

Novel Genetic Loci Identified for the Pathophysiology of Childhood Obesity in the Hispanic Population

Anthony G. Comuzzie¹, Shelley A. Cole¹, Sandra L. Laston¹, V. Saroja Voruganti¹, Karin Haack¹, Richard A. Gibbs², Nancy F. Butte^{3*}

1 Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas, United States of America, **2** Human Genome Sequencing Center, Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, Texas, United States of America, **3** USDA/ARS Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, United States of America

Abstract

Genetic variants responsible for susceptibility to obesity and its comorbidities among Hispanic children have not been identified. The VIVA LA FAMILIA Study was designed to genetically map childhood obesity and associated biological processes in the Hispanic population. A genome-wide association study (GWAS) entailed genotyping 1.1 million single nucleotide polymorphisms (SNPs) using the Illumina Infinium technology in 815 children. Measured genotype analysis was performed between genetic markers and obesity-related traits i.e., anthropometry, body composition, growth, metabolites, hormones, inflammation, diet, energy expenditure, substrate utilization and physical activity. Identified genome-wide significant loci: 1) corroborated genes implicated in other studies (*MTNR1B*, *ZNF259/APOA5*, *XPA/FOXE1 (TTF-2)*, *DARC*, *CCR3*, *ABO*); 2) localized novel genes in plausible biological pathways (*PCSK2*, *ARHGAP11A*, *CHRNA3*); and 3) revealed novel genes with unknown function in obesity pathogenesis (*MATK*, *COL4A1*). Salient findings include a nonsynonymous SNP (rs1056513) in *INADL* ($p = 1.2E-07$) for weight; an intronic variant in *MTNR1B* associated with fasting glucose ($p = 3.7E-08$); variants in the *APOA5-ZNF259* region associated with triglycerides ($p = 2.5-4.8E-08$); an intronic variant in *PCSK2* associated with total antioxidants ($p = 7.6E-08$); a block of 23 SNPs in *XPA/FOXE1 (TTF-2)* associated with serum TSH ($p = 5.5E-08$ to $1.0E-09$); a nonsynonymous SNP ($p = 1.3E-21$), an intronic SNP ($p = 3.6E-13$) in *DARC* identified for MCP-1; an intronic variant in *ARHGAP11A* associated with sleep duration ($p = 5.0E-08$); and, after adjusting for body weight, variants in *MATK* for total energy expenditure ($p = 2.7E-08$) and in *CHRNA3* for sleeping energy expenditure ($p = 6.0E-08$). Unprecedented phenotyping and high-density SNP genotyping enabled localization of novel genetic loci associated with the pathophysiology of childhood obesity.

Citation: Comuzzie AG, Cole SA, Laston SL, Voruganti VS, Haack K, et al. (2012) Novel Genetic Loci Identified for the Pathophysiology of Childhood Obesity in the Hispanic Population. PLoS ONE 7(12): e51954. doi:10.1371/journal.pone.0051954

Editor: Dana C. Crawford, Vanderbilt University, United States of America

Received: August 15, 2012; **Accepted:** November 7, 2012; **Published:** December 14, 2012

Copyright: © 2012 Comuzzie et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants from the National Institutes of Health (NIH) (DK080457), and the USDA/ARS (Cooperative Agreement 6250-51000-053). Work performed at the Texas Biomedical Research Institute in San Antonio, Texas was conducted in facilities constructed with support from the Research Facilities Improvement Program of the National Center for Research Resources, NIH (C06 RR013556, C06 RR017515). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: nbutte@bcm.edu

Introduction

Obesity is a complex disease influenced by genetic and environmental factors and their interactions. The current surge in childhood obesity in the U.S. is attributable to an interaction between a genetic predisposition toward efficient energy storage and a permissive environment of readily available food and sedentary behaviors [1]. Genetic architecture of common polygenic childhood obesity remains largely unknown. In genetic studies, the phenotypic description of the obese child usually has been limited to body mass index (BMI). BMI represents a composite trait of fat free mass (FFM) and fat mass (FM) and thus loci influencing BMI may differ from more direct measures of adiposity. In addition, markers of biological processes underlying the development of obesity such as dietary intake, energy expenditure and nutrient partitioning may be more effectual in identifying causal genetic variants [2]. In epidemiology studies, childhood obesity has been shown to be genetically correlated with glucose intolerance, hypertension, dyslipidemia, insulin resistance,

chronic inflammation, and risk for fatty liver disease [3,4]. Identification of genes underlying these distinct patterns of association also may unravel important biological pathways involved in the pathophysiology of childhood obesity.

Genome-wide association studies (GWAS) have the potential to localize genetic loci contributing to obesity down to a few 100 kb [5]. In fact, a recent meta-analysis of the adult GIANT Consortium established 32 susceptibility BMI loci [6], several of which were confirmed in French and German children with extreme obesity [7] and European adults with early-onset obesity [8]. Two novel loci near *OLFM4* and within *HOOX5* were recently reported based on a meta-analysis of 14 pediatric studies of BMI [9]. These pediatric GWAS were confined to BMI and cohorts of European ancestry.

Here, we present findings from a GWAS designed to identify genetic variants influencing childhood obesity and its comorbidities in the Hispanic population. We have published evidence of heritability, pleiotropy amongst traits, and chromosomal regions implicated in obesity among Hispanic children in our VIVA LA

FAMILIA Study [4,10–15] In-depth phenotyping was performed to characterize the children, including anthropometry, body composition, growth, metabolites, hormones, inflammation, diet, energy expenditure and substrate utilization and physical activity. Our high-density SNP genotyping and phenotypes representing not only adiposity, but also biological processes associated with the development and consequences of childhood obesity enabled localization of novel genetic loci associated with the pathophysiology of childhood obesity.

Materials and Methods

The VIVA LA FAMILIA Study was designed to identify genetic variants influencing pediatric obesity and its comorbidities. Family recruitment and phenotyping were conducted in 2000–2005 in Houston, TX. 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.

The VIVA LA FAMILIA study design and methodology have been described in detail elsewhere⁴. GWAS was performed on 815 children from 263 Hispanic families. The number of families by sibships was: 8 (one child), 40 (two children), 155 (three children), 48 (four children), 6 (five children), 3 (six children), 2 (seven children) and 1 (eight children). Each family was ascertained on an obese proband, defined as a BMI > 95th percentile, between the ages 4–19 y. Once identified, the obese proband and all siblings, 4 to 19 y of age, and their parents were invited to the Children's Nutrition Research Center for a tour and full explanation of the study prior to consenting. The cross-sectional, longitudinal study design consisted of baseline measurements, with a one-year follow-up to track children's growth and body compositional changes. In-depth baseline phenotyping included vital signs, anthropometry and body composition, diet and physical fitness, 24-h calorimetry, eating behavior, physical activity, fasting blood sampling for DNA and other biochemistries.

Briefly, blood pressure, heart rate and temperature were taken using an automated monitor. Anthropometric measurements were performed using standardized techniques according to Lohman [16]. Body composition was determined by dual-energy x-ray absorptiometry. Repeated measures after one year were used to compute growth velocities and changes in FM, FFM and energy storage [17]. Methods used to measure fasting blood and 24-h urinary biochemistries are described elsewhere [14,18,19]. A multiple-pass 24-h dietary recall was recorded on two occasions using Nutrition Data System (NDS) [20]. Eating behavior was assessed with a dinner meal and eating in the absence of hunger [21]. Room respiration calorimetry was used to make 24-h measurements of energy expenditure and substrate oxidation [13]. Maximum oxygen consumption (VO_2max) and heart rate maximum (HRmax) were measured on a treadmill [22]. Actiwatch accelerometers were used to measure frequency, duration and intensity of physical activity [23].

Genotyping

The Illumina HumanOmni1-Quad v1.0 BeadChips were used to genotype 1.1 million single nucleotide polymorphisms (SNPs) in 815 children enrolled in the VIVA LA FAMILIA Study. Genotype calls were obtained after scanning on the Illumina BeadStation 500GX and analysis using the GenomeStudio software. Our genotyping error rate (based on duplicates) was 2 per 100,000 genotypes. Illumina has included sample-dependent and -independent controls on their BeadChips to test for accuracy of the

procedure. The average call rate for all SNPs per individual sample was 97%.

SNP genotypes were checked for Mendelian consistency using the program SimWalk2 [24]. The estimates of the allele frequencies and their standard errors were obtained using SOLAR [25]. Specific SNPs were removed from analysis if they had call rates < 95% ($n = 6,596$), were uncommon alleles in less than 5 participants (26,537), deviated from Hardy-Weinberg equilibrium ($n = 0$), or were monoallelic ($n = 56,448$). The number of SNPs that passed quality control and were included in the GWA analysis was 899,892.

Genome-wide Association Analysis

Measured genotype analysis (MGA) was performed using the SOLAR program [25]. All phenotypes were transformed by inverse normalization to meet assumptions of normality. We obtained residuals using linear regression models adjusted for age, sex, their interaction and higher order terms. Because energy expenditure is strongly influenced by body weight across the age range of 4 to 19 years of age, total energy expenditure and sleeping energy expenditure were additionally adjusted for body weight. Also, observed energy intakes consumed in a snack and dinner were adjusted for total energy expenditure or estimated energy requirement, again to compensate for the wide range of ages in our cohort.

Each SNP genotype was converted in SOLAR to a covariate measure equal to 0, 1, or 2 copies of the minor allele (or, for missing genotypes, the weighted covariate based on imputation). These covariates were included in the variance-components mixed models for measured genotype analyses [26] versus null models that incorporated the random effect of kinship and fixed effects such as age, sex, their interaction and higher order terms. For the initial GWA screen, we tested each SNP covariate independently as a 1 degree of freedom likelihood ratio test. The p-value threshold for genome-wide significance ($\alpha = 0.05$) was set at 1.01×10^{-7} . The p-value threshold for genome-wide significance was computed for our family-based cohort that takes into account pedigree structure. The effective number of SNPs given linkage disequilibrium (LD) was calculated by the method of Moskvina and Schmidt [27] as implemented in SOLAR. LD was computed in SOLAR using all available information (all genotyped SNPs on all individuals). The average ratio of SNP effective number/actual number obtained from analysis of 1,989 non-overlapping bins of SNPs was used to calculate the genome-wide effective number of tests and thus the significance threshold for genome-wide association. We performed quantitative transmission disequilibrium test (implemented in SOLAR) to test for population stratification.

Results

GWAS was performed on 815 children from 263 Hispanic families. Mean \pm SD (range) was 25.2 ± 7.5 (13.4 to 61.9) for the children's BMI, 85.6 ± 20.8 (4 to 100) for BMI percentile and 1.52 ± 1.01 (-3.0 to 4.5) for BMI z-score. Measured genotype analysis examined 129 obesity-related traits including BMI and adiposity as well as biological processes associated with the pathophysiology of childhood obesity; a description of the phenotypes is provided in **Table S1**. In our GWAS, population stratification was not significant and therefore did not confound our associations. A complete listing of all suggestive ($p < 1.0E-06$) and genome-wide significant ($p < 1.0E-07$) genetic variants and their associated traits are presented by chromosomal position in **Table S2**.

Anthropometry and Body Composition

BMI status, body composition and the growth process were assessed from repeated measurements of body weight, height, FFM and FM at baseline and after one-year (**Table 1**). A nonsynonymous SNP (rs1056513; G1178S (NP_005790.2)) in *INADL* on chromosome 1 attained near genome-wide significance ($p = 1.2E-07$) for weight, and BMI, FFM, FM, trunk FM, and hip circumference ($p = 8.3E-06$ to $1.6E-07$). SNP rs1056513 is common (MAF = 0.50) and accounted for 3% of the variance in body weight and body composition in this cohort. Weight z-score change was significantly associated with an intronic variant in *COL4A1* on chromosome 13 ($p = 4.7E-08$) (Supporting Information **Figure S1**). Linear growth (height change) was associated with a variant in the 5'UTR region of *TSEN34* on chromosome 19 ($p = 4.5E-08$).

Endometabolic Traits

Genome-wide significant variants were identified for several endometabolic traits (**Table 2**). An intronic variant in *MTNR1B* on chromosome 11 was strongly associated with fasting glucose ($p = 3.7E-08$). Intronic and 3'UTR variants in the *APOA5-ζNF259* region on chromosome 11 were associated with triglycerides ($p = 2.5-4.8E-08$). An intronic variant in *PCSK2* on chromosome 20 was associated with total antioxidants ($p = 7.6E-08$). Variants in the flanking 3'UTR regions of *RNASE1* on chromosome 14 and *ASSIP11* on chromosome 7 were associated with 24-h urinary nitrogen excretion and creatinine excretion, respectively ($p = 8.2-8.4E-08$). An intronic SNP in *GCHI* on chromosome 14 was associated with 24-h urinary dopamine: creatinine ratio ($p = 6.3E-08$). A nonsynonymous SNP (rs3733402; S143T (NP_000883.2)) in *KLKB1* on chromosome 4 was identified for serum free IGF-1. An intronic SNP in *MPRIP* on chromosome 17 was associated with serum IGFBP-3 ($p = 7.2E-08$). A block of twenty-three SNPs in the flanking 5'UTR region of *XPA/FOXEL (TTF-2)* on chromosome 9 were highly associated with serum TSH ($p = 5.5E-08$ to $1.0E-09$) and found to be in strong linkage disequilibrium with one another ($R = 0.91-1.00$).

Inflammation Markers

Genome-wide significant variants were identified for inflammation markers (**Table 3**). Highly significant associations for a nonsynonymous SNP (rs12075; G42D (NP_002027.2)) ($p = 1.3E-21$) and an intronic SNP ($p = 3.6E-13$) in *DARC* on chromosome 1 were identified for MCP-1. Coding variants in *GREB1* on chromosome 2 ($p = 6.5E-08$) and *DFNB31* on chromosome 9 ($p = 2.0E-08$) were also identified for MCP-1. A variant in the 3'UTR for *CCR3* was highly associated with MCP1, as well as an intronic SNP in *RASGEF1A* ($p = 4.6-9.6E-08$). A variant in the intronic region of *ABO* was strongly associated with IL-6 ($p = 2.0E-08$).

Diet, Energy Expenditure, Substrate Utilization and Physical Activity

Genome-wide significant variants were associated with energy intake, energy expenditure and substrate utilization, and physical activity (**Table 4**). An intronic variant in *TMEM229B* on chromosome 14 was associated with *ad libitum* energy intake at dinner ($p = 5.1E-08$). An intronic variant in *ARHGAP11A* on chromosome 15 was associated with sleep duration ($p = 5.0E-08$). A SNP in the intronic region of *C21orf34* was detected for respiratory quotient (RQ) during sleep ($p = 5.3E-08$). A variant was identified for accelerometer-measured light activity in the 5'UTR region of *RPL7P3* on chromosome 9 ($p = 7.5E-08$) and sedentary-

Table 1. Measured genotype analysis for anthropometric and body composition traits.

SNP	Chr	Coordinate (GB 36.2)	Location (GB 36.2)	SNP (GB 36.2)	Trait	MGA p-value	Effect Size	Minor Allele	Minor Allele Frequency	Gene Symbol	Gene Name
rs1056513	1	62152886	coding-nonsyn	[A/G]	Weight (kg)	1.18E-07	0.031	A	0.495	<i>INADL</i>	InaD-likelike (Drosophila)
rs1056513	1	62152886	coding-nonsyn	[A/G]	BMI (kg/m ²)	8.34E-06	0.021	A	0.495	<i>INADL</i>	InaD-likelike (Drosophila)
rs1056513	1	62152886	coding-nonsyn	[A/G]	Fat mass (kg)	1.59E-07	0.035	A	0.495	<i>INADL</i>	InaD-likelike (Drosophila)
rs1056513	1	62152886	coding-nonsyn	[A/G]	Trunk fat mass (kg)	2.36E-07	0.035	A	0.495	<i>INADL</i>	InaD-likelike (Drosophila)
rs1056513	1	62152886	coding-nonsyn	[A/G]	Fat free mass (kg)	2.80E-07	0.034	A	0.495	<i>INADL</i>	InaD-likelike (Drosophila)
rs1056513	1	62152886	coding-nonsyn	[A/G]	Hip circumference (cm)	2.47E-06	0.022	A	0.495	<i>INADL</i>	InaD-likelike (Drosophila)
rs494558	13	109727163	intron	[T/C]	Weight z-score change (SD/y)	4.66E-08*	0.045	G	0.078	<i>COL4A1</i>	collagen, type IV, alpha 1
rs40357	19	59385339	5UTR	[T/C]	Height change (cm/y)	4.49E-08*	0.045	G	0.391	<i>TSEN34</i>	tRNA splicing endonuclease 34 homolog (S. cerevisiae)

Abbreviations: SNP, single nucleotide polymorphism; chr, chromosome; MGA, measured genotype analysis; nonsyn, nonsynonymous; BMI, body mass index; UTR, untranslated region. *Significant according to the widely used significance threshold p-value of $<5 \times 10^{-8}$. doi:10.1371/journal.pone.0051954.t001

Table 2. Measured genotype analysis for endometabolic traits.

SNP	Chr	Coordinate (GB 36.2)	Location (GB 36.2)	SNP (GB 36.2)	Trait	MGA p-value	Effect Size	Minor Allele	Minor Allele Frequency	Gene Symbol	Gene Name
rs3733402	4	187395028	coding-nonsym	[A/G]	IGF-1 free (ng/mL)	9.01E-08	0.037	G	0.337	KLKB1	kalikrein B, plasma (Fletcher factor) 1
rs11974269	7	21114203	flanking_3UTR	[T/G]	Urinary creatinine (mmol/d)	8.38E-08**	0.051	C	0.128	ASS1P11	argininosuccinate synthetase 1 pseudogene 11
rs7030241	9	99590196	flanking_5UTR	[T/A]	TSH (μ U/mL)	1.03E-09	0.038	T	0.253	XPA/FOXE1	xeroderma pigmentosum, complementation group A/forkhead box E1 (thyroid transcription factor 2)
rs10830963	11	92348358	intron	[G/C]	Glucose (mg/dL)	3.72E-08*	0.048	G	0.205	MTNRTB	melatonin receptor 1B
rs3741298	11	116162771	intron	[A/G]	Triglycerides (mg/dL)	2.47E-08*	0.043	G	0.480	ZNF259	zinc finger protein 259
rs2266788	11	116165896	3UTR	[A/G]	Triglycerides (mg/dL)	4.82E-08	0.042	G	0.163	APOA5	apolipoprotein A-V
rs10131141	14	20331573	flanking_3UTR	[A/G]	Urinary nitrogen (g/8.19E-08** d)	0.040	0.040	G	0.280	RNASE1	ribonuclease, RNase A family, 1 (pancreatic)
rs3783637	14	54417868	intron	[T/C]	Urinary free dopamine: creatinine	6.29E-08**	0.051	A	0.121	GCHI	GTP cyclohydrolase 1
rs61744862	17	17008907	intron	[A/G]	IGFBP-3 (ng/mL)	7.24E-08**	0.037	A	0.041	MPRIP	myosin phosphatase Rho interacting protein
rs6044834	20	17384473	intron	[A/C]	Total antioxidants (mM)	7.60E-08	0.036	C	0.069	PCSK2	proprotein convertase subtilisin/kexin type 2

Abbreviations: SNP, single nucleotide polymorphism; chr, chromosome; MGA, measured genotype analysis; nonsyn, nonsynonymous.

IGF-1, insulin-like growth factor-1; TSH, thyroid stimulating hormone; UTR, untranslated region; IGFBP3, binding protein 3.

*Significant according to the widely used significance threshold p value of $<5 \times 10^{-6}$.

**Significant according to the our population-specific significance threshold p value of $<1.01 \times 10^{-7}$.

doi:10.1371/journal.pone.0051954.t002

Table 3. Measured genotype analysis for inflammation markers.

SNP	Chr	Coordinate (GB 36.2)	Location (GB 36.2)	SNP (GB 36.2)	Trait	MGA p-value	Effect Size	Minor Allele	Minor Allele Frequency	Gene Symbol	Gene Name
rs863002	1	157441544	intron	[A/G]	MCP-1 (pg/mL)	3.59E-13*	0.062	A	0.244	DARC	Duffy blood group, chemokine receptor
rs12075	1	157441978	coding-nonsyn	[A/G]	MCP-1 (pg/mL)	1.31E-21*	0.103	A	0.436	DARC	Duffy blood group, chemokine receptor
rs73175262	2	11675882	coding	[A/G]	MCP-1 (pg/mL)	6.46E-08**	0.049	A	0.053	GREB1	growth regulation by estrogen in breast cancer 1
rs7645716	3	46311785	flanking_3UTR	[A/G]	MCP-1 (pg/mL)	9.56E-08**	0.041	A	0.381	CCR3	chemokine (C-C motif) receptor 3
rs79509430	9	116280727	coding	[T/C]	MCP-1 (pg/mL)	2.03E-08*	0.056	A	0.068	DFNB31	deafness, autosomal recessive 31
rs657152	9	135129086	intron	[A/C]	IL-6 (pg/mL)	2.03E-08*	0.041	A	0.254	ABO	ABO blood group (transferase A, alpha 1-3-N-acetylgalactosaminyltransferase; transferase B, alpha 1-3-galactosyltransferase)
rs28461806	10	43075760	intron	[A/G]	MCP-1 (pg/mL)	4.58E-08*	0.053	G	0.049	RASGEF1A	RasGEF domain family, member 1A

Abbreviations: SNP, single nucleotide polymorphism; chr, chromosome; MGA, measured genotype analysis; nonsyn, nonsynonymous;

MCP-1, monocyte chemoattractant protein-1; UTR, untranslated region; IL-6, interleukin-6.

*Significant according to the widely used significance threshold p value of $<5 \times 10^{-8}$.

**Significant according to the our population-specific significance threshold p value of $<1.01 \times 10^{-7}$.

doi:10.1371/journal.pone.0051954.t003

light activity in 3'UTR of *CTCF* on chromosome 20 ($p = 3.6E-08$). In addition, adjusting for body weight, highly significant variants were identified for total energy expenditure ($p = 2.7E-08$) for *MATK* on chromosome 19 and sleeping energy expenditure ($p = 6.0E-08$) for *CHRNA3* on chromosome 15.

Discussion

Extensive phenotyping and high-density SNP genotyping enabled localization of novel genetic loci associated with the pathophysiology of obesity in Hispanic children. Our unprecedented phenotypes represent not only adiposity, but also biological processes underlying the development of childhood obesity. The number of genome-wide significant genetic variants detected substantiates our analytical strategy and statistical power to identify variants.

Genome-wide significant and suggestive genetic variants were associated with anthropometric indices, body composition and growth rate. The common nonsynonymous SNP rs1056513 in *INADL* accounted for 3% of the variance in body weight and body composition, and is highly conserved across mammalian species. *INADL* encodes a PDZ domain-containing protein thought to play a role in tight junctions and adipocyte differentiation [28]. Weight z-score change observed over one-year was associated with an intronic variant in *COL4A1*, which encodes a basement membrane collagen. Height change was associated with a 5'UTR variant in *TSEN34*, which is involved in tRNA splicing, a fundamental process required for cell growth and division [29].

Fasting serum glucose was associated with an intronic variant (MAF = 0.20; effect size 4.8%) in *MTNR1B* which encodes one of the melatonin receptors expressed in the retina and brain. Corroborating our findings, this variant (rs10830963) has been strongly associated with fasting glucose levels in adults [30,31] and children [32–35]. Melatonin, the ligand to *MTNR1B*, has an inhibitory effect on insulin secretion resulting in elevated fasting glucose.

Fasting serum triglycerides were associated with variants in the intron and 3'UTRs within the *APOA5-ζNF259* region. *APOA5* is an important determinant of circulating triglyceride levels [36]. SNPs detected in our GWAS have been associated with triglycerides in other populations [37]. In the Kosrae population, triglyceride levels were associated with seven SNPs near *APOC3/A5* [38]. Variants in the *APOA5-ζNF259* region (including rs3741298) were associated with HDL-C and ApoA-1 response to therapy with statins and fenofibric acid in patients with dyslipidemia [39].

Association with total antioxidants was shown for a variant in *PCSK2*, a proprotein convertase that is involved in proteolytic processing of neuropeptide and hormone precursors. *PCSK2* is highly expressed in the islets of Langerhans, where it plays a role in the conversion of proinsulin to insulin.

A LD block of 23 SNPs in the *XPA/FOXE1 (TTF-2)* region was highly associated with fasting serum TSH (MAF = 0.25-0.29, effect size = 2.9-3.8%). The association is likely attributed to *FOXE1* rather than *XPA* which is involved in DNA excision repair [40] and the skin disease xeroderma pigmentosum [41]. *FOXE1* or *thyroid transcription factor 2 (TTF-2)* belongs to the 'forkhead' gene family and is involved in promoting the migration process or in repressing differentiation of the thyroid follicular cells until migration has occurred [42] and has been associated with thyroid cancer [43–46]. In the Kosrae population, plasma TSH levels were strongly associated with 10 SNPs in a region encompassing *TTF-2* on chromosome 9 [38]. In a cohort of Caucasian adults, genetic variation in *FOXE1 (TTF-2)* showed significant effects on

Table 4. Measured genotype analysis for diet, energy expenditure, substrate utilization and physical activity.

SNP	Chr	Coordinate (GB 36.2)	Location (GB 36.2)	SNP (GB 36.2)	Trait	MGA p-value	Effect Size	Minor Allele	Minor Allele Frequency	Gene Symbol	Gene Name
rs16933006	9	15325914	flanking_5UTR	[A/C]	Light activity (min/d)	7.49E-08**	0.035	C	0.110	RPL7P3	ribosomal protein L7 pseudogene 33
rs17104363	14	67009236	intron	[T/C]	Dinner intake, adj EER (kcal)	5.13E-08**	0.044	G	0.050	TMEM229B	transmembrane protein 229B
rs8037818	15	30714768	intron	[A/G]	Sleep duration (min/d)	4.95E-08*	0.042	G	0.181	ARHGAP11A	Rho GTPase activating protein 11A
rs8040868	15	76698236	coding	[A/G]	Sleep energy expenditure adj weight (kcal/d)	5.95E-08**	0.048	G	0.239	CHRNA3	cholinergic receptor, nicotinic, alpha 3 (neuronal)
rs12104221	19	3748100	intron	[A/G]	Total energy expenditure adj weight (kcal/d)	2.65E-08*	0.054	A	0.381	MATK	megakaryocyte-associated tyrosine kinase
rs6025590	20	55503911	flanking_3UTR	[T/C]	Sedentary&light activity (min/d)	3.61E-08*	0.039	A	0.308	CTCF	CCCTC-binding factor (zinc finger protein)-like
rs2823615	21	16405004	intron	[T/A]	Sleep RQ	5.28E-08**	0.044	A	0.178	C21orf34	long intergenic non-protein coding RNA 478 (LINC00478)

Abbreviations: SNP, single nucleotide polymorphism; chr, chromosome; MGA, measured genotype analysis;

EER, estimated energy expenditure; UTR, untranslated region; RQ, respiratory quotient.

*Significant according to the widely used significance threshold p value of $<5 \times 10^{-8}$.

**Significant according to the our population-specific significance threshold p value of $<1.01 \times 10^{-7}$.

doi:10.1371/journal.pone.0051954.t004

free T4 levels and borderline effects on serum TSH [47]. Three of the SNPs (rs925488, rs1877431, rs1588635) detected in the VIVA cohort was also reported by Lowe et.al [38]; there was no overlap with Medici et al (43). Also, three of the SNPs (rs2805809, rs2668804, rs2808693) reported by Lowe et al. (38) was in high linkage disequilibrium (1.00) with rs2805771, rs2808699, rs7875482 detected in the VIVA cohort. This is a region of high LD, located in the 5'UTR-region of the gene which has been shown to influence transcriptional regulation of *FOXE1(TTF-2)*. The functional variant is likely to be situated at this locus, but its exact localization remains to be elucidated in future studies, involving in-depth resequencing.

A unifying role of inflammation in chronic diseases including cardiovascular disease, diabetes, hypertension and obesity is emerging. In the VIVA GWAS, variants in six genes were associated with proinflammatory markers. Our findings support a major role of Duffy antigen receptor for chemokines (*DARC*) in the regulation of the circulating levels of the cysteine-cysteine (CC) chemokine, MCP-1 (effect size = ~10%). In circulation, MCP-1 is bound to erythrocyte DARC that acts as a chemokine receptor/reservoir of proinflammatory cytokines [48]. Our strongest association for MCP-1 was with a nonsynonymous, highly conserved SNP in *DARC* (rs12075; MAF = 0.44) which replicated results from the Framingham Heart Study GWAS in Caucasian adults [49]. MCP-1 was also associated with SNPs in *GREB1*, *DEFB31*, *RASGEF1A*, and *CCR3*. Huber et al. found increased expression of *CCR3* in subcutaneous and visceral adipose tissue in obese patients compared to lean controls [50]. Studies by Schnabel et al [49] and Naitza et al. [51] found suggestive associations of serum MCP-1 with *CCR2* which is also a MCP-1 receptor and is in the same region of chromosome 3 as *CCR3*, and referred to as the *CCR2/CCR3* cytokine receptor gene cluster.

Fasting serum IL-6 levels were associated with variants in the *ABO* gene that determines blood group. *ABO* blood group has been found to be associated with a number of biomarkers such as von Willebrand factor levels, Factor VIII levels, thrombomodulin, TNF- α and ICAM-1 [52,53], and in our case IL-6. The mechanism by which the A and B alleles affect these biomarkers is uncertain. In Caucasians with and without type 1 diabetes, a variant rs579459 near the *ABO* blood group gene accounted for 19% of the variance in E-selectin levels [52]; in our GWAS, this same variant was associated with IL-6 levels ($p = 1.7E-07$; **Table S2**).

Total energy expenditure, adjusted for body weight, was significantly associated with rs12104221 in *MATK* which encodes a protein-tyrosine kinase involved in signal transduction pathways [54]. Sleeping energy expenditure, adjusted for body weight, was associated with rs8040868 in *CHRNA3* (cholinergic receptor, neuronal nicotinic, alpha polypeptide 3), a member of a superfamily of ligand-gated ion channels that mediate fast signal transmission at synapses. After binding to acetylcholine, the receptor responds by opening ion-conducting channels across the plasma membrane, suggesting a plausible role of the coding variant (rs8040868) in energy metabolism [55]. Acetylcholine receptors activate proopiomelanocortin neurons that in turn activate melanocortin-4 receptors that are involved in the regulation of energy intake and expenditure [56], [57].

Sleep duration was associated with an intronic SNP in *ARHGAP11A* (rho GTPase activating protein 11A) that encodes a 1,023-amino acid protein that has a rhoGAP domain and tyrosine phosphorylation site. Evidence is emerging that obesity affects sleep, and that sleep patterns and disorders may have an effect on weight [58]. Although the mechanism is unclear, sleep

disturbances are characteristic of Prader Willi Syndrome, caused by a deletion in 15q11-q13 that encompasses *ARHGAP11A* [59].

Sedentary-light physical activity was associated with a variant in *CTCF*, an 11-zinc-finger factor involved in gene regulation. CTCF forms methylation-sensitive insulators that regulate X-chromosome inactivation which may play a role in epigenetic regulation [60].

Variants in the 11 genes known to cause extreme early-onset obesity also may contribute to milder forms of obesity [61,62]. None of the genotyped variants in genes for monogenic obesity reached genome-wide significance in our GWAS, although several variants in *CRHR1*, *CRHR2*, *MCHR1*, *MC3R*, *MC4R* and *POMC* were nominally associated. Similarly, variants in or nearby the susceptibility genes for obesity did not attain genome-wide significance but several - *CHST3*, *KCTD15*, *MTCH2*, *SFRS10*, *SH2B1* and *TMEM18* - were nominally associated with obesity-related traits, consistent with GWAS in children of European-American [63] or European ancestry [9]. The lack of genome-wide significant findings for monogenic causes of obesity or susceptibility genes may be a function of our sample size and statistical power, or the presence of rare variants in the Hispanic population not represented in the Illumina platform.

The VIVA LA FAMILIA Study is unique in its consideration of a pediatric Hispanic population. We believe that our extensive phenotyping and genotyping enabled localization of novel genetic loci associated with obesity in Hispanic children, despite our relatively small sample size. Our phenotypes represent not only adiposity, but also biological processes underlying the development and consequences of childhood obesity. We applied a standard and stringent approach to our measured genotype analysis, but do agree replication is desirable. We believe we have identified genes that warrant further investigation.

In conclusion, unprecedented in-depth phenotyping and high-density SNP genotyping enabled the localization of novel genetic loci associated with the pathophysiology of obesity in Hispanic children. Identified genome-wide significant loci: 1) corroborated genes implicated in other studies (*MTNR1B*, *ZNF259/APOA5*, *XPA/FOXE1 (TTF-2)*, *DARC*, *CCR3*, *ABO*); 2) localized novel genes in plausible biological pathways (*PCSK2*, *ARHGAP11A*, *CHRNA3*); and 3) revealed novel genes with unknown function in obesity pathogenesis (*MATK*, *COL4A1*). As with other GWAS, the variants identified are likely not the actual causal variants but rather markers for genomic regions or loci in which the causal variants lie. Characterization of the underlying functional genetic variants contributing to this serious public health problem in Hispanic children will involve additional study.

Supporting Information

Figure S1 GWAS Manhattan plots are displayed for three phenotypes: total energy expenditure, adjusted for body weight, measured by 24-h room calorimetry; fasting serum thyroid stimulating hormone; and 1-y change in weight z-score. The genomic coordinates are shown along the X-axis, and the negative logarithm of the association p-value for each SNP on the Y-axis. (TIFF)

Table S1 Description of the phenotypes used in the genome-wide association study. (DOCX)

Table S2 Suggestive and genome-wide significant genetic variants identified by measured genotype analysis. (DOCX)

Acknowledgments

We thank all the families who participated in the VIVA LA FAMILIA Study. The authors wish to acknowledge the contributions of Grace-Ellen Meixner, B.S. and Maria del Pilar Villegas, B.S., M.S. for technical assistance while working on this project at the Texas Biomedical Research Institute.

This work is a publication of the U.S. Department of Agriculture (USDA)/Agricultural Research Service (ARS) Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine and Texas Children's Hospital, Houston, Texas. The contents of this

publication do not necessarily reflect the views or policies of the USDA, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Author Contributions

Conceived and designed the experiments: AGC SAC RAG NFB. Performed the experiments: SAC KH NFB. Analyzed the data: SLL VSV KH. Contributed reagents/materials/analysis tools: SAC KH. Wrote the paper: AGC SAC SLL VSV KH RAG NFB.

References

- O'Rahilly S, Farooqi IS (2008) Human obesity as a heritable disorder of the central control of energy balance. *Int J Obes (Lond)* 32 Suppl 7: S55-S61.
- Muller MJ, Bosty-Westphal A, Krawczak M (2010) Genetic studies of common types of obesity: a critique of the current use of phenotypes. *Obes Rev* 11: 612–618.
- Barlow SE, Dietz WH (1998) Obesity evaluation and treatment: expert committee recommendations. *Pediatrics* 102: e29.
- Butte NF, Cai G, Cole SA, Comuzzie AG (2006) VIVA LA FAMILIA Study: genetic and environmental contributions to childhood obesity and its comorbidities in the Hispanic population. *Am J Clin Nutr* 84: 646–654.
- Hakonarson H, Grant SF (2011) GWAS and its impact on elucidating the etiology of diabetes. *Diabetes Metab Res Rev* 10.1002/dmrr.1221. [Epub ahead of print].
- Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42: 937–948.
- Scherag A, Dina C, Hinney A, Vatn V, Scherag S, et al. (2010) 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* 6: e1000916.
- Meyre D, Delplanque J, Chevre JC, Lecoecur C, Lobbens S, et al. (2009) Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat Genet* 41: 157–159.
- Bradfield JP, Taal HR, Timpson NJ, Scherag A, Lecoecur C, et al. (2012) A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat Genet* 10.1038/ng.2247. [Epub ahead of print].
- Cai G, Cole SA, Freeland-Graves JH, MacCluer JW, Blangero J, et al. (2004) Genome-wide scans reveal quantitative trait loci on 8p and 13q related to insulin action and glucose metabolism: the San Antonio Family Heart Study. *Diabetes* 53: 1369–1374.
- Cai G, Cole SA, Butte NF, Bacino CA, Diego V, et al. (2006) A quantitative trait locus on chromosome 18q for physical activity and dietary intake in Hispanic children. *Obesity* 14: 1596–1604.
- Cai G, Cole SA, Butte NF, Voruganti VS, Comuzzie AG (2007) A quantitative trait locus (QTL) on chromosome 13q affects fasting glucose levels in Hispanic children. *J Clin Endocrinol Metab* 92: 4893–4896.
- Cai G, Cole SA, Butte NF, Voruganti VS, Comuzzie AG (2008) Genome-wide scan revealed genetic loci for energy metabolism in Hispanic children and adolescents. *Int J Obes (Lond)* 32: 579–585.
- Cai G, Cole SA, Butte NF, Smith CW, Mehta NR, et al. (2008) A genetic contribution to circulating cytokines and obesity in children. *Cytokine* 44: 242–247.
- Voruganti VS, Goring HH, Diego V, Cai G, Mehta N, et al. (2007) Genome-wide scan for plasma ghrelin detects linkage on chromosome 1p36 in Hispanic children: results from the Viva la Familia study. *Pediatr Res* 62: 445–450.
- Lohman TG, Roche AF, Martorell R (1988) Anthropometric Standardization Reference Manual. Champaign: Human Kinetics.
- Butte NF, Christiansen E, Sorensen TI (2007) Energy imbalance underlying the development of childhood obesity. *Obesity* 15: 3056–3066.
- Butte NF, Comuzzie AG, Cole SA, Mehta NR, Cai G, et al. (2005) Quantitative genetic analysis of the metabolic syndrome in Hispanic children. *Pediatr Res* 58: 1243–1248.
- Butte NF, Cai G, Cole SA, Wilson TA, Fisher JO, et al. (2007) Metabolic and behavioral predictors of weight gain in Hispanic children: the VIVA LA FAMILIA Study. *Am J Clin Nutr* 85: 1478–1485.
- Johnson RK, Driscoll P, Goran MI (1996) Comparison of multiple-pass 24-hour recall estimates of energy intake with total energy expenditure determined by the doubly labeled water method in young children. *J Am Diet Assn* 96: 1140–1144.
- Fisher JO, Cai G, Jaramillo S, Cole SA, Comuzzie AG, et al. (2007) Heritability of hyperphagic eating behavior and appetite-related hormones among Hispanic children. *Obesity* 15: 1484–1495.
- Butte NF, Puyau MR, Adolph AL, Vohra FA, Zakeri I (2007) Physical activity in nonoverweight and overweight Hispanic children and adolescents. *Med Sci Sports Exerc* 39: 1257–1266.
- Puyau MR, Adolph AL, Vohra FA, Zakeri I, Butte NF (2004) Prediction of activity energy expenditure using accelerometers in children. *Med Sci Sports Exerc* 36: 1625–1631.
- Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet* 58: 1323–1337.
- Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62: 1198–1211.
- Boerwinkle E, Chakraborty R, Sing CF (1986) The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. *Ann Hum Genet* 50: 181–194.
- Moskvina V, Schmidt KM (2008) On multiple-testing correction in genome-wide association studies. *Genet Epidemiol* 32: 567–573.
- Zhong J, Krawczyk SA, Chaerkady R, Huang H, Goel R, et al. (2010) Temporal profiling of the secretome during adipogenesis in humans. *J Proteome Res* 9: 5228–5238.
- Wende H, Volz A, Ziegler A (2000) Extensive gene duplications and a large inversion characterize the human leukocyte receptor cluster. *Immunogenetics* 51: 703–713.
- Prokopenko I, Langenberg C, Florez JC, Saxena R, Soranzo N, et al. (2009) Variants in MTNR1B influence fasting glucose levels. *Nat Genet* 41: 77–81.
- Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, et al. (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 42: 105–116.
- Kelliny C, Ekelund U, Andersen LB, Brage S, Loos RJ, et al. (2009) Common genetic determinants of glucose homeostasis in healthy children: the European Youth Heart Study. *Diabetes* 58: 2939–2945.
- Bouatia-Najji N, Bonnefond A, Froguel P (2009) Insights from the genetics of fasting glucose: lessons for diabetes. *Med Sci* 25: 897–902.
- Reinehr T, Scherag A, Wang HJ, Roth CL, Kleber M, et al. (2011) Relationship between MTNR1B (melatonin receptor 1B gene) polymorphism rs10830963 and glucose levels in overweight children and adolescents. *Pediatr Diabetes* 12: 435–441.
- Holzappel C, Siegrist M, Rank M, Langhof H, Grallert H, et al. (2011) Association of a MTNR1B gene variant with fasting glucose and HOMA-B in children and adolescents with high BMI-SDS. *Eur J Endocrinol* 164: 205–212.
- Pennacchio LA, Olivier M, Hubacek JA, Cohen JC, Cox DR, et al. (2001) An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science* 294: 169–173.
- Sarwar N, Sandhu MS, Ricketts SL, Butterworth AS, Di AE, et al. (2010) Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. *Lancet* 375: 1634–1639.
- Lowe JK, Maller JB, Pe'er I, Neale BM, Salit J, et al. (2009) Genome-wide association studies in an isolated founder population from the Pacific Island of Kosrae. *PLoS Genet* 5: e1000365.
- Brautbar A, Covarrubias D, Belmont J, Lara-Garduno F, Virani SS, et al. (2011) Variants in the APOA5 gene region and the response to combination therapy with statins and fenofibrate acid in a randomized clinical trial of individuals with mixed dyslipidemia. *Atherosclerosis* 219: 737–742.
- Tanaka K, Miura N, Satokata I, Miyamoto I, Yoshida MC, et al. (1990) Analysis of a human DNA excision repair gene involved in group A xeroderma pigmentosum and containing a zinc-finger domain. *Nature* 348: 73–76.
- States JC, McDuffie ER, Myrand SP, McDowell M, Cleaver JE (1998) Distribution of mutations in the human xeroderma pigmentosum group A gene and their relationships to the functional regions of the DNA damage recognition protein. *Hum Mutat* 12: 103–113.
- Manley NR, Capecchi MR (1998) Hox group 3 paralogs regulate the development and migration of the thymus, thyroid, and parathyroid glands. *Dev Biol* 195: 1–15.
- Eriksson N, Tung JY, Kiefer AK, Hinds DA, Francke U, et al. (2012) Novel associations for hypothyroidism include known autoimmune risk loci. *PLoS One* 7: e34442.
- Denny JC, Crawford DC, Ritchie MD, Bielinski SJ, Basford MA, et al. (2011) Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. *Am J Hum Genet* 89: 529–542.
- Landa I, Ruiz-Llorente S, Montero-Conde C, Inglada-Perez L, Schiavi F, et al. (2009) The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. *PLoS Genet* 5: e1000637.

46. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, et al. (2009) Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat Genet* 41: 460–464.
47. Medici M, van der Deure WM, Verbiest M, Vermeulen SH, Hansen PS, et al. (2011) A large-scale association analysis of 68 thyroid hormone pathway genes with serum TSH and FT4 levels. *Eur J Endocrinol* 164: 781–788.
48. Rull A, Camps J, onso-Villaverde C, Joven J (2010) Insulin resistance, inflammation, and obesity: role of monocyte chemoattractant protein-1 (or CCL2) in the regulation of metabolism. *Mediators Inflamm* 2010: 326580. Epub 2010 Sep 23.
49. Schnabel RB, Baumert J, Barbalic M, Dupuis J, Ellinor PT et al. (2010) Duffy antigen receptor for chemokines (Darc) polymorphism regulates circulating concentrations of monocyte chemoattractant protein-1 and other inflammatory mediators. *Blood* 115: 5289–5299.
50. Huber J, Kiefer FW, Zeyda M, Ludvik B, Silberhumer GR, et al. (2008) CC chemokine and CC chemokine receptor profiles in visceral and subcutaneous adipose tissue are altered in human obesity. *J Clin Endocrinol Metab* 93: 3215–3221.
51. Naitza S, Porcu E, Steri M, Taub DD, Mulas A, et al. (2012) A genome-wide association scan on the levels of markers of inflammation in Sardinians reveals associations that underpin its complex regulation. *PLoS Genet* 8: e1002480.
52. Paterson AD, Lopes-Virella MF, Waggott D, Boright AP, Hossaini SM, et al. (2009) Genome-wide association identifies the ABO blood group as a major locus associated with serum levels of soluble E-selectin. *Arterioscler Thromb Vasc Biol* 29: 1958–1967.
53. Pare G, Ridker PM, Rose L, Barbalic M, Dupuis J, et al. (2011) Genome-wide association analysis of soluble ICAM-1 concentration reveals novel associations at the NFKB1K, PNPLA3, RELA, and SH2B3 loci. *PLoS Genet* 7: e1001374.
54. Bennett BD, Cowley S, Jiang S, London R, Deng B, et al. (1994) Identification and characterization of a novel tyrosine kinase from megakaryocytes. *J Biol Chem* 269: 1068–1074.
55. Woolf NJ, Butcher LL (1986) Cholinergic systems in the rat brain: III. Projections from the pontomesencephalic tegmentum to the thalamus, tectum, basal ganglia, and basal forebrain. *Brain Res Bull* 16: 603–637.
56. Mineur YS, Abizaid A, Rao Y, Salas R, DiLeone RJ, et al. (2011) Nicotine decreases food intake through activation of POMC neurons. *Science* 332: 1330–1332.
57. Barsh GS, Schwartz MW (2002) Genetic approaches to studying energy balance: perception and integration. *Nat Rev Genet* 3: 589–600.
58. Kelly-Pieper K, Lamm C, Fennoy I (2011) Sleep and obesity in children: a clinical perspective. *Minerva Pediatr* 63: 473–481.
59. Torrado M, Araoz V, Baialardo E, Abraldes K, Mazza C, et al. (2007) Clinical-etiologic correlation in children with Prader-Willi syndrome (PWS): an interdisciplinary study. *Am J Med Genet A* 143: 460–468.
60. Klenova EM, Morse HC, III, Ohlsson R, Lobanov VV (2002) The novel BORIS+CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer. *Semin Cancer Biol* 12: 399–414.
61. Cole SA, Butte NF, Voruganti VS, Cai G, Haack K, et al. (2010) Evidence that multiple genetic variants of MC4R play a functional role in the regulation of energy expenditure and appetite in Hispanic children. *Am J Clin Nutr* 91: 191–199.
62. Cotsapas C, Speliotes EK, Hatoum IJ, Greenawald DM, Dobrin R, et al. (2009) Common body mass index-associated variants confer risk of extreme obesity. *Hum Mol Genet* 18: 3502–3507.
63. Zhao J, Grant SF (2011) Genetics of childhood obesity. *J Obes* 2011: 845148. Epub 2011 May 26.