



Published in final edited form as:

Obesity (Silver Spring). 2017 July ; 25(7): 1270–1276. doi:10.1002/oby.21869.

Exome sequencing reveals novel genetic loci influencing obesity related traits in Hispanic children

Aniko Sabo¹, Pamela Mishra¹, Shannon Dugan-Perez¹, V. Saroja Voruganti², Jack W. Kent Jr.³, Divya Kalra¹, Shelley A. Cole³, Anthony G. Comuzzie³, Donna M. Muzny¹, Richard A. Gibbs¹, and Nancy F. Butte⁴

¹Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX

²Department of Nutrition and UNC Nutrition Research Institute, University of North Carolina, Chapel Hill, NC, USA

³Department of Genetics, Texas Biomedical Research institute, San Antonio, TX, USA

⁴USDA/ARS Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, Houston, TX

Abstract

Objective—Perform whole exome sequencing in 928 Hispanic children and identify variants and genes associated with childhood obesity.

Methods—Single nucleotide variants were identified from Illumina whole exome sequencing data using integrated read mapping, variant calling and annotation pipeline (Mercury). Association analyses of 74 obesity related traits and exonic variants were performed using SeqMeta software. Rare autosomal variants were analyzed using gene-based association analyses, and common autosomal variants were analyzed at single nucleotide variant (SNV) level.

Results—1) Identified rare exonic variants in 10 genes and 16 common SNVs in 11 genes that were associated with obesity traits in a cohort of Hispanic children 2) Discovered novel rare variants in peroxisome biogenesis factor 1 (*PEXI*) associated with several obesity traits (weight, weight z-score, BMI, BMI z-score, waist circumference, fat mass, trunk fat mass), and 3) Replicated previously reported SNVs associated with childhood obesity.

Conclusions—Convergence of whole exome sequencing, a family-based design, and extensive phenotyping discovered novel rare and common variants associated with childhood obesity. Linking *PEXI* to obesity phenotypes poses a novel mechanism of peroxisomal biogenesis and metabolism underlying the development of childhood obesity.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

CONTACT INFO: Aniko Sabo, sabo@bcm.edu, Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030 and Nancy F. Butte, nbutte@bcm.edu, USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, 1100 Bates Street, Houston, TX 77030.

DISCLOSURE: The authors declared no conflict of interest.

Keywords

Hispanics; Childhood Obesity; Genetics

Introduction

Obesity has reached epidemic proportions in the US Hispanic population. The prevalence of obesity in Hispanics has doubled in the past 20 years with 42.5% of adults and 22.4% of children classified with obesity, increasing their risks for cardiometabolic diseases (1). The genetic influence on the transmission of obesity and its comorbidities is indisputable. Quantitative genetic analyses of body mass index (BMI), adiposity and related traits indicate heritability on the order of 0.4–0.7 (2). Genome-wide linkage and genome-wide association studies (GWAS) have identified a number of putative chromosomal regions and loci, however, they account for a small fraction and even jointly only a modest proportion of the familial risk. A recent GWAS meta-analysis of BMI in 339,224 adults identified 97 loci that accounted for about 2.7% of the BMI variation (3). Most GWAS focus on polymorphic variants with relatively high frequency; however, data are rapidly accumulating that rare variants have a large cumulative effect on normal phenotypic variation and are extremely important to complex diseases (4). Fewer GWAS have been conducted on childhood obesity (5–9). Although novel variants were identified, major common variants related to obesity overlapped to a substantial degree between children and adults. GWAS results explained only a few percent of the apparent genetic variance contributing to childhood obesity, as found in adult obesity. Furthermore, most variants found in GWAS are in the intronic or intergenic regions, making it difficult to elucidate the underlying mechanisms of obesity.

Thus, in this study, we aimed to explore the contribution of rare and common exonic variants to childhood obesity. Whole exome and whole genome sequencing are revealing the enormous extent of rare variation, much of which is unique to that individual's family (4). Major contributors to inherited disease susceptibility are likely to be alleles that arose recently in extended pedigrees (10). Purifying natural selection eliminates highly deleterious variants before they reach high frequency, therefore disease risk alleles with large effects should be enriched at lower frequencies (11).

The VIVA LA FAMILIA (VIVA) cohort was established to identify genetic variants influencing childhood obesity in the Hispanic population. Our formalized quantitative genetic analyses found the heritability of BMI and obesity-related traits to be significant (1). GWA approaches used in VIVA (6) identified specific variants related to childhood obesity; however, they did not nearly account for its high heritability indicating that a relatively large number of variants remain to be discovered. Recent advances in DNA sequencing allow us to efficiently and effectively identify both common and rare coding variants and assess their impact on obesity and associated phenotypes (10). Stop-gain, stop-loss, and a subset of non-synonymous coding variants are most likely to have direct functional consequences on disease susceptibility. In the VIVA study we take advantage of a family-based study design that will allow us to detect functional variants present at higher frequency among relatives than in the general population.

The aims of this study were 1) to generate and annotate variants from whole exome sequencing data in 928 Hispanic children from the VIVA cohort; and 2) to perform statistical analyses to identify rare as well as common variants in exonic regions associated with childhood obesity.

Methods

VIVA study participants and phenotype measurements

VIVA cohort used for exome sequencing comprised 293 Hispanic families with an average family size of 5 persons (range 3 to 10) (1). All enrolled children and parents gave written informed consent or assent. The protocol was approved by the Institutional Review Boards for Human Subject Research for Baylor College of Medicine and Affiliated Hospitals and for Texas Biomedical Research Institute. Each family was ascertained on a proband with obesity between the ages 4–19 y. The VIVA cohort was highly enriched for obesity: the majority of the parents were either classified with overweight (34%) or obesity (57%) and 52% of the children were classified as with obesity. Among the obese children, 62% were above the 99th BMI percentile, indicating severe obesity. Among these nuclear families, there were 32 extended family groupings, resulting in a large number of sib-pairs and first-degree cousins (1457 related-pairs).

Phenotyping has been described in detail elsewhere (1, 6). Briefly, the phenotyping included standard anthropometry and body composition by dual energy x-ray absorptiometry (DXA); birth weights from Texas birth records; diet by 24-h recalls; total energy expenditure and substrate utilization by 24-h room calorimetry; physical activity by accelerometry; and fasting biochemistries analyzed by standard techniques.

Whole Exome Sequencing and genome variant identification

For each DNA sample, the entire exome was captured using the custom NimbleGen VCRome 2.1 capture reagent that targets coding exons from Consensus Coding Sequence (CCDS), NCBI RNA reference sequences (RefSeq) and Vega Human Genome Annotations. Capture enrichment was followed by sequencing on Illumina platform using previously described standard protocols (12). Illumina sequence analysis was performed using the Human Genome Sequencing Center's integrated Mercury pipeline (13). Briefly, the sequencing reads were mapped to the GRCh37 human reference sequence using BWA. The Atlas suite was used to call single nucleotide variants (SNVs). Lastly, variant annotation was accomplished through the Cassandra annotation suite utilizing different databases to predict functional consequences of genomic variants and place the variants in a biological framework. Individual genotype, variant and sample level QC procedure was performed using a custom pipeline following the guidelines established for the CHARGE-S project (14). The QC procedure details are provided in the supplement.

Genotype-phenotype association analysis

Phenotypes were transformed using the inverse normalization procedure implemented in Sequential Oligogenic Linkage Analysis Routines (SOLAR)(15) and residualized using sex, age and age² as covariates. For the birthweight phenotype, sex and gestational age were used

as covariates in the normalization/residualization procedure. Family relationships were reconstructed and verified using Primus (16) software utilizing the SNVs from exome sequencing data. Principal component analysis was conducted using PCAir method as implemented in the GENESIS package (17). Unlike standard PCA methods (e.g. Eigenstrat), the method implemented in PCAir accounts for relatedness of the individuals and identifies PCs that accurately capture population structure and are robust to familial relatives in the sample set.

Both single variant and gene-based association analyses were conducted for 74 phenotypes using SeqMeta (18) software. Only variants on autosomal chromosomes were analyzed in this study, and all analyses used additive genetic models. The empirical pedigree information was transformed into a kinship matrix using kinship2 software (19), and included in the association analysis following the method of Chen *et al.* (20) as implemented in SeqMeta software. First ten PCs were included as covariates in the association analysis. Single variant analysis included common variants with minor allele frequency (MAF) greater or equal to 0.01. The significance level was calculated for each phenotype using Bonferroni correction, dividing the nominal significance level 0.05 by the number of SNV tests performed.

Analysis of exome data has potential to identify rare causal variants that are not genotyped in GWA studies. However, the low frequencies of rare variants will negatively impact the statistical power of an association test unless the sample size is very large. To ameliorate this problem several methods that test for the collective effect of a group of rare variants have been proposed, including burden and variance-component test. The burden tests (e.g. T1, T5 and weighted-sum method) collapse information for multiple genetic variants into a single genetic score and test for association between this score and a trait. The Sequence Kernel Association Test (SKAT), on the other hand, aggregates score test statistics of individual variants in a SNV-set to compute p-values for a gene, and is robust when variants with both positive and negative effects are included.

Gene-based analysis included rare (MAF < 0.01), predicted protein-altering variants: stop-gain, stop-loss, missense and splice site variants. The association analysis for gene-based tests was conducted using SKAT and weighted-sums burden tests (WST) using Madsen-Browning weights ($1/(MAF^*(1-MAF))$). The significance level for the gene based tests was calculated using Bonferroni correction, dividing the nominal significance level 0.05 by the number of gene tests performed. This included all genes that had at least two predicted protein-altering variants, multiplied by two to account for testing with both SKAT and WST.

Results

Whole exome sequencing - identified variants

For the whole exome sequencing of the 928 VIVA samples, the average coverage per sample across the targeted bases by the capture design was 130X, and on average 92% of the bases targeted had at least 20X read coverage. Genotype and variant level QC analyses removed 12% of variants that did not pass our QC metrics. We removed two samples that did not pass the sample level QC metrics, and 10 samples that failed pedigree information verification. The final dataset comprised 283,587 SNVs identified in 916 samples. The dataset included

150,718 SNVs that have potential to affect the protein function, by either affecting a splice site (1,213 splice site variants), amino acid change (146,708 missense variants) or resulting in a stop gain (2,568) or stop loss (229). We identified 146,708 missense variants, of these, 123,250 were rare (MAF <0.01) and 33,720 were predicted to be damaging by Polyphen2. A total of 31,136 missense mutations were considered rare and damaging. The frequency of SNVs shown in Table 1 is similar to patterns and rates observed in an independent cohort (21), and demonstrates that the number of variants most likely to affect protein function (nonsense and damaging missense) is enriched in singletons and doubletons as compared to more common variants (tripletons and more). The data has been submitted to dbGAP Study Accession phs000616.

Gene-based association analysis for rare variants (MAF < 0.01)

Gene level association analysis for rare (MAF < 0.01), potentially functional SNVs (stop-gain, stop-loss, splicing and missense) was performed across the 74 phenotypes using WST and SKAT. The full list of phenotypes is provided in Table S1 (supplement). The p-value for significance was established at 1.72×10^{-6} based on the number of gene tests performed (29,110 tests: 14,555 genes, multiplied by two tests each). We evaluated all genes that had at least two functional, rare SNVs, and discovered 16 significant associations, three based on WST, and 13 based on SKAT (Table 2).

The strongest gene-based association discovered was for variants in adiponectin (*ADIPOQ*) for serum adiponectin levels (pskat = 2.1×10^{-11}). The association was based on four rare predicted protein-altering variants (cumulative MAF = 0.009). We explored in more detail this gene-based finding by looking at the single variant level associations for each of the four underlying SNVs. The strongest single SNP association ($p = 6.15 \times 10^{-11}$, effect size $\beta = -2.7809$) was for missense variant rs200573126 (p.Gly45Arg, MAF = 0.003). To investigate the contribution of other variants we removed rs200573126 from the analysis, and analyzed the remaining three variants by SKAT. The association was only nominally significant ($P = 0.02$), indicating that the rs200573126 variant is driving the association.

We discovered significant association of variants in cystathionine gamma-lyase (*CTH*) with serum cystathionine levels (pskat = 2.12×10^{-7}). The association was based on five rare predicted protein-altering variants (cumulative MAF = 0.01). We explored in more detail the gene-based finding by looking at the single variant level associations for each of the five underlying SNVs. Only one variant (rs28941785, p.Thr67Ile, MAF = 0.004,) passed the nominal significance value ($P = 1.7 \times 10^{-7}$, effect size $\beta = 1.9769$). Similarly to *ADIPOQ*, we removed rs28941785 variant, and the association based on remaining SNVs was not significant ($P = 0.29$), indicating that the rs28941785 variant is underlying the association.

We also discovered a significant association of peroxisome biogenesis factor 1 (*PEXI*) with seven different obesity-related phenotypes. The association was based on 17 rare predicted protein-altering variants (cumulative MAF = 0.022), with the strongest association with weight for age Z-score (pskat = 2.26×10^{-7}), followed by trunk fat mass (pskat = 6.55×10^{-7}), weight (pskat = 6.59×10^{-7}), and BMI (pskat = 1.14×10^{-6}) (Figure 1). To further investigate the association of *PEXI* and weight for age z-score we looked at the single variant level association results for each of the 17 rare, potentially functional variants. We

discovered that only one missense variant (rs141510219, pVal734Ile, MAF = 0.005) of the 17 SNVs had nominally significant association ($P = 8.97 \times 10^{-8}$, effect size $\beta = 1.8638$) indicating that this single variant is responsible for driving the signal (Figure 2). After the removal of the rs141510219 variant, and reanalysis of the seven phenotypes using SKAT, none of the seven traits met the p-value significance threshold (p-values = 0.03 – 0.46), implicating the rs141510219 as the causal variant.

We discovered significant association ($pskat = 4.23 \times 10^{-7}$) of variants in suppressor of cytokine signaling 4 (*SOCS4*) with sleeping energy expenditure, as measured by 24-h room calorimetry. The association was based on two rare predicted protein-altering variants (cumulative MAF = 0.01). We looked at the single variant level associations for the two underlying SNVs. Only one variant (rs146421724, p.Gly54Ser, MAF = 0.0096) passed the nominal significance value ($P = 3.59 \times 10^{-7}$, effect size $\beta = 1.4778$). The complete list of significant gene-based findings, along with p-values, number of underlying SNVs and cumulative MAFs, are listed in Table 2.

Single variant association analysis for common variants (MAF 0.01)

We performed single variant association analysis for 54,360 common variants with MAF 0.01 across the 74 phenotypes. Following Bonferroni correction, we established the p-value significance level at 9.2×10^{-7} . We discovered 16 common SNV with significant association with at least one of the phenotypes (Table 3). We have successfully confirmed three findings from previous VIVA GWAS studies (6, 22, 23) and report several novel associations. We have replicated association of four variants in *SLC2A9* with serum levels of uric acid, with the strongest association for missense variant rs16890979 ($P = 6.1 \times 10^{-13}$). We have replicated associations of missense SNV (rs12075) in *DARC* ($P = 1.46 \times 10^{-27}$), and missense SNV (rs58037016) in *OR10J3* ($P = 3.2 \times 10^{-7}$) with levels of monocyte chemotactic protein 1 (MCP-1). We have also confirmed the association discovered in the VIVA GWAS study, of the missense SNV (rs3733402) in *KLKB1* with levels of insulin-like growth factor 1 ($P = 7.85 \times 10^{-10}$). Important to note is that 87% of the samples were in common between the VIVA GWAS and this exome study, therefore *KLKB1* association is a demonstration of consistency of results across different platforms, and not an independent replication.

We discovered eight novel associations. The strongest was association of a missense variant (rs1800234, MAF = 0.041) in *PPAR- α* with increased diet protein levels ($P = 1.05 \times 10^{-7}$, effect size $\beta = 0.632$). We also discovered association of two synonymous SNVs (rs2236261 and rs2236260) in *NGDN* with an increase in respiratory quotient as measured by 24-h room calorimetry. The complete list of single SNVs that met the significance threshold, along with gene names, p-values, MAFs, and effect sizes are listed in Table 3.

Discussion

Whole exome sequencing revealed a novel obesity gene, *PEXI*, in the VIVA LA FAMILIA cohort of 916 Hispanic children. *PEXI* variants were strongly associated with multiple indices of obesity – weight, BMI, waist circumference, fat mass and trunk fat mass. The non-synonymous SNV in *PEXI* (rs141510219) found to be responsible for the significant

gene-based associations is rare (MAF = 0.005) and highly penetrant in the ten affected children from three pedigrees. These children would be classified with severe obesity based on their BMI z-scores (2.3 to 4.1) and percent fat mass (33 to 51%).

Defects in *PEX1* present a novel mechanism for the development of childhood obesity. *PEX1* is a member of the ATPase family, a large group of ATPases Associated with diverse cellular Activities (AAA). This protein is often anchored to a peroxisomal membrane and plays a role in the import of proteins into peroxisomes and *de novo* formation of peroxisomes (24). Mutations in *PEX1* are responsible for several peroxisome biogenesis disorders (PBD), with a spectrum of phenotypes from the most severe Zellweger syndrome, characterized by severe neurologic dysfunction, craniofacial abnormalities, and liver dysfunction to Heimler syndrome, at the mildest end of the PBD spectrum, characterized by hearing loss, enamel hypoplasia, and nail abnormalities.

Peroxisomes are involved in critical metabolic pathways in nearly all cells in the body required for health and development (25). *PEX* genes encode peroxins, which are proteins involved in peroxisome assembly. The peroxisome matrix of mammalian cells has over 70 enzymes required for lipid metabolism including α -, β - and ω -oxidation of fatty acids (FA), and other biochemical processes. While peroxisomal β -oxidation does not play a significant role in energy production, it does play a critical role in intermediate metabolism of bile acids, very long chain FA, pristanic acid, polyunsaturated FA, dicarboxylic FA, xenobiotics, prostaglandins and leukotriens (26). Phytanic acid and pristanic acid influence transcription by activating nuclear receptors RXR- α and PPAR- α . Disruption of hypothalamic peroxisomes and reactive oxygen species (ROS) can affect central regulation of energy metabolism (27). Peroxisomal proliferation in POMC neurons by a PPAR- δ agonist decreased ROS levels and increased food intake in lean mice fed a high fat diet; the opposite occurred with PPAR- δ antagonist.

Most recently, a gene-based meta-analysis applied to the GIANT data replicated known variants (*FTO*, *TMEM18*, *MC4R*, *ADCY3*), but also identified six novel variants including *PEX2* for BMI, substantiating the possible role of *PEX* genes in the regulation of body weight (28). The *PEX1* variant (rs141510219) was not assayed in the GIANT GWAS study, most likely because it is a very rare variant. In the latest ExAC dataset comprising 60,706 unrelated individuals (29), the average minor allele frequency for rs141510219 was 0.00028, with the highest frequency in the Latino population (0.00139), and lower frequencies in the European population (0.00023) and others. In contrast, in our VIVA cohort this variant was present at frequency of 0.005. Although the functionality of the non-synonymous SNV in *PEX1* will require further study, the strong associations seen with obesity phenotypes in these Hispanic children and the critical metabolic pathways involving *PEX1* present a novel mechanism underlying the development of obesity. We recognize that the lack of independent replication and functional validation of the *PEX1* findings are limitations of this study, but hope this result will spur further investigation of the potential link between peroxisomal biogenesis and metabolism and obesity.

Several other rare variants associated with obesity-related traits were identified. Because the genetic architecture of rare variants acting together in complex diseases is variable, we used

two gene level tests to examine different scenarios (30). No gene was significant in both tests, pointing to the complementary nature, and usefulness of applying several tests because the effect and directionality of aggregated variants are not known *a priori*. Ultimately, we also found that most gene level tests were driven by one rare SNV.

Unique to the VIVA study design was the measurement of 24-h energy expenditure in room respiration calorimeters. We found that the rate of sleeping energy expenditure was significantly associated with rare variants in *SOCS4*. The protein encoded by the *SOCS4* gene belongs to the suppressor of cytokine signaling (SOCS) protein family. Obesity is accompanied by increased proinflammatory and decreased anti-inflammatory cytokines which contribute to local and systemic inflammation and disturbances in glucose homeostasis (31). Animal studies have demonstrated the effect of proinflammatory cytokines on energy expenditure (32, 33). Variants in the gene *CTH* were associated with increased levels of plasma cystathionine. The *CTH* gene encodes a cytoplasmic enzyme that converts cystathionine into cysteine. The rs28941785 variant was associated with cystathionine levels in an exome-chip based study of the plasma metabolome (34). In VIVA, plasma cysteine was positively associated with obesity and insulin resistance, highlighting a link between the sulfur amino acid metabolic pathway and obesity and cardiometabolic risk (35). The association of rs200573126, a rare variant in the *ADIPOQ* gene with serum adiponectin levels has been reported previously in an exome sequencing study of Hispanic Americans (36). Previous GWAS studies have reported association of common variants in the *ADIPOQ* gene with adiponectin levels (37). In our exome dataset we assayed two common variants: synonymous variant rs2241766 and missense variant rs17366743, and neither was associated with adiponectin levels in the VIVA cohort.

We also identified common variants that have not been identified by previous GWAS of BMI (3, 5–9). A missense variant in *PPAR- α* was associated with dietary protein (%energy). Dietary protein (%energy) which was higher among the obese children may have been a marker for obesity (38). Respiratory quotient (RQ) was associated with two synonymous SNV in neuroguidin (*NGDN*), an EIF4E binding protein involved with the development of the vertebrate nervous system (39). In addition, we replicated previous VIVA GWAS findings for *KLKB1*, *DARC*, *OR10J3* and *SLC2A9* (6, 22, 23).

In summary, the convergence of whole exome sequencing of a large cohort, a family-based design, and extensive phenotyping resulted in the discovery of novel rare and common variants and replication of previously reported variants associated with childhood obesity. Our major finding linking *PEXI* to obesity phenotypes poses a novel mechanism of peroxisomal biogenesis and metabolism underlying the development of childhood obesity. Replication of the *PEXI* variants as well as other variants associated with related endometabolic traits and functional studies are necessary to validate these findings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

FUNDING: Whole exome sequencing, variant identification, and association analyses were supported by NIH/NHGRI (U54 HG003273-12 and 1UM1HG008898). The VIVA cohort collection and phenotyping was supported by NIH (R01 DK59264, DK092238 and DK080457) and USDA/ARS under Cooperative Agreement 58-6250-51000. This investigation was partly conducted in facilities constructed with support from NIH/NCRR (C06 RR013556 and C06 RR017515).

We thank the families who participated in the VIVA LA FAMILIA study. The authors wish to acknowledge the contributions of Jennifer E. Below with pedigree reconstruction, and Andrew Carroll with SNV identification.

References

- Butte NF, Cai G, Cole SA, Comuzzie AG. Viva la Familia Study: genetic and environmental contributions to childhood obesity and its comorbidities in the Hispanic population. *The American journal of clinical nutrition*. 2006; 84(3):646–54. quiz 73–4. Epub 2006/09/09. [PubMed: 16960181]
- Herrera BM, Lindgren CM. The genetics of obesity. *Current diabetes reports*. 2010; 10(6):498–505. Epub 2010/10/12. DOI: 10.1007/s11892-010-0153-z [PubMed: 20931363]
- Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015; 518(7538):197–206. Epub 2015/02/13. DOI: 10.1038/nature14177 [PubMed: 25673413]
- Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. *Annual review of medicine*. 2012; 63:35–61. Epub 2012/01/18. DOI: 10.1146/annurev-med-051010-162644
- Scherag A, Dina C, Hinney A, Vatin V, Scherag S, Vogel CI, et al. Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and German study groups. *PLoS genetics*. 2010; 6(4):e1000916. Epub 2010/04/28. doi: 10.1371/journal.pgen.1000916 [PubMed: 20421936]
- Comuzzie AG, Cole SA, Laston SL, Voruganti VS, Haack K, Gibbs RA, et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PloS one*. 2012; 7(12):e51954. Epub 2012/12/20. doi: 10.1371/journal.pone.0051954 [PubMed: 23251661]
- Bradfield JP, Taal HR, Timpson NJ, Scherag A, Lecoeur C, Warrington NM, et al. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nature genetics*. 2012; 44(5):526–31. Epub 2012/04/10. DOI: 10.1038/ng.2247 [PubMed: 22484627]
- Namjou B, Keddache M, Marsolo K, Wagner M, Lingren T, Cobb B, et al. EMR-linked GWAS study: investigation of variation landscape of loci for body mass index in children. *Frontiers in genetics*. 2013; 4:268. Epub 2013/12/19. doi: 10.3389/fgene.2013.00268 [PubMed: 24348519]
- Felix JF, Bradfield JP, Monnereau C, van der Valk RJ, Stergiakouli E, Chesi A, et al. Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Human molecular genetics*. 2016; 25(2):389–403. Epub 2015/11/26. DOI: 10.1093/hmg/ddv472 [PubMed: 26604143]
- Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. *Cell*. 2011; 147(1):32–43. Epub 2011/10/04. DOI: 10.1016/j.cell.2011.09.008 [PubMed: 21962505]
- Marth GT, Yu F, Indap AR, Garimella K, Gravel S, Leong WF, et al. The functional spectrum of low-frequency coding variation. *Genome biology*. 2011; 12(9):R84. Epub 2011/09/16. doi: 10.1186/gb-2011-12-9-r84 [PubMed: 21917140]
- Bainbridge MN, Wang M, Wu Y, Newsham I, Muzny DM, Jefferies JL, et al. Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. *Genome biology*. 2011; 12(7):R68. Epub 2011/07/27. doi: 10.1186/gb-2011-12-7-r68 [PubMed: 21787409]
- Reid JG, Carroll A, Veeraraghavan N, Dahdouli M, Sundquist A, English A, et al. Launching genomics into the cloud: deployment of Mercury, a next generation sequence analysis pipeline. *BMC bioinformatics*. 2014; 15:30. Epub 2014/01/31. doi: 10.1186/1471-2105-15-30 [PubMed: 24475911]

14. Yu B, Pulit SL, Hwang SJ, Brody JA, Amin N, Auer PL, et al. Rare Exome Sequence Variants in CLCN6 Reduce Blood Pressure Levels and Hypertension Risk. *Circulation Cardiovascular genetics*. 2015; Epub 2015/12/15. doi: 10.1161/circgenetics.115.001215
15. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *American journal of human genetics*. 1998; 62(5):1198–211. Epub 1998/05/23. DOI: 10.1086/301844 [PubMed: 9545414]
16. Staples J, Qiao D, Cho MH, Silverman EK, Nickerson DA, Below JE. PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *American journal of human genetics*. 2014; 95(5):553–64. Epub 2014/12/03. DOI: 10.1016/j.ajhg.2014.10.005 [PubMed: 25439724]
17. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genetic epidemiology*. 2015; 39(4):276–93. Epub 2015/03/27. DOI: 10.1002/gepi.21896 [PubMed: 25810074]
18. Arie Voorman, JB., Chen, Han, Lumley, Thomas. 2015. Available from: <https://cran.r-project.org/web/packages/seqMeta/index.html>
19. Sinnwell JP, Therneau TM, Schaid DJ. The kinship2 R package for pedigree data. *Human heredity*. 2014; 78(2):91–3. Epub 2014/07/31. DOI: 10.1159/000363105 [PubMed: 25074474]
20. Chen H, Meigs JB, Dupuis J. Sequence kernel association test for quantitative traits in family samples. *Genetic epidemiology*. 2013; 37(2):196–204. Epub 2013/01/03. DOI: 10.1002/gepi.21703 [PubMed: 23280576]
21. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*. 2012; 485(7397):242–5. DOI: 10.1038/nature11011 [PubMed: 22495311]
22. Voruganti VS, Laston S, Haack K, Mehta NR, Cole SA, Butte NF, et al. Serum uric acid concentrations and SLC2A9 genetic variation in Hispanic children: the Viva La Familia Study. *The American journal of clinical nutrition*. 2015; 101(4):725–32. Epub 2015/04/04. DOI: 10.3945/ajcn.114.095364 [PubMed: 25833971]
23. Voruganti VS, Laston S, Haack K, Mehta NR, Smith CW, Cole SA, et al. Genome-wide association replicates the association of Duffy antigen receptor for chemokines (DARC) polymorphisms with serum monocyte chemoattractant protein-1 (MCP-1) levels in Hispanic children. *Cytokine*. 2012; 60(3):634–8. Epub 2012/09/29. DOI: 10.1016/j.cyto.2012.08.029 [PubMed: 23017229]
24. Smith JJ, Aitchison JD. Peroxisomes take shape. *Nature reviews Molecular cell biology*. 2013; 14(12):803–17. Epub 2013/11/23. DOI: 10.1038/nrm3700 [PubMed: 24263361]
25. Braverman NE, Raymond GV, Rizzo WB, Moser AB, Wilkinson ME, Stone EM, et al. Peroxisome biogenesis disorders in the Zellweger spectrum: An overview of current diagnosis, clinical manifestations, and treatment guidelines. *Molecular genetics and metabolism*. 2016; 117(3):313–21. Epub 2016/01/12. DOI: 10.1016/j.ymgme.2015.12.009 [PubMed: 26750748]
26. Van Veldhoven PP. Biochemistry and genetics of inherited disorders of peroxisomal fatty acid metabolism. *Journal of lipid research*. 2010; 51(10):2863–95. Epub 2010/06/19. DOI: 10.1194/jlr.R005959 [PubMed: 20558530]
27. Diano S, Liu ZW, Jeong JK, Dietrich MO, Ruan HB, Kim E, et al. Peroxisome proliferation-associated control of reactive oxygen species sets melanocortin tone and feeding in diet-induced obesity. *Nature medicine*. 2011; 17(9):1121–7. Epub 2011/08/30. DOI: 10.1038/nm.2421
28. Hagg S, Ganna A, Van Der Laan SW, Esko T, Pers TH, Locke AE, et al. Gene-based meta-analysis of genome-wide association studies implicates new loci involved in obesity. *Human molecular genetics*. 2015; 24(23):6849–60. Epub 2015/09/18. DOI: 10.1093/hmg/ddv379 [PubMed: 26376864]
29. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016; 536(7616):285–91. Epub 2016/08/19. DOI: 10.1038/nature19057 [PubMed: 27535533]
30. Lange LA, Hu Y, Zhang H, Xue C, Schmidt EM, Tang ZZ, et al. Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *American journal of human genetics*. 2014; 94(2):233–45. Epub 2014/02/11. DOI: 10.1016/j.ajhg.2014.01.010 [PubMed: 24507775]

31. Mancuso P. The role of adipokines in chronic inflammation. *ImmunoTargets and therapy*. 2016; 5:47–56. Epub 2016/08/17. DOI: 10.2147/itt.s73223 [PubMed: 27529061]
32. Li F, Li Y, Tan B, Wang J, Duan Y, Guo Q, et al. Alteration of inflammatory cytokines, energy metabolic regulators, and muscle fiber type in the skeletal muscle of postweaning piglets. *Journal of animal science*. 2016; 94(3):1064–72. Epub 2016/04/12. DOI: 10.2527/jas.2015-9646 [PubMed: 27065268]
33. Wernstedt I, Edgley A, Berndtsson A, Faldt J, Bergstrom G, Wallenius V, et al. Reduced stress- and cold-induced increase in energy expenditure in interleukin-6-deficient mice. *American journal of physiology Regulatory, integrative and comparative physiology*. 2006; 291(3):R551–7. Epub 2006/02/04. DOI: 10.1152/ajpregu.00514.2005
34. Rhee EP, Yang Q, Yu B, Liu X, Cheng S, Deik A, et al. An exome array study of the plasma metabolome. *Nature Communications*. 2016; 7:12360. <http://www.nature.com/articles/ncomms12360#supplementary-information>. doi: 10.1038/ncomms12360
35. Elshorbagy AK, Valdivia-Garcia M, Refsum H, Butte N. The association of cysteine with obesity, inflammatory cytokines and insulin resistance in Hispanic children and adolescents. *PloS one*. 2012; 7(9):e44166. Epub 2012/09/18. doi: 10.1371/journal.pone.0044166 [PubMed: 22984471]
36. Bowden DW, An SS, Palmer ND, Brown WM, Norris JM, Haffner SM, et al. Molecular basis of a linkage peak: exome sequencing and family-based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study. *Human molecular genetics*. 2010; 19(20):4112–20. Epub 2010/08/07. DOI: 10.1093/hmg/ddq327 [PubMed: 20688759]
37. Hivert MF, Manning AK, McAteer JB, Florez JC, Dupuis J, Fox CS, et al. Common variants in the adiponectin gene (ADIPOQ) associated with plasma adiponectin levels, type 2 diabetes, and diabetes-related quantitative traits: the Framingham Offspring Study. *Diabetes*. 2008; 57(12):3353–9. Epub 2008/09/09. DOI: 10.2337/db08-0700 [PubMed: 18776141]
38. Wilson TA, Adolph AL, Butte NF. Nutrient adequacy and diet quality in non-overweight and overweight Hispanic children of low socioeconomic status: the Viva la Familia Study. *Journal of the American Dietetic Association*. 2009; 109(6):1012–21. Epub 2009/05/26. DOI: 10.1016/j.jada.2009.03.007 [PubMed: 19465183]
39. Jung MY, Lorenz L, Richter JD. Translational control by neuroguidin, a eukaryotic initiation factor 4E and CPEB binding protein. *Molecular and cellular biology*. 2006; 26(11):4277–87. Epub 2006/05/18. DOI: 10.1128/mcb.02470-05 [PubMed: 16705177]

Study Importance

- Genome-wide linkage and genome-wide association studies (GWAS) have identified a number of putative chromosomal regions and loci, but they account for a small fraction and even jointly only a modest proportion of the familial risk for obesity
- convergence of whole exome sequencing, a family-based design, and extensive phenotyping resulted in discovery of novel rare and common variants associated with childhood obesity in a cohort of 928 Hispanic children
- major finding is linking peroxisome biogenesis factor 1 (*PEXI*) to obesity phenotypes, a novel mechanism of peroxisomal biogenesis and metabolism underlying the development of childhood obesity

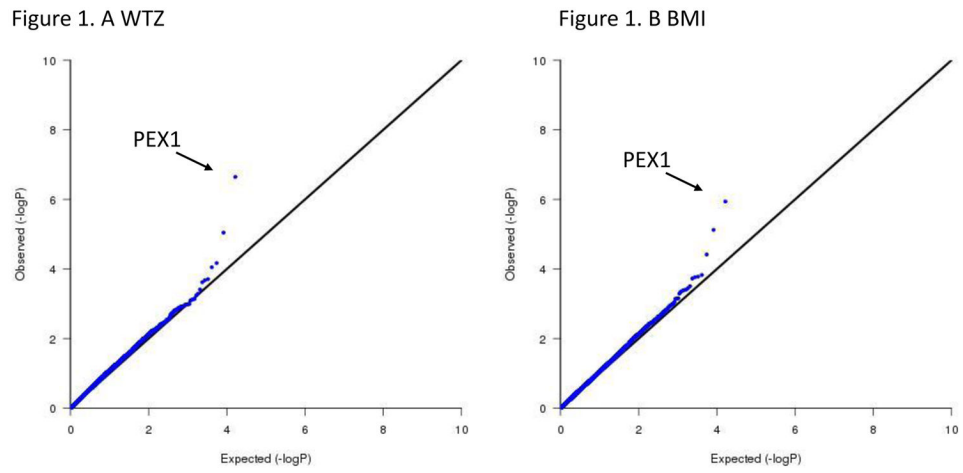


Figure 1. Weight z-score (A) and BMI (B) gene-based association results with *PEX1* gene p-value highlighted; observed vs expected $-\log p$ values (QQ plot)

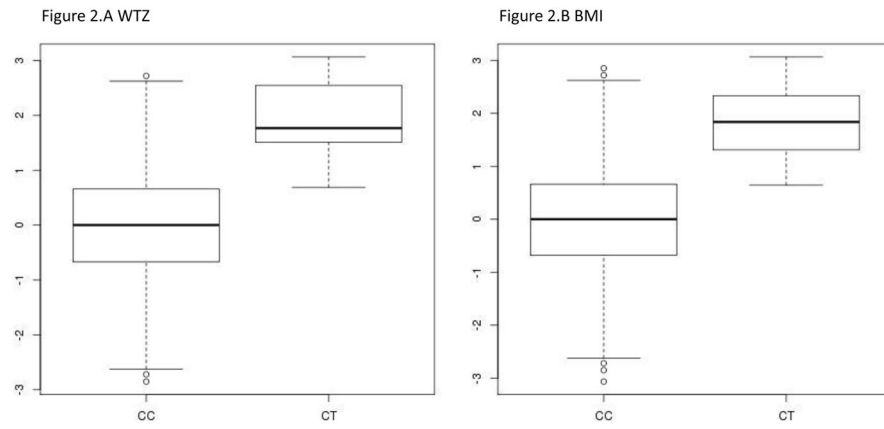


Figure 2. Weight Z (A) and BMI (B) normalized phenotype distribution by *PEXI* rs141510219 genotype: homozygous reference allele (C/C), heterozygous variant (C/T).

Table 1

Frequencies of SNVs in VIVA children (N=916)

| Type of Mutation | Singletons | Singletons (%) | Doubletons | Doubletons (%) | Triple or more | Triple or more (%) | Total |
|--|------------|----------------|------------|----------------|----------------|--------------------|---------|
| Missense | 40,704 | 61.3 | 32,944 | 59.8 | 73,060 | 52.9 | 146,708 |
| Nonsense | 973 | 1.5 | 755 | 1.4 | 1,069 | 0.8 | 2,797 |
| Synonymous | 24,712 | 37.2 | 21,398 | 38.8 | 64,002 | 46.3 | 110,112 |
| Total | 66,389 | 100 | 55,097 | 100 | 13,8131 | 100 | 259,617 |
| Polyphen2 missense classification | | | | | | | |
| Benign * | 22,186 | 55.3 | 18,158 | 56.0 | 47,253 | 65.9 | 87,597 |
| Possibly damaging * | 6,774 | 16.9 | 5,615 | 17.3 | 10,486 | 14.6 | 22,875 |
| Probably damaging * | 11,130 | 27.8 | 8,671 | 26.7 | 13,919 | 19.4 | 33,720 |
| Total ** | 40,090 | 100 | 32,444 | 100 | 71,658 | 100 | 144,192 |

* Functional prediction according to Polyphen2. For variants where different splice forms had different predictions, the most damaging was chosen

** Total not matching above missense count because not all variants had Polyphen2 prediction

Table 2

Significant gene level associations

| Phenotype | Gene | SKAT p-value | WST p-value | Number of SNVs | Cumulative MAF |
|---|---------------|-------------------------------|-------------------------------|----------------|----------------|
| Anthropometry and Body Composition | | | | | |
| Weight (kg) | <i>PEXI</i> | 6.59 × 10⁻⁷ | 0.4458 | 17 | 0.022 |
| Weight for age z-score | <i>PEXI</i> | 2.26 × 10⁻⁷ | 0.4275 | 17 | 0.022 |
| BMI (kg/m ²) | <i>PEXI</i> | 1.14 × 10⁻⁶ | 0.4655 | 17 | 0.022 |
| BMI z-score | <i>PEXI</i> | 1.34 × 10⁻⁶ | 0.6811 | 17 | 0.022 |
| Waist circumference (cm) | <i>PEXI</i> | 1.49 × 10⁻⁶ | 0.3672 | 17 | 0.022 |
| Fat mass (kg) | <i>PEXI</i> | 1.66 × 10⁻⁶ | 0.6009 | 17 | 0.023 |
| Trunk fat mass (kg) | <i>PEXI</i> | 6.55 × 10⁻⁷ | 0.4814 | 17 | 0.024 |
| Systolic blood pressure (mmHg) | <i>CLIC6</i> | 0.0004 | 1.01 × 10⁻⁶ | 6 | 0.007 |
| Energy Expenditure | | | | | |
| Sleeping energy expenditure (kcal/d) | <i>SOCS4</i> | 4.23 × 10⁻⁷ | 0.2712 | 2 | 0.01 |
| Biochemistries | | | | | |
| Adiponectin (ng/mL) | <i>ADIPOQ</i> | 2.1 × 10⁻¹¹ | 2.85 × 10 ⁻⁵ | 4 | 0.009 |
| Plasma cystathionine (μmol/L) | <i>CTH</i> | 2.12 × 10⁻⁷ | 0.1539 | 5 | 0.01 |
| Free T4 (ng/dL) | <i>LINGO1</i> | 1.6 × 10⁻⁷ | 0.4982 | 4 | 0.01 |
| Serum IGF-1 free (ng/mL) | <i>ABI3</i> | 8.75 × 10⁻⁷ | 0.2054 | 6 | 0.012 |
| Serum IGF-1 free (ng/mL) | <i>SALL3</i> | 1.32 × 10⁻⁶ | 0.1043 | 4 | 0.007 |
| Urinary dopamine/creatinine ratio | <i>TARSL2</i> | 3.16 × 10 ⁻⁵ | 7.13 × 10⁻⁷ | 4 | 0.003 |
| HDL (mg/dL) | <i>FHDC1</i> | 0.0241 | 1.5 × 10⁻⁶ | 23 | 0.04 |

Abbreviations: SKAT, Sequence Kernel Association Test; WST, Weighted Sums Test; SNV, single nucleotide variant; MAF, minor allele frequency; BMI, body mass index; T4, thyroxine; IGF-1, insulin-like growth factor-1; HDL, high density lipoproteins.

For each significant association both Sequence Kernel Association Test (SKAT) and Weighted Sums Test (WST) p-values are provided, with significant findings highlighted in **bold**. The number of SNVs used in the test, and the cumulative MAF is listed. Gene level significance cutoff p-value 1.72 × 10⁻⁶

Table 3

Significant single variant level associations

| Trait | Gene | RS ID | SNV type | p-value | Minor allele frequency | Effect size |
|--------------------------------|---------------|-------------|------------|------------------------|------------------------|-------------|
| <i>Novel associations</i> | | | | | | |
| Diet protein (% energy) | <i>PPARA</i> | rs1800234 | missense | 1.05×10^{-7} | 0.041 | 0.632 |
| Total energy expenditure RQ | <i>NGDN</i> | rs2236261 | synonymous | 3.37×10^{-7} | 0.31 | 0.274 |
| Total energy expenditure RQ | <i>NGDN</i> | rs2236260 | synonymous | 8.94×10^{-7} | 0.35 | 0.258 |
| Serum AST (U/L) | <i>CYB5R2</i> | rs61729556 | missense | 6.44×10^{-7} | 0.017 | -0.924 |
| Testosterone (ng/mL) | <i>MUC21</i> | rs112415706 | missense | 6.75×10^{-7} | 0.11 | -0.408 |
| HOMA-IR | <i>UBXN11</i> | rs7522905 | synonymous | 7.04×10^{-7} | 0.40 | 0.2512 |
| HDL (mg/dL) | <i>ATPIA4</i> | rs76528638 | missense | 7.13×10^{-7} | 0.016 | -1.003 |
| Eotaxin (pg/mL) | <i>ACKR2</i> | rs1366046 | 3' UTR | 7.25×10^{-7} | 0.34 | -0.252 |
| <i>Replicated associations</i> | | | | | | |
| Serum MCP-1 (pg/mL) | <i>DARC</i> | rs12075 | missense | 8.52×10^{-27} | 0.44 | 0.533 |
| Serum MCP-1 (pg/mL) | <i>OR10J3</i> | rs58037016 | missense | 3.2×10^{-7} | 0.13 | -0.363 |
| Serum IGF-1 free (ng/mL) | <i>KLKB1</i> | rs3733402 | missense | 7.85×10^{-10} | 0.34 | -0.309 |
| Serum IGF-1 free (ng/mL) | <i>KLKB1</i> | rs925453 | synonymous | 1.68×10^{-8} | 0.26 | -0.304 |
| Serum uric acid (mg/dL) | <i>SLC2A9</i> | rs16890979 | missense | 6.1×10^{-13} | 0.39 | -0.364 |
| Serum uric acid (mg/dL) | <i>SLC2A9</i> | rs13125646 | synonymous | 1.53×10^{-11} | 0.39 | 0.340 |
| Serum uric acid (mg/dL) | <i>SLC2A9</i> | rs13113918 | synonymous | 1.93×10^{-11} | 0.24 | 0.387 |
| Serum uric acid (mg/dL) | <i>SLC2A9</i> | rs10939650 | synonymous | 8.42×10^{-10} | 0.32 | 0.330 |

Abbreviations: RQ, respiratory quotient; AST, aspartate aminotransferase; HOMA-IR, homeostatic model assessment-insulin resistance; HDL, high density lipoproteins; Eotaxin, fasting serum eotaxin; MCP-1, monocyte chemoattractant protein-1; IGF-1, insulin-like growth factor-1. Single variant significance cutoff p-value 9.2×10^{-7} .