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

Objective: To perform whole exome sequencing in 928 Hispanic children and identify variants and genes associated with childhood obesity.

Methods: Single-nucleotide variants (SNVs) were identified from Illumina whole exome sequencing data using integrated read mapping, variant calling, and an 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 the SNV level.

Results: (1) 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 were identified, (2) novel rare variants in peroxisome biogenesis factor 1 (PEX1) associated with several obesity traits (weight, weight zscore, BMI, BMI zscore, waist circumference, fat mass, trunk fat mass) were discovered, and (3) previously reported SNVs associated with childhood obesity were replicated.

Conclusions: Convergence of whole exome sequencing, a family-based design, and extensive phenotyping discovered novel rare and common variants associated with childhood obesity. Linking PEX1 to obesity phenotypes poses a novel mechanism of peroxisomal biogenesis and metabolism underlying the development of childhood obesity.

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 BMI, adiposity, and related traits indicate heritability on the order of 0.4 to 0.7 (2). Genome-wide linkage and genome-wide association studies (GWAS) have identified several putative chromosomal regions and loci; however, they accounted for a small fraction and, even jointly, only a modest proportion of familial risk. A recent GWAS meta-analysis of BMI in 339,224 adults identified 97 loci that accounted for about 2.7% of BMI variation (3). Most GWAS have focused 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 have been identified, major common variants related to obesity overlapped to a substantial degree between children and adults. GWAS results have 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.

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 nonsynonymous 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**

The VIVA cohort used for exome sequencing comprised 293 Hispanic families with an average family size of 5 persons (range 3-10) (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 of 4 and 19 years. 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 with obesity. Among the children with obesity, 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 (1,457 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, birth weights from Texas birth records, diet by 24-hour recalls, total energy expenditure and substrate utilization by 24-hour 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 the consensus coding sequence, NCBI RNA reference sequences, and Vega human genome annotations. Capture enrichment was followed by sequencing on the 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 Burrows-Wheeler Aligner (BWA). The Atlas suite was used to call single-nucleotide variants (SNVs). Lastly, variant annotation was accomplished through the Casandra 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 quality control (QC) procedures were performed using a custom pipeline following the guidelines established for the CHARGE-S project (14). The QC procedure details are provided in online Supporting Information.

**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 age^2 as covariates. For the birth weight 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 (PCA) was conducted using the PC-AiR method as implemented in the GENESIS package (17). Unlike standard PCA methods (e.g., EIGENSTRAT), the method implemented in PC-AiR 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 10 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 GWAS. 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 tests. 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 Test (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 130 times, and on average 92% of the bases targeted had at least 20 reads coverage. Genotype and variant-level QC analyses removed 12% of variants that did not pass our QC metrics, and 10 samples that failed pedigree information verification. The final data set comprised 283,587 SNVs identified in 916 samples. The data set included 150,718 SNVs that have the potential to affect the protein function by either affecting a splice site (1,213 splice site variants) or 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 PolyPhen-2. 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 (triplets and more). The data have been submitted to dbGAP Study Accession phs000616.

Gene-based association analysis for rare variants

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 Supporting Information Table S1. The $P$ value for significance was established at $\leq 1.72 \times 10^{-6}$ 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, 3 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 ($pskat = 2.1 \times 10^{-11}$). 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 a significant association of variants in cystathionine gamma-lyase (CTH) with serum cystathionine levels ($pskat = 2.12 \times 10^{-7}$). 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, pThr67Ile, MAF = 0.004) passed the nominal significance value ($P = 1.7 \times 10^{-7}$, effect size $\beta = 1.9769$). Similar to ADIPOQ, we removed the 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 (PEX1) 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 ($pskat = 2.26 \times 10^{-7}$), followed by trunk fat mass ($pskat = 6.55 \times 10^{-7}$), weight ($pskat = 6.59 \times 10^{-7}$), and BMI ($pskat = 1.14 \times 10^{-6}$) (Figure 1). To further investigate the association of PEX1 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 1 missense variant (rs141510219, p.Val734Ile, MAF = 0.005) of the 17 SNVs had a nominally significant association ($P = 8.97 \times 10^{-8}$, effect size

| Table 1 Frequencies of SNVs in VIVA children (N = 916) |
|---------------------------------------------|
| **Type of mutation** | **Singletons (%)** | **Doubletons (%)** | **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 | 138,131 | 100 | 259,617 |

| PolyPhen-2 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 |

$^{a}$Functional prediction according to PolyPhen-2. For variants in which different splice forms had different predictions, the most damaging was chosen.
$^{b}$Total not matching above missense count because not all variants had PolyPhen-2 prediction.
\( \beta = 1.8638 \), 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 a 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-hour room calorimetry. The association was based on two rare predicted protein-altering variants (cumulative MAF = 0.01). We looked at the single-variant level

| Phenotype                              | Gene     | SKAT \( P \) value | WST \( P \) value | Number of SNVs | Cumulative MAF |
|----------------------------------------|----------|---------------------|-------------------|----------------|---------------|
| Anthropometry and body composition     |          |                     |                   |                |               |
| Weight (kg)                            | \( PEX1 \) | \( 6.59 \times 10^{-7} \) | 0.4458            | 17             | 0.022         |
| Weight for age \( z \) score           | \( PEX1 \) | \( 2.26 \times 10^{-7} \) | 0.4275            | 17             | 0.022         |
| BMI (kg/m\(^2\))                       | \( PEX1 \) | \( 1.14 \times 10^{-6} \) | 0.4655            | 17             | 0.022         |
| BMI \( z \) score                      | \( PEX1 \) | \( 1.34 \times 10^{-6} \) | 0.6811            | 17             | 0.022         |
| Waist circumference (cm)               | \( PEX1 \) | \( 1.49 \times 10^{-6} \) | 0.3672            | 17             | 0.022         |
| Fat mass (kg)                          | \( PEX1 \) | \( 1.66 \times 10^{-6} \) | 0.6009            | 17             | 0.023         |
| Trunk fat mass (kg)                    | \( PEX1 \) | \( 6.55 \times 10^{-7} \) | 0.4814            | 17             | 0.024         |
| Systolic blood pressure (mmHg)         | \( CLIC6 \) | 0.0004              | \( 1.01 \times 10^{-6} \) | 6              | 0.007         |
| Energy expenditure                     |          |                     |                   |                |               |
| Sleeping energy expenditure (kcal/d)   | \( SOCS4 \) | \( 4.23 \times 10^{-7} \) | 0.2712            | 2              | 0.01          |
| Biochemistries                          |          |                     |                   |                |               |
| Adiponectin (ng/mL)                    | \( ADIPOQ \) | \( 2.1 \times 10^{-11} \) | \( 2.85 \times 10^{-5} \) | 4              | 0.009         |
| Plasma cystathionine (\( \mu \)mol/L) | \( CTH \) | \( 2.12 \times 10^{-7} \) | 0.1539            | 5              | 0.01          |
| Free T4 (ng/dL)                        | \( LINGO1 \) | \( 1.6 \times 10^{-7} \) | 0.4982            | 4              | 0.01          |
| Serum IGF-1 free (ng/mL)               | \( ABI3 \) | \( 8.75 \times 10^{-7} \) | 0.2054            | 6              | 0.012         |
| Serum IGF-1 free (ng/mL)               | \( SALL3 \) | \( 1.32 \times 10^{-6} \) | 0.1043            | 4              | 0.007         |
| Urinary dopamine/creatinine ratio      | \( TARSL2 \) | \( 3.16 \times 10^{-5} \) | \( 7.13 \times 10^{-7} \) | 4              | 0.003         |
| HDL (mg/dL)                            | \( FHDC1 \) | 0.0241              | \( 1.5 \times 10^{-6} \) | 23             | 0.04          |

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 are listed. Gene level significance cutoff \( P \) value \( \leq 1.72 \times 10^{-6} \).

Abbreviations: SNV, single-nucleotide variant; MAF, minor allele frequency; T4, thyroxine; IGF-1, insulin-like growth factor-1; HDL, high-density lipoproteins.

Figure 1 (A) Weight for age \( z \) score and (B) BMI gene-based association results, with \( P \) value highlighted; observed versus expected \(-\log P\) values (QQ plot).
associations for the two underlying SNVs. Only one variant (rs146421724, p.Gly54Ser, MAF = 0.0096) passed the nominal significance value (P = 3.59 × 10⁻⁷, effect size β = 1.4778). The complete list of significant gene-based findings, along with P values, number of underlying SNVs, and cumulative MAFs, is provided in Table 2.

Single-variant association analysis for common variants

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 SNVs 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 × 10⁻¹³). We have replicated associations of missense SNV (rs12075) in Duffy antigen receptor for chemokines (DARC) (P = 1.46 × 10⁻²⁷) and missense SNV (rs58037016) in olfactory receptor family 10 subfamily J member 3 (OR10J3) (P = 3.2 × 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 Kallikrein B1 (KLKB1) with levels of insulin-like growth factor-1 (P = 7.85 × 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 peroxisome proliferator-activated receptor alpha (PPAR-α) with increased diet protein levels (P = 1.05 × 10⁻⁷, effect size β = 0.632). We also discovered association of two synonymous SNVs (rs2236261 and rs2236260) in neuroguidin (NGDN) with an increase in respiratory quotient as measured by 24-hour room calorimetry. The complete list of single SNVs that met the significance threshold, along with names, P values, MAFs, and effect sizes, is provided in Table 3.

Discussion

Whole exome sequencing revealed a novel obesity gene, PEX1, in the VIVA LA FAMILIA cohort of 916 Hispanic children. PEX1 variants were strongly associated with multiple indices of obesity—weight, BMI, waist circumference, fat mass, and trunk fat mass. The nonsynonymous SNV in PEX1 (rs141510219) found to be responsible for the significant gene-based associations is rare (MAF = 0.005) and highly penetrant in the 10 affected children from three pedigrees. These children would be classified with severe obesity based on their BMI z scores (2.3-4.1) and percent fat mass (33%-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. 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, 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 peroxisome biogenesis disorders 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 more than 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, proteoglycans, 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 can affect central regulation of energy metabolism (27). Peroxisomal proliferation in POMC neurons by a PPAR-δ agonist decreased reactive oxygen species levels and increased food intake in lean mice fed a high-fat diet; the opposite occurred with a 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 data set comprising 60,706 unrelated individuals (29), the average MAF 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 a frequency of 0.005. Although the functionality of the nonsynonymous 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 is 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-hour 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 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).
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 data set, 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 children with obesity, may have been a marker for obesity (38). Respiratory quotient was associated with two synonymous SNVs in NUDN, an EIL4E binding protein involved with the development of the vertebrate nervous system (39). In addition, we replicated previous VIVA GWAS findings for KLF11, DARC, OR10U3, 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 PEX1 to obesity phenotypes poses a novel mechanism of peroxisomal biogenesis and metabolism underlying the development of childhood obesity. Replication of the PEX1 variants, as well as other variants associated with related endometabolic traits and functional studies, is necessary to validate these findings.

Acknowledgments

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.

© 2017 The Obesity Society

References

1. 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. Am J Clin Nutr 2006;84:646-54; quiz 673-674.
2. Herrera BM, Lindgren CM. The genetics of obesity. Curr Diab Rep 2010;10:498-505.
3. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 2015;518:197-206.
4. Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. Annu Rev Med 2012;63:35-61.
5. Scherag A, Dina C, Hinney A, 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 Genet 2010;6:e1000916. doi:10.1371/journal.pgen.1000916.
6. Comuzzie AG, Cole SA, Laston SL, et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. PLoS One 2012;7:e51954. doi:10.1371/journal.pone.0051954.
7. Bradfield JP, Taal HR, Timpson NJ, et al. A genome-wide association meta-analysis identifies new childhood obesity loci. Nature Genet 2012;44:526-531.
8. Namjou B, Keddache M, Marsolo K, et al. EMR-linked GWAS study: investigation of variation landscape of loci for body mass index in children. Front Genet 2013;4:268. doi:10.3389/fgene.2013.00268.
9. Felix JF, Bradfield JP, Monnereau C, et al. Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. Hum Mol Genet 2016;25:389-403.
10. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. Cell 2011;147:32-43.