



EMR-linked GWAS study: investigation of variation landscape of loci for body mass index in children

Bahram Namjou^{1*}, Mehdi Keddache^{2,3}, Keith Marsolo^{3,4}, Michael Wagner^{3,4}, Todd Lingren^{3,4}, Beth Cobb¹, Cassandra Perry⁵, Stephanie Kennebeck^{2,3}, Ingrid A. Holm⁶, Rongling Li⁷, Nancy A. Crimmins^{2,3}, Lisa Martin^{2,3}, Imre Solti^{3,4}, Isaac S. Kohane⁸ and John B. Harley^{1,2,3,9}

¹ Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

² Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

³ School of Medicine, University of Cincinnati, Cincinnati, OH, USA

⁴ Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

⁵ Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA

⁶ Division of Genetics and Genomics, Department of Pediatrics, The Manton Center for Orphan Disease Research, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

⁷ National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

⁸ Center for Biomedical Informatics, Harvard Medical School and Children's Hospital Informatics Program, Boston, MA, USA

⁹ Department of Veteran Affairs Medical Center, Cincinnati, OH, USA

Edited by:

Mariza de Andrade, Mayo Clinic, USA

Reviewed by:

Ayo Priscille Doumatey, National Institute of Health, USA

Kelli K. Ryckman, University of Iowa, USA

*Correspondence:

Bahram Namjou, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229, USA
e-mail: Bahram.namjou@cchmc.org

Common variations at the loci harboring the fat mass and obesity gene (FTO), MC4R, and TMEM18 are consistently reported as being associated with obesity and body mass index (BMI) especially in adult population. In order to confirm this effect in pediatric population five European ancestry cohorts from pediatric eMERGE-II network (CCHMC-BCH) were evaluated.

Method: Data on 5049 samples of European ancestry were obtained from the Electronic Medical Records (EMRs) of two large academic centers in five different genotyped cohorts. For all available samples, gender, age, height, and weight were collected and BMI was calculated. To account for age and sex differences in BMI, BMI z-scores were generated using 2000 Centers of Disease Control and Prevention (CDC) growth charts. A Genome-wide association study (GWAS) was performed with BMI z-score. After removing missing data and outliers based on principal components (PC) analyses, 2860 samples were used for the GWAS study. The association between each single nucleotide polymorphism (SNP) and BMI was tested using linear regression adjusting for age, gender, and PC by cohort. The effects of SNPs were modeled assuming additive, recessive, and dominant effects of the minor allele. Meta-analysis was conducted using a weighted z-score approach.

Results: The mean age of subjects was 9.8 years (range 2–19). The proportion of male subjects was 56%. In these cohorts, 14% of samples had a BMI ≥ 95 and 28 $\geq 85\%$. Meta analyses produced a signal at 16q12 genomic region with the best result of $p = 1.43 \times 10^{-7}$ [$p_{(rec)} = 7.34 \times 10^{-8}$] for the SNP rs8050136 at the first intron of FTO gene ($z = 5.26$) and with no heterogeneity between cohorts ($p = 0.77$). Under a recessive model, another published SNP at this locus, rs1421085, generates the best result [$z = 5.782$, $p_{(rec)} = 8.21 \times 10^{-9}$]. Imputation in this region using dense 1000-Genome and Hapmap CEU samples revealed 71 SNPs with $p < 10^{-6}$, all at the first intron of FTO locus. When heterogeneity was permitted between cohorts, signals were also obtained in other previously identified loci, including MC4R (rs12964056, $p = 6.87 \times 10^{-7}$, $z = -4.98$), cholecystokinin CCK (rs8192472, $p = 1.33 \times 10^{-6}$, $z = -4.85$), Interleukin 15 (rs2099884, $p = 1.27 \times 10^{-5}$, $z = 4.34$), low density lipoprotein receptor-related protein 1B [LRP1B (rs7583748, $p = 0.00013$, $z = -3.81$)] and near transmembrane protein 18 (TMEM18) (rs7561317, $p = 0.001$, $z = -3.17$). We also detected a novel locus at chromosome 3 at COL6A5 [best SNP = rs1542829, minor allele frequency (MAF) of 5% $p = 4.35 \times 10^{-9}$, $z = 5.89$].

Conclusion: An EMR linked cohort study demonstrates that the BMI-Z measurements can be successfully extracted and linked to genomic data with meaningful confirmatory results. We verified the high prevalence of childhood rate of overweight and obesity in our cohort (28%). In addition, our data indicate that genetic variants in the first intron of FTO, a known adult genetic risk factor for BMI, are also robustly associated with BMI in pediatric population.

Keywords: BMI, obesity, polymorphism, GWAS

INTRODUCTION

The electronic MEDical Records and GENomics (eMERGE) Network, founded in 2007, is a consortium of multiple adult and pediatric institutions developed to explore the utility of DNA biorepositories linked to electronic medical records (EMR) in advancing genomic medicine (McCarty et al., 2011). For each site, the primary site-specific phenotypes have undergone genome-wide association studies (GWAS) with data and results shared through the network. In the pediatric population, however, genetic studies are challenging due to different developmental phase and growth patterns, different spectrums of disease, and unusual rare genetic or congenital abnormalities.

In both adults and children, obesity is a major risk factor for a number of chronic diseases, a steady and continuous rise in prevalence over the past four decades holds serious and ominous medical and economic burdens (Kopelman, 2000). The phenotype is highly heritable. Family and twin studies have shown that between 40 and 70% of the inter-individual variation in obesity can be attributable to genetic factors (Maes et al., 1997). In recent years, large-scale GWAS have identified many loci associated with Body Mass Index (BMI), the most common measure of obesity, but these loci combined explain only 2–4% of the heritability (Speliotes et al., 2010). Thus far, four waves of GWAS studies for BMI identified 32 loci that reached genome-wide significance and unequivocally were associated with BMI in a large meta-analysis performed by the GIANT (Genetic Investigation of ANthropometric Traits) consortium (Frayling et al., 2007; Scuteri et al., 2007; Loos et al., 2008; Willer et al., 2009; Speliotes et al., 2010; Loos, 2012; Mägi et al., 2013). The firstly identified locus, FTO (fat mass and obesity associated gene), has the largest effect on obesity-susceptibility with obesity at a risk of 1.20 fold. Moreover, the frequency of the BMI-increasing allele is high in white Europeans (i.e., 40%). As a consequence, of all 32 BMI-associated loci, the FTO locus explains the largest proportion of the inter-individual variation in BMI (0.34%; Speliotes et al., 2010; Loos, 2012).

In this study, we investigated the genetic association of pediatric BMI using anthropomorphic measures extracted from medical records in a collection of already genotyped samples from two large pediatric cohort repositories (CCHMC and CHB) in order to confirm and identify additional genetic loci.

MATERIALS AND METHODS

STUDY SUBJECTS

Protocols for this study were approved by the Institutional Review Boards (IRBs) at the institutions where participants were recruited. Only those self-reported to have European ancestry were selected for study. The anthropometric measurements of height and weight, as well as age of measurement and gender, were extracted from the EMR. All enrolled participants with measured weight and height on the same day were included. All inconsistent or out of range values were excluded. Out of range was defined as any height or weight values higher or lower than is considered biologically possible according to Centers of Disease Control and Prevention (CDC) growth charts. Children and teens, aged 2 through 19 years old were included based on CDC growth chart requirements. In addition, three patients with the ICD-9 code for Prader Willi syndrome were excluded from final results.

After removing the missing data and outliers, out of a total of 5,049 individuals, 2860 samples were included in the study. The demographic distributions of these samples are shown in **Table 1**.

GENOTYPING

High throughput single nucleotide polymorphism (SNP) genotyping was carried out previously in CCHMC and BCH using different Illumina™ or Affymetrix™ platforms (**Table 1**). Quality control (QC) of the data was performed before imputation. In each genotyped cohort, standard QC criteria were met and SNPs were removed if (a) >10% missing genotyping, (b) out of Hardy–Weinberg equilibrium (HWE, $P < 0.001$), or a minor allele frequency (MAF) <1%. Samples with call rate <98% were excluded. Principle component analysis (PCA) was performed to identify outliers and hidden population structure using EIGENSTRAT (Price et al., 2006). Based on examination of the scree plot, the first two PCs were retained and used as covariates during the association analysis in order to adjust for population stratification.

PHENOTYPING

We obtained height and weight measurements from EMR in order to calculate BMI [$\text{wt (kg)}/(\text{ht(m)}^2)$]. When multiple measurements were available for a subject, the most recent measurement was selected and all inconsistent measures were excluded. BMI z-scores and percentiles were generated using the 2000 CDC growth charts (study¹). These z-scores and percentiles account for the age and sex differences in BMI throughout childhood. All data for BMI-z scores (−3 to +3, mean = 0) were scaled to positive value (+4, 1–7, mean = 4) to be used as a quantitative trait for the GWAS study. In order to assess the burden of increased body weight on health and estimate the effect size, standard cut-offs were also used, and the tail BMI distribution was considered as a binary phenotype ($\geq 95\%$ as case and $\leq 20\%$ as control). For the published FTO locus, Phenome wide association study (PheWas) was also performed in which presence or absence of each ICD-9 codes were considered as binary phenotype. Only ICD-9 codes with 50 or more available samples were included (143 codes) in the analysis.

STATISTICAL ANALYSIS

Genome-wide association studies analysis was performed by cohort in PLINK (Purcell et al., 2007) using regression models and adjusting for age, sex, and the first two principal components (PC). For BMI z-score, the primary analysis was performed using an additive effect of the minor alleles. However, as previous BMI associations have reported better model fit with different models, recessive and dominant models were subsequently evaluated. For binary phenotypes (dichotomization of BMI using a tails approach and PheWas analyses), allelic association was assessed between cases and controls by chi-square with 1 degree of freedom (df). Allelic odds ratio (OR) and 95% confidence intervals (95% CIs) were obtained. In PheWas analyses, permutation procedure was performed using sample randomization strategy in which case

¹http://www.cdc.gov/healthyweight/downloads/BMI_group_calculator_English.xls

Table 1 | Demographic distribution of pediatric cohorts under study.

	# Europeans	# After removing outliers and missing values	M/F	Mean age (95% CI)	Array
CHB	741	613	387/226	13.30 (12.97–13.66)	Affymetrix-Axiom
CCHMC	829	696	338/358	10.80 (10.53–11.12)	Omni-5
	657	405	261/144	7.18 (6.73–7.63)	Omni-1
	1270	942	589/353	7.32 (7.03–7.62)	Illumina-610
	1552	204	28/176	13.70 (13.13–14.23)	Affymetrix-6
Total	5049	2860	1603/1257	9.8 (8.67–10.85)	

and control labels are permuted randomly ($\times 10000$) in order to obtain empirical p values and to correct for multiple testing.

For specific target regions, imputation-based analyses were performed using the impute2-Gtool pipeline and the publicly available 1000 Genomes Project as the reference haplotype panel composed of 1092 samples (release version 2 of the 1000 Genomes Project Phase I² (Howie et al., 2011)). For each batch of imputation runs, the standard Markov Chain Monte Carlo (MCMC) algorithm implemented in impute-2, was used with the following threshold criteria (burnin = 10, iteration = 30, and $N_e = 20000$, buffer = 250 kb). A threshold of 0.90 for the posterior probability of each genotype was then applied for genotype calling and conversion using Gtool. For each imputation run, the overall genotype concordance rate was more than 95%. Additional post imputation filtering were also implemented to remove poorly imputed variants with low concordance rate according to the impute-2 standard protocol (info >0.4; Howie et al., 2011). To graphically display the results, LocusZoom was used (Pruim et al., 2010).

META-ANALYSES

The results from primary analysis in each cohort were assembled to conduct a fixed effects weighted Z meta-analysis using Metal (Willer et al., 2010). This approach controls the differences in phenotype scaling across the studies and weights the signed Z statistics from each study by its sample size (i.e., weighted sum), from which a probability is calculated. The program also applies the genomic control correction to control type I error rates using summary statistics from each cohort. After QC filtrations in each cohort (as described above), meta-analyses were performed on SNP markers that were overlap among all five cohorts. 92670 SNP markers were in this category. At the next step in order to allow heterogeneity between cohorts, we applied a minimum weight of at least 1000 samples for analyses and identify additional effects. 583824 SNP markers were evaluated in this mode. We considered genome-wide significance thresholds of nominal p -value $<10^{-8}$ for any new findings and report all significant results ($p < 0.001$) of previously known loci that concurred with previous publications in terms of strand direction, MAF, and supporting evidence from nearby region. In addition, to describe the presence or absence of excess variation between cohorts, we

evaluated the Q -statistic and I^2 as a measure of heterogeneity (Willer et al., 2010).

RESULTS

The demographic distribution of the European ancestry population under study (Table 1) shows that the overall mean age of participants was 9.8 (95% CI = 8.67–10.85) years old with 56% being male. Table 2 shows the estimated prevalence of overweight and obesity in these pediatric cohorts with a rate of 28% overweight (≥ 85 th %ile) and 14% of obesity (≥ 95 th %ile). This distribution was consistent across all cohorts (Table 2).

Genome-wide analyses were conducted within each cohort. Associations between SNPs and BMI assumed an additive genetic model and summary statistics were subsequently used for meta-analysis using a weighted z -score method. After cleaning the data by applying our QC criteria, the ratio of the observed to expected χ^2 test statistic (lambda) was $\lambda = 1.007$ (Figure 1B). The results of the meta-analyses of all studies revealed a significant signal of association at 16q12 (Manhattan plot, Figure 1A).

The typed SNP rs8050136 at first intron of FTO gene produced consistent evidence of association in all cohorts with the best overall result of ($P = 1.43 \times 10^{-7}$, $z = 5.26$) and with no heterogeneity between study cohorts ($p = 0.77$; Figure 1A; Table 3). In addition, the allele frequencies and strand alignment are similar across cohorts and consistent with European ancestry. In consistent with previous publications, when mean of BMI- z was stratified by genotype (AA, AC, CC) for SNP (rs8050136), the additive association with risk allele was observed and it is

Table 2 | Summary of BMI-for-age and prevalence of overweight and obese children from the CCHMC-BCH cohorts.

	Male	Female	Total
Number of children assessed	1603	1257	2860
Underweight (<5th %ile)	10	9	10
Normal BMI (5th–85th %ile)	62	63	62
Overweight or obese (≥ 85 th %ile)*	28	28	28
Obese (≥ 95 th %ile)	15	12	14

Summary of children's BMI-for-age and prevalence of overweight and obesity. *Terminology based on Barlow and the Expert Committee (2007).

²ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/20110521

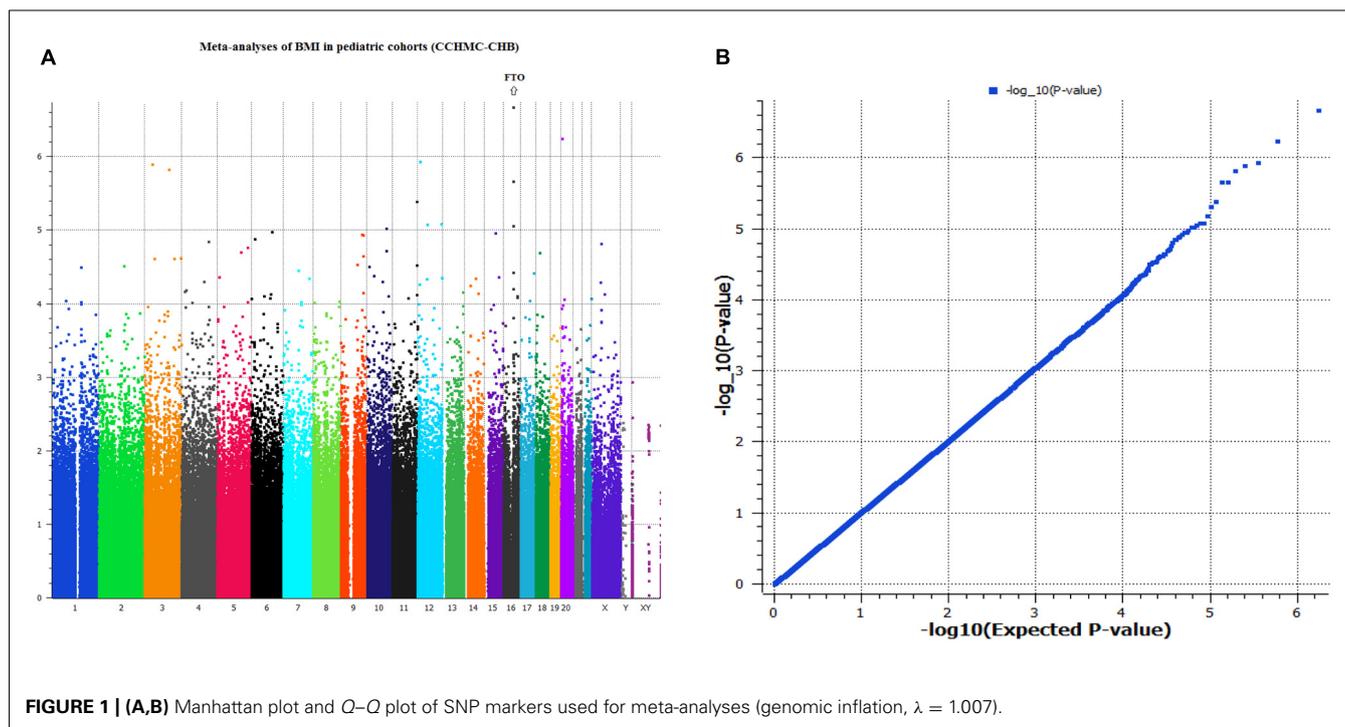


FIGURE 1 | (A,B) Manhattan plot and Q–Q plot of SNP markers used for meta-analyses (genomic inflation, $\lambda = 1.007$).

Table 3 | Most associated SNPs with BMI-z in CCHMC-BCH pediatric cohorts.

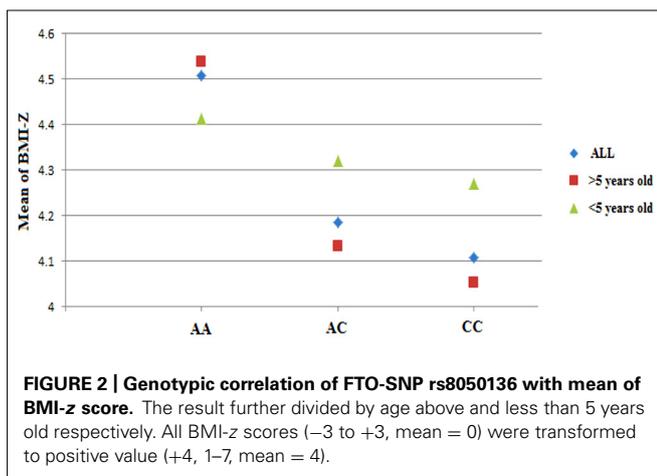
Chr	Position	Gene	SNP	Minor allele	MAF	Z score*	$P_{(add)}$	$P_{(rec)}$	P_{Dom}	Cochran-q	I^2 (%)
16	53816275	FTO	rs8050136	A	0.39	5.26	1.43e-07	7.34e-08	0.0008	0.71	0
16	53813367	FTO	rs17817449	G	0.39	5.02	5.56e-07	5.10e-08	0.001	0.49	0
16	53820527	FTO	rs9939609	A	0.39	5.01	6.07e-07	8.53e-08	0.001	0.75	0
16	53800954	FTO	rs1421085	C	0.39	4.65	3.45e-06	8.21e-09	0.02	0.61	0
18	57673799	Near MC4R	rs12964056	A	0.24	-4.98	6.87e-07	0.0004	4.96e-06	0.95	0
3	42299870	CCK	rs8192472	A	0.37	-4.85	1.33e-06	0.0007	0.0002	0.89	0
4	142763570	IL15	rs2099884	T	0.15	4.29	1.27e-05	0.0006	0.0004	0.82	0
2	142855291	LRP1B	rs7583748	G	0.10	-3.81	0.0001	0.03	8.08e-005	0.19	0.41
2	644953	TMEM18	rs7561317	A	0.17	-3.17	0.001	0.09	0.02	0.74	0

*The direction of all effect (weighted z scores) are for the minor alleles. The I^2 inconsistency metric was null to small for all of the markers ($I^2 = 0-42\%$).

shown in **Figure 2** (risk allele, A). There was 0.4 z-score-unit difference in mean of BMI-z score between homozygotes with risk and non-risk genotype in our pediatric cohorts (**Figure 2**). We further subdivided all cohorts into two age strata of less than 5 and above 5 years old. In meta-analyses of both strata, the minor allele (A) was associated additively with a higher BMI (**Figure 2**).

Next, we performed imputation-based association followed by conditional analysis to identify independent association in the FTO locus. We identified 71 SNPs at the first intronic region of FTO that were significantly associated with BMI in European ancestry, all of which were of similar magnitude (Z score $4.4 < z < 5.26$, p -values between $10^{-7} < p < 10^{-6}$) and with high linkage disequilibrium (LD) with each other ($r^2 > 0.8$; **Figure 3A**). The haplotype boundaries and recombination rate in this ancestry are

also shown in **Figure 3**. Indeed, the SNP rs8050136 was in proxy with other well-known variants in this region including rs9939609 ($r^2 = 0.98$), rs17817449 ($r^2 = 0.99$), rs1421085 ($r^2 = 0.93$) and all resided in the same haplotype and are associated with obesity in adults (4–9). Therefore, no independent effect has been identified. **Table 3** shows the summary of SNP results after regression analyses under additive and recessive models adjusted for age, gender, and PC. No significant difference was observed when the cohort site is included as another covariate (Data not shown). Of note, and only in this intron, more than 10 polymorphic indels were also associated with BMI-z, in particular A/AT at chr16:53822169 (MAF = 36%) and TTTC/T at chr16:53829962 (MAF = 35%; $p = 2.41 \times 10^{-5}$, $p = 8.09 \times 10^{-5}$, respectively). We noticed a subtle improvement of overall results using recessive model in our cohorts (**Figure 3B**; **Table 3**). In particular, another published SNP



rs1421085 generated the best result $z = 5.782$, $p_{(\text{Rec})} = 8.21 \times 10^{-9}$ (Table 2).

Although BMI is a continuous trait, standard cut-offs were also used to assess the burden of increased body weight on health as a binary phenotype and estimate the effect size. When the tail distribution was considered (>95% as case and <20% as control), an OR of 1.61 (95% CI = 1.31–1.97, $p = 4.04 \times 10^{-6}$) was detected for the best surrogate marker rs8050136, adjusted by age, sex, and PC. The strongest effect size was observed using a recessive model for rs1421085 [(OR = 2.79, 95% CI = 1.89–4.10), $p = 1.83 \times 10^{-7}$].

When heterogeneity was allowed between cohorts, weaker signals were also obtained in loci, such as near the MC4R region (rs12964056, $p = 6.87 \times 10^{-7}$, $z = -4.98$), cholecystokinin CCK (rs8192472, $p = 1.33 \times 10^{-6}$, $z = -4.85$), Interleukin 15 (rs2099884, $p = 1.27 \times 10^{-5}$, $z = 4.34$), low density lipoprotein receptor-related protein 1B [LRP1B (rs7583748, $p = 0.00013$, $z = -3.81$), and near transmembrane protein 18 (TMEM18; rs7561317, $p = 0.001$, $z = -3.17$)], all of which have been previously reported to be associated with obesity or BMI (Table 3; 4–10). The imputation result for the MC4R region shows multiple association markers (Figure 3C).

We also performed imputation in additional regions of interest. Of note, one new locus that has not been previously reported to be associated with BMI, passed the GWAS significance level in our cohorts ($p < 1.0E-10^{-8}$). The best marker was a relatively infrequent intronic SNP, rs1542829 in the COL6A5 gene in chromosome 3, with overall MAF of 5% that produced a $p = 4.35 \times 10^{-9}$, $z = 5.89$ under an additive model adjusted for age, gender, and PC. The allele frequency of this marker was consistent among cohorts and with CEU-Hapmap data, and it was in HWE. Additional SNP markers in this region, after imputation, produced probabilities at the level of 10^{-5} (Figure 3D). Of note, the majority of these SNPs had MAF of less than 10% with less cohesive haplotype boundaries (Figure 3D). When we considered the tail distribution (>95% as case and <20% as control) of BMI-z as a binary phenotype, an OR of 2.90 (95% CI = 1.93–4.34), $p = 9.03 \times 10^{-8}$ was obtained for this marker with MAF of 9% in cases vs. 3% in controls. Other unreported loci with suggestive associations to pediatric BMI ($10^{-7} < p < 10^{-5}$),

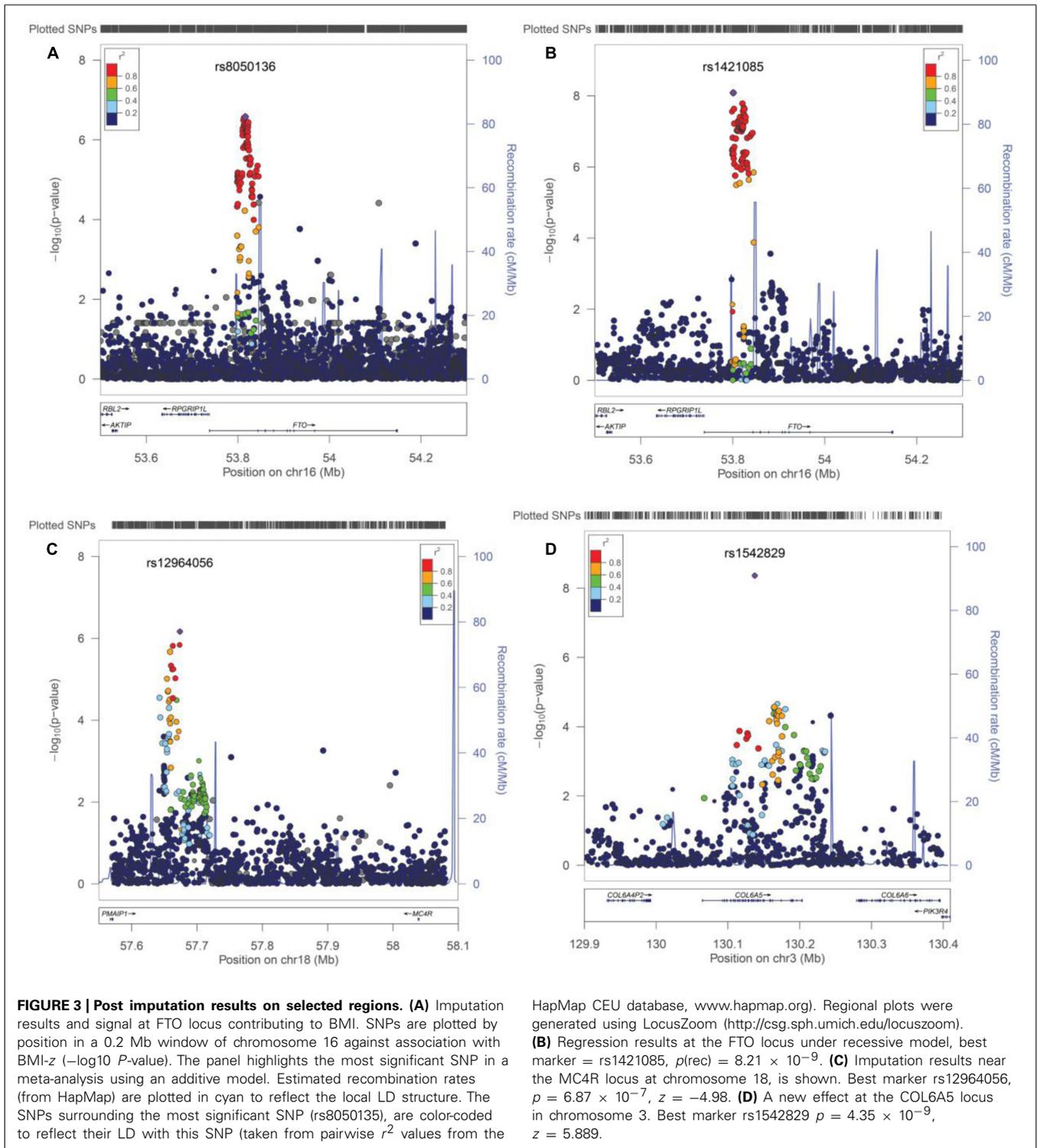
include KCNH5 (rs10136789, $p = 4.62 \times 10^{-7}$, $z = 5.05$), APOL5 (rs2016586, $p = 3.26 \times 10^{-6}$, $z = -4.67$), LRR7 (rs10889850, $p = 1.77 \times 10^{-6}$, $z = -4.78$), and GALNT13 (rs12693973, $p = 1.65 \times 10^{-6}$, $z = -4.80$).

Finally, using ICD-9 diagnostic codes in our collections, we have also performed a Phewas study for the best identified markers in the FTO locus. In this approach, presence or absence of each diagnostic ICD-9 code was included as a binary phenotype, allelic associations were assessed between cases and controls, and the final results were corrected for multiple testing. By this approach, a negative association was detected between cases with hypertrophic cardiomyopathy or valvular structural heart disease and the FTO common risk alleles (ICD-9 codes = 424.0–424.3 and 425) in a subset of 81 cases and 2259 controls. This effect remained significant after 10000 permutations [$p(\text{perm}) = 0.0009$, OR = 0.53 (95% CI 0.37–0.77)]. Suggestive positive associations were also observed between the risk alleles and ICD-9 codes for impaired glucose tolerance test (ICD-9 = 790.2) and myopia (ICD-9 = 367.1; $p < 0.05$), however, this effect did not remain significant after permutation and correcting for multiple testing. It is noteworthy to mention that in our pediatric cohorts, the number of patients with diabetes related diagnostic codes (one of the BMI-related phenotype) was small and not sufficient for independent analysis.

DISCUSSION

In this study, we evaluated BMI in five European ancestry pediatric cohorts with available EMR-linked genotyped data from the CCHMC-BCH eMERGE-II Network site. We successfully utilized anthropomorphic measures to calculate BMI-for-age percentile, derived BMI-z scores according to CDC growth charts, performed quantitative trait locus GWAS study and conducted meta analyses. The overall adjusted meta-analysis result of 2860 European samples produced the best signal in the 16q12 genomic region at the first intron of the FTO locus for a cluster of SNPs. The best typed marker rs8050136 produced [$p_{(\text{Rec})} = 7.34 \times 10^{-8}$], $z = 5.26$] and with no heterogeneity between cohorts ($p = 0.77$). When the tail distribution was considered ($\geq 95\%$ as case and $\leq 20\%$ as control), an OR of 1.61 (95% CI = 1.313–1.965, $p = 4.04 \times 10^{-6}$) was detected. Notably, this OR estimate in our data was relatively higher than previous adult studies (OR = 1.2; Loos, 2012). In any case, considering the effect size (or OR) at the range of 1.2 and the high MAF of the FTO loci (0.40) in Europeans, 700 samples were sufficient for us to achieve an optimum power of 0.8 with a type 1 error level of 0.05. In fact, genetic variants in the first intron of FTO present as the strongest BMI-associated GWAS locus in humans. Over the past few years, the association of the FTO locus has been repeatedly replicated, not only for BMI, but also for obesity risk, body fat percentage, waist circumference, and other obesity-related traits, in particular type II diabetes (Scuteri et al., 2007; Frayling et al., 2007). Recent association with dyslipidemia, hypertension, reduced brain volume, Alzheimer's disease, and dementia has also been reported that could be confounded with obesity and vascular complications (Pausova et al., 2009; Keller et al., 2011).

The functional mechanism of FTO however is still elusive and is currently the subject of intense interest. It is an AlkB-like, Fe(II)-



and 2-oxoglutarate-dependent nucleic acid demethylase that has been shown to demethylate 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA, respectively (Gerken et al., 2007). A link between FTO demethylase activity and increased fat mass was suggested by recent animal studies. Notably, homozygous mutant *fto*^{-/-} mice show postnatal growth retardation and

a significant reduction in adipose tissue and lean body mass, an observation that also was supported by the deleterious mutation Ile367Phe in mouse FTO protein with an impaired activity from a separate study (Church et al., 2009; Gao et al., 2010). In both studies, the leanness of *fto*-deficient mice seems to be the result of increased energy expenditure and systemic sympathetic

activation. Interpreting energy expenditure from these results, however, was challenging because of body composition differences and growth retardation. Overall, these experimental data suggest that inactivation of the FTO gene protects from obesity. On the other hand, mice globally overexpressing FTO are obese, hyperphagic, and exhibit normal energy expenditure when corrected for lean tissue mass (Church et al., 2010). Most human studies also report that obesity-predisposing FTO alleles are associated with increased food intake, but not energy expenditure (Speakman et al., 2008; Haupt et al., 2009). Although the detailed mechanisms are still not clear, this is suggested to be the result of increased expression of FTO in humans, due to cis-regulatory variation in intron 1 of this gene which is consistent with the mouse models, in which decreased levels of FTO cause a lean body habitus (Zabena et al., 2009; Berulava and Horsthemke, 2010). Moreover, a positive correlation of FTO gene expression with other adipocytokine gene expressions, including leptin, perilipin, and visfatin, has also been shown (Haupt et al., 2009).

Consistent with previous studies and because of high LD between FTO intron-1 markers in Europeans (Figure 2), conditional analyses didn't produce independent effect; however, we observed a subtle improvement of FTO association using recessive models in comparison to additive models (Figure 1; Table 2). This effect was unique to the FTO locus, while use of a dominant model was best near the MC4R region (Table 3). Indeed, between-study variations in regard to the optimal inheritance model of the FTO polymorphisms have been noticed previously. In one study, eight different meta-analysis results of BMI in women with polycystic ovary syndromes were systematically reviewed and the recessive model was found to fit best in half of them and the additive model worked best in the rest (Wojciechowski et al., 2012). In our pediatrics population in which we had an enrichment of earlier age of obesity with a prevalence rate of 28%, a recessive pattern may be more relevant in comparison to the general population. Under this model, another intron-1 marker, rs1421085, produced the best result $p_{(Rec)} = 8.21 \times 10^{-9}$, $z = 5.78$ (Table 3; $r^2 = 0.93$ with rs8050136). Indeed, the variant rs1421085 is particularly interesting. It is located within a highly conserved element and the risk allele C has been predicted to substantially reduce binding affinity for CUX1, a transcription factor implicated in the regulation of FTO (Stratigopoulos et al., 2011; Peters et al., 2013).

Limited evidence suggests that the cross-sectional FTO association with BMI varies by age. Specifically, at early ages (up to 5–7 years), the association between common variation at FTO and BMI appears to be reduced in magnitude (Hardy et al., 2010). Longitudinal twin studies also suggest that with increasing age loci such as FTO may be able to exert a greater effect on BMI which depends on the influence of shared environmental effect and the timing of adiposity rebounds (Haworth et al., 2008; Sovio et al., 2011). In our cohorts when we divide all samples into two strata of less than 5 and above 5 years old, we didn't observe any opposing effect, however a higher magnitude was identified for strata above 5 years old (OR = 1.72) (Figure 2). Larger sample size with detailed longitudinal data would seem to be needed to fully elucidate this correlation.

Furthermore, in the context of rare and severe phenotypes, recently, in a large Palestinian Arab consanguineous multiplex family with nine affected, a homozygous R316Q enzyme inactivating mutation in the FTO gene, resulted in a broad spectrum of clinical manifestations including severe intrauterine growth retardation, severe microcephaly, and death from infection before the age of three (Boissel et al., 2009). Of note, six out of eight cases had structural heart defect with cardiomyopathy. This would appear consistent with our unique observation of a negative association of obesity-predisposing FTO alleles with cardiomyopathy [$p_{(corr)} = 0.0009$, OR = 0.53 (95% CI 0.37–0.77)] that could result in lower expression of FTO; although further studies with larger sample sizes are necessary to confirm or refute this finding. In fact, to our knowledge, no SNP in intron 1 of FTO has been previously associated with any trait unrelated to BMI. Recently, an independent effect at the eighth intron of the FTO locus has been reported to be associated with melanoma (GenoMEL Consortium et al., 2013). The best marker, rs16953002, was replicated using 12,313 cases and 55,667 controls of European ancestry in a study conducted by the GenoMEL consortium (combined $P = 3.6 \times 10^{-12}$, OR = 1.16). Notably, in their study, none of the BMI related SNPs in intron 1 were associated with melanoma. Similarly, in our collection, there was no effect observed at the eighth intron of the FTO gene with BMI ($p = 0.54$ for rs16953002, $r^2 < 0.01$ with rs8050136). This suggests independent functions and genetic risks for FTO that broaden the existing paradigm and identify distinct pathogenic effects (GenoMEL Consortium et al., 2013).

In this report, we have supported association in other previously reported BMI loci, in particular loci near the MC4R region (rs12964056, $p = 6.87 \times 10^{-7}$, $z = -4.98$), cholecystokinin CCK (rs8192472, $p = 1.33 \times 10^{-6}$, $z = -4.85$), Interleukin 15 (rs2099884, $p = 1.27 \times 10^{-5}$, $z = 4.34$), low density lipoprotein receptor-related protein 1B [LRP1B (rs7583748, $p = 0.00013$, $z = -3.81$)] and near transmembrane protein 18 (TMEM18; rs7561317, $p = 0.001$, $z = -3.17$; Table 2). Because of their lower effect on BMI and obesity risk (OR ~1.10, in adult meta-data) and lower allele frequency, the identification of these loci requires a quadrupling of the sample size in a random population. Finding all of these loci in our pediatric collections, despite limited sample size, indicate the enrichment of the genetic signal to noise ratio given the shorter amount of time that environment has had an effect.

Additionally, we have detected a new unreported signal at chromosome 3 (Col6A5) (best SNP is rs1542829, MAF of 5% $p = 4.35 \times 10^{-9}$, $z = 5.89$). This marker produced an OR of 2.90 when the tail distributions (>95% as case and <20% as control, 386 cases, and 572 controls) was considered as binary phenotype. Considering this level of OR (2.90), even with MAF of 5%, 500 samples were sufficient for us to achieve the extraordinary power (0.99) with a type 1 error level of 0.05. This could be considered as one of the rare obesity risk loci that we expect to detect in these special cohorts. The $\alpha 5$ -containing collagen VI (Col6A5, COL29A1), belongs to the class of collagens containing von Willebrand factor type A domains. These collagens form filaments with globular domains containing vWA motifs, which are involved in protein-ligand interactions for the organization of tissue architecture and

cell adhesion. Collagen VI is a major extracellular matrix (ECM) protein with a critical role in maintaining skeletal muscle functional integrity. It has been suggested that type VI collagen is a fibrotic component that restricts adipose tissue expandability. In humans, Col6A3 gene expression in adipose tissues was found to correlate with visceral adipose tissue mass and pro-inflammatory gene expression (Pasarica et al., 2009). Mutations in different families of this gene have also been associated with myopathy and muscular dystrophy. Recently, it has been shown that COL6A5 is involved in adhesion at myotendinous and dermal–epidermal junctions (Sabatelli et al., 2012). Different polymorphisms in or near this gene have been linked to atopic dermatitis and eczema, but with contradictory reports (Söderhäll et al., 2007; Naumann et al., 2011). In our cohorts, the number of samples with Atopic dermatitis and related conditions (ICD-9 = 691) was only 49 with a trend of association for published SNP rs7629719 ($p = 0.14$); adjusting the results based on presence or absence of atopy, didn't have any effect on overall BMI associations. A larger sample size is necessary to further elucidate this coexistent condition and to determine whether COL6A5 has any role in obesity related conditions.

Four additional novel loci with homogenous but suggestive associations ($10^{-7} < p < 10^{-5}$) to childhood BMI were also identified in this study. These include KCN5 (rs10136789, $p = 4.62 \times 10^{-7}$, $z = 5.05$), a voltage-gated potassium channel with various function in neurotransmitter regulation, hormone release, cardiac function, and cell volume; APOL5 (rs2016586, $p = 3.26 \times 10^{-6}$, $z = -4.67$), a component of high-density lipoprotein with a potential role in lipid metabolism; LRRC7 or Densin (rs10889850, $p = 1.77 \times 10^{-6}$, $z = -4.78$) a core component of post-synaptic densities and GALNT13 (rs12693973, $p = 1.65 \times 10^{-6}$, $z = -4.80$) a member of the UDP-*N*-acetyl- α -D-galactosamine and a major enzyme responsible for the synthesis of O-glycan. Independent cohorts are necessary to confirm these preliminary suggestive findings and their importance in childhood obesity.

Despite the limitations of using an EMR-derived data set for analysis of secondary phenotypes including errors in data extractions, discordant time of sampling, and underlying coexistent conditions of our pediatric cohorts, we demonstrate that a strong signal, larger than seen in adult populations is detectable. We have removed all inconsistent data and outliers to the best of our ability. We have also excluded infants and those less than 2 years old because of the complexity of growth chart pattern and many potential maternal effects on infants from perinatal periods. In addition we assessed the distribution of BMI-*z* in the whole population with a large sample size as a quantitative trait rather than attempting to identify limited cases and controls. From the statistical standpoint, quantitative traits usually are preferred in meta-GWAS studies because they improve power to detect a genetic effect and often have a more interpretable outcome (Bush and Moore, 2012). Furthermore, BMI is a highly heritable trait in humans and, as mentioned above, up to 70% of the inter-individual variation in obesity can be attributable to genetic factors *per se* (Maes et al., 1997); therefore, given the strong genetic confirmation described here, indeed, we managed to repurpose the genotyping data collected for the analyses of another phenotype

and successfully find association between the new phenotype and genotypic data.

In summary, using the EMR-linked genotyped data, we have confirmed association of several previously known BMI loci, in particular with the FTO gene [OR of 1.61 (95% CI = 1.31–1.97)]. Our data also support the importance of variants at the FTO locus in childhood obesity and with saturation of an earlier age of onset, these data point to a closer functional variant in this locus.

ACKNOWLEDGMENTS

We are grateful to the individuals who participated in this study. We thank the genotyping core facilities in both academic centers (CCHMC-BCH) and our colleagues who facilitated the genotyping and recruitment of subjects. This work was supported by a grant from the National Human Genomic Research Institute: 1U01HG006828. “TL and IS were partially supported by NIH grants 5R00LM010227-04 and 1R21HD072883-01.”

REFERENCES

- Boissel, S., Reish, O., Proulx, K., Kawagoe-Takaki, H., Sedgwick, B., Yeo, G. S., et al. (2009). Loss-of-function mutation in the dioxygenase-encoding FTO gene causes severe growth retardation and multiple malformations. *Am. J. Hum. Genet.* 85, 106–111. doi: 10.1016/j.ajhg.2009.06.002
- Berulava, T., and Horsthemke B. (2010). The obesity-associated SNPs in intron 1 of the FTO gene affect primary transcript levels. *Eur. J. Hum. Genet.* 18, 1054–1056. doi: 10.1038/ejhg.2010.71
- Bush, W. S., and Moore, J. H. (2012). Chapter 11: genome-wide association studies. *PLoS Comput. Biol.* 8:e1002822. doi: 10.1371/journal.pcbi.1002822
- Barlow, S. E., and the Expert Committee. (2007). Expert committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: summary report. *Pediatrics* 120, S164–S192. doi: 10.1542/peds.2007-2329C
- Church, C., Moir, L., McMurray, F., Girard, C., Banks, G. T., Teboul, L., et al. (2010). Overexpression of Fto leads to increased food intake and results in obesity. *Nat. Genet.* 42, 1086–1092. doi: 10.1038/ng.713
- Church, C., Lee, S., Bagg, E. A., McTaggart, J. S., Deacon, R., Gerken, T., et al. (2009). A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. *PLoS Genet.* 5:e1000599. doi: 10.1371/journal.pgen.1000599
- Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., et al. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316, 889–894. doi: 10.1126/science.1141634
- Gerken, T., Girard, C. A., Tung, Y. C. L., Webby, C. J., Saudek, V., Hewitson, K. S., et al. (2007). The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* 318, 1469–1472. doi: 10.1126/science.1151710
- Gao, X., Shin, Y. H., Li, M., Wang, F., Tong, Q., and Zhang, P. (2010). The fat mass and obesity associated gene FTO functions in the brain to regulate postnatal growth in mice. *PLoS ONE* 5: e14005. doi: 10.1371/journal.pone.0014005
- GenoMEL Consortium, Iles, M. M., Law, M. H., Stacey, S. N., Han, J., Fang, S., et al. (2013). A variant in FTO shows association with melanoma risk not due to BMI. *Nat. Genet.* 45, 428–432. doi: 10.1038/ng.2571
- Haupt, A., Thamer, C., Staiger, H., Tschritter, O., Kirchhoff, K., Machicao, F., et al. (2009). Variation in the FTO gene influences food intake but not energy expenditure. *Exp. Clin. Endocrinol. Diabetes* 117, 194–197. doi: 10.1055/s-0028-1087176
- Hardy, R., Wills, A. K., Wong, A., Elks, C. E., Wareham, N. J., Loos, R. J., et al. (2010). Life course variations in the associations between FTO and MC4R gene variants and body size. *Hum. Mol. Genet.* 19, 545–552. doi: 10.1093/hmg/ddp504
- Haworth, C. M., Carnell, S., Meaburn, E. L., Davis, O. S., Plomin, R., and Wardle, J. (2008). Increasing heritability of BMI and stronger associations with the FTO gene over childhood. *Obesity* 16, 2663–2668. doi: 10.1038/oby.2008.434
- Howie, B., Marchini, J., and Stephens, M. (2011). Genotype imputation with thousands of genomes. *G3 (Bethesda)* 1, 457–470. doi: 10.1534/g3.111.001198
- Keller, L., Xu, W., Wang, H. X., Winblad, B., Fratiglioni, L., and Graff, C. (2011). The obesity related gene, FTO, interacts with APOE, and is associated with Alzheimer's

- disease risk: a prospective cohort study. *J. Alzheimers Dis.* 23, 461–469. doi: 10.3233/JAD-2010-101068.
- Kopelman, P. G. (2000). Obesity as a medical problem. *Nature* 404, 635–643.
- Loos, R. J. (2012). Genetic determinants of common obesity and their value in prediction. *Best Pract. Res. Clin. Endocrinol. Metab.* 26, 211–26. doi: 10.1016/j.beem.2011.11.003
- Loos, R. J., Lindgren, C. M., Li, S., Wheeler, E., Zhao, J. H., Prokopenko, I., et al. (2008). Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat. Genet.* 40, 768–775. doi: 10.1038/ng.140
- Maes, H. H., Neale, M. C., and Eaves, L. J. (1997). Genetic and environmental factors in relative body weight and human obesity. *Behav. Genet.* 27, 325–351. doi: 10.1023/A:1025635913927
- McCarty, C. A., Chisholm, R. L., Chute, C. G., Kullo, I. J., Jarvik, G. P., Larson, E. B., et al. (2011). The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med. Genomics* 4:13. doi: 10.1186/1755-8794-4-13.
- Mägi, R., Manning, S., Youssef, A., Pucci, A., Santini, F., Karra, E., et al. (2013). Contribution of 32 GWAS-identified common variants to severe obesity in European adults referred for Bariatric surgery. *PLoS ONE* 7:e70735. doi: 10.1371/journal.pone.0070735
- Naumann, A., Söderhäll, C., Fölster-Holst, R., Baurecht, H., Harde, V., Müller-Wehling, K., et al. (2011). A comprehensive analysis of the COL29A1 gene does not support a role in eczema. *J. Allergy Clin. Immunol.* 127, 1187.e7–1194.e7. doi: 10.1016/j.jaci.2010.12.1123
- Pasarica, M., Gowronska-Kozak, B., Burk, D., Remedios, I., Hymel, D., Gimble, J., et al. (2009). Adipose tissue collagen VI in obesity. *J. Clin. Endocrinol. Metab.* 94, 5155–5162. doi: 10.1210/jc.2009-0947
- Pausova, Z., Syme, C., Abrahamowicz, M., Xiao, Y., Leonard, G. T., Perron, M., et al. (2009). A common variant of the FTO gene is associated with not only increased adiposity but also elevated blood pressure in French Canadians. *Circ. Cardiovasc. Genet.* 2, 260–269. doi: 10.1161/CIRCGENETICS.109.857359
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., et al. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26, 2336–2337. doi: 10.1093/bioinformatics/btq419
- Peters, U., North, K. E., Sethupathy, P., Buyske, S., Haessler, J., Jiao, S., et al. (2013). A systematic mapping approach of 16q12.2/FTO and BMI in more than 20,000 African Americans narrows in on the underlying functional variation: results from the Population Architecture using Genomics and Epidemiology (PAGE) study. *PLoS Genet.* 9:e1003171. doi: 10.1371/journal.pgen.1003171
- Stratigopoulos, G., LeDuc, C. A., Cremona, M. L., Chung, W. K., and Leibel, R. L. (2011). Cut-like homeobox 1 (CUX1) regulates expression of the fat mass and obesity-associated and retinitis pigmentosa GTPase regulator-interacting protein-1-like (RPGRIPL) genes and coordinates leptin receptor signaling. *J. Biol. Chem.* 286, 2155–2170. doi: 10.1074/jbc.M110.188482
- Söderhäll, C., Marenholz, I., Kerscher, T., Rüschemdorf, F., Esparza-Gordillo, J., Worm, M., et al. (2007). Variants in a novel epidermal collagen gene (COL29A1) are associated with atopic dermatitis. *PLoS Biol.* 5:e242. doi: 10.1371/journal.pbio.0050242
- Sabatelli, P., Gualandi, F., Gara, S. K., Grumati, P., Zamparelli, A., Martoni, E., et al. (2012). Expression of collagen VI a5 and a6 chains in human muscle and in Duchenne muscular dystrophy-related muscle fibrosis. *Matrix Biol.* 31, 187–196. doi: 10.1016/j.matbio.2011.12.003
- Sovio, U., Mook-Kanamori, D. O., Warrington, N. M., Lawrence, R., Briollais, L., Palmer, C. N., et al. (2011). Association between common variation at the FTO locus and changes in body mass index from infancy to late childhood: the complex nature of genetic association through growth and development. *PLoS Genet.* 7:e1001307. doi: 10.1371/journal.pgen.1001307
- Speakman, J. R., Rance, K. A., and Johnstone, A. M. (2008). Polymorphisms of the FTO gene are associated with variation in energy intake, but not energy expenditure. *Obesity (Silver Spring)* 16, 1961–1965. doi: 10.1038/oby.2008.318
- Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* 42, 937–948. doi: 10.1038/ng.686
- Scuteri, A., Sanna, S., Chen, W. M., Uda, M., Albai, G., Strait, J., et al. (2007). Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet.* 3:e115. doi: 10.1371/journal.pgen.0030115
- Wojcickowski, P., Lipowska, A., Rys, P., Ewens, K. G., Franks, S., Tan, S., et al. (2012). Impact of FTO genotypes on BMI and weight in polycystic ovary syndrome: a systematic review and meta-analysis. *Diabetologia* 55, 2636–2645. doi: 10.1007/s00125-012-2638-6
- Willer, C. J., Li, Y., Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191. doi: 10.1093/bioinformatics/btq340
- Willer, C. J., Speliotes, E. K., Loos, R. J., Li, S., Lindgren, C. M., Heid, I. M., et al. (2009). Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* 41, 25–34. doi: 10.1038/ng.287
- Zabena, C., González-Sánchez, J. L., Martínez-Larrad, M. T., Torres-García, A., Alvarez-Fernández-Represa, J., Corbatón-Anchuelo, A., et al. (2009). The FTO obesity gene. Genotyping and gene expression analysis in morbidly obese patients. *Obes. Surg.* 19, 87–95. doi: 10.1007/s11695-008-9727-0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 October 2013; paper pending published: 31 October 2013; accepted: 16 November 2013; published online: 03 December 2013.

Citation: Namjou B, Keddache M, Marsolo K, Wagner M, Lingren T, Cobb B, Perry C, Kennebeck S, Holm IA, Li R, Crimmins NA, Martin L, Solti I, Kohane IS and Harley JB (2013) EMR-linked GWAS study: investigation of variation landscape of loci for body mass index in children. *Front. Genet.* 4:268. doi: 10.3389/fgene.2013.00268

This article was submitted to *Applied Genetic Epidemiology*, a section of the journal *Frontiers in Genetics*.

Copyright © 2013 Namjou, Keddache, Marsolo, Wagner, Lingren, Cobb, Perry, Kennebeck, Holm, Li, Crimmins, Martin, Solti, Kohane and Harley. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.