New insights from GWAS on longitudinal and cross-sectional BMI and related phenotypes in admixed children with Native American and European ancestries

Lucas Vicuña1,2, Esteban Barrientos1,2, Tomás Norambuena1,2, Danilo Alvares3, Juan Cristobal Gana4, Valeria Leiva3, Verónica Mericq5, Cristian Meza6, Ana Pereira7, José L. Santos8, Susana Eyheramendy1,2*

1 Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Santiago, Chile.
2 Instituto Milenio de Investigación Sobre los Fundamentos de los Datos (IMFD), Santiago, Chile.
3 Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Chile.
4 Department of Pediatric Gastroenterology and Nutrition. Division of Pediatrics. School of Medicine. Pontificia Universidad Católica de Chile, Santiago, Chile.
5 Institute of Maternal and Child Research, Faculty of Medicine, University of Chile, Santiago, Chile.
6 INGEMAT-CIMFAV, Faculty of Engineering, Universidad de Valparaíso, Valparaíso, Chile.
7 Institute of Nutrition and Food Technology, University of Chile, Santiago, Chile.
8 Department of Nutrition, Diabetes and Metabolism, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile.

* Corresponding author: susana.eyheramendy@uai.cl

Abstract

Body-mass index (BMI) is a well-known marker of adiposity across all ages. The genetic architecture of BMI has been thoroughly studied among adults. In contrast, there are a few genome-wide association studies (GWAS) on children. Further, GWAS on children have been performed almost exclusively in Europeans at single ages. We aimed to better understand the genetic architecture of BMI trajectory across ages and how BMI is affected by Native American genetic ancestry. We performed cross-sectional and longitudinal GWAS for BMI-related traits on 904 admixed Chilean children with mostly European and Mapuche Native American genetic ancestry. We focused on BMI and two traits that occur at the minimum of the childhood BMI growth trajectory, namely, age at adiposity rebound (Age-AR) and BMI at adiposity rebound (BMI-AR). We found several variants in the immune gene HLA-DQB3 that are strongly associated with BMI at ages 1.5-2.5 years old, but not at other ages. We also identified a variant in the sex-determining gene DMRT1 significantly associated with Age-AR (P = 9.8 × 10⁻⁹). Further, BMI was significantly higher in Mapuche than in European children at all ages between 5.5 and 16.5 years old, but not before. Finally, Age-AR was significantly lower (P = 0.013) by 1.64 years in the Mapuche children compared with Europeans.

Introduction

Childhood obesity is a major public health problem across the world. One well-studied indicator of obesity is the body mass index (BMI), whose alterations in children have been associated with risk of adult obesity, as well as related diseases, including type 2
diabetes and cardiometabolic diseases. Even though several environmental causes explain an important part of childhood obesity, there is still a lack of knowledge about the genetic factors underlying susceptibility to this disease.

BMI follows a nonlinear trajectory over time during childhood. BMI trajectory is characterized by three main phases: i) a rapid increase with an adiposity peak (AP) at ~9 month of age; ii) a decline reaching its lowest value at ~5.5 years, also called the adiposity rebound (AR); and iii) a subsequent peak at early adulthood.

The heritability of adult obesity has been estimated to be 40 – 70% [34]. Approximately 22.4% of adult BMI heritability is explained by 941 SNPs, according to a genome-wide association study (GWAS) meta-analysis performed in ~700,000 individuals of European ancestry. Such studies have been mostly performed in adults, but for most of these loci it is unknown whether or not and to which extent they also affect BMI in children. Some cross-sectional GWAS studies on European populations have shown that several loci influence BMI in adulthood and childhood, whereas other loci seem to act only during adulthood or childhood. For instance, a GWAS found variants in the FTO and MC4R genes significantly associated with BMI in children (<6 years old) as well as in adults, but a locus in MAF only associated with BMI in adults. A GWAS meta-analysis found significant associations with FTO, MC4R, TMEM18, SDCCAG8 and TNKS/MSRA loci, but associations for the later gene were limited to children and adolescents. Another GWAS meta-analysis identified 15 loci significantly associated with BMI (they considered a single BMI value at the oldest age between 2 and 10 years). Among them, 12 were previously associated with adult BMI or childhood obesity, suggesting that the 3 remaining loci, located near ELP3, RAB27B and ADAM23 genes, act only during childhood. A further GWAS meta-analysis identified 18 SNPs significantly associated with BMI in paediatric cohorts from diverse ancestries pooled together, although most of these loci only reached the genome-wide significance threshold in Europeans when same-ancestry cohorts were analyzed separately. Most of these loci had been previously associated with adult BMI, childhood BMI or childhood obesity.

A few GWAS on European children have identified loci affecting BMI in distinct phases of child growth. A recent GWAS meta-analysis identified 4 loci significantly associated with childhood BMI-related traits: one LEPR/LEPROT gene variant associated with BMI-AP; one FTO variant and one TFAP2B variant associated with Age-AR; and a GNPDA2 variant associated with BMI-AR. Among them, the TFAP2B, GNPDA2 and FTO variants are also associated with adult BMI, whereas the LEPR/LEPROT variant only associates with a BMI-related trait in the infant phase. The authors suggested that adult BMI-related variants start influencing BMI by the time of AR. A candidate gene study on a European multi-cohort found that a FTO locus is positively associated with BMI from 5.5 years onward, but is inversely associated below age 2.5 years old. They also confirmed this findings using a longitudinal linear mixed model. Warrington et al 2015 performed a longitudinal GWAS to identify associations between genotypes and BMI trajectories across childhood. They found 4 loci previously associated with BMI in adults or children as well as a novel locus in the FAM120AOS gene significantly associated with BMI at 8 years of age. To our knowledge, Warrington et al 2015 is the only longitudinal GWAS on childhood BMI.

As mentioned before, the vast majority of these GWA studies have been performed in Europeans, but we do not know to what extent these loci affect BMI in populations with other continental ancestries, in particular those with Native American ancestry. This is relevant, as obesity and obesity-related disorders are markedly affected by genetic ancestry. For instance, Latinos, which usually have mixed European, Native American and African ancestries, are more susceptible to lipid-related disorders than any other US group, in part due to their Native-American genetic heritage.
The aims of this study are twofold: i) to estimate how much does Native American ancestry affect childhood BMI trajectory, Age-AR as well as BMI-AR; and ii) to identify genetic variants associated with these traits.

Materials and methods

Sample collection and genotyping.

We analyzed individuals of the “Growth and Obesity Chilean Cohort Study” (GOCS), who were genotyped using the Infinium Multietnic Global BeadChip (Illumina), as described in [21]. After quality controls were applied, we obtained a filtered data set of 904 individuals and 774,433 autosomal SNPs. Further details can be found in [21].

Weight and height measurements.

Weight and height measurements were taken since 2006 at Instituto de Nutrición y Tecnología de los Alimentos (INTA), Santiago, Chile, on children/adolescents in barefoot and light clothes by a trained dietitian following standardized protocols; ICC for all measurements was $\geq 0.75$ [41]. Weight and height were measured twice at each individual visit, and the average was considered as the final value. Before and after puberty, weight and height were measured once per year; during puberty they were measured every 6 months. Weight was measured with a portable electronic scale (Seca 770), precision of 0.1 kg, and height was measured with a portable stadiometer (Harpenden 603) to the nearest 0.1 cm. BMI (kg/m$^2$) and Z-scores (for < 18 years old) were calculated based on the World Health Organization 2007 growth references. Weight and height data prior to 2006 were retrieved from medical records.

Local Ancestry Estimation.

We used RFMIX to infer the local ancestry of each SNP allele from our Chilean sample. We used reference populations from the 1000 Genomes Project, namely, Yoruba (YRI, n=108) for the African ancestry, Utah Residents with Northern and Western European Ancestry (CEU, n=36) for the European ancestry, and Peruvian (PEL) individuals with > 95% Native American ancestry (n=29). We excluded individuals with > 5% SNP missing rate. We inferred the gametic phase of individuals with the Beagle 5 software, using the HapMap37 human genome build 37 recombination map. To obtain the local ancestry, we used the --forward-backward parameter, as recommended in the manual. Local ancestry estimation identified three continental ancestries: European, African and Native American.

Global Ancestry Estimation.

Global ancestry proportions of Chilean children were estimated with Admixture 1.3, using K=4 ancestral populations. This K value was chosen because it was able to distinguish the two main Native American subcomponents of Chileans, namely, Mapuche and Aymara. In order to distinguish Peruvian (PEL) individuals with > 95% Native American ancestry, we used K=3 ancestral populations and the reference populations mentioned in the previous section. Global ancestry estimation identified four proportion of ancestries corresponding to European, African, Aymara and Mapuche. We used the Mapuche proportion ancestry in the GWAS to identify its effect on the phenotypes.
Variant and gene annotations.

Variant annotations [GRCh37 (hg19) assembly] were retrieved with the web tool Variant Effect Predictor (VEP) from Ensembl [25]. Upstream and downstream variants were defined as those located 10 Kb upstream or downstream of the gene, respectively. Intergenic variants were defined as those located > 100 Kb upstream or downstream of the closest gene. We used HaploReg v4.1 [26] to identify the genes located closest to associated intergenic variants. Reported GWAS associations were retrieved from the NHGRI GWAS Catalog [27]. When more than one variant in a gene has been associated with the same phenotype, we reported the strongest association.

Derivation of Age-AR and BMI-AR

Before modeling BMI trajectory over time, we applied some filters on the individuals. We excluded 51 individuals who had no sufficient measurements between 2 and 10 years old. To estimate Age-AR, we implemented a longitudinal statistical model of Age at which the minimum BMI is found.

Intergenic variants were defined as those located 10 Kb upstream or downstream of the gene, respectively. The predicted trajectory for individual \(i\) can then be written as

\[
\log(BMI)_{ij} = (\hat{\beta}_1 S_i + \hat{\beta}_0) + (\hat{\beta}_2 S_i + \hat{\beta}_3) Age_{ij} + (\hat{\beta}_4 S_i + \hat{\beta}_5) Age_{ij}^2 + (\hat{\beta}_6 S_i + \hat{\beta}_7) Age_{ij}^3 + \ldots + (\hat{\beta}_9 i + \hat{\beta}_10) Age_{ij} + \hat{\beta}_11 Age_{ij}^2 + \hat{\beta}_12 Age_{ij}^3 + \epsilon_{ij}
\]

where \(n_i\) corresponds to the number of Age/BMI measurements for individual \(i\), \(S_i\) represents the gender of the individual (0 female, 1 male). The parameters \(\beta_k\) for \(k = 1, \ldots, 7\) are fixed effects whereas the parameters \(\beta_{ki}\) for \(k = 8, \ldots, 11\) are random effects for each individual \(i\) (i.e. \(\beta_{ki} \sim N(0, \sigma^2_{\beta})\)). The errors \(\epsilon_{ij}\) are assumed independent across the different individuals \(i\) but dependent between observations for the same individual (index \(j\)), and are normally distributed with variance \(\sigma^2\).

The predicted trajectory for individual \(i\) can be written as

\[
\log(BMI)_{ij} = (\hat{\beta}_1 S_i + \hat{\beta}_0) + (\hat{\beta}_2 S_i + \hat{\beta}_3) Age_{ij} + (\hat{\beta}_4 S_i + \hat{\beta}_5) Age_{ij}^2 + (\hat{\beta}_6 S_i + \hat{\beta}_7) Age_{ij}^3 + \ldots + (\hat{\beta}_9 i + \hat{\beta}_10) Age_{ij} + \hat{\beta}_11 Age_{ij}^2 + \hat{\beta}_12 Age_{ij}^3 + \epsilon_{ij}
\]

where \(\hat{\beta}_k\) (hat) represents the estimators of the parameters. From this trajectory, the Age of Adiposity rebound (Age-AR) and the BMI at Age-AR are obtained by finding the Age at which the minimum BMI is found.

A local minimum of BMI is found by solving the following equation:

\[
3(\hat{\beta}_4 S_i + \hat{\beta}_7 + \hat{\beta}_{11})x_i^2 + 2(\hat{\beta}_4 S_i + \hat{\beta}_5 + \hat{\beta}_{10})x_i + (\hat{\beta}_2 S_i + \hat{\beta}_3 + \hat{\beta}_9 i) = 0
\]

which has as its solution:

\[
\text{Age-AR}_i = \frac{1}{3(\hat{\beta}_6 S_i + \hat{\beta}_7 + \hat{\beta}_{11})} (- (\hat{\beta}_4 S_i + \hat{\beta}_5 + \hat{\beta}_{10} i )) \pm 2\sqrt{(\hat{\beta}_4 S_i + \hat{\beta}_5 + \hat{\beta}_{10} i)^2 - 3(\hat{\beta}_6 S_i + \hat{\beta}_7 + \hat{\beta}_{11} (\hat{\beta}_2 S_i + \hat{\beta}_3 + \hat{\beta}_9 i))}
\]
where the minimum obtained above needs to satisfy:

$$6(\hat{\beta}_6 S_i + \hat{\beta}_7 + \hat{\beta}_{11i}) \text{Age-AR}_i + 2(\hat{\beta}_4 S_i + \hat{\beta}_5 + \hat{\beta}_{10i}) > 0$$  \hfill (5)

To identify BMI-AR, we replace the Age term from equation (1) by the Age-AR value obtained from equation (3).

**Cross-sectional GWAS on BMI.**

For the cross-sectional GWAS, we analyzed individuals between 1.5 and 16.5 years old, across 16 strata of 1 year interval each (SI Table). BMI was expressed as kg/m$^2$ and age in years. Individuals with no weight or height measurements in a particular stratum were removed. When > 1 measurement was taken in the same individual and time interval, we took the mean of the BMI measurements. To get a normal distribution of BMI values, we applied Box-Cox transformations (values for each stratum are shown in SI Table). These transformed values were standardized to get mean = 0 and standard deviation = 1, obtaining zBMI scores [13] [14]. These zBMI scores were used to test for genome-wide associations. We applied the following linear regression model:

$$z\text{BMI}_i = \beta_0 + \beta_1 \text{GT}_{ij} + \beta_2 S_i + \beta_3 \text{GA}_i + \beta_4 \text{LA}_ij + \epsilon_i$$  \hfill (6)

where $\beta_1$ represents the effect of the additive genotype (GT) of SNP $j$; $\beta_2$ represents the effect of gender (S); $\beta_3$ represents the effect of the global Mapuche Native American ancestry proportion (GA); and $\beta_4$ represents the effect of local Native American ancestry, where $\text{LA}_ij$ takes the values 0, 1 or 2 depending on the number of Native American alleles. Of note, while our global ancestry analysis is able to separately estimate the Mapuche and Aymara ancestral Native American subcomponents of Chileans, the local ancestry analysis only considers a general Native American population.

**GWAS on longitudinal BMI data.**

To estimate longitudinal genotype-phenotype associations, we implemented a linear mixed model for BMI. We adjusted for gender, age, global ancestry, and local ancestry. We assumed an additive inheritance model. The model equation is as follows:

$$\log(\text{BMI})_{ij} = (\beta_1 S_i + \beta_0) + (\beta_2 S_i + \beta_3) \text{Age}_{ij} + \beta_4 \text{Age}^2_{ij} + \beta_5 \text{Age}^3_{ij} + \beta_6 \text{GA}_i + \beta_7 \text{LA}_i + \beta_8 \text{GT}_i + \beta_9 i + \beta_{10i} \text{Age}_{ij} + \beta_{11i} \text{Age}^2_{ij} + \epsilon_{ij}$$  \hfill (7)

where $m_i$ corresponds to the number of Age/BMI measurements for individual $i$, $S_i$ represents the gender of the individual (0 female, 1 male).

**GWAS on Age-AR and BMI-AR.**

Similarly as with BMI, we obtained standardized Age-AR and BMI-AR values by applying Box-Cox transformations followed by standardization. We obtained estimates of $\lambda = 0.841$ for Age-AR and $\lambda = -0.759$ for BMI-AR. We performed genome-wide associations for standardized Age-AR, using the following linear regression model:
\[ z_{\text{Age-AR}}_i = (\beta_1 S_i + \beta_0) + (\beta_2 S_i + \beta_3)z_{\text{BMI-AR}}_i + (\beta_4 S_i + \beta_5)z_{\text{BMI-AR}}^2_i + (\beta_6 S_i + \beta_7)z_{\text{BMI-AR}}^3_i + \beta_8 \text{GA}_i + (\beta_9 S_i + \beta_{10})GT_i + \beta_{11} \text{LA}_i + \epsilon_i \] (8)

\[ z_{\text{BMI-AR}}_i = (\beta_1 S_i + \beta_0) + (\beta_2 S_i + \beta_3)z_{\text{Age-AR}}_i + (\beta_4 S_i + \beta_5)z_{\text{Age-AR}}^2_i + (\beta_6 S_i + \beta_7)z_{\text{Age-AR}}^3_i + \beta_8 \text{GA}_i + (\beta_9 S_i + \beta_{10})GT_i + \beta_{11} \text{LA}_i + \epsilon_i \] (9)

Results

Estimation of Age-AR and BMI-AR

We first characterized BMI trajectories in 904 children from the “Growth and Obesity Chilean Cohort Study” (GOCS) from age 2 years to age 10 years (Fig 1). We implemented longitudinal mixed models (see details in Methods) in boys and girls pooled together as well as separately. Fig 1 shows the expected BMI trajectories for the three analyses. We observed a strong fit between observed and expected curves, as revealed by a Pearson’s correlation coefficient of \( r = 0.97 \). Using the same longitudinal model, we estimated BMI-AR and Age-AR (see Methods). The mean Age-AR was 4.5 years in pooled children, 4.35 (SD = 1.35) years in boys and 4.55 (SD = 1.71) years in girls. The mean BMI-AR was 16.3 (SD = 1.4) kg/m² in pooled children, 16.4 (SD = 1.29) kg/m² in boys and 16.2 (SD = 1.42) kg/m² in girls.

Effect of global ancestry on BMI, Age-AR and BMI-AR.

The children of this cohort have on average 52.1% European, 43.8% Mapuche Native American, 2.6% Aymara Native American, and 1.5% African global ancestry proportions (21) (SI Fig). To quantify how much Mapuche and European ancestries affect BMI at each age stratum, we applied linear regressions with BMI as the dependent variable and global Mapuche ancestry as the independent variable. We estimated BMI values for hypothetical individuals with 100% Mapuche ancestry and with 100% European ancestry, hereafter referred to as “Mapuche” and “Europeans”. When compared with Europeans, Mapuche individuals have significantly higher BMI at all age strata between 5.5 and 16.5 years old, but not before 5.5 years old (Fig 2A). Indeed, intercept \( P \)-values were almost all \( < 0.01 \) except at 7.5-8.5 years old (\( P < 0.05 \); Fig 2A and SI Table). Further, BMI differences between Mapuche and European individuals tend to increase with age (Fig 2A). Similarly, global ancestry had a mild but significant effect in the whole longitudinal BMI trajectory (\( P = 0.045 \); effect size = 0.094).

We explored whether individual Mapuche and European global ancestries have an effect on Age-AR and BMI-AR. By regressing Age-AR onto Mapuche global ancestry, our model predicted that Age-AR is significantly lower by 1.64 years in a Mapuche child (mean = 3.53 years) than in a European child (mean = 5.17 years; \( P = 0.013 \); Fig 2B). On the other hand, BMI-AR was 16.9 and 15.9 kg/m² in a Mapuche and a European child, respectively (Fig 2C). However, this difference was not significant.

Identification of variants associated with BMI.

We performed GWAS on BMI in each age stratum by performing linear regression models (see Methods for details; SI Table) shows the number of individuals in each...
Fig 1. Estimation of Age-AR and BMI-AR. Fitted BMI (kg/m²) curves over time (years) in the whole cohort as well as in boys and girls considered separately. Vertical and horizontal dotted lines pinpoint the observed mean Age-AR and mean BMI-AR in boys and girls.

stratum). Besides gender and age, we adjusted for the global ancestry proportion of each individual to account for population substructure as well as for local ancestry to adjust for the ancestry of each haplotype (i.e. SNP). S3 Fig S4 Fig S5 Fig S6 Fig S7 Fig S8 Fig S9 Fig S10 Fig S11 Fig S12 Fig S13 Fig S14 Fig S15 Fig S16 Fig S17 Fig show the corresponding Manhattan plots. We found a significant association for the rs75964957 variant at 4.5 - 5.5 years old stratum (P < 3.5 × 10⁻⁸; S1 Table). The second strongest association (P < 1.6 × 10⁻⁷) was achieved by a genetic signal at chromosome 6 harboring several variants in high linkage disequilibrium, as revealed by a clear peak in the Manhattan plot of Fig 3A. This genetic signal was apparent in the 1.5 - 2.5 years old stratum, but not at other strata (Fig 3B).

The strongest variants of this peak, namely, rs9275582, rs9275593 and rs9275595, map a promoter flanking region of the HLA-DQB3 pseudogene, which is part of the human leukocyte antigen (HLA) region (see Discussion). S2 Table shows all variants from the HLA peak achieving nominal genome-wide significance (P < 1 × 10⁻⁶), with their corresponding annotations. We identified additional association peaks at chromosomes 4 and 20 (S18 Fig). The strongest peak variant at chromosome 4, rs12501266, maps the gene SORCS2, whereas the strongest peak variant at chromosome 20, rs474169, maps the gene SNAP25-AS1. Of note, rs474169 is also associated with BMI at age stratum 10.5 - 11.5 (S1 Table). In contrast to the HLA peak, peak variants at chromosomes 4 and 20 show associations across several strata during childhood and/or adolescence (S18 Fig).

In order to identify variants associated with the whole BMI trajectory during body growth, we performed a longitudinal GWAS on BMI (see Methods for details). The strongest association was achieved by rs35266519 (P = 2.9 × 10⁻⁷), which is a missense variant of the the GSDMB gene. This gene has been GWAS-associated with
Identification of variants associated with Age-AR and BMI-AR.

To identify variants associated with Age-AR, we implemented a regression model described in the Methods section. We detected one variant significantly associated with Age-AR ($P < 9.8 \times 10^{-9}$), S2 Table and S20 Fig. This variant, namely rs445398, maps an intron of the DMRT1 gene. DMRT1 is a key gene involved in sex differentiation (see Discussion). In addition, we identified 33 variants achieving nominal significance ($P < 1 \times 10^{-6}$), S2 Table.

We also performed a GWAS for BMI-AR (see Methods for details). We detected 8 variants achieving nominal significance (S3 Table). S21 Fig shows the Manhattan plot.
Table 1. Top cross-sectional associations for BMI. Shown is the SNP rs ID with the associated allele, physical location, SO consequence/feature type, gene, effect size of the genotype ($\beta_{GT}$), association $P$-value of the genotype ($P_{GT}$), age stratum and frequency of associated allele. NCT: noncoding transcript variant.

<table>
<thead>
<tr>
<th>SNP ID-Allele</th>
<th>Location</th>
<th>Consequence</th>
<th>Gene</th>
<th>$\beta_{GT}$</th>
<th>$P_{GT}$</th>
<th>Age stratum</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs75964957-G</td>
<td>1:239032385</td>
<td>intergenic</td>
<td>-</td>
<td>1.56</td>
<td>4.7E-08</td>
<td>4.5-5.5</td>
<td>0.0077</td>
</tr>
<tr>
<td>rs269511-A</td>
<td>5:114688671</td>
<td>intergenic</td>
<td>-</td>
<td>-0.27</td>
<td>2.2E-07</td>
<td>1.5-2.5</td>
<td>0.4646</td>
</tr>
<tr>
<td>rs9275595-C</td>
<td>6:32772142</td>
<td>promoter flanking region</td>
<td>HLA-DQB3</td>
<td>0.30</td>
<td>2.5E-07</td>
<td>1.5-2.5</td>
<td>0.2865</td>
</tr>
<tr>
<td>rs9275593-A</td>
<td>6:32771630</td>
<td>promoter flanking region</td>
<td>HLA-DQB3</td>
<td>0.30</td>
<td>2.5E-07</td>
<td>1.5-2.5</td>
<td>0.2865</td>
</tr>
<tr>
<td>rs9275595-T</td>
<td>6:32772142</td>
<td>promoter flanking region</td>
<td>HLA-DQB3</td>
<td>0.30</td>
<td>2.5E-07</td>
<td>1.5-2.5</td>
<td>0.2865</td>
</tr>
<tr>
<td>rs7134291-A</td>
<td>12:142832322</td>
<td>intron</td>
<td>GRIN2B</td>
<td>-0.33</td>
<td>3.0E-07</td>
<td>8.5-9.5</td>
<td>0.1726</td>
</tr>
<tr>
<td>rs7896870-T</td>
<td>10:127827684</td>
<td>intron</td>
<td>ADAM12</td>
<td>0.25</td>
<td>4.0E-07</td>
<td>0.5-1.5</td>
<td>0.4906</td>
</tr>
<tr>
<td>rs4741659-G</td>
<td>10:127827684</td>
<td>intron/NCT</td>
<td>RP11-124813.1</td>
<td>0.28</td>
<td>7.2E-07</td>
<td>14.5-15.5</td>
<td>0.3335</td>
</tr>
<tr>
<td>rs77468591-T</td>
<td>10:12937704</td>
<td>downstream</td>
<td>CCDC3</td>
<td>-1.84</td>
<td>9.2E-07</td>
<td>4.5-5.5</td>
<td>0.0039</td>
</tr>
</tbody>
</table>

Table 2. Strongest associations for Age-AR. Shown is the SNP rs ID with the associated allele, physical position in the chromosome, consequence type, gene, effect size of the interaction between genotype and sex ($\beta_{GT\timesSex}$) with the corresponding association $P$-value ($P_{GT\timesSex}$), and the frequency of the associated allele. NCT: noncoding transcript variant.

<table>
<thead>
<tr>
<th>SNP ID-Allele</th>
<th>Location</th>
<th>Consequence</th>
<th>Gene</th>
<th>$\beta_{GT\timesSex}$</th>
<th>$P_{GT\timesSex}$</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs445398-C</td>
<td>9:954336</td>
<td>intron</td>
<td>DMRT1</td>
<td>5.81</td>
<td>9.8E-09</td>
<td>0.165</td>
</tr>
<tr>
<td>rs7395201-G</td>
<td>18:37721370</td>
<td>downstream</td>
<td>RPL17P45</td>
<td>-5.45</td>
<td>7.3E-08</td>
<td>0.007</td>
</tr>
<tr>
<td>rs8706261-G</td>
<td>6:110655312</td>
<td>intron</td>
<td>METTL24</td>
<td>-5.13</td>
<td>3.8E-07</td>
<td>0.008</td>
</tr>
<tr>
<td>rs73139599-T</td>
<td>3:13815160</td>
<td>intergenic</td>
<td>RPL17P45</td>
<td>-4.97</td>
<td>8.4E-07</td>
<td>0.003</td>
</tr>
<tr>
<td>rs969092-G</td>
<td>12:42234778</td>
<td>intron/NCT</td>
<td>RP11-630C16.2</td>
<td>4.96</td>
<td>9.8E-07</td>
<td>0.238</td>
</tr>
<tr>
<td>rs11066997-T</td>
<td>12:114661524</td>
<td>intergenic</td>
<td>RPL17P45</td>
<td>4.94</td>
<td>9.9E-07</td>
<td>0.358</td>
</tr>
</tbody>
</table>

Discussion

In this study, we investigated the genetic architecture of BMI-related traits on admixed children with mainly European and Mapuche Native American ancestry. There are a few studies that have analyzed the genetic architecture of longitudinal growth traits during childhood and adolescence [6] [7] [8] [9] [14], and most of them have been performed in European populations. To our knowledge, Vicuña et al., 2021 [21] is the only study that has quantified the relationship between Native American genetic ancestry and growth traits. The authors showed that the pubertal age where maximum height growth occurs (peak height velocity), is significantly older by 0.73 years in Europeans than in Mapuche adolescents on average. We are the first to quantify how Native American ancestry affects childhood growth traits. One of our most important findings is the observation that increases in Mapuche ancestry associate with significant increases in BMI between 5.5 and 16.5 years old as well as with BMI trajectory between 2 – 10 years old. Moreover, the difference in BMI between Mapuche and Europeans...
Fig 3. Genome-wide associations with BMI. A: GWAS at age stratum 1.5 – 2.5. B: Association peak of chromosome 6 across age strata. The region shown has physical coordinates 6: 32676320 – 32683820 and was centered at the variant with the strongest association.

increases steadily after 5.5 years old. This makes sense, since it is known that the contribution of heritable genetic variation to BMI increases from childhood (4 years) to young adulthood (19 years) [11]. We did not find significant differences at age stratum 4.5 – 5.5 years old, likely due to a lack of statistical power derived from the sample size of our cohort. Another important finding was the significantly lower Age-AR in the Mapuche compared with Europeans on average. A higher adiposity in the Mapuche during childhood could explain in part why this population is particularly susceptible to developing metabolic disorders such as insulin resistance, obesity, cholesterol gallstones and metabolic syndrome during adulthood, since all of these disorders are strongly associated with elevated lipid levels [20]. Moreover, Mapuche ancestry among Chileans is distinctively associated with heart diseases, hypertension and diabetes mellitus [19]. Further, the prevalence of type II diabetes and obesity among the Mapuche increased significantly following the change from a rural to an urban life style [33], suggesting that...
genotype-environment interactions may lead to a higher genetic susceptibility to developing cardiometabolic diseases.

It is unknown whether the effect of Mapuche genetic ancestry on BMI-related traits was originated by random genetic drift in ancient Native American or European populations, by adaptation to selective pressures (see [16]) or by a combination of adaptation, drift and/or other evolutionary forces. It is also possible that environmental factors could partially contribute to such differences. However, the children of our cohort belong to the same urban district in Santiago and to the same middle-lower socioeconomic group. Therefore, we do not expect high variation in environmental factors such as nutrition, exposure to pollutants or incidence of pathogenic diseases. However, it is worth noting that in this same cohort, mothers of boys with high Mapuche ancestry (≥ 3 Mapuche last names: 4.8% of total boys) have a lower educational level than mothers of boys with low Mapuche ancestry (0 Mapuche last names: 80% of boys) [35], which in turn could potentially have an effect on their son’s BMI.

Our GWAS on BMI detected a significant association for the intergenic variant rs75964957 at age stratum 4.5 – 5.5 years. This was despite the sample size at this stratum (n = 708). The second stronger association was for the rs9275582, rs9275593 and rs9275595 variants, which map a promoter flanking region of the HLA-DQB3 unprocessed pseudogene. Noteworthy, rs9275595 has been previously associated with BMI in adults (P = 6 × 10^{-6}, 0.016 kg/m^2 increase; [23]). These three variants constitute the top of an association peak that was observed at age stratum 1.5 – 2.5 years, but not at other age strata. The HLA-DQB3 gene belongs to the human leucocyte antigen (HLA) gene cluster, which harbors hundreds of genes that are fundamental for immune function. HLA genes encode proteins of the major histocompatibility complex, which present antigenic peptides to immune cells, in order to distinguish between “self” and “non-self” agents [29]. It is unknown why these HLA-DQB3 variants have such a distinct effect on BMI at 1.5 – 2.5 years old, but the observation that they map a promoter flanking region suggests that they affect early BMI by regulating the expression of HLA-DQB3 or additional HLA genes in high linkage disequilibrium with HLA-DQB3. Future studies will be needed to validate this finding. It is worth noting that the HLA region shows high variability among ethnic groups [37] and that the HLA-DR/DQ region is the major determinant of susceptibility to childhood type 1 diabetes [38]. The strongest association of the longitudinal GWAS on BMI was for rs35266519, a missense variant of the GSDMB gene. Among other phenotypes, GSDMB variants have been GWAS-associated with type 1 diabetes (P < 6 × 10^{-13}) [39] and subcutaneous adipose tissue (P < 3 × 10^{-8}) [40], two traits that are strongly related with obesity.

We identified a variant of the DMRT1 gene significantly associated with Age-AR. Remarkably, this association was captured using a sample size of n = 696. DMRT1 is a hallmark gene involved in sex differentiation by maintaining the fates of testes or ovaries in adult mammals [30]. Indeed, deletion or inactivation of DMRT1 in humans causes XY male-to-female sex reversal [31]. Thus, it is possible that genetic variation in this gene could affect endocrine mechanisms of infant growth. Moreover, this effect on infant growth may lead to an altered pubertal timing, since it has been observed that Mapuche boys have a higher incidence of precocious puberty [35].

We could not replicate our findings in an independent cohort, due to the lack of longitudinal paediatric growth cohorts with enough longitudinal BMI measurements and relatively high Native American ancestry. To our knowledge, the Salvador-SACALA cohort from Brazil [32] and the COIPIS cohort from Mexico [10] are the only paediatric cohorts with genome-wide genotype information and Native American admixture. However, these cohorts present severe limitations for replication. In the
Salvador-SCAALA cohort, the mean Native American global ancestry is too low (6%), in contrast to high mean African and European ancestries (51% and 46%, respectively; [2]). Also, in this cohort only a single BMI measurement per individual was taken in a broad age window (4 – 11 years old). Regarding the COIPIS cohort, while the level of Native American ancestry is appropriate (36%; [10]), it has a similar problem, namely, that only a single BMI measurement per individual was taken in children aged 3 – 18 years old.

Conclusion

Our study contributes to better understand the biological mechanisms underlying BMI and Age-AR differences between Native Americans and Europeans. Also, it highlights the importance of performing serial cross-sectional GWAS for complex longitudinal traits like BMI trajectory, since longitudinal studies alone might not be well-suited to capture effects that occur at short time intervals. Finally, our results have medical relevance, since they could be useful to identify genetic risk factors for infant obesity based on the individual’s proportion of Native American genetic ancestry.

Consent to Participate

Informed consents from participants were obtained from parents or guardians. Children agreed to participate when they turned 7 years old. This study was approved by the Scientific Ethics Committees of Instituto de Nutrición y Tecnología en Alimentos (INTA) and Pontificia Universidad Católica de Chile.

Acknowledgments

This work was supported by the Fondo Nacional de Ciencia y Tecnología (FONDECYT) [1200146 to S.E., L.V. and E.B.; 1190346 to V.M.; 1150416 and 1150486 to JLS., 1190801 to C.M.]. S.E., T.N. and L.V. were additionally supported by the Instituto Milenio de Investigación Sobre los Fundamentos de los Datos (IMFD). S.E. conceived the project. S.E. and L.V. designed experiments. E.B., L.V., T.N., C.M., D.A. and V.L. analyzed the data. A.P. and V.M. collected phenotype data. S.E., JLS. and JC. G. raised funds for the genotyping data. L.V., S.E., A.P, E.B. wrote the manuscript. All authors critically reviewed and accepted the final version.

References


32. Scliar, M. O. et al. Admixture/fine-mapping in brazilians reveals a west african associated potential regulatory variant (rs114066381) with a strong female-specific effect on body mass and fat mass indexes. *Int J Obes (Lond)* **45**, 1017–1029 (2021).


