





OPEN

## Evidence for correlations between BMI-associated SNPs and circRNAs

Luisa Sophie Rajcsanyi<sup>1,2</sup>, Inga Diebels<sup>1</sup>, Lydia Pastoors<sup>1</sup>, Deniz Kanber<sup>3</sup>, Triinu Peters<sup>1,2</sup>, Anna-Lena Volckmar<sup>4</sup>, Yiran Zheng<sup>1,2</sup>, Martin Grosse<sup>3</sup>, Christoph Dieterich<sup>5,6</sup>, Johannes Hebebrand<sup>1,2</sup>, Frank J. Kaiser<sup>3</sup>, Bernhard Horsthemke<sup>3</sup> & Anke Hinney<sup>1,2</sup>

Circular RNAs (circRNAs) are regulators of processes like adipogenesis. Their expression can be modulated by SNPs. We analysed links between BMI-associated SNPs and circRNAs. First, we detected an enrichment of BMI-associated SNPs on circRNA genomic loci in comparison to non-significant variants. Analysis of sex-stratified GWAS data revealed that circRNA genomic loci encompassed more genome-wide significant BMI-SNPs in females than in males. To explore whether the enrichment is restricted to BMI, we investigated nine additional GWAS studies. We showed an enrichment of trait-associated SNPs in circRNAs for four analysed phenotypes (body height, chronic kidney disease, anorexia nervosa and autism spectrum disorder). To analyse the influence of BMI-affecting SNPs on circRNA levels *in vitro*, we examined rs4752856 located on hsa\_circ\_0022025. The analysis of heterozygous individuals revealed an increased level of circRNA derived from the BMI-increasing SNP allele. We conclude that genetic variation may affect the BMI partly through circRNAs.

The prevalence of obesity is increasing globally<sup>1–3</sup>. Due to an elevated body mass index (BMI), the risk to develop comorbidities, like cardiovascular diseases, diabetes, cancers, and Alzheimer's disease, increases concomitantly, causing a greater risk for a premature death<sup>3–5</sup>. Genetic variants associated with BMI have been identified in genome-wide association studies (GWAS)<sup>4,6–8</sup>. The most recent meta-analysis of BMI GWAS detected 941 genetic loci associated with the BMI<sup>9</sup>. The functional characterisation of these single nucleotide polymorphisms (SNPs) remains challenging, as the majority of GWAS-identified signals reside within non-coding regions and are thus located in enhancers or sequences harbouring non-coding RNAs (ncRNAs), such as circular RNAs (circRNAs)<sup>6,10,11</sup>.

CircRNAs derive from a backsplicing process linking a downstream splice donor to an upstream splice acceptor forming covalently closed loops<sup>12–14</sup>. These circular transcripts are expressed in a cell-type-, tissue-, and developmental stage-specific manner, exhibiting the highest enrichment in the brain and during neurogenesis<sup>12,14–17</sup>. Previous studies linked circRNAs to various diseases and cellular processes, such as the preadipocyte differentiation, known as adipogenesis<sup>18–21</sup>. A knockdown of circSAMD4A, for instance, inhibited the adipogenesis and eventually led to a reversion of the murine weight gain when fed a high-fat diet. In humans, overexpression of circSAMD4A was associated with an increased BMI<sup>19</sup>. Dynamic regulation of circRNAs during adipogenesis and in obesity provided insights in the emerging regulatory roles of these transcripts. Two circRNAs, circTshz2-1 and circArhgap5-2, were found to be substantial regulators of the preadipocyte maturation as depletions of the circRNAs inhibited the respective mechanism<sup>18</sup>.

Evidence has emerged that genetic variations can affect circRNAs. Accordingly, the level of circRNAs is influenced in a genotype-specific manner<sup>22,23</sup>. For example, carriers of the C-allele of a multiple-sclerosis (MS)-associated SNP showed increased levels of a STAT3-derived circRNA<sup>22</sup>. Additionally, SNPs located within the flanking sites of circRNAs showed higher correlations with the respective circular transcript than SNPs located within the circRNA sequence<sup>24</sup>.

<sup>1</sup>Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Essen, University of Duisburg-Essen, Essen, Germany. <sup>2</sup>Center for Translational Neuro- and Behavioural Sciences, University Hospital Essen, Essen, Germany. <sup>3</sup>Institute of Human Genetics, University Hospital Essen, Essen, Germany. <sup>4</sup>Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany. <sup>5</sup>Department of Internal Medicine III, University Hospital Heidelberg, Heidelberg, Germany. <sup>6</sup>German Center for Cardiovascular Research (DZHK), Partner site Heidelberg/Mannheim, Heidelberg, Germany. ✉email: luisa.rajcsanyi@uni-due.de; anke.hinney@uni-due.de

Analysis	Group classification	Number of SNPs
P-value based analysis	Significant ( $P < 5 \times 10^{-8}$ )	40,835
	Non-significant ( $P \geq 5 \times 10^{-8}$ )	2,283,734
Sensitivity analysis	Significant ( $P < 5 \times 10^{-8}$ )	40,835
	Non-significant ( $P \geq 5 \times 10^{-7}$ ; excluded SNPs with $5 \times 10^{-8} < P < 5 \times 10^{-7}$ )	2,269,795
	Non-significant ( $P \geq 5 \times 10^{-6}$ ; excluded SNPs with $5 \times 10^{-8} < P < 5 \times 10^{-6}$ )	2,248,179
	Non-significant ( $P \geq 5 \times 10^{-5}$ ; excluded SNPs with $5 \times 10^{-8} < P < 5 \times 10^{-5}$ )	2,215,048
Linkage disequilibrium (LD) analysis	Significant (within a 1 Mb region)	805,908
	Non-significant (not within a 1 Mb region)	1,518,661
	Significant (in regions of high LD)	77,104
	Non-significant (not within regions of high LD)	2,247,465

**Table 1.** SNP classification for the analyses of the BMI GWAS data. Following the processing of data by removing SNPs with ambiguous information and mappings, the analyses of the BMI GWAS data<sup>9</sup> were split into three sub-analyses. The initial analysis and its SNP set assignment was based on the general GWAS P-value threshold of  $5 \times 10^{-8}$ . The ensued sensitivity analysis aimed for a more stringent group delineation between significant and non-significant SNPs by excluding SNPs whose P-value fell within a certain range. Each non-significant group was separately analysed in relation to the initial significant SNP group ( $P < 5 \times 10^{-8}$ ). Considering two distinct approximations for the LD of SNPs, non-significant variants were added to the respective significant group if they were either located within a 1 Mb region adjacent to a significant SNP or if they are located in known regions of high LD (as extracted from plinkQC<sup>70</sup>).

To date, there is a lack of studies exploring GWAS-identified SNPs and their impact on circRNAs, particularly in the context of BMI. Given that circRNAs are highly expressed in the brain<sup>16</sup> and energy homeostasis is associated with the hypothalamus<sup>25</sup>, we speculate there may be links between SNPs that are associated with BMI alterations and circRNAs. Consequently, the present study aimed to investigate interactions of genome-wide significant SNPs ( $P < 5 \times 10^{-8}$ ) identified in the most recent meta-analysis of BMI GWAS<sup>9</sup> with circRNAs. We initially examined whether these significant SNPs are more frequently located on circRNA genomic loci compared to non-significant variants. Investigation of additional GWAS studies pertaining to psychiatric, neurological, anthropometric, and peripheral traits ensued. Subsequently, we selected one genome-wide significant SNP (rs4752856) for BMI, which is located on a circRNA derived from an obesity-associated gene (*MTCH2*; hsa\_circ\_0022025), and assessed the relative levels of the circRNA derived from the two alleles in vitro.

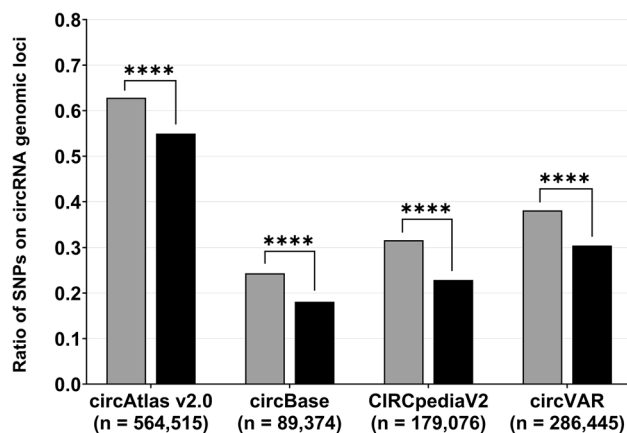
## Results

**Enrichment of genome-wide significant BMI-SNPs.** Initially, we investigated if the proportion of genome-wide significant SNPs ( $P < 5 \times 10^{-8}$ ) on circRNA genomic loci deviates from the ratio of non-significant SNPs ( $P \geq 5 \times 10^{-8}$ ) located within these regions. Therefore, we used data from the most recent BMI GWAS meta-analysis<sup>9</sup> (see Table 1 and Supplementary Table S2) and four publicly available circRNA databases (circAtlas v2.0, circBase, CIRCpediaV2 and circVAR; see Supplementary Table S1). Due to different reference genome versions between circRNA and the majority of SNP data (see Supplementary Table S1 and S2), we subjected the SNP data to a genome version lift over to ensure compatibility with all circRNA databases. Notably, due to ambiguous mappings, over 10,000 SNPs had to be removed from our SNP pool. In total, 2,324,569 BMI-analysed SNPs remained in the study (see Table 1).

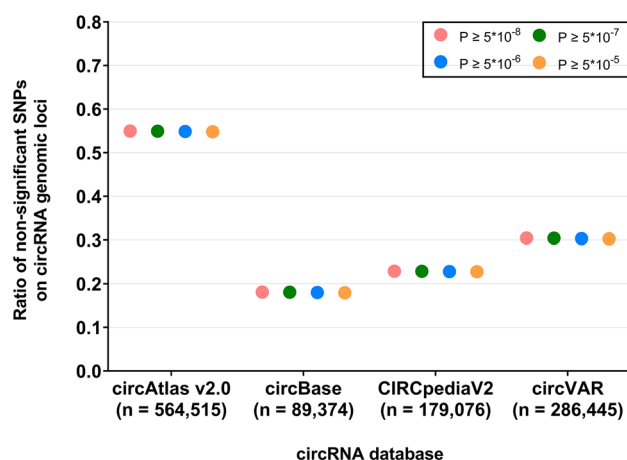
We observed a significant and consistent enrichment of genome-wide significant SNPs ( $P < 5 \times 10^{-8}$ , see Table 1) for BMI on circRNA genomic loci throughout all circRNA datasets (see Fig. 1 and Supplementary Table S5). Thus, 62.86% (25,669 SNPs) of all significant SNPs were present on at least one circRNA genomic locus included in circAtlas v2.0 (GRCh38), while the proportion of non-significant SNPs on circRNAs was significantly lower (54.96%;  $P < 0.0001$ ; OR = 1.39; 95% CI = [1.36, 1.42]). Analogous findings were detected in the datasets of circBase, CIRCpediaV2 and circVAR. Genomic loci of circRNAs included in circBase harboured 24.36% of significant SNPs (9,949 SNPs;  $P < 0.0001$ ; OR = 1.46; 95% CI = [1.43, 1.49]), while circRNAs extracted from CIRCpediaV2 and circVAR encompassed 31.62% (12,913 SNPs;  $P < 0.0001$ ; OR = 1.56; 95% CI = [1.53, 1.59]) and 38.11% (15,561 SNPs;  $P < 0.0001$ ; OR = 1.41; 95% CI = [1.38, 1.44]) of significant BMI-SNPs, respectively. The ratio of non-significant SNPs within the genomic coordinates of circRNAs remained significantly lower (see Fig. 1 and Supplementary Table S5).

**Sensitivity analysis reveals marginal deviations.** Next, we tested the sensitivity of our analysis by classifying the non-significant SNPs in accordance with three newly defined thresholds ( $P \geq 5 \times 10^{-7}$ ,  $P \geq 5 \times 10^{-6}$  and  $P \geq 5 \times 10^{-5}$ ). Despite a stricter definitional delineation, we observed solely marginal deviations in ratios of non-significant SNPs located on circRNA genomic loci (see Fig. 2 and Supplementary Table S6). Accordingly, we still detected a statistically significant enrichment of genome-wide significant BMI variants on the genomic loci of circRNAs of all databases when re-classifying the group of non-significant SNPs (see Supplementary Table S6).

**Defining approximations of the linkage disequilibrium of SNPs does not affect the outcome.** Non-significant SNPs can be located within linkage disequilibrium (LD) regions comprising signifi-



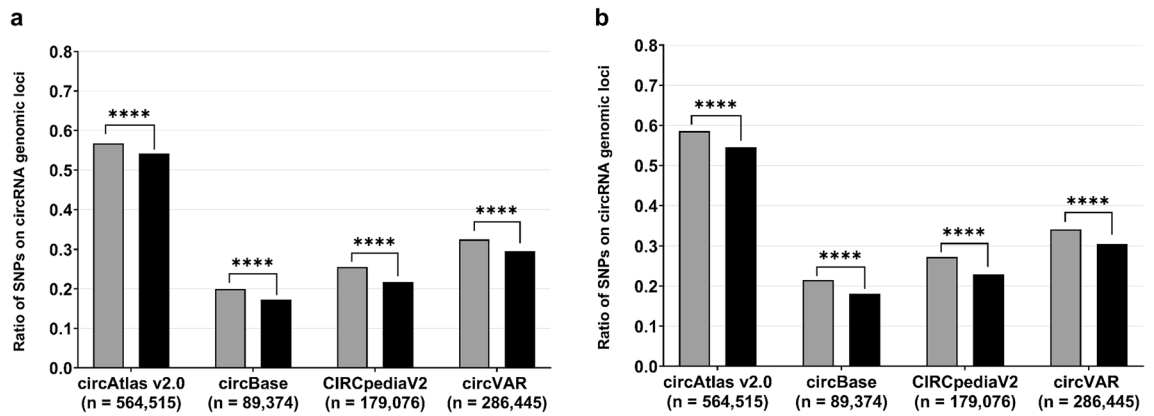
**Figure 1. Genome-wide significant SNPs for BMI are enriched in genomic loci of circRNAs.** SNPs were classified as significant and non-significant based on the genome-wide P-value threshold of  $5 \times 10^{-8}$ . Correspondingly, SNPs with a lower P-value than this threshold were considered significant (grey; total of 40,835 SNPs), while SNPs with P-values exceeding this cut-off were defined as non-significant (black; total of 2,283,734 SNPs). Genomic positions of these SNPs were checked if they matched the genomic coordinates of circRNAs extracted from the databases circAtlas v2.0 (GRCh38), circBase, CIRCpediaV2 and circVAR using a custom R script. The numbers shown in the parentheses indicate the number of circRNAs included in the respective dataset (see also Supplementary Table S1). The results of the statistical tests can be found in Supplementary Table S5. \*\*\*\* $P < 0.0001$ ; SNP single nucleotide polymorphism.



**Figure 2. Sensitivity analysis reveals deviations of ratios of circRNA-located non-significant SNPs.** To obtain a more distinct group delineation between significant and non-significant BMI-SNPs within the sensitivity analysis, the non-significant SNPs ( $P \geq 5 \times 10^{-8}$ ; pink) were re-classified applying the additional P-value cut-offs of  $5 \times 10^{-7}$  (green),  $5 \times 10^{-6}$  (blue) and  $5 \times 10^{-5}$  (orange). Non-significant SNPs with P-values ranging between these thresholds and the initial threshold of  $5 \times 10^{-8}$  were removed from the respective group. The results were obtained by applying the custom R script to the newly defined dataset of non-significant SNPs and were compared to the initial group of significant SNPs ( $P < 5 \times 10^{-8}$ ). The respective statistical outcomes are stated in Supplementary Table S6. SNP single nucleotide polymorphism.

cant variants<sup>26,27</sup>. By defining approximations for the LD structures of the BMI-associated SNPs, we obtained an overall increase in the number of 'significant' variants and a decrease in the quantity of non-significant SNPs (see Table 1). Remarkably, by applying these newly defined groups to our R script, we detected similar results as in our previous examinations. Again, the genome-wide significant SNPs were more frequently located on circRNA genomic loci than non-significant variants, regardless of whether the LD was approximated by a 1 Mb region adjacent to significant SNPs or by known regions of high LD (see Fig. 3 and Supplementary Table S7).

**Enrichment of genome-wide significant SNPs is not a general characteristic.** Given the observed enrichment of genome-wide significant BMI-SNPs on genomic loci of circRNAs, we aimed to uncover



**Figure 3. Consideration of the linkage disequilibrium of BMI-SNPs located on circRNA genomic loci.**

(a) To correct for the LD of SNPs, we have defined non-significant SNPs (black; total of 1,518,661 SNPs) that were located within a 1 Mb region adjacent to a significant SNP ( $P < 5 \times 10^{-8}$ ) as a 'significant region' (grey; total of 805,908 SNPs) as well. (b) Further, in an additional LD approximation, non-significant variants (black; total of 2,247,465 SNPs) located within an already described region of high LD (as stated in plinkQC<sup>70</sup>) were classified as 'significant' (grey; total of 77,104 SNPs). The corresponding statistical results can be obtained from Supplementary Table S7. \*\*\*\* $P < 0.0001$ ; SNP single nucleotide polymorphism.

whether this localisation is a BMI-specific feature or if it is a general phenomenon. Therefore, we extracted data of additional GWAS studies analysing arbitrary, but distinct phenotypes (see Supplementary Tables S2 and S3). As circRNAs are generally highly expressed in the brain and especially during neurogenesis<sup>16</sup>, we assumed that this SNP enrichment might be an effect exclusively seen for traits associated with the central nervous system (CNS). To investigate a widespread range of traits, we have chosen GWAS datasets pertaining to one anthropometric trait (body height<sup>9</sup>), two neurological conditions (amyotrophic lateral sclerosis<sup>28</sup> and epilepsy<sup>29</sup>), four peripheral diseases (chronic kidney disease<sup>30</sup>, heart failure<sup>31</sup>, pernicious anemia<sup>32</sup> and ulcerative colitis<sup>33</sup>) and two psychiatric disorders (anorexia nervosa<sup>34</sup> and autism spectrum disorder<sup>35</sup>; see Supplementary Tables S2 and S3). As we identified consistent results throughout the four circRNA datasets for BMI and since GWAS data is predominantly provided in the GRCh37 genome version, we decided to exclusively analyse the additional SNP datasets with the circRNAs extracted from circBase. Therefore, all prior performed processing steps are omitted.

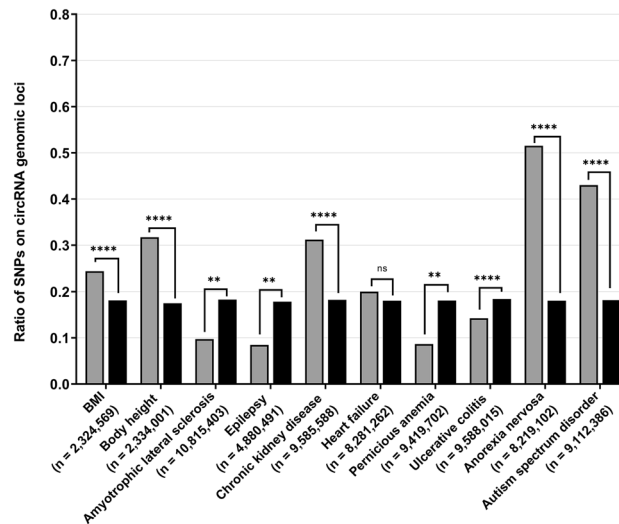
A statistically significant enrichment of genome-wide significant SNPs ( $P < 5 \times 10^{-8}$ ) was determined for SNPs from GWAS pertaining to both psychiatric disorders, namely anorexia nervosa (AN;  $P < 0.0001$ ; OR = 4.83; 95% CI = [3.88, 6.00]) and autism spectrum disorder ( $P < 0.0001$ ; OR = 3.40; 95% CI = [2.26, 5.12]), as well as the anthropometric trait of the body height ( $P < 0.0001$ ; OR = 2.19; 95% CI = [2.17, 2.22]) and the peripheral chronic kidney disease ( $P < 0.0001$ ; OR = 2.03; 95% CI = [1.82, 2.27]; see Fig. 4 and Supplementary Table S8). No significant difference of circRNA genomic loci localisation was detected for SNPs extracted from a heart failure GWAS ( $P = 0.38$ ; OR = 1.14; 95% CI = [0.85, 1.51]; see Fig. 4 and Supplementary Table S8). The inverse circRNA enrichment of non-significant SNPs was detected for SNPs from GWAS studies analysing the neurological traits of amyotrophic lateral sclerosis (ALS;  $P = 0.003$ ; OR = 0.48; 95% CI = [0.30, 0.78]) and epilepsy ( $P = 0.008$ ; OR = 0.43; 95% CI = [0.22, 0.82]) as well as the peripheral diseases pernicious anemia ( $P = 0.003$ ; OR = 0.43; 95% CI = [0.24, 0.76]) and ulcerative colitis ( $P < 0.0001$ ; OR = 0.74; 95% CI = [0.69, 0.78]; see Fig. 4 and Supplementary Table S8).

Considering that a low BMI is also a characteristic of AN<sup>36</sup>, we additionally wanted to assess if significant SNPs localised on circRNAs overlap between AN and BMI. Of the 9949 significant BMI-SNPs located on circBase-extracted circRNAs (see Supplementary Table S5) and the 168 significant and circRNA-located AN-SNPs (see Supplementary Table S8), none overlapped (not shown).

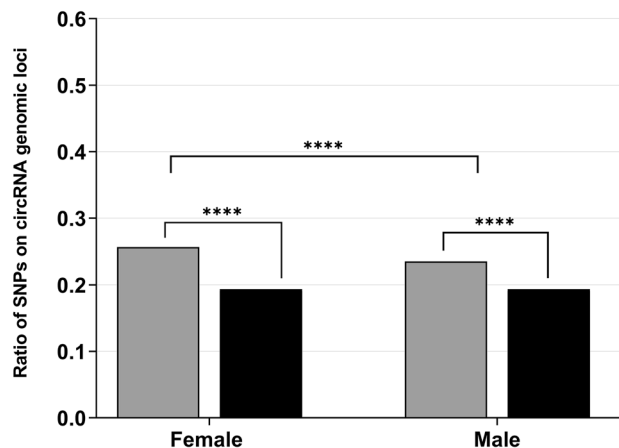
### Deviations of significant BMI-SNPs on circRNAs genomic loci between females and males.

To assess if the enrichment of significant BMI-SNPs on circRNA genomic loci applies equally to both sexes, we extracted data from an additional GWAS for BMI<sup>37</sup> examining both sexes separately. More than 25% of the significant BMI SNPs for females are localised in at least one circRNA genomic locus included in circBase (7109 SNPs; see Fig. 5 and Supplementary Table S9). The proportion of significant SNPs for males was lower with 23.53% (5372 SNPs). For both sexes, the fraction of non-significant SNPs was identical at 19.35% (females: 5,292,550 SNPs; males: 5,294,287 SNPs). We were thus able to detect a significant enrichment of genome-wide significant variants for both, females ( $P < 0.0001$ ; OR = 1.44; 95% CI = [1.40, 1.48]) and males ( $P < 0.0001$ ; OR = 1.28; 95% CI = [1.24, 1.32]; see Fig. 5 and Supplementary Table S9). By comparing both sexes with each other, we detected a significant sex difference for the circRNA localisation of the genome-wide significant SNPs. BMI-associated SNPs significant in females are more abundant on circRNA genomic loci than the significant SNPs for BMI in males ( $P < 0.0001$ ; OR = 1.13; 95% CI = [1.08, 1.17]; see Fig. 5 and Supplementary Table S10). Additionally, we were able to confirm this enrichment of BMI-associated SNPs in females in the largest dataset of circAtlas v2.0 (GRCh37; see Supplementary Fig. S1 and Supplementary Table S10).

Subsequently, we assessed whether circRNA-located SNPs significant in females and males overlap. We have found 2780 SNPs that were significant for both sexes and were localised on at least one circRNA (not shown). To



**Figure 4. Enrichment of genome-wide significant SNPs is not a cross-phenotype feature.** To test whether the detected enrichment of genome-wide significant SNPs ( $P < 5 \times 10^{-8}$ ) is BMI-specific or a general characteristic, we extracted datasets of additional GWAS studies analysing various phenotypes. We have selected one anthropometric trait (body height<sup>9</sup>), two neurological diseases (amyotrophic lateral sclerosis<sup>28</sup> and epilepsy<sup>29</sup>), four peripheral diseases (chronic kidney disease<sup>30</sup>, heart failure<sup>31</sup>, pernicious anemia<sup>32</sup> and ulcerative colitis<sup>33</sup>) as well as two psychiatric disorders (anorexia nervosa<sup>34</sup> and autism spectrum disorder<sup>35</sup>). The results of the BMI GWAS are presented as a reference and are exclusively based on the analysis with the circBase circRNA dataset. The corresponding statistical results are stated in Supplementary Table S8. \*\*\*\* $P < 0.0001$ ; \*\* $P < 0.01$ ; *ns* not significant, *SNP* single nucleotide polymorphism.



**Figure 5. Deviations of genome-wide significant BMI-SNPs located on circRNAs between females and males.** To explore whether the enrichment of significant BMI-associated SNPs deviates between females and males, we extracted data of a BMI-GWAS<sup>37</sup> analysing both sexes separately. We applied the custom R script to the significant (grey; total of 27,653 SNPs in females; total of 22,833 SNPs in males) and non-significant (black; total of 27,352,598 SNPs in females; total of 27,357,420 SNPs in males) SNP data concomitantly with the circRNA data extracted from circBase separately for both sexes. Subsequently, we compared the number of circRNA-located significant SNPs for females against the quantity of SNPs encompassed in circRNA loci for males (Chi-square test). The statistical results are stated in Supplementary Table S9. \*\*\*\* $P < 0.0001$ ; *SNP* single nucleotide polymorphism.

obtain information about overrepresented functions and cellular compartments via a gene ontology (GO) analysis, we first mapped those overlapping SNPs to their corresponding genes using the Variant Effect Predictor (VEP) by Ensembl. The SNPs mapped to 198 unambiguous Ensembl gene IDs. The GO analyses revealed no significant overrepresentation for any biological process. By analysing the GO molecular functions terms, 160 of those genes were shown to be involved in protein binding (GO:0005515, FDR  $P = 4.83 \times 10^{-8}$ ). Further, the 198 genes seem to be enriched in the nucleus (GO:0005634, FDR  $P = 5.89 \times 10^{-4}$ ), membrane-bounded organelles (GO:0043227,

FDR  $P = 6.36 \times 10^{-4}$ ; GO:0043231, FDR  $P = 6.82 \times 10^{-4}$ ), nuclear lumen (GO:0031981, FDR  $P = 8.25 \times 10^{-4}$ ), intracellular organelles (GO:0043229, FDR  $P = 8.29 \times 10^{-4}$ ), cytosol (GO:0005829, FDR  $P = 8.51 \times 10^{-4}$ ), nucleoplasm (GO:0005654, FDR  $P = 1.27 \times 10^{-3}$ ), intracellular anatomical structure (GO:0005622, FDR  $P = 1.54 \times 10^{-3}$ ), organelles (GO:0043226, FDR  $P = 2.20 \times 10^{-3}$ ), cytoplasm (GO:0005737, FDR  $P = 6.30 \times 10^{-3}$ ), membrane-enclosed lumen (GO:0031974, FDR  $P = 1.23 \times 10^{-2}$ ), organelle lumen (GO:0070013, FDR  $P = 1.34 \times 10^{-2}$ ; GO:0043233, FDR  $P = 1.46 \times 10^{-2}$ ) and cytoplasmic stress granules (GO:0010494, FDR  $P = 2.59 \times 10^{-2}$ ).

**Interactions of microRNAs (miRNAs) and circRNAs.** As it is known that circRNAs often function as miRNA sponges<sup>14,38</sup>, we aimed to explore potential interactions between circRNAs harbouring significant BMI-SNPs and miRNAs. Thus, we downloaded data from the Encyclopedia of RNA Interactions (ENCORI)<sup>39</sup> containing 475,341 miRNA target sites (GRCh37) of 643 miRNAs that were correlated with various circBase circRNAs. In total, 5717 circRNAs from the circBase database contained at least one of the 9949 significant SNPs for BMI (see Supplementary Table S5). Of these, 4100 (71.72%) were linked to at least one miRNA target site (not shown). The 20 most miRNA-comprising circRNAs by raw numbers were on average 1,243,002 bp long. To be able to estimate more precisely the number of miRNA binding sites presumably present in a given circRNA, we have normalised the number of miRNA targets found for a specific circRNAs on its length (see Supplementary Fig. S2b). Thereby, hsa\_circ\_0005103 showed the highest ratio of miRNAs based on its length (0.1449; see Supplementary Fig. S2b), while hsa\_circ\_0007319 showed the highest raw count of miRNA target sites (4275; see Supplementary Fig. S2a). 641 miRNAs (99.69%) could be linked to a circRNA containing at least one BMI-SNP. Of these, two of the miRNAs with the highest numbers of circRNAs linked (hsa\_miR-15a-5p and hsa\_miR-16-5p; see Supplementary Fig. S3) have been implicated in obesity-related traits<sup>40–43</sup>.

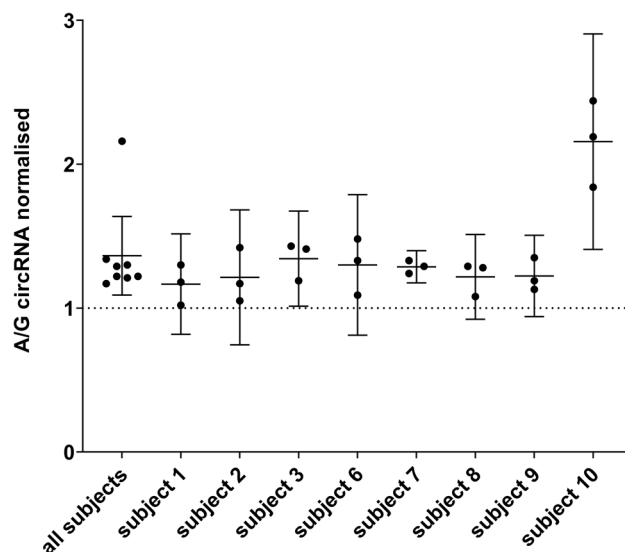
**Experimental validation of the SNP and a circRNA candidate.** Next, to assess whether BMI-associated SNPs might affect the level of circRNAs, we selected one genome-wide significant SNP for BMI located on a circRNA. Accordingly, we screened for a genome-wide significant SNP detected in the BMI GWAS<sup>9,44</sup> that is prospectively localised on a circRNA extracted from circBase, while further fulfilling predefined requirements. 9949 BMI-associated SNPs were found to be located on circRNAs included in circBase (see Supplementary Table S5). Of these, 5471 variants exhibited a MAF of at least 30% and 1749 SNPs were further detected in regions of circRNAs previously confirmed in multiple cell lines and tissues. After the validation of the set distance between SNP and circRNAs' backsplice junction (BSJ) and LD structures, the SNP rs4752856 (chr11: 47,648,402; GRCh37; MAF = 0.3526;  $P = 9.4 \times 10^{-46}$ )<sup>9</sup> whose risk allele (A) predisposes to an increased BMI ( $\beta = 0.0242$ )<sup>9</sup> was selected. It is located on the mitochondrial carrier homologue 2 gene (*MTCH2*)-derived circRNA hsa\_circ\_0022025 (chr11: 47,647,226–47,648,679; GRCh37). By genotyping 21 recruited participants (age:  $24.95 \pm 3.46$  years old; BMI:  $20.77 \pm 1.77$  kg/m<sup>2</sup>), we identified that ten probands were heterozygous (G/A), ten were homozygous for G/G and one was homozygous for A/A. The Hardy–Weinberg-Equilibrium was thus fulfilled.

**Allelic levels of rs4752856-containing circRNAs are skewed in favor of the risk allele.** To determine the relative levels of the circRNAs derived from the G- and A-allele of rs4752856, we performed highly sensitive primer-extension assays (SNaPshot)<sup>45</sup> on the genomic DNA (gDNA) and circRNA-derived complementary DNA (cDNA) of heterozygous individuals. The gDNA was used to normalize the circRNA values. For each sample, the assay was performed in triplicates. Homozygous samples acted as assay controls and were not further analysed (see Supplementary Figs. S4 and S5). Two samples of heterozygous individuals were excluded from subsequent statistical analysis due to low signals and high background noise. Thus, eight heterozygous samples remained in the study (for exemplary output see Supplementary Fig. S6). All of these samples revealed a distinct skew in favor of the BMI-increasing allele (A-allele; see Fig. 6 and Table 2). The one-sample Wilcoxon test showed that the normalised circRNA value was significantly higher (observed median = 1.25; asymptotic two-sided  $P$ -value = 0.012; effect size  $r = 0.63$ ; see Supplementary Table S11) than the hypothesized median value of 1. Collectively, we ascertained that the A-allele of rs4752856 was 25% more abundant on the circRNA than the G-allele (see Fig. 6 and Table 2).

## Discussion

Our study aimed to analyse the putative interactions of SNPs affecting BMI variation and circRNAs. Therefore, we initially investigated whether genome-wide significant SNPs ( $P < 5 \times 10^{-8}$ ) for BMI are more frequently located on genomic loci harbouring circRNAs than non-significant SNPs ( $P \geq 5 \times 10^{-8}$ ). We indeed identified a significant enrichment of genome-wide significant BMI-associated SNPs on circRNA genomic loci extracted from the databases circAtlas v2.0, circBase, CIRCpediaV2 and circVAR. This enrichment was consistent throughout the four circRNA datasets. Concurrently, our sensitivity analysis revealed analogous findings. Despite the exclusion of SNPs with  $P$ -values within a certain range and a resultant stricter delineation between the definitions of significant and non-significant SNPs, we still detected the previously shown enrichment of significant SNPs. Equally, this clustering was still evident when considering the LD structures of SNPs, regardless of whether a 1 Mb region was defined as an LD block or whether already known regions of high LD were analysed.

A previous study observed that SNPs with the highest impact on circRNAs, so-called circRNA quantitative trait loci (circQTLs), were mainly located in the flanking sites of the respective circRNAs, while merely a limited proportion reside inside the circRNAs. By analysing various sequence-defined elements, an intron-based enrichment was detected<sup>24</sup>. As GWAS signals predominantly map to non-coding regions<sup>6,7,10,11</sup> and as we exclusively analysed the genomic loci of circRNAs based on the start and stop positions deposited in the respective databases, and thereby neglecting splice isoforms of circRNAs, our results might be driven by this intron-specific enrichment of the significant SNPs. Consequently, we may have overestimated the number of SNPs that are



**Figure 6.** Allelic ratios of *hsa\_circ\_0022025* in eight heterozygous individuals for SNP rs4752856. Primer extension analysis of the allelic levels was performed with the ABI Prism SNaPshot Multiplex kit (Applied Biosystems, Foster City, CA, USA) as previously described<sup>45</sup>. The assay was performed in triplicates for each subject indicated by the single data points. For ‘all subjects’, each data point refers to the mean value of one subject. The allelic circRNA ratios were normalised with the allelic gDNA ratios. Besides the mean value (middle bar), the upper and lower limits of the 95% confidence interval are shown. The descriptive statistics of the normalised results were analysed with SPSS (version 28.0.0.0) and are stated in Supplementary Table S11. *circRNA* circular RNA.

Subject	Triplicate I			Triplicate II			Triplicate III		
	A/G circRNA	A/G gDNA	Normalised circRNA/gDNA	A/G circRNA	A/G gDNA	Normalised circRNA/gDNA	A/G circRNA	A/G gDNA	Normalised circRNA/gDNA
Subject 1	0.61	0.59	1.02	0.65	0.55	1.18	0.76	0.58	1.30
Subject 2	0.62	0.59	1.05	0.66	0.56	1.17	0.80	0.56	1.42
Subject 3	0.70	0.59	1.19	0.78	0.55	1.41	0.80	0.56	1.43
Subject 6	0.65	0.60	1.09	0.84	0.57	1.48	0.75	0.56	1.33
Subject 7	0.74	0.57	1.29	0.71	0.57	1.24	0.79	0.59	1.33
Subject 8	0.62	0.57	1.08	0.70	0.54	1.29	0.75	0.58	1.28
Subject 9	0.69	0.62	1.13	0.80	0.60	1.35	0.72	0.60	1.19
Subject 10	1.08	0.59	1.84	1.37	0.56	2.44	1.25	0.57	2.19

**Table 2.** Allelic ratios of the SNP rs4752856 on the circRNA *hsa\_circ\_0022025* and the genomic fragment of *MTCH2*. The A/G ratios of the primer extension products were determined for both the circRNA-derived cDNA as well as genomic DNA (for details see “Materials and methods” section). The circRNA ratios were normalized with regard to the genomic DNA. A risk allele, G major allele, *circRNA* circular RNA, *gDNA* genomic DNA.

potentially located on the matured circRNAs. However, this equally applies to significant and non-significant variants. Yet, SNPs in the flanking introns exerting their effects on circRNAs as proposed by Liu et al.<sup>24</sup> might at least be partially covered by our genomic level-based analysis. Still, since we cannot utilize the circRNA database data to infer which splice forms are generally formed, our approach can potentially account for the majority of isoforms putatively generated. Conversely to the intronic circQTL enrichment, an additional report showed that circQTLs were less likely to be intronic variants and instead tended to be localised within 5' untranslated regions and exonic regions<sup>46</sup>, concurring with the fact that the majority of circRNAs are comprised of exons<sup>24</sup>. Nevertheless, both studies have demonstrated that circQTLs with a high influence on circRNAs are more frequently located on disease-associated loci than SNPs that do not influence the circRNA<sup>24,46</sup>. Although we assigned the SNPs significance level based on the P-value and we thus might have encompassed non-causal variants in the significant group, it is certainly feasible that the majority of our significant SNPs localized to circRNAs may have a detectable influence on these circular transcripts. Generally, genetic variants located within genomic loci of circRNAs could putatively alter sequences of binding sites for miRNAs and RNA-binding proteins (RBPs), albeit Thomas and Sæstrom<sup>47</sup> ascertained that the SNP density at these binding sites is generally diminished in circRNAs. Yet, as circRNAs exhibit a cell type specific expression<sup>17</sup>, it is still plausible that while there is generally a reduced enrichment of SNPs in these regions, this may vary depending on the cell type from which the

circRNA was extracted. Still, an altered binding affinity of circRNAs towards their targets can have profound consequences, which remain to be validated in subsequent *in vitro* and *in vivo* studies.

As we observed a significant accumulation of the genome-wide significant SNPs for BMI on the genomic loci of circRNAs, we subsequently aimed to explore whether this enrichment was evident for SNPs associated with additional diseases and traits. As a previous report has demonstrated that circRNAs are highly enriched in the brain and synapses<sup>16</sup>, we postulated that the detected enrichment might be associated with CNS-mediated diseases and traits. Replication of our analyses with GWAS data pertaining to anthropometric, neurological, peripheral, and psychiatric phenotypes revealed a significant clustering of genome-wide significant SNPs for the psychiatric disorders (AN and autism spectrum disorder), the anthropometric measure of the body height and the peripheral CKD. The inverse enrichment of non-significant variants was detected for SNPs extracted from GWAS pertaining ulcerative colitis, pernicious anemia and the two neurological diseases (ALS and epilepsy). No significant deviation in circRNA localisation of significant and non-significant variants was ascertained for variants analysed in a GWAS for heart failure. Consequently, we could not confirm our hypothesis. Apparently, this enrichment is not a general feature of circRNAs. Clearly, the results may be biased due to the characteristics and implementations of the GWAS. The GWAS for BMI and body height, for instance, were conducted by the same authors<sup>9</sup> and given the definition of the BMI (kg/m<sup>2</sup>), we cannot exclude the possibility that the results obtained for the BMI GWAS were driven by the results of the dataset pertaining the body height. Generally, a high polygenicity for BMI and body height has been reported<sup>48</sup>. Further, genetic correlations between the phenotypes may have impacted our results. For example, past studies of our research group identified nine SNP alleles at three independent genetic loci that were associated with both, AN, and BMI. We were able to show that all nine AN susceptibility alleles concomitantly accounted for a lower BMI<sup>49</sup>. This correlation was confirmed by additional studies<sup>50–52</sup>. Yet, we did not assess any overlapping significant SNPs for BMI and AN located on circRNAs and consequently, could exclude SNP-based bias. Still, correlation and reciprocal interference between our results regarding BMI and AN cannot be fully excluded as our sex-specific analysis based on data from the BMI GWAS of Pulit et al.<sup>37</sup> implies that genome-wide significant BMI-SNPs in females are more abundant on circRNA genomic loci than significant BMI-SNPs in males. Since it is known that AN patients are predominantly female<sup>53</sup> and SNPs associated with a higher risk of AN predispose to a lower BMI<sup>49–52</sup>, we can assume that the BMI-SNPs on circRNAs mainly predispose to a lower BMI and that the effect on the etiology of obesity is marginal. In addition, most phenotypes analysed have already been associated with circRNAs<sup>54–59</sup>. For example, circRNA expression profiling of post-mortem brains of autistic patients and controls identified 60 circRNAs and more than 8000 autism-associated circRNA-miRNA-mRNA interactions. It was shown that targets of these axes were mainly risk genes for autism. Some of these high-risk genes were even modulated by upregulated circRNAs acting as a miR-204-3p sponge in human neuronal cells<sup>54</sup>. Merely studies on AN and pernicious anemia and their circRNA implication are lacking. Yet, it is unknown how disease-associated SNPs affect these mechanisms.

Previously, most circRNAs were functionally characterised as miRNA sponges<sup>14,38</sup>. Consequently, we have explored putative miRNA-circRNAs interactions based on a dataset extracted from the ENCORI database<sup>39</sup>. More than 70% of all circBase-extracted circRNAs were found to be linked to a least one miRNA. Yet, the circRNAs with the highest raw counts of miRNA target sites tended to be large in size. We assume that these circRNAs do not in fact span the entire range between predicted start and stop position stated in the database, but are indeed smaller. Therefore, we performed a normalisation based on the circRNA length. This yielded low ratios of miRNA target sites in contrast to the circRNAs length.

Almost all miRNAs included in the ENCORI dataset were determined to be correlated to SNP-harboured circRNAs. Within the five miRNAs linked to the highest number of circRNAs, two were associated with an obesity-related trait. Interestingly, hsa-miR-15a-5p has been reported to be involved in the process of adipocyte differentiation<sup>40</sup> and was differentially expressed in response to a low-fat diet<sup>41</sup>. Similarly, hsa-miR-16-5p was found to be downregulated after bariatric surgery<sup>42</sup> and after aerobic exercise training<sup>43</sup>. It is thus feasible, that certain circRNAs linked to BMI-associated SNPs might act on miRNAs as well.

Notably, due to predictions of circRNAs being based on computational algorithms and a lack of experimental validation, the possibility arises that the circRNAs from the databases represent false positives<sup>60</sup>. In general, given the high number of SNPs included in this study, the statistical power increased substantially and thus might exaggerate the potentially clinically negligible effects<sup>61</sup>.

To analyse a potential functional effect of SNPs on circRNAs obtained by our *in silico* analyses, we selected a BMI-increasing SNP (rs4752856) located on a *MTCH2*-derived circRNA (hsa\_circ\_0022025) for follow-up *in vitro* analyses. Using a highly sensitive primer extension assay, we determined the relative levels of the circRNAs derived from the G- and A-allele of heterozygous probands. We detected a significant skewing in favour of the risk allele (A). Normally, when investigating *cis*-regulatory effects, a large sample pool is required. But as we exclusively analysed heterozygous individuals, the respective other allele acted as the internal control<sup>45</sup> warranting our small number of eight probands.

Previous studies have already demonstrated that SNPs have a considerable impact on the circRNAs' levels<sup>22,23</sup>. Zhou et al. showed that SNP rs12196996, which was significantly correlated with an increased risk to develop coronary artery disease (CAD), was associated with decreased circFOXO3 levels in individuals with the GG genotype. It has been suggested that the increased CAD risk is caused by the effect of the SNP on circFOXO3<sup>23</sup>. Further, it is known that the levels of circRNAs in adipose tissue are different between obese and lean individuals<sup>18,19</sup>. One circRNA, circSAMD4A, exhibited a positive correlation with BMI in obese patients which was assumed to exert its effects through interactions with miR-138-5p, ultimately regulating the expression of *EZH2*<sup>19</sup>. Further, a deep sequencing analysis discovered thousands of circRNAs within the adipose tissue, which were dynamically regulated during adipogenesis and in obesity. A downregulation of circRNAs in contrast to the linear mRNAs was detected in obese mice<sup>18</sup>.



*MTCH2* is known to be an obesity susceptibility gene<sup>62,63</sup>. Previously, it has been positively correlated with BMI variance and was reported to be upregulated during adipogenesis<sup>62,64</sup>. Elevated *MTCH2* expression levels have been determined in obese women<sup>62</sup>. Given that circRNAs can regulate mRNA levels due to their role as miRNA sponges<sup>14,38</sup> and given that we have detected increased levels of the hsa\_circ\_0022025 derived from the BMI-increasing allele (A) of rs4752856, it is feasible that this circRNA may act as a miRNA sponge indirectly affecting the *MTCH2* mRNA levels. Additionally, *MTCH2* protein levels are increased in obese individuals<sup>62</sup>, which might potentially reflect the increased mRNA levels. Still, it has been reported that circRNAs can undergo cap-independent translation<sup>14</sup>. Thus, our detected increase of hsa\_circ\_0022025 derived from the risk allele (A) might hint at such a cap-independent translation, consequently raising the protein levels of *MTCH2*. Further, animal models have shown that a knockout of *Mtch2* leads to a lower susceptibility for weight gain<sup>63</sup>. These transgenic mice exhibited elevated levels of energy expenditure<sup>63</sup>, while a loss of *Mtch2* was shown to be protective for diet-induced obesity<sup>65</sup>. Hence, rs4752856's BMI-increasing effect<sup>9</sup> might be caused by the here detected increased levels of hsa\_circ\_0022025.

Accordingly, it is indeed conceivable that BMI-affecting SNPs, such as the here examined rs4752856, might exert an impact on circRNAs and ultimately affect the BMI of the risk allele carriers. Yet, as circRNAs are cell-type specifically expressed<sup>17</sup>, a replication of the analysis using circRNAs from different sources could yield in different outcomes and since we solely examined one SNP on one circRNA, the obtained results merely serve as an indication for putative SNP-mediated mechanisms and should be confirmed in additional analyses of multiple circRNA-located and BMI-affecting variants.

Taken together, we have determined a significant enrichment of genome-wide significant SNPs for BMI in circRNAs in comparison to non-significant variants. A sensitivity analysis as well as approximations of LD structures revealed similar outcomes. Yet, we did not ascertain evidence that this enrichment is consistent throughout various phenotypes. Our analysis of the BMI-increasing SNP rs4752856 located on the *MTCH2*-derived circRNA hsa\_circ\_0022025 revealed a higher abundance of the risk (A) allele on the circular transcript compared to the major (G) allele. Thus, our analyses extend the current knowledge of interactions of SNPs with circRNAs providing evidence for effect of BMI-affecting SNPs on circRNAs and thus implications in obesity. Further analyses need to determine the extent of the here detected findings.

## Materials and methods

**Datasets of circRNAs and GWAS SNPs.** The circRNA datasets were downloaded from four publicly available circRNA databases, namely circAtlas v2.0<sup>66</sup>, circBase<sup>67</sup>, CIRCpediaV2<sup>68</sup> and circVAR<sup>69</sup> (see Supplementary Table S1). All data was extracted from the download section of the respective database and contained exclusively human data. Data with ambiguous or incomplete annotations and circRNAs derived from genomic loci of the sex chromosomes were excised. The data was further checked for internal database duplicates.

The SNP summary statistics of the meta-analysis of BMI GWAS conducted by Yengo et al.<sup>9</sup> was downloaded from the Genetic Investigation of Anthropometric Traits (GIANT) consortiums' website (see Supplementary Table S2). Given the inconsistent assignment of reference genomes amongst the circRNA databases (see Supplementary Table S1), the SNP coordinates were shifted from GRCh37 to GRCh38 using the remapping service of the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/genome/tools/remap>). As the lift over of certain SNPs was inconclusive, the respective variants were excluded. In total, 2,324,569 SNPs remained in the analysis (see Supplementary Tables S2 and Table 1).

In order to replicate the analysis for further phenotypes, additional GWAS data for ALS<sup>28</sup>, AN<sup>34</sup>, autism spectrum disorder<sup>35</sup>, body height<sup>9</sup>, CKD<sup>30</sup>, epilepsy<sup>29</sup>, heart failure<sup>31</sup>, pernicious anemia<sup>32</sup> and ulcerative colitis<sup>33</sup> were downloaded (see Supplementary Table S2). Putative gender deviations were analysed with data of an additional BMI GWAS<sup>37</sup>. All GWAS data was acquired in the GRCh37 genome version and was exclusively analysed pertaining the autosomal chromosomes.

**SNP classification.** To assess whether genome-wide significant SNPs for BMI show a higher abundance on the genomic loci of circRNAs than non-significant variants, SNPs were assigned as significant ( $P < 5 \times 10^{-8}$ ) or non-significant ( $P \geq 5 \times 10^{-8}$ ) based on their P-value (see Table 1).

Next, a sensitivity analysis was implemented intended to provide a more stringent delineation between the groups of SNPs from the BMI GWAS. Three additional P-value thresholds were introduced ( $5 \times 10^{-7}$ ,  $5 \times 10^{-6}$ ,  $5 \times 10^{-5}$ ). The SNPs with P-values between  $5 \times 10^{-8}$  and the corresponding novel cut-off were excluded (see Table 1). Accordingly, while the group of significant SNPs remained unchanged, we obtained three diminished sets of non-significant variants ( $P \geq 5 \times 10^{-7}$ ,  $P \geq 5 \times 10^{-6}$  or  $P \geq 5 \times 10^{-5}$ ).

Subsequently, approximations of the LD structures of the BMI-SNPs were defined. Hence, all non-significant SNPs within a 1 Mb region adjacent to significant SNPs were assigned as significant as well (see Table 1). Since this definition of a 1 Mb region as a LD block was expected to yield a large number of false positives, the analysis of known regions of high LD as represented in plinkQC<sup>70</sup> (<https://github.com/meyer-lab-cshl/plinkQC/blob/master/inst/extdata/high-LD-regions-hg38-GRCh38.txt>) was ensued. If a region of high LD contained one of the significant SNPs (based on the P-value), all non-significant SNPs within this region were also classified as significant (see Table 1).

The group assignment of SNPs derived from GWAS of additional phenotypes is shown in Supplementary Table S3.

**Enrichment analysis.** A custom R script (R version: 4.0.5; RStudio version 1.4.1106) was applied to match the genomic coordinates of the circRNAs with those of the GWAS SNPs. The genomic coordinates of the circRNAs were constructed based on the predicted start and stop positions provided in the respective circRNA

database. The total number of unique SNPs located on the genomic loci of circRNAs was extracted from the generated output. Subsequently, a two-sided Chi-square test and the odds ratio (OR) were calculated using GraphPad Prism (version: 9.2.0). The confidence intervals (CI) of the OR were computed using the Woolf logit intervals. The confidence level was set to 95%.

**Gene mapping and gene ontology analyses.** To perform a functional assignment of SNPs, the rsIDs of these were initially allocated to the corresponding Ensembl gene IDs using the Ensembl Variant Effect Predictor (<https://ensembl.org/Tools/VEP>). Next, the gene IDs were analysed by the means of the Gene Ontology (<http://www.geneontology.org>) to yield functional classifications. Therefore, the PANTHER overrepresentation test was subjected applying the GO Ontology database. All 20,589 human genes listed in the database were used as a reference. As a statistical test, the Fisher's exact test with a False Discovery Rate (FDR) correction was applied. Significance was given with FDR  $P < 0.05$ .

**In silico analyses of interactions between miRNAs and circRNAs.** To analyse whether the circRNAs harbouring significant SNPs for BMI might interact with miRNAs, a dataset containing information regarding miRNA-circRNA interactions supported by Ago CLIP-sequencing data was downloaded from the ENCORI database (<https://starbase.sysu.edu.cn>)<sup>39</sup>. This contained 475,341 target sites of 643 miRNAs known to be linked to circRNAs. The data was based on the GRCh37 genome version. Using a marginally modified version of the custom R script implemented before, it was checked whether the circBase ID of circRNAs containing BMI-SNPs was included in the ENCORI dataset.

**Experimental validation of SNP effects.** To determine the relevance of BMI-affecting SNPs on circRNAs, a genome-wide significant SNP for BMI ( $P < 5 \times 10^{-8}$ )<sup>9,44</sup>, whose circRNA localisation was predicted by the preceding in silico analysis, was selected for subsequent in vitro studies. Following conditions needed to be met: (1) the SNP is located on a circRNA that was verified to be expressed in multiple cell lines and tissues; and (2) the SNP exhibits a MAF of at least 30% to ensure the detection of heterozygous carriers required for the allelic expression assay (as described in<sup>45</sup>); and (3) the SNP is located in close proximity ( $< 2$  kb) to the BSJ of the circRNA in order to amplify both BSJ and SNP on one fragment to exclude potential unspecific amplicons; and (4) is either a lead SNP or in high LD to a lead SNP.

**Study group.** We collected blood of 21 healthy individuals (age:  $24.95 \pm 3.46$  years old; BMI:  $20.77 \pm 1.77$  kg/m<sup>2</sup>; 52.38% female). Written informed consent was given by each participant. This study was approved by the Ethics Committee of the Medical Faculty of the University Duisburg-Essen (15-6534-BO) and was performed in accordance with the Declaration of Helsinki.

**Genotyping.** The DNA of each participant was isolated from whole-blood. Subjects were genotyped performing a PCR (Veriti 96-well thermal cycler; Applied Biosystems, Foster City, CA, USA) with SNP-specific gDNA primers (see Supplementary Table S4). Samples were purified with the QIAquick PCR Purification kit (Qiagen, Hilden, Germany) and sent for Sanger sequencing to Microsynth Seqlab GmbH in Göttingen, Germany. Sequence analysis and genotype assignment were performed by at least two experienced scientists using the SeqMan Pro software (version: v.10.1.0). Discrepancies were solved by either reaching consensus or by re-sequencing. The fulfilment of the Hardy-Weinberg-Equilibrium was checked.

**Experimental validation of the circRNA.** Peripheral blood mononuclear cells (PBMCs) were isolated from the participants' whole-blood applying a density gradient centrifugation utilizing Lymphoprep (Stemcell Technologies, Vancouver, BC, Canada) pre-filled Leucosep tubes (Greiner Bio-One GmbH, Frickenhausen, Germany) according to the manufacturer's instructions. Total RNA was extracted from the isolated PBMCs using TRIzol (Invitrogen AG, Carlsbad, CA, USA). The manufacturer's protocol was modified to increase the cell number and to minimize the dilution factor. Next, the total RNA samples were treated with at least 1 U RNase R (Epicentre Biotechnologies, Madison, WI, USA) per 1  $\mu$ g RNA for 15 min at 37 °C. Subsequently, the samples were subjected to at least 2 U DNase I (New England Biolabs GmbH, Ipswich, MA, USA) per 10  $\mu$ g RNA for 10 min at 37 °C followed by an EDTA (5 mM) heat inactivation step for 10 min at 75 °C. The remaining RNA was reversely transcribed to cDNA using the qScript cDNA SuperMix (Quanta Biosciences Inc., Beverly, MA, USA) as described in the manufacturer's instructions. Residual linear contaminations were excluded for each sample by performing a PCR with primers spanning two exons including their in-between intron of the sex hormone binding globulin (*SHBG*; see Supplementary Table S4). Furthermore, the successful circRNA isolation was validated conducting a PCR with divergent primers amplifying the yippee-like 2 (*YPEL2*)-derived circRNA, hsa\_circ\_0005600. Divergent primers spanning the BSJ and the SNP (see Supplementary Table S4) were applied in a PCR (Veriti 96-well thermal cycler; Applied Biosystems, Foster City, CA, USA) on the circRNA-derived cDNA template to confirm the presence of both. A 2.5%-agarose gel electrophoresis validated the expected fragment size and was followed by purification of the PCR products (QIAquick PCR purification kit; Qiagen, Hilden, Germany) and Sanger sequencing (performed by Microsynth Seqlab GmbH, Göttingen, Germany). Sequence analysis and circRNA confirmation were conducted by at least two experienced scientists with the SeqMan Pro software (version: v.11.0.0). Discrepancies were solved by either reaching consensus or by re-sequencing. The desired fragments were successfully ascertained for all examined samples.

**Primer extension analysis.** To determine the relative levels of hsa\_circ\_0022025, the amplification of the genomic DNA and circRNA-derived cDNA fragments for ten heterozygous individuals (G/A), two homozygous G/G probands, and one homozygous A/A participant was repeated in triplicates. The fragments were confirmed by a 2.5%-agarose gel and were subsequently purified (QIAquick PCR purification kit; Qiagen, Hilden, Germany). Next, the samples were subjected to a primer extension assay using the ABI Prism SNaPshot Multiplex kit (Applied Biosystems, Foster City, CA, USA) and a specific SNaPshot primer (see Supplementary Table S4) following the manufacturer's instructions. Equal amounts of cDNA and gDNA of each investigated participant were applied. Samples were heated for 3 min to 96 °C, followed by 25 cycles of 96 °C for 10 s, 53 °C for 5 s and 60 °C for 30 s. The SNaPshot reaction products were purified with shrimp alkaline phosphatase (GE Healthcare, Waukesha, WI, USA) and were analysed by gel capillary electrophoresis on the ABI Prism 3700 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). The resultant electropherograms were examined using the GeneMapper 4.0 software by Applied Biosystem (Foster City, CA, USA). Next, allelic ratios of the circRNA-derived cDNA were normalised with the gDNA allelic ratios as in Eq. (1). The descriptive statistics with confidence intervals were based on the t-test and a one-sample Wilcoxon test and was conducted with SPSS (version: 28.0.0.0). The confidence level was set to 95%.

$$\frac{A\text{-allele circRNA}/G\text{-allele circRNA}}{A\text{-allele gDNA}/G\text{-allele gDNA}} \quad (1)$$

### Data availability

The datasets generated and analysed during the current study are available in the Zenodo repository, <https://doi.org/10.5281/zenodo.6726258>. The sources of the used GWAS and circRNA datasets are stated in Supplementary Tables S1 and S2. Biological material can be obtained for research purposes. The R script can be obtained by the authors upon reasonable request.

Received: 17 February 2022; Accepted: 11 July 2022

Published online: 25 July 2022

### References

- Janssen, F., Bardoutsos, A. & Vidra, N. Obesity prevalence in the long-term future in 18 European countries and in the USA. *Obes. Facts* **13**, 514–527. <https://doi.org/10.1159/000511023> (2020).
- Schienkiewitz, A., Mensink, G., Kuhnert, R. & Lange, C. Übergewicht und Adipositas bei Erwachsenen in Deutschland. *J. Health Monit.* **2**(2). <https://doi.org/10.17886/RKI-GBE-2017-025> (2017).
- Williams, E. P., Mesidor, M., Winters, K., Dubbert, P. M. & Wyatt, S. B. Overweight and obesity: Prevalence, consequences, and causes of a growing public health problem. *Curr. Obes. Rep.* **4**, 363–370. <https://doi.org/10.1007/s13679-015-0169-4> (2015).
- Blüher, M. Obesity: Global epidemiology and pathogenesis. *Nat. Rev. Endocrinol.* **15**, 288–298. <https://doi.org/10.1038/s41574-019-0176-8> (2019).
- World Health Organisation. *Obesity: Preventing and Managing the Global Epidemic* (WHO, 2000).
- Giral, H., Landmesser, U. & Kratzer, A. Into the wild: GWAS exploration of non-coding RNAs. *Front. Cardiovasc. Med.* **5**, 181. <https://doi.org/10.3389/fcvm.2018.00181> (2018).
- Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**, 467–484. <https://doi.org/10.1038/s41576-019-0127-1> (2019).
- Hinney, A. *et al.* Genome wide association (GWA) study for early onset extreme obesity supports the role of fat mass and obesity associated gene (FTO) variants. *PLoS One* **2**, e1361. <https://doi.org/10.1371/journal.pone.0001361> (2007).
- Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649. <https://doi.org/10.1093/hmg/ddy271> (2018).
- Bhartiya, D. & Scaria, V. Genomic variations in non-coding RNAs: Structure, function and regulation. *Genomics* **107**, 59–68. <https://doi.org/10.1016/j.ygeno.2016.01.005> (2016).
- Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102–110. <https://doi.org/10.1093/hmg/ddv259> (2015).
- Barrett, S. P. & Salzman, J. Circular RNAs: Analysis, expression and potential functions. *Development* **143**, 1838–1847. <https://doi.org/10.1242/dev.128074> (2016).
- Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. & Brown, P. O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* **7**, e30733. <https://doi.org/10.1371/journal.pone.0030733> (2012).
- Kristensen, L. S. *et al.* The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.* **20**, 675–691. <https://doi.org/10.1038/s41576-019-0158-7> (2019).
- Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338. <https://doi.org/10.1038/nature11928> (2013).
- Rybak-Wolf, A. *et al.* Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol. Cell* **58**, 870–885. <https://doi.org/10.1016/j.molcel.2015.03.027> (2015).
- Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L. & Brown, P. O. Cell-type specific features of circular RNA expression. *PLoS Genet.* **9**, e1003777. <https://doi.org/10.1371/journal.pgen.1003777> (2013).
- Arcinas, C. *et al.* Adipose circular RNAs exhibit dynamic regulation in obesity and functional role in adipogenesis. *Nat. Metab.* **1**, 688–703. <https://doi.org/10.1038/s42255-019-0078-z> (2019).
- Liu, Y. *et al.* Circular RNA SAMD4A controls adipogenesis in obesity through the miR-138-5p/EZH2 axis. *Theranostics* **10**, 4705–4719. <https://doi.org/10.7150/thno.42417> (2020).
- Verduci, L., Tarcitano, E., Strano, S., Yarden, Y. & Blandino, G. CircRNAs: Role in human diseases and potential use as biomarkers. *Cell Death Dis.* **12**, 468. <https://doi.org/10.1038/s41419-021-03743-3> (2021).
- Sun, W. *et al.* CircRNA expression profiles in human visceral preadipocytes and adipocytes. *Mol. Med. Rep.* **21**, 815–821. <https://doi.org/10.3892/mmr.2019.10886> (2020).
- Paraboschi, E. M., Cardamone, G., Solda, G., Duga, S. & Asselta, R. Interpreting non-coding genetic variation in multiple sclerosis genome-wide associated regions. *Front. Genet.* **9**, 647. <https://doi.org/10.3389/fgene.2018.00647> (2018).
- Zhou, Y. L. *et al.* CircFOXO3 rs12196996, a polymorphism at the gene flanking intron, is associated with circFOXO3 levels and the risk of coronary artery disease. *Aging (Albany, NY)* **12**, 13076–13089. <https://doi.org/10.18632/aging.103398> (2020).

24. Liu, Z. *et al.* Detection of circular RNA expression and related quantitative trait loci in the human dorsolateral prefrontal cortex. *Genome Biol.* **20**, 99. <https://doi.org/10.1186/s13059-019-1701-8> (2019).
25. Friedman, J. M. Leptin and the endocrine control of energy balance. *Nat. Metab.* **1**, 754–764. <https://doi.org/10.1038/s42255-019-0095-y> (2019).
26. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504. <https://doi.org/10.1038/s41576-018-0016-z> (2018).
27. Uffelmann, E. *et al.* Genome-wide association studies. *Nat. Rev. Methods Primers* **1**, 59. <https://doi.org/10.1038/s43586-021-00056-9> (2021).
28. Iacoangeli, A. *et al.* Genome-wide meta-analysis finds the ACSL5-ZDHHC6 locus is associated with ALS and links weight loss to the disease genetics. *Cell Rep.* **33**, 108323. <https://doi.org/10.1016/j.celrep.2020.108323> (2020).
29. International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. *Nat. Commun.* **9**, 5269. <https://doi.org/10.1038/s41467-018-07524-z> (2018).
30. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat. Genet.* **51**, 957–972. <https://doi.org/10.1038/s41588-019-0407-x> (2019).
31. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163. <https://doi.org/10.1038/s41467-019-13690-5> (2020).
32. Glanville, K. P., Coleman, J. R. I., O'Reilly, P. F., Galloway, J. & Lewis, C. M. Investigating pleiotropy between depression and autoimmune diseases using the UK biobank. *Biol. Psychiatry Glob. Open Sci.* **1**, 48–58. <https://doi.org/10.1016/j.bpsgos.2021.03.002> (2021).
33. de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261. <https://doi.org/10.1038/ng.3760> (2017).
34. Watson, H. J. *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214. <https://doi.org/10.1038/s41588-019-0439-2> (2019).
35. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444. <https://doi.org/10.1038/s41588-019-0344-8> (2019).
36. Hirtz, R. & Hinney, A. Genetic and epigenetic findings in anorexia nervosa. *Med. Gen.* **32**, 25–29 (2020).
37. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174. <https://doi.org/10.1093/hmg/ddy327> (2019).
38. Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388. <https://doi.org/10.1038/nature11993> (2013).
39. Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **42**, D92–D97. <https://doi.org/10.1093/nar/gkt1248> (2014).
40. Shi, C. *et al.* Adipogenic miRNA and meta-signature miRNAs involved in human adipocyte differentiation and obesity. *Oncotarget* **7**, 40830–40845. <https://doi.org/10.18632/oncotarget.8518> (2016).
41. Assmann, T. S., Riezu-Boj, J. I., Milagro, F. I. & Martinez, J. A. Circulating adiposity-related microRNAs as predictors of the response to a low-fat diet in subjects with obesity. *J. Cell. Mol. Med.* **24**, 2956–2967. <https://doi.org/10.1111/jcmm.14920> (2020).
42. Langi, G., Szczerbinski, L. & Kretowski, A. Meta-analysis of differential miRNA expression after bariatric surgery. *J. Clin. Med.* **8**, 1220. <https://doi.org/10.3390/jcm8081220> (2019).
43. Zhang, T. *et al.* Circulating MiRNAs as biomarkers of gait speed responses to aerobic exercise training in obese older adults. *Aging (Albany, NY)* **9**, 900–913. <https://doi.org/10.18632/aging.101199> (2017).
44. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206. <https://doi.org/10.1038/nature14177> (2015).
45. Berulava, T. & Horsthemke, B. The obesity-associated SNPs in intron 1 of the FTO gene affect primary transcript levels. *Eur. J. Hum. Genet.* **18**, 1054–1056. <https://doi.org/10.1038/ejhg.2010.71> (2010).
46. Ahmed, I., Karedath, T., Al-Dasim, F. M. & Malek, J. A. Identification of human genetic variants controlling circular RNA expression. *RNA* **25**, 1765–1778. <https://doi.org/10.1261/rna.071654.119> (2019).
47. Thomas, L. F. & Sætrom, P. Circular RNAs are depleted of polymorphisms at microRNA binding sites. *Bioinform.* **30**, 2243–2246. <https://doi.org/10.1093/bioinformatics/btu257> (2014).
48. Hemani, G. *et al.* Inference of the genetic architecture underlying BMI and height with the use of 20,240 sibling pairs. *Am. J. Hum. Genet.* **93**, 865–875. <https://doi.org/10.1016/j.ajhg.2013.10.005> (2013).
49. Hinney, A. *et al.* Evidence for three genetic loci involved in both anorexia nervosa risk and variation of body mass index. *Mol. Psychiatry* **22**, 321–322. <https://doi.org/10.1038/mp.2016.126> (2017).
50. Hübel, C. *et al.* Genetic correlations of psychiatric traits with body composition and glycemic traits are sex- and age-dependent. *Nat. Commun.* **10**, 5765. <https://doi.org/10.1038/s41467-019-13544-0> (2019).
51. Baker, J. H., Schaumberg, K. & Munn-Chernoff, M. A. Genetics of anorexia nervosa. *Curr. Psychiatry Rep.* **19**, 84. <https://doi.org/10.1007/s11920-017-0842-2> (2017).
52. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241. <https://doi.org/10.1038/ng.3406> (2015).
53. Mitchell, J. E. & Peterson, C. B. Anorexia nervosa. *N. Engl. J. Med.* **382**, 1343–1351. <https://doi.org/10.1056/NEJMcp1803175> (2020).
54. Chen, Y. J. *et al.* Genome-wide, integrative analysis of circular RNA dysregulation and the corresponding circular RNA-microRNA-mRNA regulatory axes in autism. *Genome Res.* **30**, 375–391. <https://doi.org/10.1101/gr.255463.119> (2020).
55. Ravnik-Glavac, M. & Glavac, D. Circulating RNAs as potential biomarkers in amyotrophic lateral sclerosis. *Int. J. Mol. Sci.* **21**, 1714. <https://doi.org/10.3390/ijms21051714> (2020).
56. Liu, X. *et al.* Circular RNA circANAPC2 mediates the impairment of endochondral ossification by miR-874-3p/SMAD3 signalling pathway in idiopathic short stature. *J. Cell. Mol. Med.* **25**, 3408–3426. <https://doi.org/10.1111/jcmm.16419> (2021).
57. van Zonneveld, A. J., Kolling, M., Bijkerk, R. & Lorenzen, J. M. Circular RNAs in kidney disease and cancer. *Nat. Rev. Nephrol.* **17**, 814–826. <https://doi.org/10.1038/s41581-021-00465-9> (2021).
58. Shao, L. *et al.* Silencing of circIgf1r plays a protective role in neuronal injury via regulating astrocyte polarization during epilepsy. *FASEB J.* **35**, e21330. <https://doi.org/10.1096/fj.202001737RR> (2021).
59. Wang, K. *et al.* A circular RNA protects the heart from pathological hypertrophy and heart failure by targeting miR-223. *Eur. Heart J.* **37**, 2602–2611. <https://doi.org/10.1093/eurheartj/ehv713> (2016).
60. Vromman, M., Vandesompele, J. & Volders, P. J. Closing the circle: Current state and perspectives of circular RNA databases. *Brief. Bioinform.* **22**, 288–297. <https://doi.org/10.1093/bib/bbz175> (2021).
61. Faber, J. & Fonseca, L. M. How sample size influences research outcomes. *Dental Press J. Orthod.* **19**, 27–29. <https://doi.org/10.1590/2176-9451.19.4.027-029.ebo> (2014).
62. Kulyte, A. *et al.* MTCH2 in human white adipose tissue and obesity. *J. Clin. Endocrinol. Metab.* **96**, E1661–E1665. <https://doi.org/10.1210/jc.2010-3050> (2011).
63. Buzaglo-Azriel, L. *et al.* Loss of muscle MTCH2 increases whole-body energy utilization and protects from diet-induced obesity. *Cell Rep.* **14**, 1602–1610. <https://doi.org/10.1016/j.celrep.2016.01.046> (2016).

64. Bernhard, F. *et al.* Functional relevance of genes implicated by obesity genome-wide association study signals for human adipocyte biology. *Diabetologia* **56**, 311–322. <https://doi.org/10.1007/s00125-012-2773-0> (2013).
65. Bar-Lev, Y. *et al.* Mimp/Mtch2, an obesity susceptibility gene, induces alteration of fatty acid metabolism in transgenic mice. *PLoS One* **11**, e0157850. <https://doi.org/10.1371/journal.pone.0157850> (2016).
66. Wu, W., Ji, P. & Zhao, F. CircAtlas: An integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol.* **21**, 101. <https://doi.org/10.1186/s13059-020-02018-y> (2020).
67. Glažar, P., Papavasileiou, P. & Rajewsky, N. circBase: A database for circular RNAs. *RNA* **20**, 1666–1670. <https://doi.org/10.1261/rna.043687.113> (2014).
68. Dong, R., Ma, X. K., Li, G. W. & Yang, L. CIRCpedia v2: An updated database for comprehensive circular RNA annotation and expression comparison. *Genomics Proteomics Bioinform.* **16**, 226–233. <https://doi.org/10.1016/j.gpb.2018.08.001> (2018).
69. Zhao, M. & Qu, H. circVAR database: Genome-wide archive of genetic variants for human circular RNAs. *BMC Genomics* **21**, 750. <https://doi.org/10.1186/s12864-020-07172-y> (2020).
70. plinkQC: Genotype quality control in genetic association studies (2020).

## Acknowledgements

We thank all participants for their participation. We are further indebted to Sieglinde Dürkop for her excellent technical support. This study was funded by the Deutsche Forschungsgemeinschaft (DFG; HI 86512-1, DFG Research Unit FOR2488), the BMBF (01GS0820; PALGER 2017-33: 01DH19010) and the Stiftung Universitätsmedizin Essen. We further acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen.

## Author contributions

L.S.R., I.D., L.P., A.L.V., and A.H. designed the study. L.P. and A.H. recruited the probands. A.H., C.D., B.H. and J.H. were responsible with the experimental design. I.D., D.K., B.H. and F.J.K. performed the molecular genetic experiments. L.S.R. and Y.Z. assembled the datasets. L.S.R. performed the bioinformatic analysis. L.S.R. and T.P. performed the statistical analysis. L.S.R., I.D., A.H., and J.H. interpreted the data. L.S.R. wrote the draft of the manuscript and included input of all authors. All authors approved of the final version of the manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-16495-7>.

**Correspondence** and requests for materials should be addressed to L.S.R. or A.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

# DuEPublico

Duisburg-Essen Publications online

UNIVERSITÄT  
DUISBURG  
ESSEN

*Offen im Denken*

ub | universitäts  
bibliothek

This text is made available via DuEPublico, the institutional repository of the University of Duisburg-Essen. This version may eventually differ from another version distributed by a commercial publisher.

**DOI:** 10.1038/s41598-022-16495-7

**URN:** urn:nbn:de:hbz:465-20240110-063229-1



This work may be used under a Creative Commons Attribution 4.0 License (CC BY 4.0).