50**:1505-1513 https://doi.org/10.1038/s41588-018-0241-6 | PubMed

73. **Patsopoulos NA,** Baranzini SE, Santaniello A, Shoostari P, Cotsapas C, Wong G, Beecham AH, James T, Replogle J, Vlachos IS, *et al.* (2019) Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **365** https://doi.org/10.1126/science.aav7188 | PubMed

74. **Nalls MA,** Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, Tan M, Kia DA, Noyce AJ, Xue A, *et al.* (2019) Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* **18**:1091-1102 https://doi.org/10.1016/S1474-4422(19)30320-5

75. **Wightman DP,** Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, Rongve A, Børte S, Winsvold BS, Drange OK, *et al.* (2021) A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat. Genet* **53**:1276-1282 https://doi.org/10.1038/s41588-021-00921-z | PubMed

76. **Lee JJ,** Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, Sidorenko J, Karlsson Linnér R, *et al.* (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics* **50**:1112-1121 https://doi.org/10.17863/cam.30493

77. **Jansen PR,** Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, de Leeuw CA, Benjamins JS, Muñoz-Manchado AB, Nagel M, *et al.* (2019) Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nature Genetics* **51**:394-403 https://doi.org/10.1038/s41588-018-0333-3 | PubMed

78. **Bycroft C,** Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, *et al.* (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**:203-209 https://doi.org/10.1038/s41586-018-0579-z | PubMed

Mycroft (2018) UKBiobank. UKBiobank. https://www.ukbiobank.ac.uk

Van Essen (2013) Human Connectome Project. Human Connectome Project. https://www.humanconnectome.org

Zuo (2014) Consortium for Reliability and Reproducibility. Consortium for Reliability and Reproducibility. ID CoRR/html/hnu_1 http://fcon_1000.projects.nitrc.org/indi/CoRR/html/hnu_1.html

Harding (2023) MICA-MICs. Canadian Open Neuroscience Platform's data portal. https://portal.conp.ca

Zhong (2020) BM mesenchymal lineage cells. Single Cell Portal. ID SCP1017 https://singlecell.broadinstitute.org/single_cell/study/SCP1017

Peer reviews

Reviewer #1 (Public review):

The authors of this study developed a method to quantify calvarial bone marrow from MRI head scans, enabling study of its composition in large datasets of adults, usually collected to study the brain. Bone marrow intensity can be semi-quantitatively measured in T1-weighted MRI scans due to the greater signal intensity of fat than watery red marrow. This is an ingenious use of the MRI-produced information for other important phenotypes, such as bone structure and marrow content. Different head types were tested for complying to the model, which is notable.

The model was also successfully validated using several publicly available MRI resources - real data - in (1) dataset consisting of 30 individuals that were scanned 10 times each at 3-day intervals, and (2) the monozygotic (MZ) twin data from the Human Connectome Project cohort. Then the authors applied this validated method to head-MRI scans from the UK Biobank (n=33,042) to extract information on spatial distribution of bone marrow adiposity (BMA) in the calvaria, allowing a GWAS to identify associated genes.

The authors revealed high heritability and identified 41 genetic loci significantly associated with the BMA trait, including six sex-specific loci. Of note, statistics estimate that 99% of BMA trait-influencing variants are shared with BMD (497 of 500 variants), which may mean these results demonstrate the biological relevance to bone health. Some of the BMA genes were found related to the Wnt pathway, including WNT16, WNT4, NXN; this is a "positive control", since the Wnt/β-catenin signaling pathway was suggested as an important determinant of BMA. Also, associations in genes (BMP4, DLX5, LGR4, LRP4, SFRP4) that are known to specifically influence adiposity, are encouraging. Integrating mapped genes with bone marrow single-cell RNA-seq data revealed patterns of adipogenic lineage differentiation and lipid loading.

The study also investigated genetic overlap between BMA and twelve (or 13) "brain and body" traits, and identified significant genetic correlations with BMI, cognitive ability and Parkinson's disease.

In sum, since MRI head scans present a hitherto unexplored opportunity to address unresolved aspects of bone marrow biology, this study is both timely and innovative.

Comments on revised version:

The authors responded most of this reviewer's comments. Their explanations are convincing. Yet, upon re-reading the revised version of this paper, I still have concerns about the clarity of mostly analysis presentation, e.g.:

Line 130-133: the sentence is still unclear: "To obtain the BM signal intensity for an individual datapoint of the calvarium, we ... averaged these BM intensities to get the (average?) BM intensity for that datapoint. Then, we averaged (again?) these datapoint intensities across the calvarium to produce the global BMA measure for the scan."

Also, I still cannot understand whether the "overlap between the true and predicted bone marrow ...below 0.7" is concerning or not, - whether this threshold of 0.7 is arbitrary.

Genetic correlation: pls. make sure it's clear that the Rg was calculated using SNP "effect sizes".

https://doi.org/10.7554/eLife.101499.2.sa3

Reviewer #2 (Public review):

Summary:

The authors set out to enable large-scale measurement of fat in skull bone marrow using routine structural brain MRI scans. They present a neural-network pipeline trained largely on realistic simulated examples and show that the resulting skull marrow measure is highly repeatable in test-retest data and consistent in monozygotic twins. Applying it to ~33,000 UK Biobank participants, they report expected population patterns (including sex- and menopause-related differences) and identify genetic and health-related associations, creating an important resource that can be built upon by researchers interested in BMA, imaging, bone, metabolism, neuroscience, ageing, haematology, and other fields.

Major strengths:

A notable methodological strength is the training strategy: by using a large, simulated dataset that captures plausible variation in skull-layer thickness and MRI intensity, the authors reduce reliance on scarce expert-labelled images. The modelling choice (using 1D intensity profiles through the skull rather than analysing the full 3D volume) appears well matched to the anatomy and offers an efficient approach for thin, layered structures. Multiple validation steps (including test-retest reliability and twin concordance) support the robustness of the measurement pipeline.

On the biological and genetic side, the study demonstrates that the skull BMA estimate relates to known correlates of marrow fat (e.g., age/sex/menopause patterns/bone density) and integrates population imaging with large-scale genetic analysis to highlight loci and candidate genes with plausible relevance to skeletal and marrow biology. The inclusion of cross-ancestry analyses and integration with cell-type-resolved gene-expression resources further improves interpretability and usability for the community.

Major limitations:

The main limitation is conceptual rather than technical: the phenotype is derived from T1-weighted MRI intensity, which does not directly separate fat and water signals and can vary with scanner and sequence settings. The manuscript provides convincing evidence that the measure is reproducible and biologically meaningful, but it should still be interpreted as a semi-quantitative proxy for marrow fat rather than a direct fat-fraction measurement. Accordingly, the genetic and phenotypic associations are likely informative, but the most direct claims about "adiposity" would be stronger if anchored to established quantitative fat-measurement imaging or spectroscopy in the skull.

The genetic "replication" analysis in a smaller, ancestrally heterogeneous non-European-ancestry sample is useful as a test of transferability, but it is not equivalent to replication in an independent cohort of similar ancestry and is expected to show reduced SNP-level reproducibility because of differences in sample size and genetic background. This should be clearly framed so readers understand what level of generalisation is supported by the current evidence.

Likely impact and utility:

Overall, the work provides a practical method for extracting new biological information from widely available brain MRI scans and should be particularly useful to researchers working with large imaging biobanks and those studying connections between bone, blood, metabolism, and brain ageing. The combination of a scalable measurement approach and openly reported genetic results is likely to accelerate follow-up studies, including cross-cohort comparisons and mechanistic work on candidate pathways.

https://doi.org/10.7554/eLife.101499.2.sa2

Reviewer #3 (Public review):

Summary:

This paper addresses a fundamental gap in bone biology: our near-complete ignorance of the in vivo dynamics of calvarial bone marrow adiposity (BMA) at population scale. The authors developed an elegant artificial neural network trained on simulated data to automatically localize and quantify the bone marrow layer within standard T1-weighted MRI head scans; scans originally acquired to study the brain but harboring rich, unexploited information about adjacent bone. Applying this method to over 33,000 individuals from the UK Biobank, they accomplished three things that had never been done before: (1) they precisely quantified the sex-dimorphic age trajectory of calvarial BMA, including the dramatic post-menopausal rise and the protective role of hormone replacement therapy; (2) they performed the first well-powered GWAS of this trait, identifying 41 genome-wide significant loci including six sex-specific ones, with SNP heritability of 31.5%; and (3) they revealed significant genetic correlations and overlap between BMA and traits including bone mineral density, Parkinson's disease, and general cognitive ability, a finding made all the more intriguing by the recently described direct vascular channels connecting calvarial bone marrow to the meninges. Integration of GWAS genes with single-cell RNA-sequencing data from mesenchymal lineage cells further illuminated which genes govern lineage commitment to the adipogenic pathway versus lipid loading in mature adipocytes.

Comments on revised version.

The reviews raised substantive points across three domains, and the authors engaged with every one of them seriously and thoroughly.

On the validation of T1-weighted MRI as a measure of BMA: Reviewer 2 raised the strongest concern, arguing that T1-weighted signal intensity had never been formally validated as a quantitative fat-fraction measure in the calvarium. The authors responded with both a principled scientific argument and new data. They assembled existing literature demonstrating that T1-weighted signal is an established semi-quantitative proxy for marrow fat in multiple skeletal sites (Loevner et al. 2002, Shen et al. 2013, Zhang et al. 2020), and provided additional comparative analyses against quantitative T1 relaxation maps, multiple intensity normalization strategies (KDE, WhiteStripe, GMM, FCM, Z-score), DEXA-derived bone mineral density, and osteoporosis status. The biological coherence of their findings, recapitulating known sex and age profiles, identifying genes already established in cell and animal models of BMA biology, and estimating heritabilities consistent with twin data constitutes powerful, convergent evidence for construct validity. Their point that a semi-quantitative measure of a highly variable, well-demarcated biological signal can outperform a perfectly precise measure of a poorly defined entity is methodologically sound and well-argued.

On sex differences and the role of Hyperostosis frontalis interna: Reviewer 1 raised the clinically astute concern that Hyperostosis frontalis interna (HFI), a condition of inner table thickening prevalent in up to 49% of postmenopausal women, could confound calvarial BMA measurements and drive apparent sex differences. The authors performed a dedicated new analysis, stratifying BMA-BMD associations by sex and age group. They demonstrated that (1) the BMA-BMD association is robust in both males and females, (2) it remains stable across age groups, and (3) the neural network trained on simulations incorporating wide anatomical variation including inner table thickness is inherently resistant to moderate inner table thickening. Given that HFI is restricted to the frontal bone, which represents only a fraction of the calvarial surface, and that severe cases are rare (ICD-10 prevalence ~0.02% in the UK

Biobank), the authors make a convincing case that this does not materially bias their results. Their suggestion that the method could itself be used in future work to study the genetic architecture of HFI is a nice forward-looking addition.

On genetic correlation interpretation and cross-trait pleiotropy: Reviewer 1 asked for clarification of the vertical versus horizontal pleiotropy distinction and for formal Mendelian randomization to support the possible causal effect of BMA on cognition. The authors appropriately clarified the conceptual framework in the revised text and, rather than overstating a causal claim without the supporting analysis, responsibly softened the language to "may be consistent with the hypothesis that BMA could have a causal effect on cognition." This is scientifically honest and appropriate.

On mouse scRNAseq and its relevance to humans: The authors acknowledged that the results section had not explicitly stated the mouse origin of the scRNAseq data, corrected this, and provided a well-justified rationale for the relevance of mouse mesenchymal lineage data to human BMA biology, which is a well-established and widely accepted model system in this field.

On GWAS replication: The claim that the study lacked replication was addressed by clarifying the a *priori* separation of discovery (white British, n=33,042) and replication (non-white British, n=4,958) samples, with 62% of significant discovery SNPs and 95% of lead SNPs replicating in the correct direction.

Overall Assessment:

This is a technically innovative, scientifically rigorous, and biologically meaningful paper. The method is genuinely novel, the sample size is among the largest ever applied to this phenotype, the genetic findings are well-powered and well-replicated, and the integration across imaging, genetics, and single-cell transcriptomics is exemplary. The authors have engaged with every substantive reviewer criticism in good faith, producing new analyses where appropriate and defending, and convincingly, with findings that were challenged without adequate basis. The revised manuscript is strengthened throughout.

This paper opens a new window quite literally, through the skull - into bone marrow biology at a scale and resolution that has never been achieved before.

https://doi.org/10.7554/eLife.101499.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public review):***
>
> *The authors of this study developed a method to quantify calvarial bone marrow from MRI head scans, enabling the study of its composition in large datasets of adults, usually collected to study the brain. Bone marrow intensity can be semi-quantitatively measured in T1-weighted MRI scans due to the greater signal intensity of fat than watery red marrow. This is an ingenious use of the MRI-produced information for other important phenotypes, such as bone structure and marrow content. Different head types were tested for complying with the model, which is notable.*
>
> *The model was also successfully validated using several publicly available MRI resources - real data - in (1) a dataset consisting of 30 individuals that were scanned 10 times each at 3-day intervals, and (2) the monozygotic (MZ) twin data from the Human Connectome*

> *Project cohort. Then the authors applied this validated method to head-MRI scans from the UK Biobank (n=33,042) to extract information on the spatial distribution of bone marrow adiposity (BMA) in the calvaria, allowing a GWAS to identify associated genes.*
>
> *The authors revealed high heritability and identified 41 genetic loci significantly associated with the BMA trait, including six sex-specific loci. Of note, statistics estimate that 99% of BMA trait-influencing variants are shared with BMD (497 of 500 variants), which may mean these results demonstrate the biological relevance to bone health. Some of the BMA genes were found related to the Wnt pathway, including WNT16, WNT4, NXN; this is a "positive control", since the Wnt/β-catenin signaling pathway was suggested as an important determinant of BMA. Also, associations in genes (BMP4, DLX5, LGR4, LRP4, SFRP4) that are known to specifically influence adiposity, are encouraging. Integrating mapped genes with bone marrow single-cell RNA-seq data revealed patterns of adipogenic lineage differentiation and lipid loading.*

With regards to the reviewer's comment on the overlap between BMA and BMD trait-influencing variants, we would like to add that the correlation of effect sizes within the overlap is -0.95 as we would expect from bifurcating differentiation of mesenchymal stem cells into osteoblasts or adipocytes: the underlying common biology driving these traits results in shared traits with a negative correlation in the effects.

> *The study also investigated the genetic overlap between BMA and twelve (or 13) "brain and body" traits and identified significant genetic correlations with BMI, cognitive ability, and Parkinson's disease.*
>
> *In sum, since MRI head scans present a hitherto unexplored opportunity to address unresolved aspects of bone marrow biology, this study is both timely and innovative.*
>
> *There are, however, some assumptions, findings, and their interpretation, which require more critical focus.*
>
> *Sex-specificity is well described and studied here. Men have higher BMA than women, but post-menopausal women catch up in the BMA values. The authors believe that calvarial marrow has a number of features that make it particularly well-suited to the study of BMA process - which is clinically important in other bone sites. It has a simple "sandwiched" structure that they are able to model. This is true only to some extent: a condition called "Hyperostosis frontalis interna", of unknown etiology (described by Smith & Hemphill in 1956) - is characterized by irregular overgrowth of the inner table of the frontal bone (symmetric/bilateral). Although not of clinical significance, typically benign, studies report a prevalence of 12%; However, it's most common in postmenopausal women - where prevalences up to 49% in women over the age of 65 - have been reported. Thus, sexual dimorphism is obvious and the effect of estrogen is likely shared with whichever bone - and marrow - age-related pathology. So, for women not using HRT, this new layer of the bone might interfere with the calvarial BMA readings and in turn, affect the BMA-related analyses.*

Thank you for bringing the "Hyperostosis frontalis interna" condition to our attention. It is particularly interesting to hear that the etiology is unknown and one may suspect that some kind of calvarial bone marrow dysregulation may be part of the cause. Our model for bone marrow location was trained on simulated data which included variation in the thickness of all anatomical layers (including the inner table), so it will be robust to some thickening of the inner table. It might not be robust to the most extreme cases of inner table thickening (as described in some case reports), but these are rare. Further, it should also be noted that the other calvarial bones, representing a much greater fraction of the calvarial surface, remain largely unaffected by the thickening and would therefore yield correct localisation of the bone marrow layers. In summary, although the severe cases of hyperostosis frontalis interna

have the potential to affect our identification of the bone marrow layer, the low frequency of such cases and the restriction of the phenotype to the frontal bone means that the potential for bias is very limited.

It would be interesting to develop a method for detection of thickened inner bone so that the condition's prevalence can be quantified in a large sample like the UK Biobank and its genetic architecture be determined. This could help elucidate the etiology.

> *The authors suspect that the effect of BMA on BMD may be biased in women; they should comment on those "with low BMD and high BMA" given that hyperostosis frontalis might be an issue. A strong effect of SNPs in the ESR1 chromosomal region might be akin to the above concern.*

Thank you for raising this point, which we have followed up with a new analysis.

According to ICD-10 data in UK Biobank there are only N=105 individuals with an M85.2-diagnosed disorder. Given the total sample size of N=446,814 individuals with ICD-10 data, this would translate to a prevalence of 0.02%, which speaks for an underdiagnosis in this sample such that we cannot simply remove diagnosed individuals to control for a potential diagnostic confound.

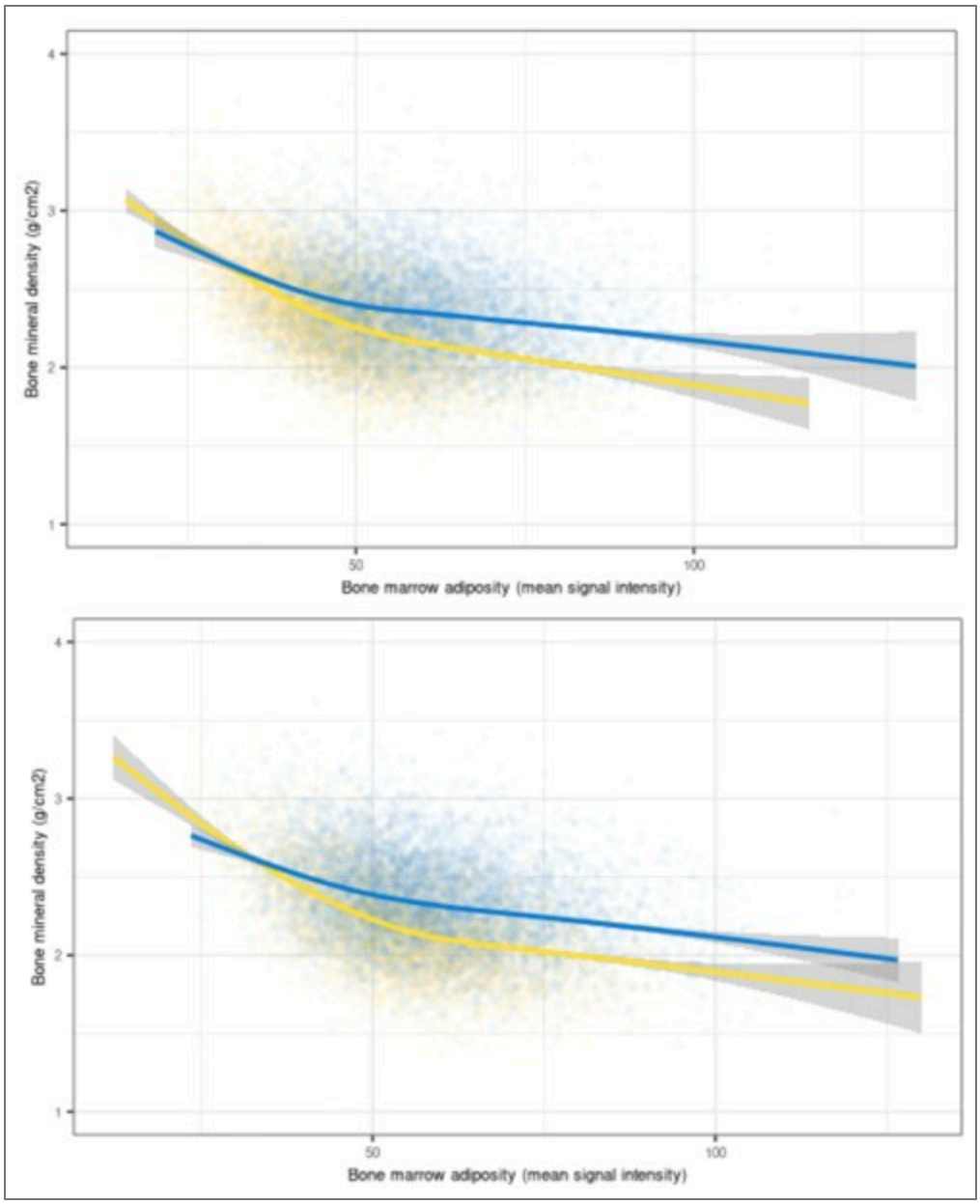
We have therefore taken a different approach to investigate this potential issue: As you elaborated in your previous comment, the prevalence of hyperostosis frontalis increases with age in females. The literature also suggests that prevalence rates do not differ between males and females in young age / prior to menopause. Therefore, we have studied the association between BMD and BMA for males and females separately, and in two age groups based on a median split of our sample (left plot: younger than 65, right plot: subjects older than 65). In these plots, the relatively large shift in female BMA and BMD is visible with the large yellow cloud at low BMA and high BMD in the left plot disappearing in the right plot. Despite this, we observe:

(1) Associations in both males (blue) and females (yellow), suggesting that the associations were not driven only by females.

(2) BMA-BMD association is largely similar across the two age groups.

If we consider that the old age group is likely to contain more cases of hyperostosis frontalis than the young group, and if we consider that old-aged females are more likely to be in this condition than men of any age, then we would expect an impact of hyperostosis frontalis on our measures to result in observable differences in Author response image 1. This is not the case. We see global age-related shifts in BMA in women, yet the association with BMD remains similar across age groups.

The technical properties of our neural network (trained on simulated data) makes it unlikely that frontal bone will contaminate the bone marrow detection globally (description above) and these results show that hyperostosis frontalis is not a considerable issue in our analysis.



Author response image 1.

> *Then, there is a perfect overlap of the BMA SNPs that are shared with BMD (497 of 500 variants), which may prove a "face validity" of the MRI-derived BMA. However, the BMD in the study was heel-derived eBMD - which is a good proxy for osteoporosis and is mostly driven by trabecular bone. Thus, there might be a concern that the BMA metrics capture some trabecular BMD.*

The reviewer is correct in pointing out that the BMA causal variants are a near-perfect subset of the BMD causal variants. The reviewer raises the concern that the BMA measurements may capture some trabecular BMD, however it should be noted that the correlation of effect sizes for the BMA/BMD overlapping causal SNPs is negative (-0.95). If our measure of BMA had been erroneously capturing trabecular BMD then we would expect to see a positive correlation of effect sizes for the BMA/BMD overlapping causal SNPs, not a negative one.

> *Next, integrating mapped genes with existing bone marrow single-cell RNA-sequencing data revealed patterns of adipogenic lineage differentiation and lipid loading. The problem here is that the scRNAseq studies of the Bone Marrow niche are overwhelmingly*

> *mouse. The authors might wish to justify why they are relevant to humans (in the absence of the human-specific scRNAseq).*

We thank the reviewer for pointing this out. We noticed that, although Figure 4 and the Methods do explicitly state that the scRNAseq data is from mouse, it is not stated in the text of the Results. This is now corrected.

The mouse is commonly used as the model organism for *in vivo* investigation of human phenotypes and bone marrow adiposity is no exception because, although mice have lower bone marrow adiposity than humans, the timing and sequence in bone marrow adiposity development are similar. BMA research makes extensive use of mouse models literature as exemplified by this review of research within the field (*https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2016.00127/full*) and this article recent article (Koh et al. 2024. "Adult skull bone marrow is an expanding and resilient haematopoietic reservoir". *https://www.nature.com/articles/s41586-024-08163-9*)

We updated the results section (line 279):

"Mesenchymal stem cells of the BM niche commit to either the adipogenic or the osteogenic lineage (Figure 4A) and both the number committing to the adipogenic lineage and their level of lipid-loading influences the total level of BMA. This aspect of BM biology is shared between humans and mice (29), so we made use of an existing mouse scRNAseq dataset of BM mesenchymal lineage cells (30) to study variation in the expression of BMA-associated genes as cells differentiate (Figure 4B)."

> *For genetic correlation analysis, the authors selected 7 body and 6 brain traits. The latter traits reflect cognition (general cognitive ability and educational attainment) and brain-related disorders. This selection might seem arbitrary. The interpretation of genetic correlation with cognitive ability, education, and Parkinson's disease was attributed to the recently discovered vascular channels that link calvarial bone marrow to the meninges. This is a fascinating hypothesis, which requires functional proof. However, there might be simpler explanations. Thus, the diploe and the inner table of the calvarium are drained by the same veins as the dura. From the anatomy textbook, we know that diploic veins connect the pericranial and endocranial venous system through the skull.*

Whilst it is true that we did not systematically compare the results of the BMA GWAS to all potentially relevant brain and body phenotypes, we did use criteria to select the phenotypes we compared to. As stated in the manuscript (line 304):

"We selected body traits (BMD, BMI, waist-to-hip ratio, systolic and diastolic blood pressure, type-2 diabetes, coronary artery disease) that have a logical connection to BMA given the mesenchymal stem cells origin of BM adipocytes and their role in bone, fat, and vasculature (29). For the brain, we selected traits reflecting cognition (general cognitive ability and educational attainment) and disorders that are prevalent in adulthood (insomnia, multiple sclerosis, Parkinson's disease, Alzheimer's disease) since it is primarily in adulthood that the adiposity of calvarial BM experiences a substantial change"

We entirely agree that the suggestion that the genetic correlation between BMA and cerebral traits may be mediated by the vascular channels linking calvarial bone marrow to the meninges is merely a hypothesis. We have therefore updated the text of the Discussion (line 470):

"We tentatively speculate that calvarial MALPs may be involved in sensing perivascular flows of CSF from the meninges to the BM and in influencing the BM's hematopoietic response, and that this might be the basis of the observed genetic overlap between BMA and some cerebral traits. However, more conventional anatomical pathways may also be relevant, as the diploë

and inner table communicate with meningeal and dural venous systems through diploic veins."

> ***Reviewer #2 (Public review):***
>
> *Summary:*
>
> *This study develops a new artificial intelligence method for high-throughput analysis of skull bone marrow from MRI data, which may be useful for large-scale biological analyses. Using this method, the authors then attempt to estimate skull bone marrow adiposity (BMA) using T1-weighted signal intensity from MRI scans of ~33,000 people, followed by genome-wide association analysis; however, the approach is inadequate because T1-weighted signal intensity is not validated for measurement of bone marrow adiposity. If it could be validated, the study would be an important advance in understanding of bone marrow adiposity and skeletal biology.*
>
> *Strengths:*
>
> *This paper is well-written, and the figures are nicely presented. The neural network method used for analysing skull bone marrow is innovative, and the authors validate this through several approaches. Therefore, the authors have achieved the aim of developing a method for large-scale analysis of skull bone marrow from MRI data.*
>
> *The GWAS is reasonably well-powered and addresses potential ethnicity differences, with one GWAS done across white males and females, and a separate GWAS in non-white participants. The methodology also conforms to common GWAS standards, including for mapping genetic variants to candidate genes. Moreover, the study further investigates the biological roles of these genes by analysing their expression in single-cell RNA sequencing data.*
>
> *Weaknesses:*
>
> *The fundamental weakness is that T1-weighted MRI signal intensity (T1W) is used as an estimate of BMA, but it has never been validated for this. The authors show that this T1W parameter measures something that is heritable and can be compared between subjects, but they don't show that it actually measures (or even estimates) calvarial BMA. There is an attempt to do so by comparing the T1W parameter with data from quantitative T1 images: the authors show a reasonable correlation with some of the quantitative T1 image data. However, this still does not show that the parameter is measuring BMA; it could be measuring some other biological characteristic, but this remains unclear. So, there is a need to validate the T1W parameter against an established measure of BMA, such as the bone marrow fat-fraction or proton density fat fraction measured from multi-echo MRI analysis.*
>
> *Without validating this BMA measurement method, it is not possible to interpret the GWAS or other findings reported in the study.*

We reject this criticism.

Although T1-weighted has not been validated as a quantitative measure of fat-fraction, there are several studies showing that it is a semi-quantitative measure of fat content (e.g. Loevner et al 2002, Shen et al 2013, Zhang et al 2020) and we also provide data that support this (figures S9-11).

Semi-quantitative measures are used in many biomedical GWASes for instance even highly heritable neuropsychiatric disorders (such as schizophrenia and bipolar disorder) involve assessment by clinicians where the test-retest kappas are in the range 0.4-0.6.

Further, we would suggest that the shortcoming of the imperfect correlation of T1w signal intensity with fat content is more than outweighed by our precision in identifying the calvarial BM cavity and the fact that the flat calvarial bone marrow has a wide range of adiposity in middle-aged and elderly individuals (compared to other bones). This lies at the root of why:

We clearly recapitulate the known sex and age profiles, as well as the effect of HRT.

We estimate high BMA heritabilities (43% in males and 23% in females)

We find clear sex differences (which is a known feature of BMA biology)

We identify a large number of the genes already known to affect BMA from earlier animal and cell work

A noisy measurement of an entity with strong biological signal (a well-defined bone marrow cavity with variation in BMA across subjects) will often be more informative than a highly precise measurement of a poorly defined entity with little signal.

> *A less critical weakness is that the GWAS has been done only on a single cohort, without replicating the findings in a follow-up cohort. For example, the authors could repeat their analysis on the remaining ~50,000 UK Biobank imaging participants for whom MRI data is now available. However, this would be pointless without knowing what biological characteristic(s) the T1W parameter is actually reflecting.*

We disagree with this comment. We separated the UKB data into discovery and replication sets prior to running the GWAS, so these datasets are independent:

(1) Further, we ran the discovery (white british individuals) GWAS separately for males and females (prior to combining) and reported in the results section: “We found them to have low genomic inflation (Figure 3A and Table S4) and to be significantly genetically correlated (Rg=.94, P=6e-27, Figure 3B)”

(2) We performed our replication GWAS in non-white British males and females. As noted in the results section: “Out of the 168 significant discovery SNPs, 62% replicated at P<.05, and 39% of the 41 lead SNPs replicated at P<.05 (Table S6). One locus replicated at genome-wide significance (P<5e-8). Furthermore, 92.7% of the lead SNPs of the discovery sample showed same effect direction in the replication sample (Table S5).”

> ***Reviewer #3 (Public review):***
>
> *Summary:*
>
> *This manuscript, "Estimating bone marrow adiposity from head MRI and identifying its genetic 2 architecture", brings together the groups of Drs. Kaufmann and Hughes in a tour de force work to develop an artificial neural network that localizes calvaria bone marrow in T1-weighted MRI head scans, with the goal of studying its composition in several large MRI datasets, and to model sex-dimorphic age trajectories, including the effect of menopause.*
>
> *Strengths:*

> *Bone marrow adiposity is a very active tissue with far-reaching implications for tissue crosstalk and human health than we had initially recognized. Although MRI has been used to measure BM, studies such as the one by these two groups are still lacking whereas very large datasets are analyzed using advanced AI machine learning tools coupled with genetic studies and a specific pathology. The groups had to develop new methods and new AI machine-learning tools for the imaging analyses.*
>
> *Weaknesses:*
>
> *Some aspects of the work that authors could add additional clarification.*
>
> *(1) Imaging Limitations: The authors provide an excellent overview and references supporting the use of MRI as a method for assessing marrow fat, particularly with some specific modifications. However, MRI images can be affected by various factors, including the presence of other tissues as well as specific MRI settings, which are much harder to precisely control when using different datasets.*

We thank the reviewer for his positive assessment of our review of methods.

Regarding MRI settings: We agree with the reviewer that differences in scan protocols can create substantial differences in the resulting images between samples. Different tools exist for harmonization of imaging data across sites, but they usually operate on tabulated data and there is no one-size-fits-all approach yet [1]. Here, we took a different approach to prevent confounding bias: We generated a large set of simulated data for training of the neural network. The simulations circumvented potential issues emerging from confound biases in training sets that we might have seen had we had combined multiple samples with different scan protocols. Nevertheless, applied to real data the models may still face confound issues, such as better BMA estimates for some scan protocols over others. We have addressed these issues as follows: (1) Validation analyses (10 repeat scans of 30 individuals and twin pairs, figure 1d and 1e) are based fully on data that was acquired on the same scanner with the same protocol. (2) Analysis in UK Biobank included data from different scan sites albeit harmonized protocols. Here we accounted for scan site in all statistical models (including GWAS).

(1) Dominik Kraft, Gloria Matte Bon, Édith Breton, Philipp Seidel, Tobias Kaufmann; Removing scanner effects with a multivariate latent approach: A RELIEF for the ABCD imaging data?. Imaging Neuroscience 2024; 2 1–7. doi: *https://doi.org/10.1162/imag_a_00157*

Regarding the presence of other tissues: We recognise in the existing text of the results section that sometimes inner or outer table voxels are wrongly identified as bone marrow, but we show that this does not have a major impact on the correct identification of the bone marrow cavity. The existing text reads:

“Poor overlap (below 0.7) was almost only observed in the thinnest bone and is explained by the fact that when the BM part of the bone is only a few layers thick (1 layer = 0.5 mm), an error by one layer will inevitably lead to a substantial fall in overlap. However, this did not result in a corresponding fall in the ratio of the predicted intensity of BM to its true intensity, because the typical BM intensity was only marginally higher than the neighbouring bone intensity. This property of the typical relative intensities of these anatomic structures also explains why the intensity ratio at high overlap is not centred on 1: any misidentification of cortical bone as BM, will typically result in an underestimate of true BM intensity (Figure S2). The neural network performed well and intensity ratios were in the range 0.9-1.1 for the vast majority of head types (Figure S3).”

Also note that we implement a number of QC measures to exclude scans where there is evidence that we may have failed to correctly identify the bone marrow cavity. The existing

text reads:

"We used two additional QC metrics to filter out calvaria where BM location was likely to have failed. First, we set an upper limit of 30 on the standard deviation of the intensity of the outer table as scans with higher values were clear outliers and were probably cases where the location of both outer table and BM has failed (Figure S6). Second, for each calvarium, we computed the Mahalonobis distance for all vertices in the two dimensions "first layer of the BM" and "BM intensity" (Figure S7). By manual inspection we found that data points with MD > 25 often had errors in BM layer identification, typically where the network had erroneously predicted a higher and more intense layer to be the BM. We considered a calvarium as failing this QC criterium if more than 0.5% of vertices have MD > 25. This criterium is very strict as errors on only 0.5% of data points in a calvarium would not significantly affect the average BM intensity for a calvarium."

> *(2) The specific density of cranial bones as it relates to the types of bone marrow: Cranial bones are extremely dense structures, which naturally interfere with MRI imaging. While it is thought that cranial bones have mostly "red bone marrow", this is only true for a short time in humans. How sensitive is their system in differentiating between red and yellow BM?*

We implemented several measures to ensure that our method would be robust to anatomical variation between individuals. As noted in the current version of the Methods section: "In order to train the neural network model, we generated a large synthetic dataset of intensity arrays, with known boundaries between anatomical structures, by simulating the thickness and intensity of the different structures located between the outer skin and the subarachnoid space. The simulation incorporated the following real-world complexities:

Different anatomical architectures (skin, subcutaneous fat, aponeurosis, outer table, BM, inner table, dura mater, arachnoid space), including when a structure is not present throughout the calvarium

Variation in thickness and intensity between vertices (on the same calvarium)

A wide variety of different calvarium types with different combinations of levels of BM adiposity, bone thickness, and subcutaneous adiposity."

Further, as noted in the Results section:

"We evaluated the performance of the neural network on simulated data using two metrics (Figure 1B): 1. the overlap between the predicted and the true BM location, and 2. the ratio between the predicted intensity of the BM and the true intensity of the BM". The accuracy in localising the bone marrow layers was good, with the only exception being: "Poor overlap (below 0.7) was almost only observed in the thinnest bone and is explained by the fact that when the BM part of the bone is only a few layers thick (1 layer = 0.5 mm), an error by one layer will inevitably lead to a substantial fall in overlap"

We also validated our procedure on real data (see Results section, subsection "Procedure validation on real data and heritability estimate"). Briefly, we checked the accuracy of our method using a dataset from the Consortium for Reliability and Reproducibility, a twin dataset from the Human Connectome Project and by manually checking many hundreds of UKBiobank scans.

We are thus confident that we accurately identify the bone marrow cavity irrespective of whether the bone marrow is red (low adiposity) or yellow (high adiposity).

> *(3) Both items above are further complicated by aging, but aging is not a linear event as we have learned. There are specific bursts of aging in humans around the age of 45 and*

> *early 60s. How do the system and model predict or incorporate these peaks of aging? It seems from the data shown that aging is reflected more as a linear phenomenon. Is this because additional aging datasets are needed?*

We agree with the reviewer that ageing probably occurs in bursts rather than being a linear process. We do see a non-linear relationship between age and BMA in our data (see figure 2B), with a more rapid rise in BMA between the ages of 45 and 65, than later in life (in women). As a result of this, when we model BMA using regression, we use orthogonal polynomials of degree 2 which allows for a non-linear relationship. However, we cannot observe bursts of BMA increase in our data because it is cross-sectional. Longitudinal data would be required to obtain information on the nature and timing of any bursts in bone marrow adiposity.

> *(4) The authors describe in richness of detail their AI learning programming and how it extracted the data from datasets. The authors also show some important correlations with specific genes, SNPs. What is not clear is how conditions such as anemia for example. An expected finding would be that patients with chronic anemia have lower bone marrow (BM) signal intensity on MRI scans than healthy people. This is because the signal intensity of BM depends on the fat-to-cell ratio in the tissue.*

We agree with the reviewer that conditions affecting the bone marrow niche have a potential to affect and be affected by bone marrow adiposity, with leukemia being a known example. This is why we believe that a method, such as the one we present here, has the potential to be useful in several biomedical fields (hematology and osteology).

> *Furthermore, patients with a host of musculoskeletal disorders ranging from osteopenia to osteoporosis, sarcopenia, and osteosarcopenia will also have altered MRI scans. When using such large datasets how did the authors control or exclude these pathological conditions, or were all these conditions likely present?*

We did not exclude specific pathologies. We were careful to train our NN model on a wide variety of skull thicknesses, bone marrow adiposity, and subcutaneous adiposity and to evaluate the performance of the model on simulated and real datasets (see answer to your point 2).

Reviewer 1 raised the issue of individuals displaying Hyperostosis frontalis interna (thickening of the inner table) and we recognize that in extreme case of this condition, where there is a major change in the anatomy of the calvarial bone, our method would probably not correctly localise the bone marrow. However, such extreme cases are rare and thus would not have a major impact on our results derived from over thirty thousand individuals. We demonstrate this with an extra analysis performed in response to the point about hyperostosis frontalis interna made by reviewer 1.

> *(5) Some of the genes and SNPs although significant showed very small correlations. What is their likely physiological significance?*

Bone marrow adiposity is a polygenic trait and we have identified 41 statistically significant loci. We had a discovery sample of approximately 30k individuals which is modest for a GWAS study, so these 41 loci are a lower bound on the number of genes influencing the BMA trait. When a large number of genes influence a trait, the effect size of an individual gene is typically relatively small. However, the SNP heritability estimates of 31.5% indicates that we are able to explain approximately one third of the phenotypic variation with the effect sizes estimated by our GWAS: this is quite a high fraction relative to many other GWASs of biomedical traits.

> *(6) The authors could use this excellent manuscript to expand their discussion to include the need for studies like theirs to be also complemented by multi-OMICS studies that will*

> *include proteomics and lipidomics of BM, bones, and muscles.*

We agree with the reviewer and hope that such studies will be undertaken in the future. We attempted to point in this direction in the last sentence of the Discussion (line 517): “Future studies can build on our developments to further validate the proposed measure of bone marrow composition and to study its effect on bone, blood, and brain”. Word count limits prevented us from further expanding on the specific kinds of studies that should be performed.

> ***Recommendations for the authors:***
>
> ***Reviewer #1 (Recommendations for the authors):***
>
> *More moderate concerns include:*
>
> *(1) In the "Genetic correlation and overlap" part, it is unclear why a high effect correlation with high SNP overlap is suggestive of vertical pleiotropy, while "moderate overlap and a low correlation of effect sizes … are more indicative of horizontal pleiotropy". This is not intuitive.*

In vertical pleiotropy (genetics > phenotype A > phenotype B): genetics drive phenotype A and phenotype A drives phenotype B (B has few direct genetic drivers of its own). If we perform a GWAS of phenotype A and a GWAS of phenotype B, one would expect to see a high overlap in the causal variants (because phenotype B is largely indirectly determined by the genetics of phenotype A) and the correlation should be high (either negative or positive) because there is a cause-effect relationship between A and B (whether phenotype A has a positive or negative effect on phenotype B).

In horizontal pleiotropy: the A and B phenotypes share the same genetic loci (but have no phenotypic influence on each other). In this case, we would observe high overlap in the associated loci, but one would not expect to see a high correlation in effect sizes (across loci) because there is no a *priori* reason to expect that genes associated with both phenotype A and phenotype B would have a consistent (negative or positive) effect across loci. For example, gene X may increase A and B, whereas gene Y may increase A but decrease B.

We have made a small update to the relevant part of the Results section and otherwise rely on the explanation above (which will be publicly available along with the manuscript):

Line 325: “These patterns of high overlap and high effect correlation within this overlap are suggestive of vertical pleiotropy i.e. a molecular mechanism influencing one trait, that in turn influences a second trait, such that most of the variants driving the first trait either have the same or the opposite direction of effect on the second trait.”

> *(2) "possible causal effect of BMA on cognition" asks for a formal analysis, like Mendelian randomization.*

Given the current wording, the reviewer is justified in asking for a formal analysis. Since we did not perform this analysis, we have changed the wording:

Line 434: “Since these are two highly correlated traits [36], a high overlap and correlation of genetic effects for both traits with BMA may be consistent with the hypothesis that BMA could have a causal effect on cognition (Table 1).”

> *(3) ll. 132-134: please reword this sentence for clarity: "Using the network-estimated location of the BM within … averaged these across all datapoints…". Please define threshold of desirable overlap between the predicted BM and the true BM (=0.7?).*

Background: The model predicts the BM localisation for a datapoint (which interval of layers of the 50 layers is bone marrow). We tested the model on a wide variety of simulated data and aim for the overlap to be as close to 1 as possible, but some error is inevitable. We found that the average overlap between the true and predicted bone marrow was only below 0.7 when the bone layer is only 4 mm thick (meaning that the bone marrow is only 1-2 mm thick). This demonstrates the high accuracy of our method in identifying a very small anatomical feature.

When applying our method to real data, we do not know the truth and therefore cannot compute the overlap between the predicted and true value. It is therefore not possible to identify datapoints where the overlap is poor (e.g. lower than 0.7) and filter them out.

Given the above, we struggle to understand in what way an overlap threshold is relevant to how we compute the signal intensity for a datapoint. Nevertheless, we recognize that the sentence pointed to by the reviewer is poorly formulated and have tried to make it clearer:

Line 130-133: “To obtain the BM signal intensity for an individual datapoint of the calvarium, we used the network model to estimate the location of the BM within the datapoint’s intensity array and averaged these BM intensities to get the BM intensity for that datapoint. Then, we averaged these datapoint intensities across the calvarium to produce the global BMA measure for the scan.”

> *In the GWAS Results, please clarify the phrases - what was "significantly genetically correlated (Rg=.94)" (also, l. 402, "genetic correlation between the sexes" - in what?).*

Genetic correlation is a statistical measure that quantifies the extent to which two traits (or the same trait in two different cohorts) are influenced by the same genetic factors. Simply put, it is the effect sizes of the SNPs in the two GWASs of interest that are correlated (after correcting for confounding effects, such as linkage desequilibrium). When comparing two GWASs, the standard formulation is to refer to their “genetic correlation”. We made a modification to the text to clarify this:

Line 239: “We found the male and female GWASs to have low genomic inflation (Figure 3A and Table S4) and to be significantly genetically correlated (Rg=.94, P=6e-27, Figure 3B)”

> *"a more than two-fold difference between the sexes" - in which metric?*

We feel that what is being compared is stated clearly in the original sentence:

Line 262: “A comparison of the male and female effect sizes of the top lead SNPs of each locus revealed 6 loci in which there is a more than two-fold difference between the sexes (loci 10, 18, 26, 30, 32, 37 in Table S5)”.

> *(4) Also In GWAS Results, a locus Dlx5 is called "SHFM" in the Supplementary Table.*

Background:

We identified 41 genome-wide significant loci and named the locus after the gene closest to the top lead SNP (bold in Figure 3C). Other genes in each locus for which genome-wide significant SNPs were eQTLs, are listed below the closest gene in normal font (Figure 3C).

In table S5, we report details of the top lead SNP for all 41 loci. We report only the nearest gene to the top lead SNP.

For locus 14, SHFM1 is the closest gene to the top lead SNP whereas DLX5 and DLX6 are genes in the locus for which genome-wide significant SNPs were eQTLs. This explains why DLX5 appears under SHFM1 in Figure 3C, but does not appear in Table S5.

> *(5) Please reword MRI jargon - "Dixon method", vertix - should be introduced, as well as abbreviation "KDE".*

We had recognised that the word “vertex” would be confusing and had replaced it by datapoint, but had unfortunately missed one occurrence in the text. This is now corrected.

Thank you for pointing out the lack of introduction of the term “KDE”. This was only explained in the supplementary materials, but has now been added to the main text:

Line 500-508: “To ensure between-subject comparability, we used the intensity normalised nu.mgz volume output by FreeSurfer. We validated this approach through comparison with well-established intensity normalization methods; Kernel Density Estimation (KDE), WhiteStripe (WS), Gaussian Mixture Model (GMM), Fuzzy C-Means (FCM), and Z-score normalization (ZS). We found the highest test-retest reliability with our approach (Figure S9), and, together with KDE (based on reference signal intensity in WM), the highest correlation with quantitative T1 relaxation maps (Figure S10).”

> ***Reviewer #2 (Recommendations for the authors):***
>
> *(1) This would be an extremely useful advance for the bone and BMA fields if only it could be confirmed that the T1W signal intensity is actually measuring BMA in some meaningful way. Or, even if not BMA, to confirm what other biological characteristic(s) it is in fact capturing. This is essential for interpreting the findings.*
>
> *(2) I note that you have compared the normalized T1W parameter with quantitative T1 data (e.g. Figure S10). However, this doesn't address the fundamental issue, because even these quantitative T1 data (e.g. from MP2RAGE) may not be measuring calvarial BMA. T1W sequences have been used to estimate BM cellularity (if not BMA directly) but are not nearly as precise as water-fat imaging. For example, one study found a reasonable correlation (0.71) between T1 relaxation times and BM fat (https://www.nature.com/articles/s41598-019-57030-5). So, if your normalized T1W parameter shows a correlation of -0.44 with the T1 MP2RAGE MRI signal (Figure S10), what does this mean in terms of how well your parameter reflects the actual BMA adiposity? We can't know this, because we also don't know if the T1 MP2RAGE signal reflects calvarial BMA.*
>
> *(3) I think my recommendations are clear from the public review. Ideally, you would be able to compare the skull BM normalized T1W parameter with PDFF data that have T2* correction (since the skull BM cavity is quite small and so may suffer from T2* effects relating to tissue inhomogeneity). But even if you had only dual-echo BMFF data, this would still be much more informative than relying only on T1 data. I hope this can be done so that the findings of the study can be properly interpreted.*

As explained above, we reject this reviewer’s claim that T1-weighted signal intensity cannot be used to perform a GWAS of BMA: other studies have shown that T1-weighted signal intensity is a semi-quantitative measure of fat fraction, we have performed extra analyses that confirm this, and our results further demonstrate this. For further detail on why we reject this criticism, see our response to this reviewer’s comments.

https://doi.org/10.7554/eLife.101499.2.sa0