

ARTICLE OPEN



Rare predicted loss of function alleles in Bassoon (BSN) are associated with obesity

Na Zhu^{1,2,10}, Charles A. LeDuc^{1,3,4,10}, Ilene Fennoy^{1,4}, Blandine Laferrère^{3,4,5}, Claudia A. Doege^{3,6}, Yufeng Shen^{1,2,7,8}, Wendy K. Chung^{1,3,4,5,9} and Rudolph L. Leibel^{1,3,4}

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 p-value of 3.6e-12 in the UK biobank cohort. Additionally, we 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 and report the clinical histories of these individuals with non-syndromic obesity with no history of neurobehavioral or cognitive disability. The BMI association was replicated in the All of Us whole genome sequencing data. Heterozygous pLoF *BSN* variants constitute a new etiology for obesity.

npj Genomic Medicine (2023)8:33; <https://doi.org/10.1038/s41525-023-00376-7>

INTRODUCTION

By 2030 it is estimated that roughly 50% of adults in the United States will be obese, with 25% having severe obesity¹. The prevalence of obesity in U.S. adults has increased from 30.5 to 41.9% from 1999 to 2023; the prevalence of severe obesity has increased from 4.7 to 9.2%. Approximately 18% of U.S. children currently are obese². The estimation from twin and GWAS data is that the risk of obesity is 30–50% heritable^{3–6}. Changes in underlying genetics cannot be responsible for large increases 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 genetic⁷. 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 ancestries¹¹. However, the known genome-wide significant loci only explain ~6% of variation in BMI^{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 disease¹⁶, 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 obesity¹⁹ and hyperlipidemia²⁰ 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 large population genomic data from the UKBB and All of Us and an extreme obesity cohort recruited at Columbia University. We report the association of predicted loss of function (pLoF) alleles in the gene *BSN* with body mass index (BMI).

RESULTS

Cohorts and overview of analysis workflow

We obtained data from three cohorts to identify new obesity risk genes: 1) the UK Biobank (UKBB) study^{21,22} (interim 200k release, Table 1) with exome sequence data and basic phenotype information. We excluded related individuals and individuals with a history of cancer or eating disorders and limited analysis to the 144,496 individuals of European ancestry by principal component analysis. 2) a cohort recruited at Columbia University (the Columbia cohort- summarized in Table 2) enriched in early onset and extreme obesity²³ with 1598 probands with exome sequencing data, and 3) the All of Us dataset of about 50,000 European ancestry individuals with both BMI and whole genome sequencing data.

We used the UKBB data as the discovery cohort for association of rare variants and BMI, and then sought additional support in the Columbia cohort and replication in the All of Us data (Fig. 1).

Association of rare variants with BMI in the discovery cohort

First, we performed an exome-wide scan of risk genes through a linear regression of BMI over counts of rare damaging variants using the UKBB data. For each gene, we tested seven different ways of combining rare (allele frequency < 1e-4 in the population) variants, including protein truncating variants (inclusion or not), and predicted

¹Department of Pediatrics, Columbia University Irving Medical Center, New York, NY, USA. ²Department of Systems Biology, Columbia University Irving Medical Center, New York, NY, USA. ³NY Obesity Research Center, Columbia University Irving Medical Center, New York, NY, USA. ⁴Naomi Berrie Diabetes Center, Columbia University Irving Medical Center, New York, NY, USA. ⁵Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA. ⁶Department of Pathology, Columbia University Irving Medical Center, New York, NY 10032, USA. ⁷Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, USA. ⁸JP Sulzberger Columbia Genome Center, Columbia University Irving Medical Center, New York, NY, USA. ⁹Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, NY, USA. ¹⁰These authors contributed equally: Na Zhu, Charles A. LeDuc. ✉email: wkc15@cumc.columbia.edu; rl232@cumc.columbia.edu

Table 1. UKBB cohort summary.

| | |
|----------------------------------|-----------------|
| BMI (mean, sd) kg/m ² | 27.3, 4.8 |
| age (mean, sd) years | 56.5, 8.1 |
| Female: male | 110,476: 90,153 |
| With cancer | 20,099 |
| With eating disorder | 160 |
| European ancestries | 1,67,246 |
| Non-European ancestry | 33,383 |
| Related samples to be removed | 5596 |

Correlation between age and BMI $p = 0.048$.

Correlation between age and Sex $p = 0.082$.

The UKBB cohort used in the analysis. Samples that were coded as having had 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. F is female. M is male. EUR indicates European ancestry.

Table 2. Patients in Columbia cohort.

| | Child-onset | Adult-onset |
|----------------------------------|-------------|-------------|
| BMI (mean, sd) kg/m ² | 41.4, 12.4 | 45.9, 11.9 |
| BMIz (mean, sd) | 6.6, 3.6 | |
| age (mean, sd) years | 12.2, 3.5 | 38.9, 12.3 |
| Female: male | 523:377 | 503:195 |
| European ancestry | 322 | 528 |
| Non-European ancestry | 352 | 170 |
| Total | 674 | 698 |

Clinical characteristics of the Columbia extreme or early-onset obesity cohort. BMI is body mass index. BMIz is calculated for all children as the BMI z score for gender and age.

damaging missense variants by REVEL at variable thresholds. We carried out the association test using REGENIE²⁴, with age, Townsend deprivation index at recruitment, smoking /alcohol status, sex, the first eight principal components, and genetic heterozygosity as covariates (Q-Q plot for UKBB in Supplementary Figure 1). We identified two genes associated with BMI with exome-wide significance (Table 3, one is *MC4R* ($\beta = 1.4$, p -value = 5.0×10^{-10} , genomic position for *MC4R* variants in Supplementary Fig. 2, BMI distribution in Supplementary figure 3), a known obesity gene; the other is *BSN* ($\beta = 6.2$, p value = 3.6×10^{-12}), a novel putative obesity risk gene (individual variants listed in Supplementary Tables 1 (*MC4R*) and 2 (*BSN*)).

The association of *BSN* with BMI is primarily driven by protein-truncating variants and damaging missense variants with REVEL > 0.75. Twenty-seven individuals have heterozygous genotypes for one of these variants (aggregated allele frequency = 9.3×10^{-5}). Figure 2a shows the BMI distribution of *BSN* predicted deleterious heterozygotes compared to the overall UKBB population (Kolmogorov-Smirnov p -value = 1.4×10^{-5}).

BSN pLOF carriers from the Columbia cohort

We identified two heterozygous pLoF *BSN* alleles in the Columbia cohort (Fig. 2b). RU2487 is heterozygous for a de novo p.Gln703X allele in *BSN*. At the time of the last assessment, she was a 23-year-old Latina woman with a history of severe obesity and type 2 diabetes mellitus diagnosed at age 19 years at which time her

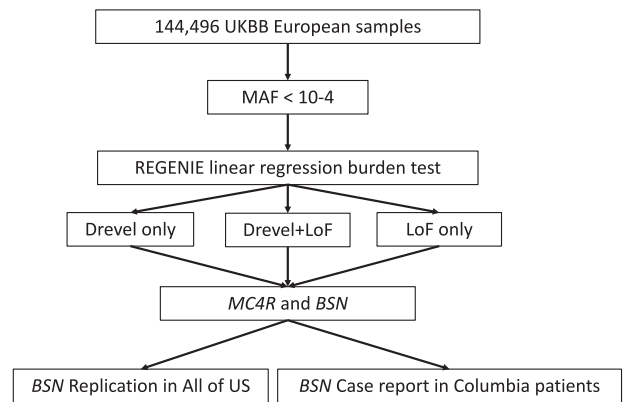


Fig. 1 Overall workflow. Only rare ($MAF < 10^{-4}$) likely damaging variants are kept for association analysis. DreveL refers to likely damaging missense variants with REVEL-predicted score above a certain threshold. In the association analysis of the discovery cohort (UKBB), 7 different combinations of variants were tested using the REGENIE linear regression gene-burden test. *BSN* reached significance and was replicated in the All of Us dataset ($N \sim 50,000$). Two *BSN* LoF heterozygotes were identified in Columbia patients.

HbA1c was 7.4%. She was amenorrheic and had extensive acanthosis nigricans, dyslipidemia, hypothyroidism, and hyperandrogenism. Her maximal weight was 113 kg at age 20. She had gastric bypass surgery for weight loss at age 21. Immediately prior to bariatric surgery, her BMI was 39.7 kg/m^2 . 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 15.7 years old with body weight of 162 kg and height of 160.9 cm ($BMI = 62.6 \text{ kg/m}^2$). Her waist circumference was 158 cm. She was 11 years old at menarche and 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 at 17.5 years. 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/m^2). $HbA1c$ normalized to 5.2%.

Association of BMI-correlated traits in BSN

The association between *BSN* and the traits correlated with BMI tested using REGENIE (Supplementary Table 3) 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 (Supplementary Table 4) using the binomial test. No diagnosis was significantly associated with *BSN* after correction for multiple testing.

Replication analysis using All of Us data

We sought to replicate the association of *BSN* using the All of Us cohort. As of February 2023, there were 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

Table 3. REGENIE linear regression for UKBB only.

| Gene | GeneDescription | MASK | A1FREQ | BETA | SE | CHISQ | p-value |
|--------------|---|-------------------|----------|------|------|-------|----------|
| <i>BSN</i> | bassoon presynaptic cytomatrix protein | LoF+Drevel >=0.75 | 9.34E-05 | 6.21 | 0.89 | 48.29 | 3.60E-12 |
| <i>MC4R</i> | melanocortin 4 receptor | LoF+Drevel >=0.25 | 1.47E-03 | 1.40 | 0.23 | 38.66 | 5.00E-10 |
| <i>PCSK1</i> | proprotein convertase subtilisin/kexin type 1 | LoF+Drevel >=0.75 | 9.34E-04 | 1.42 | 0.28 | 25.17 | 5.20E-07 |
| <i>PHF3</i> | PHD finger protein 3 | LoF+Drevel >=0.75 | 1.56E-04 | 3.26 | 0.69 | 22.18 | 2.48E-06 |
| <i>ALOX5</i> | arachidonate 5-lipoxygenase | Drevel >=0.5 | 1.56E-03 | 1.00 | 0.22 | 20.95 | 4.71E-06 |
| <i>NCR2</i> | natural cytotoxicity triggering receptor 2 | Drevel >=0.25 | 7.96E-05 | 4.29 | 0.97 | 19.63 | 9.41E-06 |

UKBB REGENIE linear regression result.

The BMI association results for UKBB cohort using REGENIE. Drevel refers to likely damaging missense variants with REVEL-predicted score about a certain threshold. Genes with p value < $e-5$ are listed with the smallest p value for each gene listed. The p value significance threshold after Bonferroni correction was $3.57e-7$. *BSN* was a novel gene that reached genome wide significance. *MC4R* and *PCSK1* are known obesity risk genes.

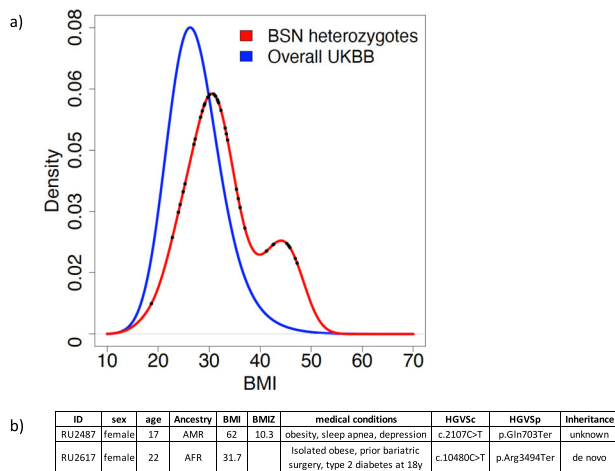


Fig. 2 BMI density distribution. **a** The BMI density distribution for pLoF *BSN* heterozygotes is shifted to a higher BMI than the overall UKBB cohort. **A.**) There is an apparent bi-modal distribution for the *BSN* pLoF heterozygotes. This distribution is not explained by putative protein functional consequences (CADD score), genomic physical location, phylogenetic conservation, epigenetic markers, or histone modification across different stages. The bimodal may be simply due to the small number of carriers and the bin size, if the bin size is increased from 2 to 5, the binomial distribution goes away. The distribution difference between overall cohort and UKBB heterozygotes was tested using the Kolmogorow-Smirnov method ($p = 1.4e-05$). The dots in the red curve represent the BMI of the individuals with predicted deleterious *BSN* variants. **b** Phenotype of *BSN* pLoF heterozygotes in the Columbia cohort.

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) with $\beta = 6.3$, a large effect size similar to the estimation from the UK Biobank cohort. Additionally, we identified 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; Supplementary Table 5). Thus, the *BSN* obesity association observed in the UKBB and Columbia cohorts was replicated in the All of Us cohort.

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, with similarly

large effect size. Additionally, we identified extremely obese individuals in the Columbia cohort of early onset and/or 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.

BSN (bassoon) is expressed primarily in the brain (including embryonic and adult brain regions that impact feeding behavior²⁵), 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 plasticity²⁶.

Bassoon is expressed during early neuronal differentiation, is selectively sorted into axons and is among the first proteins to arrive at nascent synapses²⁶. 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 vertebrates²⁷.

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 synapses²⁸. 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 cortices²⁹. These animals are not obese.

Bassoon is involved in the maintenance of the integrity of AZ³⁰. Glutamatergic synapses from *Bsn* knockout mice exhibit enhanced short-term synaptic depression with a high percentage of silent synapses but have no gross structural defects³¹, presumably due to the significant functional redundancy with Piccolo. When both proteins are absent from glutamatergic synapses, the cells undergo synapse degeneration³².

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

dysfunction³³. This study did not find coding mutations in *BSN* but first identified *BSN* as a new transcript that could be cloned 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 turnover³⁴. Human heterozygous missense variants in *BSN* have been implicated in neurodevelopmental and neurodegenerative disorders including progressive supranuclear palsy-like syndrome, a rare neurodegenerative tauopathy³⁵.

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 circuits³⁶. 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^{31,37–39}. The valence of these effects is consistent with hyperphagic obesity conveyed by hypomorphic alleles.

METHODS

We ran REGENIE for rare variants association in UKBB data to detect the risk genes associated with BMI, and then sought additional support in the Columbia cohort and replication in the All of Us data (Fig. 1). Informed consent was obtained from all human participants.

UKBB cohort

For this analysis, we included 200,643 individuals from the UK biobank²². 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 1).

Columbia cohort

The Columbia University early onset and/or 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^{23,40}. Approximately half of the probands were pediatric (either at time of 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 2). 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.

All of Us data

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-MEM⁴¹. Picard v1.93 MarkDuplicates (<http://broadinstitute.github.io/picard/>) was used to identify

and flag PCR duplicates and GATK v4.1 HaplotypeCaller⁴² 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 Peddy⁴³. Relatedness prediction in the UKBB samples, due to the large sample size, was completed with plink King⁴⁴. To ensure that private mutations carried in individual families were not over-counted, samples with a second-degree relationship or closer (a kinship coefficient greater than 0.12 in King or 0.25 in Peddy) had the relative who was more related to the overall cohort excluded.

Variant annotation: We used the Ensembl Variant Effect Predictor (VEP, Ensembl 93)⁴⁵ to annotate variant function and ANNOVAR⁴⁶ 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 WGS⁴⁷. Deleterious variants were defined as predicted loss-of-function (pLoF: including premature stop-gain, stop-loss, frameshift indels, canonical splicing variants and multi-exon deletions) or predicted damaging missense (Dmis) based on REVEL⁴⁸ score thresholds. The same annotation pipeline was used for Columbia, UKBB, and All of Us variant annotation.

Statistical analysis

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^{21,47}. We collapsed rare variants based on allele frequency and predicted variant deleteriousness. The variants were partitioned into cohort frequency < 10⁻⁴ as well as 7 variant functional groups. The variant functional groups were missense variants with REVEL ≥ x, where x is 0.25, 0.5, 0.75, with or without pLoF variants (7 combinations). Genes having less than 15 heterozygotes in a test group were removed. The significance threshold was set to 3.5e-07 (= 0.05 / (7*20,000)). We then tested the quantitative BMI for the 144 K UKBB individuals using REGENIE²⁴, 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 Fig. 1a). 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 Fig. 1b).

Columbia cohort: When there were multiple individuals with obesity in a family, the most severely affected was defined as the proband (either the child with the highest Z-score; or the adult with the highest BMI if there were only adults in the family).

We defined the threshold for genome-wide significance by Bonferroni correction for multiple testing ($n = 20,000 * 7$, threshold p -value = 3.57e-7) (workflow shown in Fig. 2).

All of Us: To attempt to replicate findings from the UKBB analysis and Columbia Cohort, we ran 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* deleterious variants using age, sex, deprivation index and median income as covariates.

All analyses were under the auspices of the Columbia University IRB “Molecular Genetic Analysis of Obesity and Non-Insulin Dependent Diabetes Mellitus” IRB #: AAAA4485. All participants provided written or electronic informed consent to take part in their respective study (Columbia, UK Biobank, or All of Us).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

UKBB genotypic and phenotypic data are available to approved investigators via the UK Biobank study (www.ukbiobank.ac.uk/). Additional information about registration for access to the data is available at www.ukbiobank.ac.uk/register-apply/. Data access for approved applications requires a data transfer agreement between the researcher's institution and UK Biobank, the terms of which are available on the UK Biobank website (www.ukbiobank.ac.uk/media/ezrderzw/applicant-mta.pdf). Original All of Us Biobank data are available to registered and approved All of Us researchers (<https://www.researchallofus.org/register/>). Genetic data requires controlled tier access, which researchers can register for through their institutions.

CODE AVAILABILITY

We used REGENIE for statistical analysis. The software is written by a team in Regeneron, available on GitHub: <https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fgithub.com%2Fregenie%2F&data=05%7C01%7Cnz2274%40cumc.columbia.edu%7Cb05e1e0311ea4ad5915308db8868e3e7%7Cb0002a9b0017404d97dc3d3b3ab09be81%7C0%7C0%7C638253756299911043%7CUnknown%7CTWFpbGZsb3d8eyJWljiMC4wLjAwMDA1LCJQIjo1V2luMzIiLCJBTiI6Ikl1haWw1CjXVCl6Mn0%3D%7C3000%7C%7C%7C&sdata=IOOVWwXc9Qpsd2pqK9p9k1oRRMPWoDnbjYdYhtFYUw%3D&reserved=0>.

Received: 24 March 2023; Accepted: 2 October 2023;

Published online: 21 October 2023

REFERENCES

- Ward, Z. J. et al. Projected U.S. state-level prevalence of adult obesity and severe obesity. *N. Engl. J. Med.* **381**, 2440–2450 (2019).
- National Health and Nutrition Examination Survey 2017–March 2020 Pre-pandemic Data Files Development of Files and Prevalence Estimates for Selected Health Outcomes. in *National Health Statistics Reports* (ed. National Center for Health, S.) (<https://doi.org/10.15620/cdc:106273>, Hyattsville, MD, 2021).
- Llewellyn, C. H., Trzaskowski, M., Plomin, R. & Wardle, J. Finding the missing heritability in pediatric obesity: the contribution of genome-wide complex trait analysis. *Int. J. Obes.* **37**, 1506–1509 (2013).
- Loos, R. J. F. & Janssens, A. C. J. W. Predicting polygenic obesity using genetic information. *Cell Metab.* **25**, 535–543 (2017).
- Luke, A. et al. Heritability of obesity-related traits among Nigerians, Jamaicans and US black people. *Int. J. Obes.* **25**, 1034–1041 (2001).
- Maes, H. H. M., Neale, M. C. & Eaves, L. J. Genetic and environmental factors in relative body weight and human adiposity. *Behav. Genet.* **27**, 325–351 (1997).
- Albuquerque, D., Nobrega, C., Manco, L. & Padez, C. The contribution of genetics and environment to obesity. *Br. Med. Bull.* **123**, 159–173 (2017).
- Wang, K. et al. A genome-wide association study on obesity and obesity-related traits. *PLoS One* **6**, e18939 (2011).
- Speakman, J. R., Loos, R. J. F., O'Rahilly, S., Hirschhorn, J. N. & Allison, D. B. GWAS for BMI: a treasure trove of fundamental insights into the genetic basis of obesity. *Int. J. Obes. (Lond.)* **42**, 1524–1531 (2018).
- Scuteri, A. et al. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet* **3**, e115 (2007).
- Khera, A. V. et al. Polygenic prediction of weight and obesity trajectories from birth to adulthood. *Cell* **177**, 587–596.e9 (2019).
- Loos, R. J. F. & Yeo, G. S. H. The genetics of obesity: from discovery to biology. *Nat. Rev. Genet* **23**, 120–133 (2022).
- Loos, R. J. The genetics of adiposity. *Curr. Opin. Genet. Dev.* **50**, 86–95 (2018).
- Marouli, E. et al. Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186–190 (2017).
- Lanktree, M. B. et al. Meta-analysis of dense gene-centric association studies reveals common and uncommon variants associated with height. *Am. J. Hum. Genet.* **88**, 6–18 (2011).
- Trynka, G. et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat. Genet.* **43**, 1193–U45 (2011).
- Stitzel, N. O. et al. Exome sequencing in suspected monogenic dyslipidemias. *Circ-Cardiovasc. Genet.* **8**, 343–50 (2015).

- Goodrich, J. K. et al. Determinants of penetrance and variable expressivity in monogenic metabolic conditions across 77,184 exomes. *Nat. Commun.* **12**, 3505 (2021).
- Wilding, J. P. H., Calanna, S. & Kushner, R. F. Once-weekly semaglutide in adults with overweight or obesity. Reply. *N. Engl. J. Med.* **385**, e4 (2021).
- Musunuru, K. et al. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N. Engl. J. Med.* **363**, 2220–7 (2010).
- Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
- Backman, J. D. et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021).
- Gill, R. et al. Whole-exome sequencing identifies novel LEPR mutations in individuals with severe early onset obesity. *Obesity* **22**, 576–584 (2014).
- Mbatchou, J. et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
- De Rosa, M. C. et al. Gene expression atlas of energy balance brain regions. *Jci Insight* **6**, e149137 (2021).
- Zhai, R. et al. Temporal appearance of the presynaptic cytomatrix protein bassoon during synaptogenesis. *Mol. Cell. Neurosci.* **15**, 417–428 (2000).
- Schoch, S. & Gundelfinger, E. D. Molecular organization of the presynaptic active zone. *Cell Tissue Res.* **326**, 379–391 (2006).
- Altrock, W. D. et al. Functional inactivation of a fraction of excitatory synapses in mice deficient for the active zone protein bassoon. *Neuron* **37**, 787–800 (2003).
- Angenstein, F. et al. Manganese-enhanced MRI reveals structural and functional changes in the cortex of bassoon mutant mice. *Cereb. Cortex* **17**, 28–36 (2007).
- Gundelfinger, E. D., Reissner, C. & Garner, C. C. Role of bassoon and piccolo in assembly and molecular organization of the active zone. *Front. Synaptic Neurosci.* **7**, 19 (2015).
- Hallermaier, S. et al. Bassoon speeds vesicle reloading at a central excitatory synapse. *Neuron* **68**, 710–723 (2010).
- Waites, C. L. et al. Bassoon and Piccolo maintain synapse integrity by regulating protein ubiquitination and degradation. *Embo J.* **32**, 954–969 (2013).
- Hashida, H. et al. Cloning and mapping of ZNF231, a novel brain-specific gene encoding neuronal double zinc finger protein whose expression is enhanced in a neurodegenerative disorder, multiple system atrophy (MSA). *Genomics* **54**, 50–58 (1998).
- Montenegro-Venegas, C., Annamneedi, A., Hoffmann-Conaway, S., Gundelfinger, E. D. & Garner, C. C. BSN (bassoon) and PRKN/parkin in concert control presynaptic vesicle autophagy. *Autophagy* **16**, 1732–1733 (2020).
- Yabe, I. et al. Mutations in bassoon in individuals with familial and sporadic progressive supranuclear palsy-like syndrome. *Sci. Rep.* **8**, 819 (2018).
- Zheng, H. & Berthoud, H. R. Neural systems controlling the drive to eat: mind versus metabolism. *Physiol. (Bethesda)* **23**, 75–83 (2008).
- Shah, B. P. et al. MC4R-expressing glutamatergic neurons in the paraventricular hypothalamus regulate feeding and are synaptically connected to the parabrachial nucleus. *Proc. Natl Acad. Sci. USA* **111**, 13193–13198 (2014).
- Fenselau, H. et al. A rapidly acting glutamatergic ARC → PVH satiety circuit postsynaptically regulated by alpha-MSH. *Nat. Neurosci.* **20**, 42–51 (2017).
- Claffin, K. E. et al. Pharmacological FGF21 signals to glutamatergic neurons to enhance leptin action and lower body weight during obesity. *Mol. Metab.* **64**, 101564 (2022).
- Li, P. et al. Genetic association analysis of 30 genes related to obesity in a European American population. *Int. J. Obes. (Lond.)* **38**, 724–9 (2014).
- Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851–8 (2008).
- DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–8 (2011).
- Pedersen, B. S. & Quinlan, A. R. Who's who? Detecting and resolving sample anomalies in human DNA sequencing studies with peddy. *Am. J. Hum. Genet* **100**, 406–413 (2017).
- Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
- McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
- Wang, K., Li, M. Y. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
- Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
- Ioannidis, N. M. et al. REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet* **99**, 877–885 (2016).

ACKNOWLEDGEMENTS

We thank the participants who generously contributed to this work and their clinicians who referred them. This work was supported by NIH grant RO1DK52431 and the NY Nutrition and Obesity Research Center: P30DK26685.

AUTHOR CONTRIBUTIONS

NZ did the computational work under the supervision of YS, WKC and RLL. NZ and CAL wrote the manuscript which was edited by BL, CAD, TS, WKC and RLL. Sequence confirmation was done by CAL. WKC and RLL collected the Columbia cohort. IF and BL clinically examined the patients. Funding was obtained by WKC and RLL.

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/s41525-023-00376-7>.

Correspondence and requests for materials should be addressed to Wendy K. Chung or Rudolph L. Leibel.

Reprints and permission information is available at <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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023