Title: Predicted loss of function alleles in Bassoon (BSN) are associated with obesity

Authors: Na Zhu* 1,2, Charles A. LeDuc* 1,3,4, Ilene Fennoy 1,4, Blandine Laferrère 3,4,5, Claudia A. Doege 3,6, Yufeng Shen 2,7,8, Wendy K. Chung** 1,3,4,5,9 Rudolph L. Leibel** 1,3,4

1. Department of Pediatrics, Columbia University Irving Medical Center, New York, NY, USA
2. Department of Systems Biology, Columbia University Irving Medical Center, New York, NY, USA
3. NY Obesity Research Center, Columbia University Irving Medical Center, New York, NY, USA
4. Naomi Berrie Diabetes Center, Columbia University Irving Medical Center, New York, NY, USA
5. Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA
6. Department of Pathology, Columbia University Irving Medical Center, New York, NY 10032, USA
7. Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, USA
8. JP Sulzberger Columbia Genome Center, Columbia University Irving Medical Center, New York, NY, USA
9. Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, NY, USA

*Co-first authors
** Co-senior/corresponding authors

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Bassoon (BSN) is a component of a hetero-dimeric presynaptic cytomatrix protein that orchestrates neurotransmitter release with Piccolo (PCLO) from glutamatergic neurons throughout the brain. Heterozygous missense variants in BSN have previously been associated with neurodegenerative disorders in humans. We performed an exome-wide association analysis of ultra-rare variants in about 140,000 unrelated individuals from the UK Biobank to search for new genes associated with obesity. We found that rare heterozygous predicted loss of function (pLoF) variants in BSN are associated with higher BMI with log10-p value of 11.78 in the UK biobank cohort. The association was replicated in the All of Us whole genome sequencing data. Additionally, we have identified two individuals (one of whom has a de novo variant) with a heterozygous pLoF variant in a cohort of early onset or extreme obesity at Columbia University. Like the individuals identified in the UKBB and All of us Cohorts, these individuals have no history of neurobehavioral or cognitive disability. Heterozygosity for pLoF BSN variants constitutes a new etiology for obesity.

Introduction:

By 2030 it is estimated that roughly 50% of adults in the United States will have obesity, with 25% having severe obesity (1). The prevalence of obesity in U.S. adults has increased from 30.5 to 41.9% from 1999-2000; the prevalence of severe obesity has increased from 4.7 to 9.2%. Approximately 18% of U.S. children currently have obesity (2). Various estimated, the risk of obesity is 30-50% heritable (3-6). Changes in the underlying genetics cannot be responsible for such changes in the prevalence of obesity over such a short period of time; however, the
propensity to gain weight in an environment with ready access to food is largely
genetic (7). Genome wide association studies have identified many common variants
associated with body weight regulation (8-10). More recently, polygenic risk scores
aggregating large numbers genetic variants, each with small contributions to energy
homeostasis can be used to predict obesity deciles in some genetic ancestries (11).
However, the genetic attributable risk for obesity remains modest at ~3% (12, 13).
Exome sequencing of large numbers of individuals has accelerated the discovery of
rare genetic contributors to quantitative phenotypes such as height (14, 15), celiac
disease (16), and dyslipidemia (17, 18). In many instances the precise
mechanistically functional relevance of these associated genetic variants remains
unknown.

Recent advances in the treatment of obesity (19) and hyperlipidemia (20) have used
human genetics to identify genes contributing to extreme phenotypes to understand
biology and molecular mechanisms and develop novel interventions. The advent of
large-scale exome/genome sequencing in the United Kingdom Biobank (UKBB) and
All of US has extended the ability to assess rare variants at large scale in addition to
prior methods of assessing common variants in GWAS. In the current study we
combine the power of exome sequence-based analysis of an extreme obesity cohort
recruited at Columbia University with data from the UKBB and All of Us. We report
the association of predicted loss of function (pLoF) alleles in the gene BSN with body
mass index (BMI).

Methods and Materials
Methods

Columbia Cohort

The Columbia University Extreme Obesity cohort was collected using protocols approved by the Institutional Review Boards at Columbia University Irving Medical Center (New York, NY) and The Rockefeller University (New York, NY). The cohort consists of 1598 individuals from 903 families. Obesity was defined as described below. Of the 903 families, 122 constitute affected child/parent trios. The remaining 781 families have 1372 affected (890 females and 482 males) and 226 unaffected family members. Cohort details have been described previously (21, 22).

Approximately half of the probands were pediatric (either recruitment or obesity onset age younger than 19 years old with 674 participants having a BMI Z score >=2; average age at enrollment 6.6 +/- 3.6 years) and half adults (obesity onset or recruitment age at least 19 years old with 698 adults with BMI >=30 kg/m²; average age 51.5 +/- 12.0 years) (Table 1). Samples were exome sequenced using xGen and SeqCap VCRome Capture. Greater than 99% of samples had depth of coverage > 10 in 80% of target regions.

Controls were the unrelated parents (without autism) from the Simons Powering Autism Research for Knowledge (SPARK) study and were exome sequenced using the XGEN-Capture (23).

UKBB Cohort

For this analysis, we included 200,643 UK biobank cases (24). The average age of this cohort is 56.4 +/- 8.1 years; mean BMI of 27.3 +/- 4.8 kg/m²; 55.1% female (Table 2).
The current release (June 2022) of the All of Us data includes whole genome sequencing for 98,622 individuals (58,190 females and 38,290 males). The average age of this cohort is 52.6 +/- 16.9 years; mean BMI is 30.9 +/- 9.0 kg/m².

**Bioinformatic analysis of exome or genome sequencing data**

**Columbia cohort**

Paired-end reads were mapped and aligned to the human reference genome (version GRCh38/hg38, accession GCA 000001405.15) using BWA-MEM (25). Picard v1.93 MarkDuplicates ([http://broadinstitute.github.io/picard/](http://broadinstitute.github.io/picard/)) was used to identify and flag PCR duplicates and GATK v4.1 HaplotypeCaller (26) in Reference Confidence Model mode to generate individual-level gVCF files from the aligned sequence data. We performed joint calling of variants from the obesity cohorts using GATK variant caller.

**Ancestry prediction and relatedness check**

We predicted the ancestry and relatedness in the Columbia cohort using Peddy (27). Relatedness prediction in the UKBB samples, due to the large sample size, was completed with plink King (28). When pairs of samples shared second degree relationship or closer (a kinship coefficient greater than 0.12 in King or 0.25 in Peddy), the sample with greater relatedness to the cohort was excluded.

**Variant annotation**

We used the Ensembl Variant Effect Predictor (VEP, Ensemble 93) (29) to annotate variant function and ANNOVAR (30) to aggregate variant population frequencies and
for in silico predictions of deleteriousness. Rare variants were defined by a population frequency < 10^{-4} in gnomAD WES and WGS (31). Deleterious variants were defined as predicted loss of function (pLoF: including premature stop-gain, frameshift indels, canonical splicing variants and multi-exon deletions) or predicted damaging missense (Dmis) based on REVEL (32) score thresholds. The same pipeline was used for Columbia, UKBB, and All of US variant annotation.

Statistical analysis

Columbia cohort

We tested the single variant association with obesity using the exact binomial test in the unrelated European participants. To identify novel risk genes for obesity in the Columbia cohort, we performed a rare variant gene burden test using the binomial test in unrelated European participants. When there were multiple individuals with obesity in a family, we defined the most severely affected as the proband (defined as the child with the highest Z-score or the adult with the highest BMI if there were only adults in the family).

A gene-based case-control association test was performed on 483 unrelated cases and 11,101 unrelated SPARK non-autism parents as population controls by comparing the frequency of rare deleterious variants in obese cases with SPARK controls. To minimize false positive variant calls and reduce batch effect, we applied additional heuristic filters in cases and controls by the following exclusion criteria (a variant was excluded if any one condition was met):

- Variants were filtered out if any of these exclusion criteria were met: (a) cohort allele frequency was > 0.01; (b) the variant was not uniquely mappable; (c) genotype allelic...
fraction was < 0.2; (d) the variant was shared in multiple cases (alternate allele count > 4) with at least half of the cases with low quality calls (allelic fraction < 0.35); (e) less than 90% of individuals (cases and controls) have ≥10x depth of average of the variant site. (f) All variants in SPARK parents were required to pass the deep variant test (33, 34). (g) All single nucleotide variants (SNVs) were high quality calls defined by GATK VQSLOD > -3 in the case cohort.

To assess the overall degree of batch effects, we compared the rare synonymous variant frequencies between cases and controls, testing the assumption that most rare synonymous variants do not have effects on obesity. A gene level burden test QQ plot for synonymous variants shows deflation with lambda=0.75 due to the limited case sample size resulting in genes that had no variants in the cases. Nevertheless, observed p-values were consistent with the expected p-value in the testable genes (Sup Figure 1).

To identify obesity-risk genes, we tested the deleterious variant burden (pLoF or Dmis) in each protein-coding gene in cases compared to controls using an exact binomial test. REVEL scores were used to predict the deleteriousness of missense variants. We performed 20 association tests for each gene, including pLoF only, Dmis only and Dmis + pLoF where Dmis was defined using 5 different REVEL score cutoffs (0.15 to 0.95 by 0.2).

UKBB cohort

After excluding related individuals and individuals with a history of cancer or eating disorder, 144,496 unrelated European individuals were selected for quantitative trait
(BMI) association analysis (31, 35). We collapsed rare variants based on allele frequency and predicted variant deleteriousness. The variants were partitioned into cohort frequency <10^{-4} and singleton population allele frequency groups as well as 10 variant functional groups. The variant functional groups were missense variants with REVEL >=x, with x ranging from 0.15 to 0.95 in 0.2 increments with or without pLoF variants (10 groups). Genes with less than 15 heterozygotes in a test group were removed. The significance threshold was set at (0.05/ (20*20,000)). We then tested the quantitative BMI for the 144K UKBB individuals using REGENIE (36), which accounts for relatedness, population structure and polygenicity. We included age, Townsend deprivation index at recruitment, smoking /alcohol status, sex, the first 8 principal components, and genetic heterozygosity as covariates. REGENIE resolved the gene-based association tests in the large UKBB dataset with no inflation or deflation in the synonymous variants with the gene-based tests (Sup Figure 2a). The type I error rate was well controlled for pLoF and Dmis variants in gene-based tests, showing minor inflation in the QQ plot (Sup Figure 2b).

Finally, we ran a meta-analysis using Fisher’s method (https://cran.r-project.org/web/packages/metap/index.html) for UKBB and Columbia samples with the same defined variant groups. We defined the threshold for genome-wide significance by Bonferroni correction for multiple testing (n=20,000*20, threshold p-value=1.3e-7) (workflow shown in Figure 1).

All of Us

To attempt to replicate findings from the UKBB analysis, we ran a linear regression on the 48,722 European ancestry individuals from the All of Us dataset using their
provided cloud-based research platform to test the association between BMI and
BSN and MEOX1 deleterious variants using age, sex, deprivation index and median
income as covariates.

Results

In the single variant association tests, we identified two exome-wide significant single
nucleotide variants (SNVs) in the Columbia cohort. rs887287256 is a c.C477A: p.
Asp159Glu variant in C6ORF52 (NM_001388310.1). The Columbia cohort had six
unrelated European individuals with obesity who were heterozygous, and no
heterozygotes or homozygotes in 11,101 SPARK controls (-log10p, 8.27, RR=276).
rs202058123, a c.G649A: p.Gly217Ser SNV in CTRC was present in five
heterozygotes with obesity in the Columbia cohort and one of the SPARK controls (-
log10p 6.1, RR =114) (Supplementary Table 1-3). We performed segregation
analysis for those Columbia families with available family members (Supplementary
Figure 3 and 4). All heterozygotes had obesity, but not all individuals with obesity in
the family had the relevant variant. However, neither variant association was
replicated when tested using the UKBB data.

We performed gene-based burden tests with 20 groups tested for each gene. Twelve
tests in three unique genes (MC4R, BSN, and MEOX1) passed Bonferroni corrected
significance (-log10 p-value >=6.9) in the combined association tests. Using a false
discovery rate < 0.1, the most significant gene-variant sets are listed in Table 3,
Supplemental Table 4 and 5.
Limiting the analysis to pLoF and Dmis variants with REVEL score >=0.25, the association test was genome-wide significant for MC4R, with a BMI effect size beta in UKBB of 1.4 kg/m² and relative risk for obesity of 5.03 in the Columbia cohort. The UKBB and Columbia heterozygotes are listed in Supplemental Table 6 and 7 and Supplemental Figure 5. Effect size was estimated with a linear regression test run on individual variants.

The combined (Columbia and UKBB) p-value (-log10P:10.33) for BSN reached genome-wide significance. This signal is primarily driven by the UKBB data since pLoF and Dmis variants with REVEL score >=0.75 are extremely rare (AF in UKBB was 9.3e-05) and few in number in the smaller Columbia cohort. The UKBB data alone have a strong signal with a BMI effect size beta of 6.21 and -log10p of 11.78. No positive effect size is observed in other missense groups. All heterozygous predicted deleterious variants in UKBB are listed in Supplemental Table 8. Figure 2a shows the BMI distribution of BSN predicted deleterious heterozygotes compared to the overall UKBB population (Kolmogorov-Smirnov pvalue 1.4e-05).

Two heterozygous pLoF BSN alleles were identified in the Columbia cohort (Figure 2b). Study IDs are known only to the study staff. RU2487 is heterozygous for a *de novo* p.Gln703X allele in BSN. At the time of the last assessment, she was a Latina woman in her 20’s with a history of severe obesity and type 2 diabetes mellitus diagnosed as a teen at which time her HbA1c was 7.4%. She was amenorrheic and had extensive acanthosis nigricans, dyslipidemia, hypothyroidism, and hyperandrogenism. Her maximal weight was 113 kg. She had gastric bypass surgery for weight loss in her 20’s. Immediately prior to bariatric surgery, her BMI was 39.7
kg/m². Her oral glucose tolerance test prior to bariatric surgery showed euglycemic
hyperinsulinemia. Her nadir body weight after surgery was 77 kg; 2 years post-
operatively she weighed 101 kg. She reports frequently feeling very hungry. She is a
college graduate with no academic or cognitive difficulties nor history of psychiatric
diagnoses. She has no family history of obesity or type 2 diabetes.

RU2617 is an African American female heterozygous for a p.R3494X variant in BSN;
the allele was not inherited from the only parent available for genetic analysis. At the
time of her initial evaluation, the patient was a teen with body weight of 162 kg and
height of 160.9 cm (BMI=62.6 kg/m²). Her waist circumference was 158 cm. She had
no history of irregular periods. She had obstructive sleep apnea requiring continuous
positive airway pressure. She initially had a normal glucose tolerance test with
normal fasting glucose and HbA1c = 6.3%; however, she subsequently developed
impaired fasting blood glucose of 105 mg/dl with persistently elevated HbA1c. She
had laparoscopic adjustable gastric banding as a teen. At 3 years post operatively,
her weight had declined to 134.2 kg and her height had increased to 163 cm (BMI of
50.5 kg/m²). HbA1c normalized to 5.2%.

The association of MEOX1 with BMI was genome-wide significant (-log10P: 7.04) in
the combined analysis of the UKBB and Columbia cohorts. In the Columbia cohort,
deleterious missense variants (REVEL >=0.15) were 10.2 times more frequent than
in the SPARK participants. The singleton deleterious variants in MEOX1 were
marginally significantly associated with BMI in UKBB (-log10P 3.21, beta 0.84). The
majority of singleton predicted deleterious MEOX1 variants in the UKBB were
associated with a higher BMI. The BMI in individuals with heterozygous MEOX1
deleterious missense variants was significantly higher than the overall UKBB (p-value 0.03 using the Kolmogorov-Smirnov test). (Figure 3a and Supplemental Table 9). In the Columbia cohort, MEOX1 predicted Dmis variants (Table 4) were enriched in the pediatric-onset compared with the adult-onset obesity cases. There were 7 heterozygotes out of 262 unrelated European ancestry obese children and 2 heterozygotes out of 362 unrelated European ancestry obese adults (p=1.2e-6 with a relative risk of 16.5 for the pediatric-onset group and p= 0.13 with a relative risk of 3.4 in adult-onset group). For the MEOX1 individual variants (Table 4 and Figure 3b), missense variant p.R213H (CADD score 27.5 and REVEL 0.926, indicating likely deleterious) was observed in 3 pediatric-onset and 1 severe adult-onset individuals in the Columbia cohort; there were none in SPARK participants. In the UKBB there were two heterozygote participants with p.R213H variants with BMI 26.4 and 29.9 kg/m2. Across the combined TOPMED and gnomAD databases p.R213H was observed only once. The missense variant p.R184Q (CADD score 28.1, REVEL score 0.662) was observed in three pediatric-onset individuals in the Columbia cohort and twice in the SPARK participants. In the UKBB, there were 10 heterozygotes: one had obesity, seven had overweight and two had normal BMI. The population frequency of the p.R184Q variant is 8.5e-05 in gnomAD and 8e-05 in TOPMED. Segregation analysis for the Columbia MEOX1 heterozygotes showed that all the heterozygotes in those families had obesity (Figure 4).

Association of BMI-correlated traits in BSN

The association between BSN and the traits correlated with BMI tested using REGENIE (Table 5) showed arm, leg and trunk fat mass and leg fat-free mass and leg predicted mass reached genome-wide significance. We also tested the
association between BSN and ICD10 diagnoses (Supplemental Table 10) using the binomial test. No diagnosis was significantly associated with BSN after correction for multiple testing.

Replication analysis using All of Us data.

We identified BSN and MEOX1 heterozygotes in the All of Us cohort. To date, there are 98,622 subjects for whom both whole genome sequencing and clinical data are available. Half of the participants (47,897) are unrelated and of European ancestry.

For each participant, we used the highest recorded BMI, giving a cohort average BMI of 30.1 +/- 7.8 kg/m². In the cohort, 12 European individuals were heterozygous for BSN pLoF variants, with an average BMI of 37.0 +/- 5.7 kg/m². Using sex, age, income, and deprivation index as covariates, we tested the association between BMI and BSN genotype using linear regression and found a significant association (p-value=0.0075, beta=6.27). Additionally, we identified an additional six BSN pLoF heterozygotes among the non-European participants (mean BMI 31.5 (SD = 8.5 kg/m²); BMI range = 22-45; 3/6 with BMI >30.0; Supplemental Table 11). Thus, the BSN obesity association observed in the UKBB and Columbia cohorts was replicated in the All of Us cohort.

MEOX1 predicted deleterious variants were not associated with higher BMI in All of Us (pvalue=0.47, beta=0.57).

**Discussion**

We have identified a gene, BSN, for which we have demonstrated an association of rare pLoF variants with obesity in two independent large cohorts: the UKBB and All of Us. Additionally, we identified extremely obese individuals in the Columbia cohort...
of extreme obesity, including an individual with extreme, early onset obesity
associated with a de novo pLoF allele. There is no evidence that these variants are
associated with intellectual disability or cognitive impairment, including direct
assessment of two individuals in the Columbia cohort. A second gene, MEox, was
identified with predicted Dmis variants associated with obesity in the UKBB and
Columbia cohorts, but this finding was not replicated in the All of Us cohort.

BSN (bassoon) is expressed primarily in the brain (including embryonic and adult
brain regions that impact feeding behavior (37)), inner hair cell ribbons, and the
retina of mammals. Bassoon is a presynaptic scaffold protein localized in the
cytomatrix at the active zone (CAZ) where it functions to orchestrate
neurotransmitter release. Bassoon participates in the formation of Golgi-derived
Piccolo-Bassoon transport vesicles that are axonally transported to newly formed
synaptic contacts. Bassoon is associated with activity-dependent short- and long-
term neuronal plasticity (38).

Bassoon is expressed during early neuronal differentiation, is selectively sorted into
axons and is among the first proteins to arrive at nascent synapses (38). The release
of neurotransmitters from the presynaptic terminal involves the active zone (AZ). The
AZ includes an electron-dense protein meshwork, the presynaptic cytomatrix.
Bassoon is one of several scaffolding proteins (along with Piccolo (PCLO), RIM,
MUNC13, and ELKS) within the presynaptic cytomatrix. BSN and PCLO are
structurally related, interact, and are the largest active-zone-specific proteins. Unlike
other the proteins in the AZ that are evolutionally conserved down to C. elegans,
Piccolo and Bassoon are only found in vertebrates (39).
Mice homozygous for LoF Bsn alleles have reduced synaptic transmission that is primarily caused by the inactivation of a significant fraction of glutamatergic synapses. These mice have spontaneous epileptic seizures. Bassoon is not essential for synapse formation but is essential for regulated neurotransmitter release from a subset of glutamatergic synapses. (40). At the ultrastructural level, these inactive synapses cannot be distinguished from functional synapses. These homozygous Bassoon mutant mice have seizures with structural brain alterations including enlarged hippocampi and cerebral cortices (41). These animals are not obese.

Bassoon is involved in the maintenance of the integrity of AZ (42). Glutamatergic synapses from Bsn knockout mice exhibit enhanced short-term synaptic depression with a high percentage of silent synapses but have no gross structural defects (43), presumably due to the significant functional redundancy with Picolo. When both proteins are absent from glutamatergic synapses, the cells undergo synapse degeneration (44).

BSN was originally identified while attempting to identify expressed cerebellar transcripts in patients with multiple system atrophy, a rare progressive neurodegenerative disease characterized by cerebellar symptoms, parkinsonism, and autonomic dysfunction (45). This study did not find coding mutations in BSN but first identified BSN as a new transcript that they could clone from the cerebellum of these patients. BSN acts in concert with Parkin RBR E3 Ubiquitin Protein Ligase (PRKN) to control presynaptic autophagy and maintain homeostatic presynaptic
proteostasis and synaptic vesicle turnover (46). Human heterozygous missense
variants in BSN have been implicated in neurodevelopmental and neurodegenerative
disorders including progressive supranuclear palsy-like syndrome, a rare
neurodegenerative tauopathy (47).

We have implicated heterozygous pLoF variants in BSN as a new genetic etiology
for human obesity that is not associated with adverse impact on cognition or other
neurobehavioral phenotypes. The expression of BSN throughout the brain suggests
that gene dosage could contribute to hyperphagia through both homeostatic and
hedonic circuits (48). Additional detailed phenotypic assessment – ideally of
individuals prior to the onset of obesity - will be required to assess this point. BSN is
expressed in the synapses of glutamatergic neurons and hypothalamic neurons
mechanistically tied to ingestive behaviors (43, 49-51). The valence of these effects
is consistent with hyperphagic obesity conveyed by hypomorphic alleles.

Declarations:
Competing interests: No author has any conflicts or competing interests rated to the
manuscript.
IRB: All studies were under the auspices of the Columbia University IRB "Molecular
Genetic Analysis of Obesity and Non-Insulin Dependent Diabetes Mellitus" IRB #:
AAAA4485 which expires on 5/1/23.

Acknowledgements
We thank the participants who generously contributed to this work and their
clinicians who referred them.
This work was supported by NIH grant NIDDK 52431 and the NY Nutrition and Obesity Research Center: P30DK26685.
Clinical characteristics of Columbia extreme or early-onset obesity cohort. BMI is body mass index. BMIZ is BMI z score. F is female. M is male. EUR indicates European ancestry, AFR is African-American ancestry, AMR is admixed Americans ancestry.

<table>
<thead>
<tr>
<th></th>
<th>Child-onset</th>
<th>Adult-onset</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI (mean, sd)</td>
<td>41.4, 12.4</td>
<td>45.9, 11.9</td>
</tr>
<tr>
<td>BMIZ (mean, sd)</td>
<td>6.6, 3.6</td>
<td></td>
</tr>
<tr>
<td>age (mean, sd)</td>
<td>12.2, 3.5</td>
<td>38.9, 12.3</td>
</tr>
<tr>
<td>F:M</td>
<td>523:377</td>
<td>503:195</td>
</tr>
<tr>
<td>EUR</td>
<td>322</td>
<td>528</td>
</tr>
<tr>
<td>AFR</td>
<td>159</td>
<td>85</td>
</tr>
<tr>
<td>AMR</td>
<td>184</td>
<td>81</td>
</tr>
<tr>
<td>other ancestries</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>total</td>
<td>674</td>
<td>688</td>
</tr>
</tbody>
</table>
Table 2. Summary of United Kingdom Biobank subjects.

<table>
<thead>
<tr>
<th>Table 2 UKBB cohort summary</th>
<th>Overall Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI (mean, sd)</td>
<td>27.3, 4.8</td>
</tr>
<tr>
<td>age (mean, sd)</td>
<td>56.5, 8.1</td>
</tr>
<tr>
<td>F:M</td>
<td>110476:90153</td>
</tr>
<tr>
<td>EUR ancestry</td>
<td>167,246</td>
</tr>
<tr>
<td>---removed for relatedness</td>
<td>4,878</td>
</tr>
<tr>
<td>---had cancer</td>
<td>16,711</td>
</tr>
<tr>
<td>---had eating disorder</td>
<td>112</td>
</tr>
<tr>
<td>---cancer and eating disorder</td>
<td>18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cohort Included in study</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR no cancer, no eating disorder</td>
<td>145,103</td>
</tr>
<tr>
<td>F:M</td>
<td>78103:67000</td>
</tr>
<tr>
<td>BMI (mean, sd)</td>
<td>27.5, 4.7</td>
</tr>
<tr>
<td>age (mean, sd)</td>
<td>56.6</td>
</tr>
</tbody>
</table>

Correlation between age and BMI =0.048, significant
correlation between age and Sex =0.082, significant

The UKBB cohort use in the analysis. Samples that were coded with, cancer, eating disorders, or both were removed from the cohort prior to analysis.

Relatedness was estimated using plink King, when sample pairs had a relatedness greater than 0.12 (second degree relative or closer) the sample that had more relatedness to the cohort was excluded.
Table 3 Meta analysis for UKBB REGENIE linear regression and Columbia binary burden test

<table>
<thead>
<tr>
<th>GeneName</th>
<th>log10(combinedPvalue)</th>
<th>FDR</th>
<th>variant_function_filter</th>
<th>MAF.UKBB</th>
<th>BETA.UKBB</th>
<th>SE.UKBB</th>
<th>LOG10.pvalue.UKBB</th>
<th>AF.Columbia</th>
<th>AF.SPARK</th>
<th>RR.Columbia</th>
<th>LOG10.Pvalue.Columbia</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC4R</td>
<td>11.08</td>
<td>1.50E-06</td>
<td>Lof+Drelvel &gt;=0.35</td>
<td>0.0001</td>
<td>1.29E-03</td>
<td>1.53</td>
<td>0.24</td>
<td>9.50</td>
<td>7.23E-03</td>
<td>1.40E-03</td>
<td>5.19</td>
</tr>
<tr>
<td>BSN</td>
<td>10.33</td>
<td>4.23E-06</td>
<td>Lof+Drelvel &gt;=0.75</td>
<td>0.0001</td>
<td>9.34E-05</td>
<td>6.36</td>
<td>0.99</td>
<td>11.78</td>
<td>0</td>
<td>3.15E-04</td>
<td>0</td>
</tr>
<tr>
<td>MEFOX1</td>
<td>2.04</td>
<td>2.03E-03</td>
<td>Drivel&gt;=0.15</td>
<td>singleton</td>
<td>1.07E-04</td>
<td>2.88</td>
<td>0.84</td>
<td>3.21</td>
<td>8.28E-03</td>
<td>8.13E-04</td>
<td>10.21</td>
</tr>
<tr>
<td>PCSK1</td>
<td>5.48</td>
<td>0.04</td>
<td>Lof+Drelvel &gt;=0.35</td>
<td>0.0001</td>
<td>1.88E-03</td>
<td>0.99</td>
<td>0.20</td>
<td>6.08</td>
<td>3.11E-03</td>
<td>1.71E-03</td>
<td>1.81</td>
</tr>
<tr>
<td>HECTD4</td>
<td>5.13</td>
<td>0.07</td>
<td>Lof+Drelvel &gt;=0.75</td>
<td>singleton</td>
<td>2.21E-04</td>
<td>2.76</td>
<td>0.58</td>
<td>5.61</td>
<td>1.04E-03</td>
<td>1.80E-04</td>
<td>5.75</td>
</tr>
<tr>
<td>ACAP3</td>
<td>5.06</td>
<td>0.07</td>
<td>Lof+Drelvel &gt;=0.95</td>
<td>singleton</td>
<td>7.61E-05</td>
<td>4.99</td>
<td>1.00</td>
<td>6.25</td>
<td>0</td>
<td>0.00E+00</td>
<td>NA</td>
</tr>
</tbody>
</table>

The statistical results from the UKBB and CUIMC cohorts were combined using the Fisher method. Genes with a meta p-value < 0.00001 and False discovery rate (FDR) < 0.1 are listed. FDR for meta p-value were calculated using Benjamin – Hochberg method. The smallest p-value for each gene are listed. The log10 significance threshold after Bonferroni correction was 6.9. BSN and MEOX1 were novel genes reached the genome wide significance. MC4R and PCSK1 are known obesity risk genes.
Table 4. Predicted deleterious variants in *MEOX1* in the Columbia cohort.

<table>
<thead>
<tr>
<th>Family_ID</th>
<th>Relationship with proband</th>
<th>Affected</th>
<th>Sex</th>
<th>Ancestry</th>
<th>OnsetGroup</th>
<th>Age at recorded BMI</th>
<th>BMI (kg/m²)</th>
<th>BMIZ</th>
<th>HGVSc</th>
<th>HGVSp</th>
<th>CA3D14_Raw</th>
<th>CADD14_PHRED</th>
<th>REVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>223</td>
<td>daughter</td>
<td>unaffected</td>
<td>female</td>
<td>EUR</td>
<td>Child</td>
<td>10-19</td>
<td>19.5</td>
<td>c.109C&gt;G</td>
<td>p.His37Asp</td>
<td>3.17</td>
<td>23.3</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>635</td>
<td>Proband</td>
<td>Obese</td>
<td>female</td>
<td>EUR</td>
<td>Child</td>
<td>10-19</td>
<td>38.5</td>
<td>4.2</td>
<td>c.109C&gt;G</td>
<td>p.His37Asp</td>
<td>3.17</td>
<td>23.3</td>
<td>0.53</td>
</tr>
<tr>
<td>757</td>
<td>Proband</td>
<td>Obese</td>
<td>female</td>
<td>EUR</td>
<td>Child [onset at 8y]</td>
<td>60-69</td>
<td>68.6</td>
<td>NA</td>
<td>c.32G&gt;T</td>
<td>p.Pro109Leu</td>
<td>1.94</td>
<td>19.22</td>
<td>0.17</td>
</tr>
<tr>
<td>X14</td>
<td>Proband</td>
<td>Obese</td>
<td>male</td>
<td>AFR</td>
<td>Child</td>
<td>10-19</td>
<td>33.7</td>
<td>5.6</td>
<td>c.451C&gt;T</td>
<td>p.Arg151Tnp</td>
<td>4.51</td>
<td>31</td>
<td>0.54</td>
</tr>
<tr>
<td>392</td>
<td>Father</td>
<td>Obese</td>
<td>male</td>
<td>AFR</td>
<td>Adult</td>
<td>30-39</td>
<td>30.8</td>
<td>NA</td>
<td>c.505A&gt;G</td>
<td>p.Ser169Gly</td>
<td>2.49</td>
<td>22</td>
<td>0.28</td>
</tr>
<tr>
<td>392</td>
<td>Proband</td>
<td>Obese</td>
<td>female</td>
<td>AFR</td>
<td>Child</td>
<td>0-8</td>
<td>30.5</td>
<td>8.2</td>
<td>c.505A&gt;G</td>
<td>p.Ser169Gly</td>
<td>2.49</td>
<td>22</td>
<td>0.28</td>
</tr>
<tr>
<td>267</td>
<td>Aunt</td>
<td>Obese</td>
<td>female</td>
<td>EUR</td>
<td>Adult</td>
<td>40-49</td>
<td>50.3</td>
<td>NA</td>
<td>c.537C&gt;T</td>
<td>p.Thr179Met</td>
<td>4.31</td>
<td>28.6</td>
<td>0.91</td>
</tr>
<tr>
<td>24</td>
<td>Proband</td>
<td>Obese</td>
<td>female</td>
<td>EUR</td>
<td>Child</td>
<td>10-19</td>
<td>38.0</td>
<td>4.1</td>
<td>c.551G&gt;A</td>
<td>p.Arg184Gln</td>
<td>4.26</td>
<td>28.1</td>
<td>0.66</td>
</tr>
<tr>
<td>662</td>
<td>Sibling</td>
<td>Obese</td>
<td>female</td>
<td>EUR</td>
<td>Child</td>
<td>10-19</td>
<td>46.9</td>
<td>6.3</td>
<td>c.551G&gt;A</td>
<td>p.Arg184Gln</td>
<td>4.26</td>
<td>28.1</td>
<td>0.66</td>
</tr>
<tr>
<td>662</td>
<td>Proband</td>
<td>Obese</td>
<td>male</td>
<td>EUR</td>
<td>Child</td>
<td>10-19</td>
<td>45.0</td>
<td>7.1</td>
<td>c.551G&gt;A</td>
<td>p.Arg184Gln</td>
<td>4.26</td>
<td>28.1</td>
<td>0.66</td>
</tr>
<tr>
<td>217</td>
<td>Proband</td>
<td>Obese</td>
<td>female</td>
<td>EUR</td>
<td>Adult</td>
<td>40-49</td>
<td>55.3</td>
<td>NA</td>
<td>c.614C&gt;T</td>
<td>p.Glu205Val</td>
<td>4.11</td>
<td>27.1</td>
<td>0.97</td>
</tr>
<tr>
<td>39_1</td>
<td>Proband</td>
<td>Obese</td>
<td>female</td>
<td>EUR</td>
<td>Child</td>
<td>10-19</td>
<td>37.3</td>
<td>3.9</td>
<td>c.638G&gt;A</td>
<td>p.Arg213His</td>
<td>4.17</td>
<td>27.5</td>
<td>0.93</td>
</tr>
<tr>
<td>374</td>
<td>Mother</td>
<td>Obese</td>
<td>female</td>
<td>EUR</td>
<td>Adult</td>
<td>20-29</td>
<td>47.3</td>
<td>NA</td>
<td>c.638G&gt;A</td>
<td>p.Arg213His</td>
<td>4.17</td>
<td>27.5</td>
<td>0.93</td>
</tr>
<tr>
<td>374</td>
<td>Proband</td>
<td>Obese</td>
<td>male</td>
<td>EUR</td>
<td>Child</td>
<td>0-9</td>
<td>44.8</td>
<td>13.4</td>
<td>c.638G&gt;A</td>
<td>p.Arg213His</td>
<td>4.17</td>
<td>27.5</td>
<td>0.93</td>
</tr>
<tr>
<td>159</td>
<td>3rd degree relative</td>
<td>Obese</td>
<td>female</td>
<td>EUR</td>
<td>Child</td>
<td>10-19</td>
<td>47.9</td>
<td>6.6</td>
<td>c.638G&gt;A</td>
<td>p.Arg213His</td>
<td>4.17</td>
<td>27.5</td>
<td>0.93</td>
</tr>
</tbody>
</table>
Figure 1. Summary of workflow

Workflow summary. Only rare (MAF < 10^{-4}) variants are filtered. For the Columbia European cohort, all rare variants are put through both a single variant test and a gene burden test. For the UKBB variants, all rare variants are tested using the regenie linear regression burden test. Genes with 19 different Dmis definitions are compared in a Meta analysis that integrates both datasets. Replication in the three genes that reached significance were attempted in the All of Us dataset, with BSN replicating.

9 Dmis definitions: REVEL=[0.15,0.25, 0.35,0.45,0.55,0.65,0.75,0.85,0.95], total of 19 test groups
Figure 2

BMI density distribution. a) The BMI density distribution for pLoF BSN heterozygotes is shifted to a higher BMI than the overall UKBB cohort. A.) There is a bi-modal distribution for the BSN pLoF heterozygotes. The distribution difference between overall cohort and UKBB heterozygotes was tested using the Kolmogorov-Smirnov method (p=1.4e-05). The dots in the blue curve represent the BSN predicted deleterious variants samples’ BMI.

b) Phenotype of BSN pLoF heterozygotes in the Columbia cohort.
Figure 3.

Lollipop plot for \textit{MEOX1}. The upper panel shows BMI-Z distribution for rare deleterious variants (population frequency $< 10^{-4}$ and revel score $\geq 0.15$) in the Columbia cohort. BMI-Z for adult samples was the normalized BMI score using UKBB mean and standard deviation and BMI-Z for the child was the raw BMI-Z score. The lower panel shows the singleton deleterious variants in the UKBB. The BMI distribution difference between \textit{MEOX1} singleton deleterious variants carriers with the overall UKBB cohort is 0.03.
Figure 4  *MEOX1* segregation

Columbia *MEOX1* heterozygous pedigrees. Shaded symbol represents obese person, red arrow indicates the proband in the family. BMI or BMIZ are indicated under the symbols.


