Received: 25 July 2018 | Accepted: 25 February 2019

DOI: 10.1111/ijpo.12530

ORIGINAL RESEARCH

WILEY pediatricobesity

Genetic risk scores demonstrate the cumulative association of single nucleotide polymorphisms in gut microbiome-related genes with obesity phenotypes in preschool age children

Anthony A. Wang[1] | Kristen Harrison[2] | Salma Musaad[3] | Sharon M. Donovan[1,4] | Margarita Teran-Garcia[1,3] | the STRONG Kids Research Team

[1] Division of Nutritional Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois

[2] Institute for Social Research, University of Michigan, Ann Arbor, Michigan

[3] Department of Human Development and Family Studies, University of Illinois at Urbana-Champaign, Urbana, Illinois

[4] Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, Urbana, Illinois

Correspondence
Margarita Teran-Garcia, MD, PhD, FTOS, Division of Nutritional Sciences, Department of Human Development and Family Studies, University of Illinois Urbana-Champaign, 904 W. Nevada St. 2005 Christopher Hall Urbana, Illinois 61801, USA.
Email: teranmd@illinois.edu

Funding information
National Institute of Food and Agriculture, Grant/Award Numbers: 2011-67001-30101, 2013-67011-21024 and Hatch 793-328; Agriculture and Food Research Initiative, Grant/Award Number: 2013-67011-21024; National Institute for Agriculture under the Illinois Transdisciplinary Obesity Prevention Program, Grant/Award Number: 2011-67001-30101; United States Department of Agriculture, Grant/Award Number: Hatch 793-328

Summary

Background: Childhood obesity is a nutrition-related disease with multiple underlying aetiologies. While genetic factors contribute to obesity, the gut microbiome is also implicated through fermentation of nondigestible polysaccharides to short-chain fatty acids (SCFA), which provide some energy to the host and are postulated to act as signalling molecules to affect expression of gut hormones.

Objective: To study the cumulative association of causal, regulatory, and tagged single nucleotide polymorphisms (SNPs) within genes involved in SCFA recognition and metabolism with obesity.

Design: Study participants were non-Hispanic White (NHW, n = 270) and non-Hispanic Black (NHB, n = 113) children (2-5 years) from the Synergistic Theory and Research on Obesity and Nutrition Group (STRONG) Kids 1 Study. SNP variables were assigned values according to the additive, dominant, or recessive inheritance models. Weighted genetic risk scores (GRS) were constructed by multiplying the reassigned values by independently generated β-coefficients or by summing the β-coefficients. Ethnicity-specific SNPs were selected for inclusion in GRS by cohort.

Results: GRS were directly associated with body mass index (BMI) z-score. The models explained 3.75%, 12.9%, and 26.7% of the variance for NHW/NHB, NHW, and NHB (β = 0.89 [CI: 0.43-1.35], P = 0.0002; β = 0.78 [CI: 0.54-1.03], $P < 0.0001$; β = 0.74 [CI: 0.51–0.97], $P < 0.0001$).

Conclusion: This analysis supports the cumulative association of several candidate genetic variants selected for their role in SCFA signalling, transport, and metabolism with early-onset obesity. These data strengthen the concept that microbiome influences obesity development through host genes interacting with SCFA.

KEYWORDS

childhood obesity, genetic risk score, gut microbiome

 © 2019 World Obesity Federation

1 | INTRODUCTION

Childhood obesity is a nutrition-related disease with multiple underlying aetiologies. Genetics play a significant role in the development of obesity, and twin studies indicate that the heritability of obesity ranges from 40% to 70%.[1] Meta-regression analyses have provided evidence that genetic influences on obesity are greater during childhood than in adulthood.[2] Genome-wide association studies (GWAS) have revealed several hundred loci within the human genome associated with obesity in adults, although only eight independent loci have been identified in children.[3,4] Building upon GWAS findings, the genetic risk score (GRS) approach has been employed to address concerns in statistical analyses over multiple testing and to account for the missing heritability of obesity.[5] GRS constructed from GWAS-identified single nucleotide polymorphisms (SNPs) and tested across prospective and retrospective cohorts support the notion that the genetic effect on obesity-related phenotypes is age-dependent with increasing influence peaking near the age of 20 (see supplementing information for additional references).[6-10] Cross-sectional studies in children also demonstrate a relationship between GRS and measures of obesity (see supplementing information for additional references).[11,12] However, the percent of variance of body mass index (BMI) and other obesity-related phenotypes explained by the GRS in studies in children remains small ranging from 1.0% to 3.4%. In fact, Le Chatelier suggested that the current obesity susceptibility genetic variants identified from GWAS are less informative in distinguishing between individuals who are lean and those who have obesity than their microbiome within the gastrointestinal tract.[13]

One possible explanation of how the gut microbiome contributes to obesity development includes the interaction between host genetics and short-chain fatty acids (SCFAs), the products of bacterial fermentation of nondigestible polysaccharides.[14] In particular, butyrate promotes overall gut health while acetate and propionate may increase host capacity for energy harvest and storage through lipogenesis and gluconeogenesis.[15,16] Through a review of the literature, we identified key genes involved in the transport and signalling of SCFAs. From these genes, four main pathways were established: (1) SCFA transport across the gut epithelium via monocarboxylate transporter 1 (*SLC16A1*) and sodium-coupled monocarboxylate transporter 1 (*SLC5A8*) (apical membrane) and via monocarboxylate transporters 4 and 5 (*SLC16A3* and *SLC16A4*) (basolateral membrane.[17,18] (2) SCFA signalling through free fatty acid receptors 2 and 3 (*FFAR2* and *3*) bind SCFAs and effect downstream regulation of appetite through downstream effectors including peptide YY (*PYY*) and glucagon-like peptide 1 (*GCG*) in enteroendocrine L cells, and leptin (*LEP*) in adipocytes.[19,20] (3) Alterations in adipose storage through angiopoietin-like 4 (*ANGPTL4*), an inhibitor of lipoprotein lipase (*LPL*).[21] Both *ANGPTL4* and *LPL* are under the transcriptional regulation of peroxisome proliferator-activated receptor gamma (*PPARG*). (4) Immunological response to lipopolysaccharide (LPS) via toll-like receptor 4 (*TLR4*) and the inflammatory response by the expression of nuclear factor kappa beta (*NFKB*) leading to the release of cytokines interleukin 6 (*IL6*) and tumour necrosis factor alpha (*TNFA*).[22]

SCFA receptors and transporters and other host responders to gut microbiome have been described, but the collective impact of common functional and regulatory variants in these genes on obesity-related phenotypes has not been studied in humans. The current study assessed the cumulative association of causal, regulatory, and tagged SNP variants within genes involved in gut microbiome and/or SCFA recognition and metabolism on obesity-related phenotypes in preschool-age children. SNP-SNP interactions within pathways were also examined. We hypothesized that genetic variation in SCFA recognition pathways would be positively associated with obesity-related phenotypes. GRS were constructed using both traditional (additive model) and nontraditional (nonadditive models) methods taking into consideration the magnitude and directionality of the effect size of each SNP on the phenotype by weighting the score.

2 | METHODS

2.1 | Study participants and anthropometric measurements

Participants for this study were preschool age children (2-5 years) combined from the Synergistic Theory and Research on Obesity and Nutrition Group (STRONG) Kids 1 Study (n = 475) cohorts from the University of Illinois (Urbana, IL; n = 265) and the University of Michigan (Ann Arbor, MI; n = 210).[23] The study protocol received approval from the Institutional Review Boards at both recruitment sites. Data regarding age, sex and ethnicity were collected from a large panel survey completed by study participant parents. Height and weight were measured to calculate BMI and related measures using a stadiometer (Peslter, USA) and electronic remote display scale (Jarden Consumer Solutions, USA) with a precision level of 0.1 cm and 0.1 kg, respectively. BMI, BMI percentile (BMIPCT), BMI z-score for age (BMIZ), weight for age z-score (WAZ), and height for age z-score (HAZ) were calculated using the standard SAS program from the Center for Disease Control and Prevention (CDC). Z-scores express the standard deviation from the mean to indicate a child's weight, height, and BMI status according to the sex and age-specific CDC growth charts from 2000.[24] Children with or without overweight were defined as having a BMIPCT greater than or equal to 85th percentile or BMIPCT less than 85th percentile, respectively. Children with BMIZ above or below four standard deviations and those with known metabolic disorders were excluded (n = 5). Only non-Hispanic White (NHW) and non-Hispanic Black (NHB) study participants were included in the following groups: NHW and NHB combined (NHW/NHB, n = 383), NHW (n = 270), and NHB (n = 113).

2.2 | Candidate gene and SNP selection

Candidate genes were identified after examination of the published literature regarding each gene's known associations with gut microbiome-related molecules including SCFAs and LPS (*FFAR2*, *FFAR3*, *ANGPTL4*, *CD36*, *SLC16A1*, *SLC16A3*, *SLC16A4*, *SLC5A8*, and

TLR4). Many of these genes have been described to be important in the recognition of gut microbiome and SCFAs.[14,25] Downstream effector genes were also included to identify potential gene-gene interactions and to test the cumulative association of carrying risk alleles in the development of obesity (*LPL*, *PYY*, *GCG*, *LEP*, *LEPR*, *NPY*, *NPY2R*, *PPARG*, *NFKB*, *IL6*, and *TNFA*).[20,22,26-33]

The SNP selection was performed using a systematic approach. SNPs within or near the genes of interest were selected for inclusion primarily for their functional or regulatory potential. The Single Nucleotide Polymorphism Database (dbSNP) and Ensembl databases were searched for SNPs located in high priority regions including the 5′ and 3′ untranslated regions (UTRs), exons (synonymous and non-synonymous variants), and the 10 base pair (bp) region within exon-intron boundaries.[34] Several in silico tools were utilized to further assess the likelihood that the SNP would impact protein function or gene regulation. Less common variants (minor allele frequency [MAF] < 10%) and, in particular, non-synonymous SNPs were included with consideration of their Sorting Tolerant from Intolerant (SIFT) and PolyPhen scores.[35,36] RegulomeDB and miRdSNP were used to identify SNPs likely to affect transcription factor and microRNA binding respectively.[37,38] In the SNP selection process, tag SNPs for the genes of interest were also identified using Haploview version 4.2 (Cambridge, MA) when functional or regulatory SNPs were not available.[39] Table S1 provides a summary of the 52 candidate SNPs.

2.3 | DNA extraction and genotyping

Genomic DNA (gDNA) was extracted and purified following the Oragene-DNA protocol for the manual purification of DNA from saliva (average yield = 9.8 μg, average OD 260/280 ratio = 1.9). Selected markers (52 total SNPs) were genotyped using either the Fluidigm SNP genotyping platform or TaqMan genotyping assays. For the Fluidigm protocol, the assay design was constructed on the Fluidigm D3 website. The Functional Genomic Unit of the W.M. Keck Center at the University of Illinois performed preamplification and genotyping using 250 ng of gDNA. Genotypes were called using Fluidigm Genotyping Analysis version 4.1.2 (San Francisco, CA, USA) at a minimum of 85% reliability. The TaqMan procedure was performed in the 7900 Real-Time machine using assays predesigned for *FFAR1*-rs10423648 and *FFAR1*-rs10422744 and a custom assay for *FFAR3*-rs424241. Fluorescent signals were detected for VIC and FAM after PCR, and genotypes were assigned using the allelic discrimination program in the sequence detection systems (SDS) 2.4 software (Applied Biosystems, Carlsbad, CA, USA).

2.4 | Statistical analysis

The MAFs, linkage disequilibrium (LD), and Hardy–Weinberg equilibrium (HWE) were calculated using the SNP & Variation Suite (SVS) software version 8 (Golden Helix, Bozeman, MT). MAFs were calculated for NHW/NHB, NHW, and NHB. D' and r^2 statistics were computed for SNPs located within the same chromosome for the NHW and NHB cohorts respectively using the composite haplotype method (CHM) in SVS. For genomic regions with multiple SNPs of interest in LD, the SNP with the strongest association with the phenotype was kept for GRS construction. χ^2 tests were used in the NHW and NHB cohorts to identify significant departures from HWE. SNPs were excluded from the analyses if the genotype call rate was less than 95%, HWE *P* value was less than 0.05, or the MAF was less than 0.05. The total number of SNPs analysed for each cohort were as follows: 15 SNPs in NHW/NHB, 47 SNPs in NHW, and 38 SNPs in NHB. SNPs were further selected for the GRS based on a *P* value < 0.3 for the association with BMIZ using the general linear model select (GLMSELECT) procedure. Although less conservative, this threshold allows for the potential for SNP combinations to have synergistic effects within the GRS. Eight total models were used where SNPs were grouped by pathway, and dominant and recessive models were tested. GRS were constructed for the NHW/NHB, NHW, and NHB using ethnicity-specific SNPs for the outcome, BMIZ, with 5, 10, and 9 SNPs selected for each cohort respectively. SNPs included for GRS in NHW/NHB cohort had similar MAF (Table S1).

A total of five GRS were constructed for each cohort: additive nonweighted (GRS-add-NW), nontraditional nonweighted (GRS-non-NW), additive weighted (GRS-add-W), nontraditional weighted (GRS-non-W), and genotypic weighted (GRS-gen-W). To construct the GRS-add-NW, each genotype group for the selected SNPs was assigned the value 0, 1, or 2 according to the additive model of risk and the values were summed. Construction of the GRS-non-NW was performed by reassigning the genotype groups as 0 or 1 according to the dominant or recessive models of risk and then summing the values. β-coefficients for each SNP were estimated in the general linear model (GLM) and were used to obtain the three weighted risk scores. Assigned values for the GRS-add-NW were multiplied by their β-coefficients and summed to generate the GRS-add-W. For the GRS-non-W, assigned values from the GRS-non-NW were multiplied by their β-coefficients and summed. The GRS-gen-W was constructed without assumption of genetic model; β-coefficients were summed without multiplication of an assigned value (the referent genotype group assigned a value of "0"). The Cochran-Armitage exact test was conducted to predict the goodness of fit to the additive model of inheritance. The β-coefficients used for the weighted scores and the *P* values for the Cochran-Armitage exact test are shown in Table S2. Normality for the GRS variables was assessed by examining skewness and kurtosis. The skewness values were greater than −1 and the kurtosis values were less than 1 for all constructed GRS.

The associations between the BMIZ outcome and each of the five types of GRS were assessed using linear regression. Logistic regression was used to generate receiver operating characteristic (ROC) curves to assess the specificity and sensitivity of each GRS to examine accuracy of each type of GRS in discriminating children with or without overweight within each cohort. Each GRS developed for BMIZ was applied to WAZ and HAZ. Gene-gene interactions were evaluated in each pathway by using SNP-SNP interaction terms in the GLM. All statistical analyses were performed with age and sex as covariates using SAS 9.4 (SAS Institute Inc., Cary, NC). An additional covariate for age (age^2)

was added to the models when age was independently associated with the phenotype. *P* values were considered significant after modification according to Bonferroni correction by dividing 0.05 by the number of pathways tested (4), the number of SNPs included in each respective cohort's GRS (5, 10, and 9), and the number of GRS (5) (NHW/NHB $P < 0.0005$, NHW $P < 0.00025$, NHB $P < 0.00028$). Bonferroni correction was also used for SNP-SNP interaction analyses, and those *P* values are provided in Figure S2. Bootstrapping analyses were conducted at 10, 100, 500, and 1000 replications to re-evaluate the observed associations between the constructed GRS and BMIZ.

Ancestry informative markers (AIMs, n = 64) were obtained from a previous report and used to generate continuous admixture scores to account for ethnicity within the combined NHW/NHB cohort (Table S3).[40] Admixture scores were generated using principal component analysis in SVS with the first three principal component scores for the 64 AIMs included as covariates. Eigenvalues for the principal components generated were 60.2, 10.4, and 10.1.

3 | RESULTS

3.1 | Participant demographics and descriptive data for genetic markers

Descriptive data of the STRONG Kids 1 study is presented in Table 1. The prevalence of children with normal weight, overweight, or obesity in the NHW/NHB cohort was 78.9%, 14.6%, and 6.5%, respectively. No significant differences in age, height, BMI, or z-scores were found between boys and girls. There were no differences in any of the anthropometric measurements or rates of overweight and obesity amongst the NHW and NHB cohorts. MAFs and HWE values for each SNP and LD tables by cohort for the genes of interest are summarized in Tables S1 and S4, respectively.

3.2 | Genetic risk scores

The SNPs selected for the construction of the GRS for each cohort and the rationale for their inclusion are in Table 2. As described earlier, all SNPs included in the GRS demonstrated a direct relationship with BMIZ according to the genetic modes of inheritance listed. Apart from the GRS-add-W in NHW/NHB, all GRS were associated with BMIZ. R^2 values and percent of BMIZ variance explained (BMIZ%) increased as the GRS progressed from additive to nonadditive and from nonweighted to weighted approaches in all three cohorts. GRS-non-W and GRS-gen-W explained the largest BMIZ% and had the highest R^2 values across and within the cohorts. GRS-gen-W explained 3.8%, 12.9%, and 26.7% of the variance in BMIZ in the NHW/NHB, NHW, and NHB cohorts respectively. A summary of the five GRS constructed is provided in Table 3 including β-coefficients and 95% CIs, and comparisons of the nonweighted and weighted scores for NHW and NHB are shown in Figures 1 and 2. Bootstrapping at 10, 100, and 1000 replications of the data set confirmed the associations between the GRS and BMIZ (Table 4). All significant GRS remained significant after Bonferroni correction except GRS-add-NW in NHH/NHB.

In general, analysis using logistic regression demonstrated similar results as observed when using linear regression. However, only the area under the curve (AUC) values for NHW improved when progressing from additive to nonadditive and from nonweighted to weighted approaches (AUC range: 0.57-0.64). The AUC values for the GRS in NHW/NHB and NHB remained relatively the same across GRS regardless of risk score construction method (Table 3 and Figure S1). While the AUC values in the NHB GRS were the highest (AUC range: 0.72-0.78), the NHW/NHB GRS performed the poorest in predicting children with overweight (AUC range 0.55-0.57). The NHW/NHB GRS had the lowest BMIZ% and further analysis into the strength of association of the GRS within NHW and NHB separately revealed a differential relationship. Whereas the GRS for NHW/NHB was associated with BMIZ in NHW, these GRS were not associated with BMIZ in NHB (data not shown).

TABLE 1 Descriptive characteristics of children in the STRONG Kids cohort stratified by sex and ethnicity

Variable	NHW/NHB n = 384	Male n = 188	Female n = 196	*P* value	NHW only n = 270	NHB only n = 114	*P* value
Age, months	47.8 ± 10.6	47.0 ± 10.2	48.6 ± 10.9	0.13	47.8 ± 10.7	47.9 ± 10.2	0.93
Height, cm	102.5 ± 7.6	102.5 ± 7.3	102.5 ± 8.0	0.99	102.2 ± 7.7	103.1 ± 7.4	0.27
Weight, kg	17.0 ± 3.0	17.1 ± 2.5	17.0 ± 3.3	0.66	17.0 ± 2.9	17.2 ± 3.1	0.54
BMI, kg/m^2	16.1 ± 1.4	16.2 ± 1.2	16.0 ± 1.6	0.21	16.2 ± 1.4	16.1 ± 1.6	0.56
BMIPCT, %	59.4 ± 27.1	59.7 ± 26.5	59.0 ± 27.7	0.80	60.6 ± 27.9	56.5 ± 27.5	0.19
BMIZ	0.30 ± 0.95	0.32 ± 0.89	0.29 ± 1.00	0.75	0.34 ± 0.93	0.22 ± 0.97	0.27
HAZ	0.28 ± 1.04	0.25 ± 1.10	0.30 ± 0.98	0.66	0.22 ± 0.93	0.41 ± 1.26	0.14
WAZ	0.33 ± 0.99	0.38 ± 1.02	0.28 ± 0.95	0.34	0.32 ± 0.90	0.37 ± 1.16	0.65
Overweight, %	14.6	17.0	12.2	0.30	22.2	18.4	0.41
Obese, %	6.5	4.8	8.2	0.19	6.1	6.7	0.81

Abbreviations: BMI, body mass index; BMIPCT, BMI percentile; BMIZ, BMI z-score; HAZ, height-for-age z-score; NHB, non-Hispanic Black; NHW, non-Hispanic White; WAZ, Weight-for-age z-score. N = 384, 188 males (49%), 196 females. Data are presented as means ± *SD*. *P* values for continuous variables were generated using Student's *t* test for continuous variables and chi-square for categorical variables.

TABLE 2 SNP inclusion in GRS for BMIZ by cohort

Gene	SNP	Risk Allele	Mode of Inheritance	*P* value	Function
NHW/NHB					
PPARG	rs12639162	A	Dominant	0.0554	Intron variant, tagged SNP
ANGPTL4	rs1044250	C	Recessive	0.1859	Missense variant
IL6	rs1554606	T	Dominant	0.0139	Intron variant, tagged SNP
SLC16A3	rs3176827	T	Dominant	0.0723	Intron variant, possible splice variant
PYY	rs2070592	G	Recessive	0.2038	5' UTR variant, possible splice variant
NHW only					
PPARG	rs12639162	A	Dominant	0.0103	Intron variant, tagged SNP
ANGPTL4	rs1044250	C	Recessive	0.2030	Missense variant
ANGPTL3	rs10889337	G	Dominant	0.0459	Intron variant
LPL	rs13702	A	Recessive	0.0031	3' UTR variant
PYY	rs2014257	A	Dominant	0.1013	TF binding motif
NPY2R	rs1047214	C	Recessive	0.2346	Synonymous variant
SLC5A8	rs7309172	G	Dominant	0.0916	3' UTR variant
SLC16A3	rs3176827	T	Dominant	0.1303	Intron variant, possible splice variant
SLC16A1	rs9429505	G	Dominant	0.0691	3' UTR variant, tagged SNP
IL6	rs1554606	T	Dominant	0.1087	Intron variant, tagged SNP
NHB only					
CD36	rs3173798	C	Dominant	0.0676	Possible splice variant
PYY	rs2070592	G	Recessive	0.0909	5' UTR variant, possible splice variant
NPY2R	rs2880415	A	Dominant	0.0181	Synonymous variant
LEP	rs11761556	C	Dominant	0.2450	Upstream variant, TF binding motif
SLC16A4	rs12062656	G	Dominant	0.0436	Intron variant, tagged SNP
SLC16A3	rs4789698	G	Recessive	0.0328	Downstream variant, tagged SNP
SLC5A8	rs1709189	C	Dominant	0.0898	Missense variant
TLR4	rs4986790	A	Dominant	0.1146	Missense variant
IL6	rs1554606	T	Dominant	0.0275	Intron variant, tagged SNP

Abbreviations: *ANGPTL3*, angiopoietin-like 3; *ANGPTL4*, angiopoietin-like 4; BMIZ, BMI z-score; *CD36*, cluster of differentiation 36; GRS, genetic risk score; *IL6*, interleukin 6; *LPL*, lipoprotein lipase; NHB, non-Hispanic Black; NHW, non-Hispanic White; *NPY2R*, neuropeptide Y receptor Y2; *PPARG*, peroxisome proliferator-activated receptor gamma; *PYY*, peptide YY; *SLC5A8*, solute carrier family 5 member 8; *SLC16A3*, solute carrier family 16 member 3; *SLC16A1*, solute carrier family 16 member 1; SNP, single nucleotide polymorphism; *TLR4*, toll-like receptor 4; TF, transcription factor; UTR, untranslated region. Determination of SNPs to be included in GRS for NHW/NHB was based on strength of association with BMIZ using general linear model (GLM). SNP inclusion for NHW and NHB GRS were based on GLMSELECT procedure of SNPs by pathway using both the dominant and recessive modes of inheritance (eight total models). Set limit entry (SLE) and set limit stay (SLS) were set to 0.3. A total of five, 10, and nine SNPs were selected for the NHW/NHB, NHW, and NHB cohorts respectively.

We next sought to determine if specific pathways were contributing more to the GRS than others by constructing scores by pathway within the NHW and NHB cohorts. In the NHW GRS-gen-W, we observed that SNPs representing genes in the adipose storage pathway (*ANGPTL4*-rs1044250, *LPL*-rs13702, and *PPARG*-rs12639162) contributed to nearly half of the BMIZ% (R^2 = 0.07, 6.44%). For NHB, both the transport and signalling pathways contributed approximately equal amounts to the GRS-gen-W (R^2 = 0.1050, 12.21% and R^2 = 0.0982, 11.27%, respectively). Although the majority of the SNPs in the GRS for NHW and NHB were different, each shared *IL6*-rs1554606 and many of the same genes (*PYY*, *SLC5A8*, and *NPY2R*). GRS-non-W and GRS-gen-W for BMIZ were also applied to WAZ and HAZ in all three cohorts (Table S5). Although the percent of variance explained was reduced, the GRS for BMIZ were associated with WAZ. The GRS for BMIZ, however; were not associated with HAZ.

3.3 | Gene–gene interactions

The SNP-SNP model and interaction term *P* values for BMIZ by pathway and by cohort (NHW and NHB separately) are shown in Figure S2. Two interactions were initially observed in NHW: *ANGPTL4*-rs1044250 and *LPL*-rs13702 (*P* = 0.0032) and *ANGPTL3*-rs10889337 and *PPARG*-rs12639162 (*P* = 0.0222). SNPs representing

TABLE 3 Comparison of GRS by cohort for BMIZ using linear and logistic** regression

	GRS-add-NW	GRS-add-W	GRS-non-NW	GRS-non-W	GRS-gen-W
NHW/NHB					
Adj R^2	0.0265	0.0179	0.0388	0.0397	0.0404
BMIZ% variance explained	2.38	1.54	3.59	3.68	3.75
Model *P* value	0.0183	0.0578	0.0030	0.0026	0.0024
GRS *P* value	0.0027	0.0160	0.0002	0.0002	0.0002
β-coefficient, 95% CI	0.10 (0.04-0.17)	0.51 (0.10-0.92)	0.19 (0.090-0.29)	0.92 (0.44-1.41)	0.89 (0.43-1.35)
AUC, CI**	0.57 (0.50-0.64)	0.55 (0.48-0.62)	0.57 (0.51-0.64)	0.57 (0.50-0.65)	0.57 (0.50-0.65)
OR, CI**	1.24 (1.03-1.48)	1.99 (0.67-5.87)	1.41 (1.07-1.85)	4.93 (1.30-18.67)	4.36 (1.24-15.39)
NHW only					
Adj R^2	0.0220	0.0657	0.0983	0.1242	0.1277
BMIZ% variance explained	2.45	6.77	10.00	12.55	12.90
Model *P* value	0.0319	<0.0001	<0.0001	<0.0001	<0.0001
GRS *P* value	0.0106	<0.0001	<0.0001	<0.0001	<0.0001
β-coefficient, 95% CI	0.07 (0.02-0.12)	0.32 (0.18-0.47)	0.20 (0.13-0.27)	0.84 (0.57-1.10)	0.78 (0.54-1.03)
AUC, CI**	0.57 (0.49-0.65)	0.60 (0.53-0.68)	0.63 (0.56-0.70)	0.64 (0.56-0.72)	0.64 (0.56-0.72)
OR, CI**	1.16 (1.00-1.35)	1.74 (1.12-2.69)	1.40 (1.12-1.74)	3.95 (1.62-9.61)	3.69 (1.62-8.42)
NHB only					
Adj R^2	0.0987	0.1532	0.2278	0.2322	0.2567
BMIZ% variance explained	11.32	16.47	23.90	24.33	26.69
Model *P* value	0.0025	<0.0001	<0.0001	<0.0001	<0.0001
GRS *P* value	0.0003	<0.0001	<0.0001	<0.0001	<0.0001
β-coefficient, 95% CI	0.18 (0.09-0.28)	0.44 (0.25-0.63)	0.29 (0.19-0.38)	0.76 (0.50-1.01)	0.74 (0.51-0.97)
AUC, CI**	0.77 (0.64-0.90)	0.72 (0.62-0.82)	0.76 (0.64-0.88)	0.74 (0.62-0.87)	0.78 (0.68-0.88)
OR, CI**	1.87 (1.32-2.65)	2.78 (1.40-5.55)	2.11 (1.43-3.12)	5.90 (2.21-15.70)	5.76 (2.23-14.87)

Abbreviations: Adj R^2, adjusted R^2; AUC, area under curve; BMIZ, BMI z-score; BMIZ%, percent of BMIZ variance explained; CI, confidence interval; GRS, genetic risk score; GRS-add-NW, additive nonweighted; GRS-non-NW, nontraditional nonweighted; GRS-add-W, additive weighted; GRS-non-W, nontraditional weighted; GRS-gen-W, genotypic weighted; NHB, non-Hispanic Black; NHW, non-Hispanic White; OR, odds ratio. The five GRS constructed for each cohort were compared. Adj R^2, BMIZ% variance explained, β-coefficients with 95% CI, model *P* value, and GRS *P* value were obtained using linear regression. AUC and ORs were obtained using logistic regression to compare children with or without overweight and are denoted with **. Age and sex were included in the statistical models as covariates. Three principal component scores derived from principal component analysis of 64 ancestry informative markers were included as covariates for the NHW/NHB cohort. All GRS were associated with BMIZ (*P* value < 0.05) with the exception of GRS-add-W in NHW/NHB (*P* = 0.0578). Adj R^2, BMIZ%, and GRS *P* value improved from GRS-add-NW to GRS-gen-W across all cohorts. The BMIZ% was highest for GRS-gen-W (3.8%, 12.9%, and 26.7%) for the NHW/NHB, NHW, and NHB cohorts, respectively.

FFAR1-3 appeared to interact with several SNPs for *PYY*, *GHRL*, *LEP*, *NPY2R*, and *NPY* in pathway 2 (SCFA signalling through free fatty acid receptors) for both the NHW and NHB cohorts, although the model *P* values did not reach statistical significance initially. In NHB, an interaction between *FFAR3*-rs424241 and *NPY2R*-rs1047214 was observed (*P* = 0.0014). None of the interactions observed remained significant after multiple testing correction. There were no other appreciable differences when comparing the SNP-SNP interactions between the NHW and NHB cohorts.

4 | DISCUSSION

The current study sought to describe the cumulative relationship between several SCFAs and LPS associated genes and obesity phenotypes in children. These results provide an example of the potential benefits of an integrative approach. Rather than selecting SNPs from GWAS, we utilized traditional and nontraditional methods of genetic analyses with a selection procedure of genetic variants rooted in the biological mechanisms underlying obesity. Specifically, we constructed genetic scores consisting of functional, regulatory, and tag SNPs in genes with known and proposed interactions with the gut microbiome. The GRS consisting of variants with putative roles in SCFA recognition and transport demonstrated significant relationships with BMIZ and WAZ in NHW and NHB children when analysed both separately and together. This approach not only provides an avenue to better understand the mechanisms driving obesity development early in life but also may identify potential targets for intervention strategies in the future.

Another advantage of the pathway-based approach is the ability to investigate gene-gene interactions. Most notably, we showed a

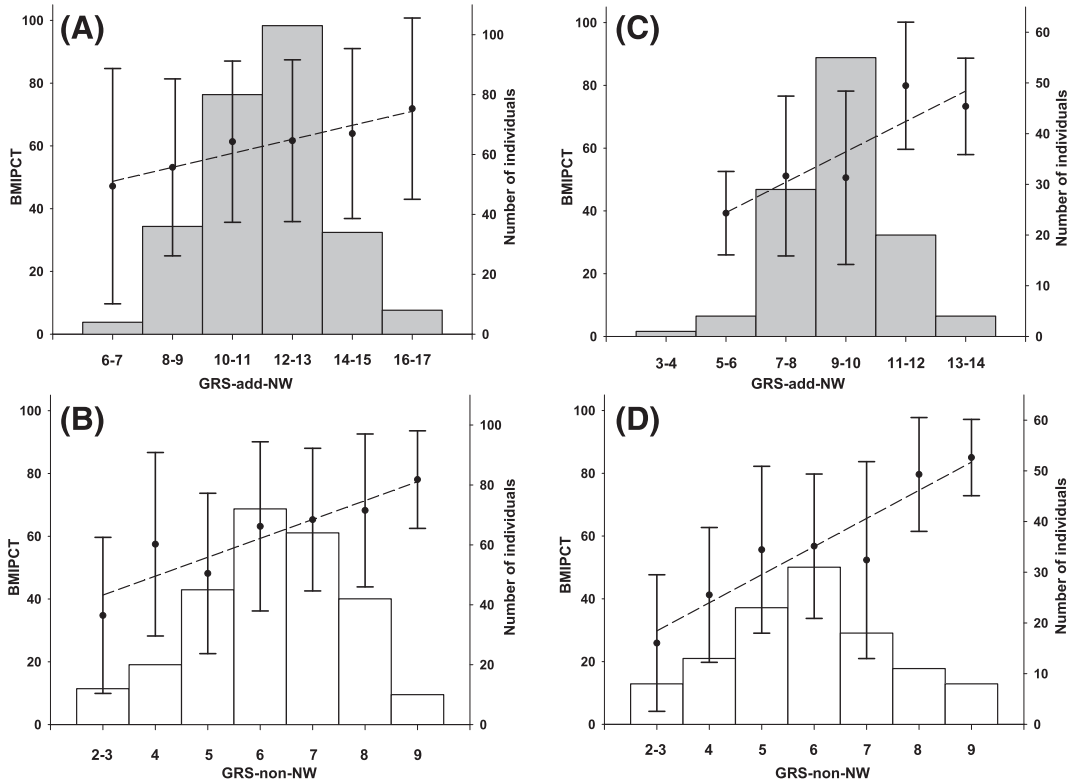

FIGURE 1 Histograms (panels A-D) in Figure 1 represent the distribution of risk alleles for the GRS-add-NW and GRS-non-NW in the NHW (n = 270) and NHB (n = 113) cohorts. Ten SNPs for the NHW GRS and nine SNPs for the NHB GRS were selected based on their independent association with BMIZ using the GLM select procedure. To construct the GRS-add-NW, genotype groups for the selected SNPs were assigned 0, 1, or 2 according to the additive model of risk and the values were summed. To construct the GRS-non-NW, genotype groups were reassigned 0 or 1 according to the dominant or recessive models of risk and the values were summed. GRS for NHW are shown in panels A and B while GRS for NHB are shown in panels C and D. All X-axes represent risk allele group categories for the GRS-add-NW (panels A and C) and GRS-non-NW (panels B and D). The left Y-axes represent the BMIPCT, and the right Y-axes represent the number of individuals in each risk allele group. Data points on the line plots imposed over the histograms represent the mean BMIPCT for individuals in each risk allele group. The dashed lines are linear regression curves. Low, medium, and high genetic risk categories were assigned for each GRS. NHW children in the low-risk category (n = 40, less than or equal to seven risk alleles) for GRS-add-NW had lower BMIZ than those in the high-risk category (n = 42, greater than or equal to 14 risk alleles (0.10 ± 0.14 vs 0.54 ± 0.14, $P = 0.0256$) (panel A). NHW children in the medium-risk category (n = 183, 10-13 risk alleles) did not differ in BMIZ from either the low-risk or high-risk groups. NHB children in the low-risk category (n = 15, less than or equal to seven risk alleles) and medium-risk category (n = 74, 8-10 risk alleles) for GRS-add-NW had lower BMIZ than those in the high-risk category ($n = 24$, greater than or equal to 11 risk alleles) (−0.31 ± 0.23 and 0.08 ± 0.10 vs 0.95 ± 0.18, $P < 0.0001$) (panel C). Abbreviations: BMIPCT, BMI percentile; BMIZ, BMI z-score; GLM, general linear model; GRS, genetic risk score; GRS-add-NW, additive nonweighted GRS; GRS-non-NW, nontraditional nonweighted GRS; NHB, non-Hispanic Black; NHW, non-Hispanic White

potential interaction between *ANGPTL4*-rs1044250 and *LPL*-rs13702 in NHW, which is consistent with the known role of *ANGPTL4* as an inhibitor of *LPL* activity.[21] Although the interactions did not survive rigorous multiple testing correction, Figure S2 illustrates potential SNP-SNP relationships that could be further explored provided an adequate sample size to conduct multiple testing. The SNP-SNP interactions examined in the present study were limited to the four selected pathways. Future studies should continue to search for functional variants and to test for genetic interactions as more evidence of these relationships becomes available. Regarding the genetic architecture of complex diseases such as obesity, additive contributions of genetic variants that are independent of each other have traditionally been assumed. However, Zuk et al suggested that this approach may generate phantom heritability as gene-gene and gene-environment interactions are not taken into consideration and estimated heritability of a phenotype may be inflated.[41] In fact, new approaches include statistical models that attempt to account for the impact of gene-environment interactions.[42]

The variance of the phenotype explained in genetic studies has been limited. Two statistical approaches in the current study strengthen the ability to describe the variance in the phenotype attributed to the GRS. First, application of β-coefficients to estimate the effect size of each SNPs' contribution to the phenotype improved the percent of variance explained by the GRS, which was anticipated based on simulation data.[43] Past studies have favoured the use of weighted values obtained from previously analysed GWAS data sets to remove bias within the sample set of interest.[9,44-46] We recognize that our use of weighted values derived from our own cohort could contribute to model overfitting; however, the use of weighted values from adult data sets may not be appropriate for child cohorts as genetic contributions to

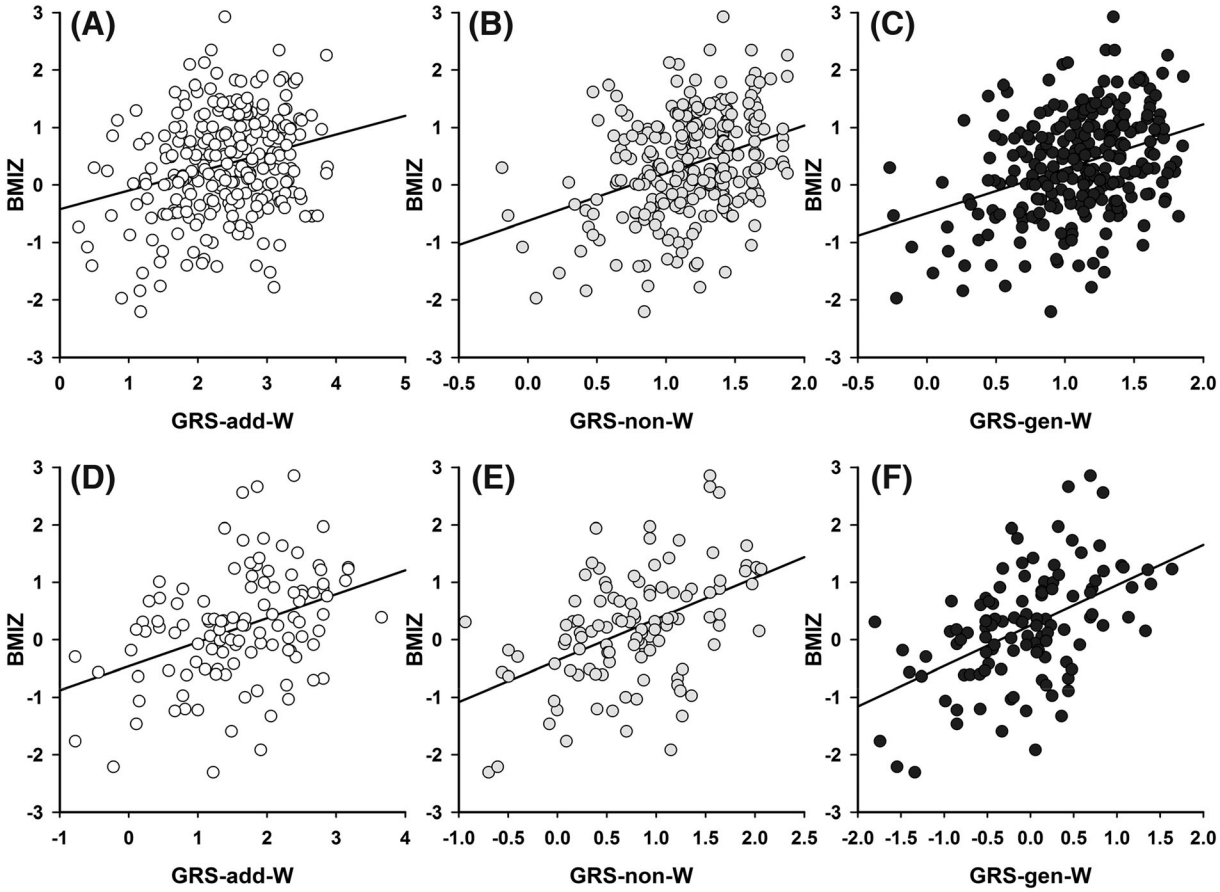

FIGURE 2 Linear regression curves for the GRS-add-W, GRS-non-W, and GRS-gen-W are displayed in panels A to F of Figure 2 for the NHW (n = 270) and NHB (n = 113) cohorts. Ten SNPs for the NHW GRS and nine SNPs for the NHB GRS were selected based on their independent association with BMIZ using the GLM select procedure. To construct the GRS-add-W, genotype groups for the selected SNPs were assigned 0, 1, or 2 according to the additive model of risk, multiplied by their respective β-coefficient, and the values were summed. To construct the GRS-non-W, genotype groups were reassigned 0 or 1 according to the dominant or recessive models of risk, multiplied by their respective β-coefficients, and the values were summed. To construct the GRS-gen-W, β-coefficients were summed for each selected SNP. GRS for NHW are shown in panels A to C while GRS for NHB are shown in panels D to F. All X-axes represent the composite GRS, while all Y-axes represent BMIZ. Adj R^2 and percent of variance explained for BMIZ increases across panels A to C for NHW and across panels D to F for NHB. GRS-gen-W had the highest percent of variance explained for BMIZ for both NHW and NHB cohorts. Abbreviations: Adj R^2, adjusted R^2; BMIPCT, BMI percentile; BMIZ, BMI z-score; GRS, genetic risk score; GRS-add-W, additive weighted GRS; GRS-non-W, nontraditional weighted GRS; GRS-gen-W, genotypic GRS; NHB, non-Hispanic Black; NHW, non-Hispanic White

obesity are known to be age-dependent.[2,47] Longitudinal assessments of the GRS are particularly needed at critical phases of development, including infant peak, adiposity rebound, and puberty. Second, comparison of the GRS-add-W to the GRS-gen-W suggests that a hypothesis-free approach to the genetic mode of inheritance is more representative of the genotype-phenotype relationship than assuming the inheritance model. Traditional methods of constructing GRS apply the additive mode of inheritance with the advantage being an improvement in the power of the analysis.[48] However, the genetic model is seldom known a priori and conforming SNPs to these models in our data set presented concerns in calculating the GRS. New statistical methods including the MAX and MERT methods have been developed to better predict the mode of inheritance from the empirical data and could be used in follow-up analyses of the GRS herein.[49]

Data using the GRS approach in a multi-ethnic cohort of children are limited. The Klimentidis study was the first to show that the mean GRS value for GWAS obesity variants was different amongst African American, Hispanic, European American, and biracial groups.[50] Similarly, we found that the NHW/NHB GRS was associated with BMIZ in NHW (n = 270) but not in NHB (n = 113). There is a possibility that certain pathways may be contributing more to the GRS than other pathways within the NHW and NHB cohorts. While SNPs in the adipose storage pathway were well represented in the NHW GRS, the NHB GRS appeared to be represented equally by both transport and signalling pathway SNPs. The underlying mechanisms driving excess adipose accumulation may be different amongst NHW and NHB children. Our data further demonstrated that similar genes were associated with obesity-related phenotypes in the NHW and NHB cohorts, but the SNPs carrying those signals often differed between the groups. This presents a challenge for researchers to find and establish shared genetic variants that can be used for association studies in multi-ethnic cohorts. We conducted PCA of AIMs to account for

TABLE 4 Comparison of GRS after bootstrapping with 10, 100, and 1000 replications

	10 replications		100 replications		1000 replications	
	GRS-non-W	GRS-gen-W	GRS-non-W	GRS-gen-W	GRS-non-W	GRS-gen-W
NHW/NHB						
Adjusted R^2	0.0589	0.0564	0.0572	0.0573	0.0577	0.0582
BMIZ% variance explained	4.15	3.90	3.59	3.60	3.72	3.77
Model *P* value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
GRS *P* value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
β-coefficient, 95% CI	0.977 (0.827-1.259)	0.900 (0.758-1.042)	0.917 (0.869-0.964)	0.875 (0.829-0.920)	0.932 (0.917-0.947)	0.893 (0.879-0.910)
NHW only						
Adjusted R^2	0.1437	0.1462	0.1341	0.1362	0.1334	0.1376
BMIZ% variance explained	13.40	13.65	12.69	12.90	12.52	12.94
Model *P* value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
GRS *P* value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
β-coefficient, 95% CI	0.833 (0.825-0.841)	0.785 (0.777-0.793)	0.834 (0.826-0.842)	0.785 (0.778-0.793)	0.834 (0.825-0.842)	0.785 (0.778-0.793)
NHB only						
Adjusted R^2	0.2372	0.2748	0.2392	0.2686	0.2543	0.2772
BMIZ% variance explained	21.55	25.28	23.18	26.11	24.43	26.72
Model *P* value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
GRS *P* value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
β-coefficient, 95% CI	0.696 (0.619-0.772)	0.714 (0.643-0.784)	0.742 (0.717-0.767)	0.731 (0.708-0.753)	0.756 (0.749-0.764)	0.738 (0.730-0.744)

Abbreviations: BMIZ, BMI z-score; BMIZ%, percent of BMIZ variance explained; GRS, genetic risk score; GRS-non-W, nontraditional weighted; GRS-gen-W, genotypic weighted; NHB, non-Hispanic Black; NHW, non-Hispanic White. Adjusted R^2, % BMIZ variance explained, β-coefficients with 95% CI, Model *P* value, and GRS *P* value were obtained using linear regression at 10, 100, and 1000 replications of the data set.

population stratification by ethnicity.[51,52] The primary advantage of this method was the creation of several continuous variables that could more accurately characterize differences amongst and within ethnic groups than a single categorical variable. Use of this technique could be valuable in future work in the field of genetic epidemiology as admixed populations increase the likelihood of false positive discovery if population stratification is not taken into consideration.[53,54]

We recognize several limitations to the findings presented in this report. Because the sample size was limited, our results need to be replicated and the statistical methods employed should be cross-validated in an independent data set. Several of the genes selected for this study including *SLC16A1* and *FFAR3* have expression in tissue beyond the gastrointestinal tract.[55-57] While the premise of this work was based on the transport and signalling pathways of SCFAs produced by gut microbes in the distal intestinal tract, the relationship between the GRS described here and obesity-related phenotypes may not be exclusively through the proposed pathways. The exact effect of host genetic variation in SCFA-associated genes coding for the solute carriers and free fatty acid receptors on SCFA uptake and recognition also has not been fully elucidated. Nevertheless, the stage has been set to further elucidate the underlying genetic and microbial mechanisms of obesity development in children. In fact, our approach is complementary to a recent review published by Dong and colleagues, which investigated the functional consequences of obesity-susceptibility loci and SNPs' identified by GWAS.[58]

In summary, the analytical methodology introduced in this study contributes to establishing a novel way by which basic research in molecular and genetic mechanisms of obesity can be utilized in population-level genetic analyses. Past works have shown relationships between many of the same genes and other disease-related phenotypes including biomarkers of cardiovascular and metabolic disease, but this is the first report to our knowledge that combined their effects into an obesity risk score in children (see supplementing information for references). While the pathway-based approach provided a biological basis for SNP selection, the statistical methods used here improved our ability to describe the genotype-phenotype relationship. Some GRS studies in infants and children have utilized the weighted score approach, but the hypothesis-free approach to the mode of inheritance for each SNP represents a departure from the traditional methods of producing GRS.

ACKNOWLEDGEMENTS

AAW, MTG, and SMD designed research; AAW, KH, and the STRONG Kids Research Group conducted research; MTG, SMD, KH, and STRONG Kids Research Group provided essential materials; AAW

analysed the data under the guidance of MTG and SM; AAW and MTG wrote the paper; MTG, SMD, and SM had primary responsibility for final content. All authors read and approved the final manuscript. This research was funded, in part, by grants from the Illinois Council for Agriculture Research to Kristen Harrison (PI) and the University of Illinois Health and Wellness Initiative to Barbara Fiese and Sharon Donovan and United States Department of Agriculture (Hatch 793-328) to Barbara Fiese (PI). The STRONG Kids Team includes Kristen Harrison, Kelly Bost, Brent McBride, Sharon Donovan, Diana Grigsby-Toussaint, Juhee Kim, Janet Liechty, Angela Wiley, Margarita Teran-Garcia and Barbara Fiese. MTG and SMD were sponsored by the National Institute for Agriculture under the Illinois Transdisciplinary Obesity Prevention Program (I-TOPP) training grant (2011-67001-30101) to the Division of Nutritional Sciences at the University of Illinois. AAW was founded by the Agriculture and Food Research Initiative (AFRI) competitive grant no. 2013-67011-21024 from the United States Department of Agriculture (USDA) National Institute of Food and Agriculture.

CONFLICT OF INTEREST

No conflict of interest was declared.

LIST OF ABBREVIATIONS

STRONG Kids	Synergistic Theory and Research on Obesity and Nutrition Group
NHW	Non-Hispanic White cohort
NHB	Non-Hispanic Black cohort
GWAS	Genome-wide association studies
SNP	Single nucleotide polymorphism
SCFA	Short-chain fatty acids
MAF	Minor allele frequency
UTR	Untranslated region
LD	Linkage disequilibrium
HWE	Hardy–Weinberg Equilibrium
GRS	Genetic risk score
GRS-add-NW	Additive nonweighted
GRS-non-NW	Nontraditional nonweighted
GRS-add-W	Additive weighted
GRS-non-W	Nontraditional weighted
GRS-gen-W	Genotypic weighted
AIMs	Ancestry informative markers
gDNA	Genomic DNA
dbSNP	Single Nucleotide Polymorphism Database
SIFT	Sorting Tolerant from Intolerant
BMI	Body mass index
BMIPCT	BMI percentile
BMIZ	BMI z-score
WAZ	Weight-for-age z-score
HAZ	Height-for-age z-score
BMIZ%	Percent of BMIZ variance explained
PCR	Polymerase chain reaction
SDS	Sequence detection system
SVS	SNP & Variation Suites
CHM	Composite haplotype method
ROC	Receiver operating curve
AUC	Area under the curve
GLM	General linear model
FFAR2	Free fatty acid receptor 2
FFAR3	Free fatty acid receptor 3
ANGPTL4	Angiopoietin-like 4
CD36	Cluster of differentiation 36
SLC16A1	Solute carrier family 16 member 1
SLC16A3	Solute carrier family 16 member 3
SLC16A4	Solute carrier family 16 member 4
SLC5A8	Solute carrier family 5 member 8
TLR4	Toll-like receptor 4
LPL	Lipoprotein lipase
PYY	Peptide YY
GCG	Glucagon
LEP	Leptin
LEPR	Leptin receptor
NPY	Neuropeptide Y
NPY2R	Neuropeptide Y receptor Y2
PPARG	Peroxisome proliferator-activated receptor gamma
NFKB	Nuclear factor kappa B
IL6	Interleukin 6
TNFA	Tumour necrosis factor alpha

ORCID

Kristen Harrison https://orcid.org/0000-0002-5630-3102
Sharon M. Donovan https://orcid.org/0000-0002-9785-4189
Margarita Teran-Garcia https://orcid.org/0000-0002-8196-7605

REFERENCES

1. Maes HHM, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet*. 1997;27(6):600-600.
2. Elks CE, den Hoed M, Zhao JH, et al. Variability in the heritability of body mass index: a systematic review and meta-regression. *Front Endocrinol*. 2012;3:29.
3. Andersen MK, Sandholt CH. Recent progress in the understanding of obesity: contributions of genome-wide association studies. *Curr Obes Rep*. 2015;4(4):401-410.
4. Bradfield JP, Taal HR, Timpson NJ, et al. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat Genet*. 2012;44(5):526-531.
5. Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM. Predictive testing for complex diseases using multiple genes: fact or fiction? *Genet Med*. 2006;8(7):395-400.
6. Elks CE, Heude B, de Zegher F, et al. Associations between genetic obesity susceptibility and early postnatal fat and lean mass: an individual participant meta-analysis. *JAMA Pediatr*. 2014;168(12):1122-1130.
7. Makela J, Lagstrom H, Pitkanen N, et al. Genetic risk clustering increases children's body weight at 2 years of age—the STEPS study. *Pediatr Obes*. 2015;11(6):459-467.

8. Steinsbekk S, Belsky D, Guzey IC, Wardle J, Wichstrom L. Polygenic risk, appetite traits, and weight gain in middle childhood: a longitudinal study. *JAMA Pediatr*. 2016;170(2):e154472.
9. Vogelezang S, Monnereau C, Gaillard R, et al. Adult adiposity susceptibility loci, early growth and general and abdominal fatness in childhood: the generation R study. *Int J Obes (Lond)*. 2015;39(6):1001-1009.
10. Warrington NM, Howe LD, Wu YY, et al. Association of a body mass index genetic risk score with growth throughout childhood and adolescence. *PLoS ONE*. 2013;8(11):e79547.
11. Beyerlein A, von Kries R, Ness AR, Ong KK. Genetic markers of obesity risk: stronger associations with body composition in overweight compared to normal-weight children. *PLoS ONE*. 2011;6(4):e19057.
12. González JR, Estévez MN, Giralt PS, et al. Genetic risk profiles for a childhood with severely overweight. *Pediatr Obes*. 2014;9(4):272-280.
13. Le Chatelier E, Nielsen T, Qin J, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature*. 2013;500 (7464):541-546.
14. Canfora EE, Jocken JW, Blaak EE. Short-chain fatty acids in control of body weight and insulin sensitivity. *Nat Rev Endocrinol*. 2015;11 (10):577-591.
15. Backhed F, Ding H, Wang T, et al. The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A*. 2004;101(44):15718-15723.
16. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*. 2006;444(7122):1027-1031.
17. Gill RK, Saksena S, Alrefai WA, et al. Expression and membrane localization of MCT isoforms along the length of the human intestine. *Am J Physiol-Cell Ph*. 2005;289(4):C846-C852.
18. Ritzhaupt A, Wood IS, Ellis A, Hosie KB, Shirazi-Beechey SP. Identification and characterization of a monocarboxylate transporter (MCT1) in pig and human colon: its potential to transport L-lactate as well as butyrate. *J Physiol-London*. 1998;513(3):719-732.
19. Samuel BS, Shaito A, Motoike T, et al. Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proc Natl Acad Sci U S A*. 2008;105 (43):16767-16772.
20. Xiong YM, Miyamoto N, Shibata K, et al. Short-chain fatty acids stimulate leptin production in adipocytes through the G-protein coupled receptor GPR41. *Proc Natl Acad Sci U S A*. 2004;101(4):1045-1050.
21. Sukonina V, Lookene A, Olivecrona T, Olivecrona G. Angiopoietin-like protein 4 converts lipoprotein lipase to inactive monomers and modulates lipase activity in adipose tissue. *Proc Natl Acad Sci U S A*. 2006; 103(46):17450-17455.
22. Lu YC, Yeh WC, Ohashi PS. LPS/TLR4 signal transduction pathway. *Cytokine*. 2008;42(2):145-151.
23. Harrison K, Bost KK, McBride BA, et al. Toward a developmental conceptualization of contributors to overweight and obesity in childhood: the six-Cs model. *Child Dev Perspect*. 2011;5(1):50-58.
24. http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm. A SAS Program for the CDC Growth Charts. Available at: http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm.
25. Delzenne NM, Neyrinck AM, Backhed F, Cani PD. Targeting gut microbiota in obesity: effects of prebiotics and probiotics. *Nat Rev Endocrinol*. 2011;7(11):639-646.
26. Barton GM, Medzhitov R. Toll-like receptor signaling pathways. *Science*. 2003;300(5625):1524-1525.
27. Bellahcene M, O'Dowd JF, Wargent ET, et al. Male mice that lack the G-protein-coupled receptor GPR41 have low energy expenditure and increased body fat content. *Brit J Nutr*. 2013;109(10):1755-1764.
28. Holzer P, Farzi A. Neuropeptides and the microbiota-gut-brain axis. *Adv Exp Med Biol*. 2014;817:195-219.
29. Hong YH, Nishimura Y, Hishikawa D, et al. Acetate and propionate short chain fatty acids stimulate adipogenesis via GPCR43. *Endocrinology*. 2005;146(12):5092-5099.
30. Kaddatz K, Adhikary T, Finkernagel F, Meissner W, Muller-Brusselbach S, Muller R. Transcriptional profiling identifies functional interactions of TGF beta and PPAR beta/delta signaling: synergistic induction of ANGPTL4 transcription. *J Biol Chem*. 2010;285(38):29469-29479.
31. Koliwad SK, Kuo TY, Shipp LE, et al. Angiopoietin-like 4 (ANGPTL4, fasting-induced adipose factor) is a direct glucocorticoid receptor target and participates in glucocorticoid-regulated triglyceride metabolism. *J Biol Chem*. 2009;284(38):25593-25601.
32. Yoshida K, Shimizugawa T, Ono M, Furukawa H. Angiopoietin-like protein 4 is a potent hyperlipidemia-inducing factor in mice and inhibitor of lipoprotein lipase. *J Lipid Res*. 2002;43(11):1770-1772.
33. Zaibi MS, Stocker CJ, O'Dowd J, et al. Roles of GPR41 and GPR43 in leptin secretory responses of murine adipocytes to short chain fatty acids. *FEBS Lett*. 2010;584(11):2381-2386.
34. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet*. 2002;3(5):391-397.
35. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4): 248-249.
36. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073-1081.
37. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22 (9):1790-1797.
38. Bruno AE, Li L, Kalabus JL, Pan YZ, Yu AM, Hu ZH. miRdSNP: a database of disease-associated SNPs and microRNA target sites on 3' UTRs of human genes. *BMC Genomics*. 2012;13(1):44.
39. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21(2):263-265.
40. Kosoy R, Nassir R, Tian C, et al. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum Mutat*. 2009;30(1):69-78.
41. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A*. 2012;109(4):1193-1198.
42. Han SS, Rosenberg PS, Ghosh A, Landi MT, Caporaso NE, Chatterjee N. An exposure-weighted score test for genetic associations integrating environmental risk factors. *Biometrics*. 2015;71(3):596-605.
43. Che R, Motsinger-Reif AA. Evaluation of genetic risk score models in the presence of interaction and linkage disequilibrium. *Front Genet*. 2013;4:138.
44. Llewellyn CH, Trzaskowski M, Plomin R, Wardle J. From modeling to measurement: developmental trends in genetic influence on adiposity in childhood. *Obesity*. 2014;22(7):1756-1761.
45. Llewellyn CH, Trzaskowski M, van Jaarsveld CH, Plomin R, Wardle J. Satiety mechanisms in genetic risk of obesity. *JAMA Pediatr*. 2014;168(4):338-344.
46. Toemen L, Gishti O, Vogelezang S, et al. Cross-sectional population associations between detailed adiposity measures and C-reactive

protein levels at age 6 years: the generation R study. *Int J Obes (Lond)*. 2015;39(7):1101-1108.

47. Cooke Bailey JN, Igo RP Jr. Genetic risk scores. *Curr Protoc Hum Genet*. 2016;91: 1.29.1-1.29.9. https://doi.org/10.1002/cphg.20
48. Bagos PG. Genetic model selection in genome-wide association studies: robust methods and the use of meta-analysis. *Stat Appl Genet Mol Biol*. 2013;12(3):285-308.
49. Freidlin B, Zheng G, Li Z, Gastwirth JL. Trend tests for case-control studies of genetic markers: power, sample size and robustness. *Hum Hered*. 2002;53(3):146-152.
50. Klimentidis YC, Chen GB, Lopez-Alarcon M, Harris JJ, Duarte CW, Fernandez JR. Associations of obesity genes with obesity-related outcomes in multiethnic children. *Arch Med Res*. 2011;42(6):509-514.
51. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12):e190.
52. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904-909.
53. Fernandez JR, Pearson KE, Kell KP, Bohan Brown MM. Genetic admixture and obesity: recent perspectives and future applications. *Hum Hered*. 2013;75(2–4):98-105.
54. Nyholt DR. Using genomic data to make indirect (and unauthorized) estimates of disease risk. *Public Health Genomics*. 2012;15(5):303-311.
55. Carneiro L, Pellerin L. Monocarboxylate transporters: new players in body weight regulation. *Obes Rev*. 2015;16(Suppl 1):55-66.
56. De Vadder F, Kovatcheva-Datchary P, Goncalves D, et al. Microbiota-generated metabolites promote metabolic benefits via gut-brain neural circuits. *Cell*. 2014;156(1–2):84-96.
57. Kimura I, Inoue D, Maeda T, et al. Short-chain fatty acids and ketones directly regulate sympathetic nervous system via G protein-coupled receptor 41 (GPR41). *Proc Natl Acad Sci U S A*. 2011;108(19): 8030-8035.
58. Dong SS, Zhang YJ, Chen YX, et al. Comprehensive review and annotation of susceptibility SNPs associated with obesity-related traits. *Obes Rev*. 2018;19(7):917-930.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Wang AA, Harrison K, Musaad S, Donovan SM, Teran-Garcia M, the STRONG Kids Research Team. Genetic risk scores demonstrate the cumulative association of single nucleotide polymorphisms in gut microbiome-related genes with obesity phenotypes in preschool age children. *Pediatric Obesity*. 2019;14:e12530. https://doi.org/10.1111/ijpo.12530