Polygenic risk score based on weight gain trajectories is a strong predictor of childhood obesity

Sarah J. C. Craig[1,2#], Ana M. Kenney[3#], Junli Lin[3], Ian M. Paul[2,4], Leann L. Birch[5], Jennifer Savage[6,7], Michele E. Marini[7], Francesca Chiaromonte[2,3,8]*, Matthew L. Reimherr[2,3]*, Kateryna D. Makova[1,2]*

1) Department of Biology, Penn State University, University Park
2) Center for Medical Genomics, Penn State University, University Park, PA
3) Department of Statistics, Penn State University, University Park, PA
4) Department of Pediatrics, Penn State College of Medicine, Hershey, PA
5) Department of Foods and Nutrition, University of Georgia, Athens, GA
6) Department of Nutritional Sciences, Penn State University, University Park, PA
7) Center for Childhood Obesity Research, Penn State University, University Park, PA
8) EMbeDS, Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà, Pisa, Italy

#Authors contributing equally
*Co-corresponding authors
Kateryna Makova (kdm16@psu.edu)
Matthew Reimherr (mreimherr@psu.edu)
Francesca Chiaromonte (fxc11@psu.edu)

Abstract

Obesity is highly heritable, yet only a small fraction of its heritability has been attributed to specific genetic variants. Missing heritability is particularly pronounced for childhood obesity. Here we studied 226 children for whom we typed almost one million single-nucleotide polymorphisms (SNPs), and collected weight and length or height at eight time points between birth and the age of three years. Leveraging longitudinal weight gain trajectory information and novel functional data analysis (FDA) techniques, we constructed a polygenic risk score (PRS) comprised of 24 SNPs. This PRS explains 56% of the variability in weight gain trajectories among the studied children. Moreover, it is significantly higher in children with (vs. without) rapid infant weight gain—a predictor of obesity later in life. We validated the constructed PRS in populations of adolescents and adults—suggesting that some genetic variants predispose to obesity at both childhood and later life stages. In contrast, PRSs from genome-wide association studies (GWAS) of adult obesity were not predictive of weight gain in our cohort of children, and did not share SNPs with our PRS. Our research provides a strong example of a successful application of FDA to a GWAS. We demonstrate that a sophisticated characterization of a longitudinal phenotype can provide increased statistical power to studies with smaller sample sizes. This has the potential of shifting the existing paradigm in GWAS.

Introduction

Obesity is a rising epidemic, and one that is increasingly affecting children. In 2018, 18% of children in the United States were obese and approximately 6% were severely obese[1]—a substantial increase from previous years[2]. Given the strong association between weight gain during childhood and obesity across the lifecourse[3], the search for early life risk factors has become a research and public health priority.

Obesity is a complex disease with an etiology influenced by environmental, behavioral, and genetic factors, which likely interact with each other[4]. For childhood obesity, dietary composition and sedentary lifestyle have often been cited as main contributors[5]. Evidence also exists for a significant role of parents' socioeconomic status[6] and maternal prenatal health factors including gestational diabetes[7] and smoking[8]. Obesity risk in children has also been associated with appetite[9] which has been shown to be partially influenced by genetics[10].

The heritability of obesity has been estimated to be between 50% and 90% (with the highest values reported for monozygotic twins and the lowest for non-twin siblings and parent-child pairs, reviewed in [11]). This is a much higher percentage than that accounted for by the genetic variants found so far[12,13]. Therefore, obesity suffers from "missing heritability"—a broad discrepancy between the estimated heritability of the phenotype and the variability explained by genetic variants discovered to date. Indeed, the search for specific genetic variants that increase the risk of obesity, in adulthood as well as in childhood, is still ongoing. Using whole-genome sequencing, researchers have found variants in individual genes that contribute to severe, early-onset obesity[14]. Moreover, genome-wide association studies (GWAS) have identified single nucleotide polymorphisms (SNPs) that are significantly associated with obesity phenotypes such as increased body mass index (BMI), high waist-to-hip ratio, etc.[15–21]. Albeit successful, these studies have some shortcomings; the individual contributions of the identified SNPs tend to be very small[12], and the prevalent focus is still on adult cohorts—with only one childhood obesity study for every 10 adult obesity studies[22].

One way to utilize the information gained from GWAS is to summarize the risk from multiple disease-causing alleles in polygenic risk scores (PRSs) that can be computed for each individual[23]. These scores are either simple counts (unweighted) or weighted sums of disease-causing alleles identified by GWAS. Notably, while several studies have constructed PRSs for childhood obesity[24–27], most have done so relying on SNPs identified by GWAS on adult BMI. Since SNPs affecting obesity risk in adults and children may differ[28,29,30], this may explain the limited[12, 31] and age-dependent[32] explanatory power of such scores for children's weight gain status.

In this study, we attempted to bridge this gap by focusing specifically on SNPs affecting obesity risk in children and by using novel, highly effective Functional Data Analysis (FDA) statistical methods developed by our group. Based on data from a deeply characterized pediatric cohort[33–35], we constructed children's growth curves and treated them as a longitudinal phenotype. FDA fully leverages

3

this longitudinal information, extracting complex signals that can be lost in standard analyses of cross-sectional or summary measurements. This increases power and specificity for assessing potentially complex and combinatorial genetic contributions. Moreover, FDA models genetic effects on the entire growth curve non-parametrically. This characterizes changes in effect size over time in a more flexible and effective manner than other statistical methods for longitudinal data. With our analyses, we identified genetic variants significantly associated with children's growth curves and combined them in a novel PRS that is strongly predictive of growth patterns and rapid infant weight gain[36,37], which is associated with obesity later in life. We also investigated how environmental and behavioral covariates compound with our novel score in affecting growth curves, and provided biological and statistical validations of our findings.

Results

Participants and DNA typing

Our study utilized 226 first-born children (out of a total of 279) enrolled in the Intervention Nurses Start Infants Growing on Healthy Trajectories (INSIGHT) study[33]. For these children, weight and length were measured at birth, 4 weeks, 16 weeks, 28 weeks, 40 weeks, and one year, and weight and height—at two and three years. Using the ratios of weight for length or height (henceforth referred to *weight-for-length/height*) at these eight time points, we constructed *growth curves* for all children (Fig. 1a; see Methods). We used weight-for-length/height ratio because it is the recommended measurement for identification of children at risk for obesity under the age of two years by the American Academy of Pediatrics (BMI is recommended afterwards)[38]. Six out of eight time points in our study fall into this category, therefore for consistency we utilized weight-for-length/height ratio for all eight time points analyzed.

In addition to growth curves, we computed *conditional weight gain* for each child (change in weight between birth and 6 months, correcting for length, see Methods). Conditional weight gain was shown to be an effective indicator of risk for developing obesity later in life in a previous study[39]. Also in our study, children who experienced *rapid infant weight gain*, i.e. those with a positive conditional weight gain, had a significantly greater weight at one ($p<2.2 \times 10^{-16}$), two ($p=9.1 \times 10^{-14}$), and three ($p=6.2 \times 10^{-13}$) years of age than children who did not (one-tailed t-tests, Fig. S1).

We isolated genomic DNA from blood samples from the 226 children and genotyped it on the Affymetrix Precision Medicine Research Array containing 920,744 SNPs across the genome. SNPs that had missing information, a minor allele frequency below 0.05, or were in the mitochondrial DNA were removed from the dataset—leaving a total of 79,498 SNPs for subsequent analyses (Fig. S2).
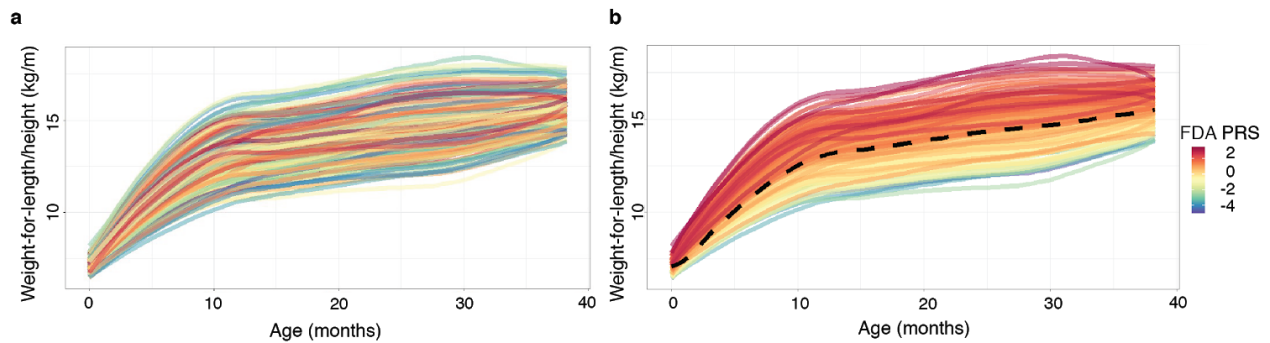
Figure 1. (a) Growth curves (color-coded by participant's ID) from birth to three years for 226 children enrolled in the INSIGHT study. (b) The same growth curves color-coded based on a gradient corresponding to our FDA-based Polygenic Risk Score. The dashed black line is the mean curve.

FDA-based Polygenic Risk Score predicts growth curves and rapid infant weight gain

The sample size of our study (n=226) is small for a traditional GWAS. However, our FDA approach allowed us to leverage the longitudinal information in growth curves to identify significant SNPs and combine them into a polygenic risk score (PRS). More specifically, we used FDA screening[40] to first reduce the analysis from 79,498 to 10,000 potentially relevant SNPs. Next, we used Functional Linear Adaptive Mixed Estimation (FLAME)[41] to identify 24 SNPs as significant predictors of children's growth curves (Table 1). Finally, we constructed our novel *FDA PRS* as a weighted sum of allele counts across the 24 selected SNPs, with weights determined with additional FDA techniques (see Methods). We found that FDA PRS is indeed a strong predictor for growth curves with a significant positive effect on weight-for-length/height ratios across time ($R^2=0.56$, $p<1\times10^{-15}$, function-on-scalar regression, see Methods), and especially between ~10 and ~30 months of age (Fig. 2a). This can also be observed noting that growth curves of children with high PRS values are concentrated above the mean curve (Fig. 1b). Moreover, FDA PRS is significantly larger for children with rapid infant weight gain compared to those without (one-tailed t-test, $p=4.2\times10^{-10}$; Fig. 2b), and is positively correlated with conditional weight gain ($R^2=0.19$, $p<1\times10^{-05}$; Fig. 2c) as well as with weight-for-length/height ratio at one ($R^2=0.50$, $p<1\times10^{-5}$), two ($R^2=0.53$, $p<1\times10^{-5}$), and three ($R^2=0.46$, $p<1\times10^{-5}$) years of age (Fig. S3).

These results are in sharp contrast with those we obtained for our cohort using a PRS based on adult obesity SNPs from another study. For each child, we calculated *Belsky PRS*—a weighted PRS based on 29 SNPs identified through adult obesity GWAS as described by Belsky and colleagues[26]. This PRS was used because it was correlated with BMI outcomes from age three to 38, so we hypothesized that it would be a good predictor of weight outcomes across the lifecourse. However, Belsky PRS is not a significant predictor of our children's growth curves from birth through age three ($R^2=0.0032$, $p=0.35$, function-on-scalar regression, Fig. 2d). Furthermore, Belsky PRS is not significantly larger for children with rapid infant weight gain compared to those without (one-tailed t-test, $p=0.22$; Fig. 2e) and does not display significant correlations with conditional weight gain ($R^2=0.0009$, $p=0.66$; Fig. 2f) and weight-for-length/height ratio at one ($R^2=0.0064$, $p=0.25$), two ($R^2=0.0036$, $p=0.37$), and three

(R^2=0.0009, p=0.71) years of age (Fig. S4). Additionally, we calculated three other previously published childhood obesity PRSs for our cohort—Elks PRS[25], den Hoed PRS[24], and Li PRS[27]. Similar to the Belsky PRS, these scores were not correlated with conditional weight gain (Fig. S5a-c) and there was not a significant difference in PRS values between children with vs. without rapid infant weight gain (Fig. S5d-f).
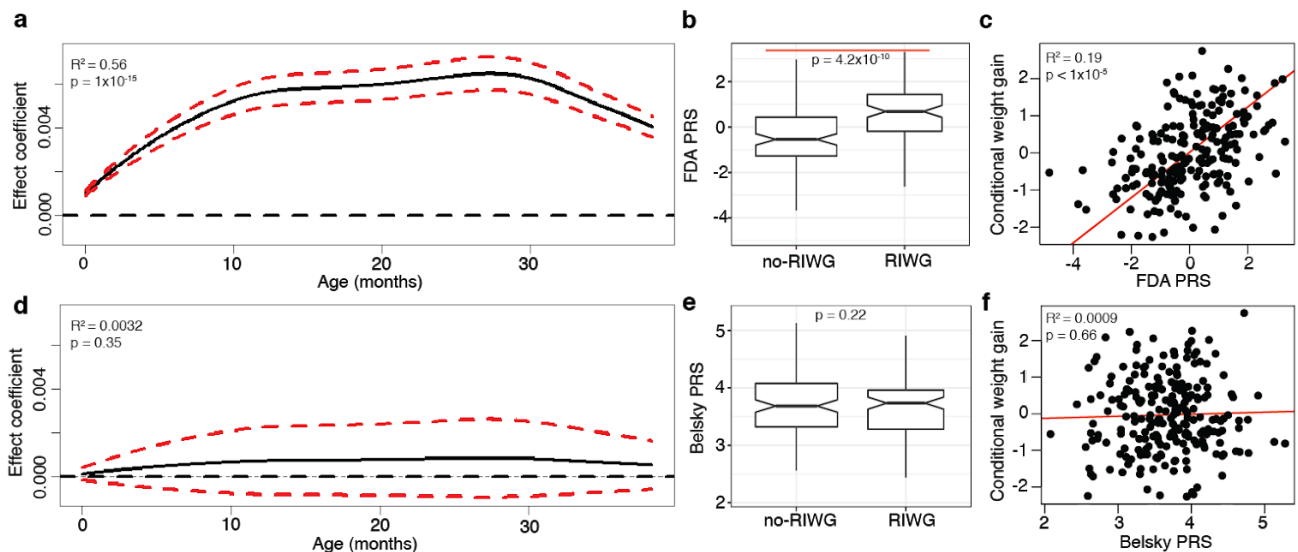

Figure 2. Polygenic risk scores (PRSs) and children's growth patterns. Estimated effect coefficient for a PRS as a predictor of children's growth curves in a function-on-scalar regression for **(a)** FDA PRS, or **(d)** Belsky PRS. Boxplots comparing a PRS between children with vs. without rapid infant weight gain (i.e. RIWG vs. no-RIWG) for **(b)** FDA PRS, or **(e)** Belsky PRS. Scatterplot of conditional weight gain vs. a PRS using **(c)** FDA PRS, or **(f)** Belsky PRS.

The 24 SNPs included in the FDA PRS (Table 1) do not appear in prior PRSs for either childhood or adult obesity, and, interestingly, are not located in genes commonly associated with obesity (e.g., *FTO*[32,42] and *MC4R*[15,42]). However, nine of the 24 SNPs in our FDA PRS can be linked directly or indirectly to obesity-related traits. In particular, using the NHGRI-EBI GWAS catalog (https://www.ebi.ac.uk/gwas/), we found that some of the SNPs are located in genes associated with BMI (rs4915535, rs10227226, rs471670), cholesterol levels (rs12039940, rs9837708, rs17626544), Type 2 diabetes (rs638348), and hypertension (rs1539759). Some SNPs we discovered are located in the vicinity of obesity-related genes. In addition to being located in *ZNF648*, a gene important for determining HDL cholesterol levels, rs12039940 is located downstream of *CACNA1E*, a gene associated with BMI change over time[43]. Another instance is rs72679478, a SNP with a high weight in the FDA PRS (Table 1). It is located within *DNAJC6*, a gene associated with Parkinson's disease, but it is also just upstream of the leptin receptor gene (*LEPR*) which has been associated with early-onset adult obesity[44]. Potential relationships between the remaining 15 SNPs and obesity should be investigated in future studies.

Contributions of environmental and behavioral covariates

Children's weight gain patterns can be affected by a variety of environmental and behavioral factors, which compound to genetic effects. To evaluate their potential effects on our results, we considered a functional regression (see Methods) of the growth curves on FDA PRS plus 11 potential confounding covariates, namely: maternal pre-pregnancy BMI, paternal BMI, child's birthweight, maternal gestational weight gain, maternal gestational diabetes, maternal smoking during pregnancy, mode of delivery, the child's sex, mother-reported child's appetite score, INSIGHT intervention group, and family socioeconomic status (Table 2). FLAME[41] applied to this regression identified FDA PRS, birthweight and appetite as significant predictors—however, the variability explained by these three predictors (R^2=0.57; Table S1) is very similar to that explained by the FDA PRS alone (R^2=0.56). Thus, genetic effects captured by our FDA PRS remain significant, and in fact strongly dominant, also when accounting for the environmental and behavioral covariates at our disposal.

To confirm these results we also considered the regression of conditional weight gain[39] on the same 12 predictors as above. Best subset selection applied to this regression identified FDA PRS (p=8.65x10^{-09}) and appetite (p=2.80x10^{-05}), but not birthweight, as significant positive predictors. The variability explained (R^2) produced by these two predictors is 0.24 (Table S1), only five percentage points higher than the one produced by FDA PRS alone (R^2=0.19). Thus, again, the majority of the explanatory power remains attributable to the FDA PRS. Group LASSO[45] applied to this regression identified FDA PRS and birthweight, but not appetite, as relevant predictors. It also selected maternal pre-pregnancy BMI and paternal BMI, but led to a lower R^2 of 0.21 (Table S1).

Notably, and not unexpectedly given its lack of association with children's growth patterns, when we reran the analyses presented above using the Belsky PRS (instead of the FDA PRS), we did not identify it as a significant predictor. For instance, best subset selection for the regression of conditional weight gain on the Belsky PRS plus the 11 environmental and behavioral covariates at our disposal retained only appetite as positive and significant predictor (p=5.53x10^{-05}); all other predictors, including the Belsky PRS itself, were eliminated.

Validation of the FDA-based Polygenic Risk Score

Biological validation of the FDA PRS. The analyses presented above assess the predictive power of our FDA PRS "in-sample"—that is, on the same data on which we selected SNPs and estimated the scoring weights.To validate the FDA PRS, we considered two independent datasets from dbGaP. It is important to note that we could not identify publicly available data from an independent cohort that matches our study design (i.e. with genome-wide SNP data and longitudinal weight and length or height measurements for children under the age of three). We thus used dbGaP data from older individuals, fully aware of the fact that these are not ideal for our purposes.

Remarkably, we were able to successfully validate FDA PRS in two independent cohorts consisting of much older individuals—adolescents and adults—as compared with our study population of

7

three-year-olds. The first dataset consists of 525 adolescents between the ages of 12 and 15 from the Philadelphia Neurodevelopment Cohort (dbGaP study phs000607.v3.p2[46–48]). Individuals are classified based on BMI-for-age percentiles as underweight (<5th percentile), normal (5th to <85th percentile), overweight (85th to <95th percentile), and obese (≥95th percentile). The distributions of our FDA PRS in these classes shift towards larger values as BMI increases from underweight to obese (Fig. 3a, upper and lower panels). While this does not translate in significant differences between all pairs of classes, the FDA PRS of obese adolescents is significantly higher than that of underweight adolescents (p=0.012, one-tailed t-test).

The second dataset consists of 3,486 adults (≥18 years of age) from the eMERGE study (dbGaP study phs000888.v1.p1) who are classified as extremely obese (BMI ≥ 40 kg/m^2) or non-obese (20 kg/m^2 ≤ BMI < 30 kg/m^2). Extremely obese individuals have significantly higher FDA PRS than non-obese individuals (one-tailed t-test, p=3.2x10^{-3}, Fig. 3b). Thus, FDA PRS based on children's weight gain patterns is predictive of extreme obesity later in life.
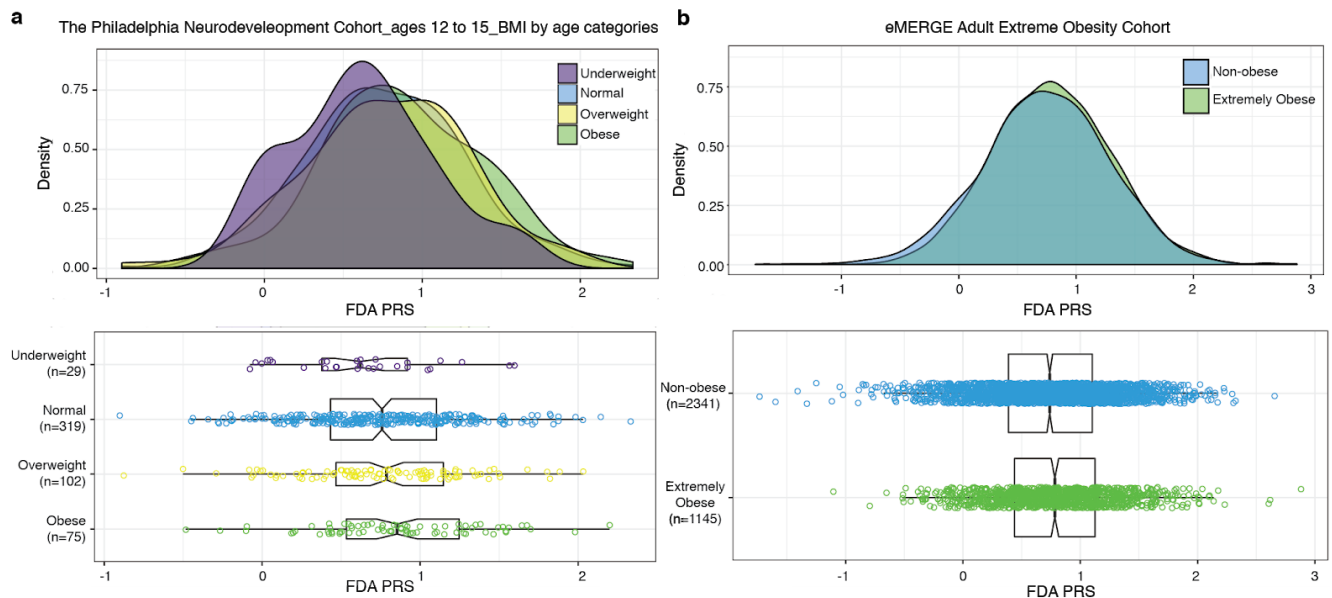

Figure 3. FDA-based Polygenic Risk Score and obesity in adolescent and adults validation cohorts. (**a**) Distributions of FDA PRS in adolescents (age 12 to 15 years) from The Philadelphia Neurodevelopment Cohort, classified as underweight, normal, overweight, and obese according to BMI-for-age percentiles (see text for details). (**b**) Distributions of FDA PRS in adults (over 18 years of age) from the eMERGE study classified as extremely obese or non-obese according to BMI values. Upper panels: density plots. Lower panel: boxplots.

Finally, as an additional form of biological validation of our FDA PRS constructed using weight-for-length/height ratio growth curves, we considered growth curves constructed using BMI. As mentioned above, weight-for-length/height ratio is recommended for children under two years of age by the American Academy of Pediatrics[38] . However, our cohort is observed at ages two and three, when BMI is recommended as the most meaningful measurement[38] . Thus, we also considered growth curves constructed using BMI measurement at all eight time points of the INSIGHT study. Notably, our weight-for-length/height FDA PRS is a strong predictor also for the BMI growth curves (R^2=0.43,

p<1x10^{-15}, function-on-scalar regression)—suggesting a reasonable consistency between the information conveyed by the two measurements, at least up to this age.

Statistical validation of the selected SNPs. We also assessed the robustness of our FDA-based SNP selection with a sub-sampling scheme akin to a 20-fold cross-validation on our original dataset. Specifically, we randomly split the data (i.e. the participants) in 20 equal parts and applied FLAME[41] to perform SNP selection 20 times, using different 19/20 of the data each time. We next counted how many times (out of 20) each SNP was selected. Notably, for the 24 SNPs included in our FDA PRS, the weights computed to construct the PRS correlate with the number of times the SNPs are selected in this sub-sampling scheme (Fig. 4). The frequency of selection captures, in a way, how stable the effect of a genetic variant is amid the complex and combinatorial signals in this type of data. Moreover, SNPs which have both the largest weights and the highest selection frequency may be the most important to interpret and validate in future studies.
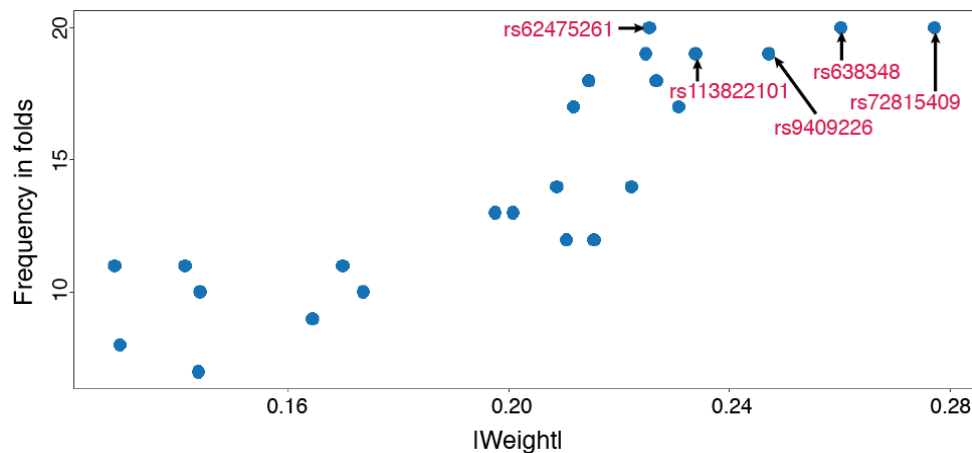

Figure 4. **Statistical validation of FDA-based SNP selection.** The frequency with which the 24 SNP included in the FDA PRS are re-selected in a 20-fold sub-sampling scheme is plotted against their absolute weights in the FDA PRS—showing a strong positive association. The SNPs with both the largest weights and the highest re-selection frequency (top five SNPs labeled by arrows) may be the most important to interpret and validate in future studies.

Discussion

A novel, highly predictive FDA-based PRS for childhood obesity. In this study, we used FDA techniques to construct a novel polygenic risk score which includes 24 SNPs selected based on children's longitudinal weight gain patterns. Among our study participants, this score explains approximately 56% of the variability in growth curves from birth to the age of three years, and approximately 19% of the variability in conditional weight gain. Moreover, our score validates on two independent datasets comprising adolescent and adults individuals, and our SNP selection shows statistical robustness.

9

While the 24 SNPs identified by our study do not appear in prior polygenic risk scores for either childhood or adult obesity, some are located in genes linked to obesity-related phenotypes in previous GWAS studies. Among the others, three SNPs are within genes previously associated with child development (puberty timing) and four within genes linked to periodontitis. Connections between obesity and puberty timing[49], as well as between obesity and periodontitis in adults[50], have been suggested, yet the functional mechanisms remain unclear. As with all GWAS-type studies, it is important to note that some of the identified SNPs may not be truly "causal", but may be in linkage disequilibrium with causal SNPs—and the genes in the immediate vicinity of such SNPs may not be those through which the phenotype is influenced (e.g., rs72679478 located upstream of the leptin receptor gene). We have also identified several SNPs with no prior associations, some of such SNPs have high statistical robustness and high weight in the PRS, and thus need to be we investigated in future functional experiments.

The power of FDA-based GWAS. Our results demonstrate a key advantage of FDA-based GWAS over traditional GWAS. Our study was set up as an ultra-high dimensional problem—with many more predictors (i.e. SNPs) than observations (i.e. individuals). By integrating FDA techniques into every step of the analysis, from the screening and selection of SNPs through the construction of the polygenic risk score, we were able to utilize a more dynamic and information-rich phenotype than the ones used in traditional cross-sectional analyses. In turn, this allowed us to unveil subtler, more complex effects with limited information. This is a valuable contribution as it expands the scope of GWAS to studies that do not comprise tens of thousands of individuals—but instead a few hundred deeply characterized participants[51].

Genetics of childhood and adult obesity. Previous studies supported a relationship between polygenic scores including adult BMI SNPs and childhood weight gain status[24–27]. However, this relationship was generally weak—and weaker the younger the age of the children[13,24,26]. In fact, Belsky and colleagues[26] themselves found no relationship between their PRS (i.e. Belsky PRS) and BMI at birth ($R^2=0.00$, $p>0.9$) and a very weak relationship at three years of age ($R^2=0.0064$, $p<0.01$). This is confirmed by our inability to detect a relationship between the Belsky PRS and growth curves or conditional weight gain measurements in our cohort. In contrast, our FDA-based PRS comprising SNPs identified from childhood growth curves was able to distinguish extreme obesity-related phenotypes in adolescents and adults from two independent validation cohorts. Thus, while SNPs with strong effects on adult obesity are minor or insignificant contributors to weight gain in childhood, the SNPs with strong effects on such gain in childhood do predict obesity later in life. This is consistent with the notion that early life weight gain, and hence its genetic underpinning, predispose to obesity across the lifecourse[3].

Other contributing factors and perspectives. Behavioral and environmental factors are important variables to consider when investigating the etiology of complex diseases. In our study we considered 11 such factors that could influence child weight gain trajectories and found that, while the FDA PRS is by far the dominant predictor, an appetite score computed on our cohort (see Methods) and birthweight have significant effects. We also found some evidence for an effect of parental BMI measurements. It has been shown that a child's appetite behavior impacts early weight gain and may have a strong genetic basis[31,10]. In agreement with this, a recent study found a positive relationship between a

childhood obesity PRS and appetite[52]. In our study child's appetite behaviour was reported by his/her mother which could have introduced some biases. Because appetite is emerging as an interesting predictor of child weight gain status, it should be explored in more detail in future studies. Birthweight has also been associated with the genetic risk for obesity, although the strength of the association between weight and genetics seems to increase as one ages[25,27,53]. Finally, parental BMI has been associated with children being overweight or obese[54–56], which could be explained by shared environment, shared genetics or their interaction.

In addition to the type of environmental and behavioral factors considered in our study, other factors may compound to and interact with genetics in shaping obesity risks. These include the microbiome, the metabolome, and the epigenome. We found previously that children's oral microbiota composition is associated with growth curves[57]. Moreover, we are collecting data on the metabolomes and epigenomes of the children in our study cohort. Our overarching goal is to develop a multi-omic model to comprehensively understand the development of childhood obesity and identify a combination of risk factors that can be used for accurate identification of children who would benefit most from early life intervention programs.

Our FDA-based polygenic risk score was computed considering the longitudinal change in weight-for-length/height ratio from birth through three years of age. An ongoing follow-up of our study participants, with weight and height collected at later time points, will allow us to further evaluate the predictive power of the FDA PRS as age progresses. Finally, we note that our children cohort (Table 2), as well as the cohorts of adolescents and adults used for validation, consisted predominantly of individuals of European ancestry. It will be of great interest to conduct similar analyses on individuals of non-European ancestries, and identify differences and commonalities in the genetic factors contributing to obesity risks among different ethnicities.

Methods

Study sample, growth curves, and conditional weight gain

We collected genetic information from 226 children recruited from the 279 families involved in the INSIGHT study[33]. These children are full-term singletons born to primiparous mothers in Central Pennsylvania. The INSIGHT study is a randomized, responsive-parenting behavioral intervention aimed at the primary prevention of childhood obesity against a home safety control. INSIGHT collected clinical, anthropometric, demographic, and behavioral variables on the children between birth and the age of three years (Table 2). In this study we utilized 11 of these variables including maternal pre-pregnancy BMI, paternal BMI, maternal pregnancy health variables (gestational weight gain, gestational diabetes, and smoking during pregnancy), family income (as a proxy for socioeconomic status), mode of delivery, child's sex, child's birth weight, INSIGHT intervention group (intervention or control), and mother-reported child's appetite at 44 weeks. The appetite score is an ordinal variable on a scale from 1-5 which summarizes the Child Eating Behavior Questionnaire (CEBQ)[58]. Domains on the

CEBQ include food responsiveness, emotional over-eating, food enjoyment, desire to drink, satiety responsiveness, slowness in eating, emotional under-eating, and food fussiness. Length was measured using a recumbent length board (Shorr Productions) for visits before two years (birth, 3-4 weeks, 16 weeks, 28 weeks, 40 weeks, and one year). Standing height was measured with a stadiometer (Seca 216) at two and three years.

To construct growth curves, we utilized the anthropometric data collected above to calculate weight-for-length/height ratio at each time point for our analysis. We used FDA to analyze these longitudinally as individual functions through the *fdapace* package in *R*. This package implements the Principal Analysis by Conditional Estimation (PACE) algorithm[59], which pools information across subjects for more accurate curve construction. We used the default settings and represented them in Fig. 1a using 51 cubic spline functions with evenly spaced knots.

Conditional weight gain z-scores were calculated as the standardized residuals from a regression of age- and sex-specific weight-for-age z-score at 6-months on the weight-for-age z-score at birth (determined using the World Health Organization sex-specific child growth standards)[34]. Length-for-age z-score at 6-months, length at birth, and precise age at the 28-week visit were considered as cofactors in this regression and thus only the change in weight between birth and 6-months was captured[34,36]. These scores are approximately normally distributed and have, by construction, a mean of 0 and a standard deviation of 1. Positive conditional weight gain z-scores correspond to a greater than average weight gain and are used to define rapid infant weight gain, which is a risk factor for developing obesity later in life[37,60,61].

Genotyping

Blood from a fingerstick was collected at the child's one year clinical research visit. Genomic DNA was isolated (Qiagen DNeasy Blood and Tissue Kit) and genotyped on the Affymetrix Precision Medicine Research Array (PMRA). Initial quality filtering was performed using the following criteria: we removed SNPs with minor allele frequency >0.05 and/or present in less than 5% of individuals. All quality filtering steps were performed in PLINK v1.9[62,63] with 79,498 SNPs remaining after quality filtering.

To obtain missing genotype calls and genotypes not included on the PMRA, we performed imputation. Individual's genotypes were first phased leveraging pedigree information (genotypes were also collected for mother and father in most cases, and some younger siblings) using SHAPEIT2[64,65]. The phased haplotypes were then used for imputation using the 1,000 Genomes Project phase 3 data[66] as a reference panel in IMPUTE2[67]. SNPs with imputation probability <90% were removed. Following imputation, we had 12,479,343 SNPs.

Functional Data Analysis techniques

First, we used an *FDA feature screening* method[40], which is an effective and fast procedure to filter out SNPs that are clearly unimportant, yielding a substantially smaller subset of SNPs that can then be used in a more advanced joint model[68]. This method is specifically designed for longitudinal GWAS and

12

can handle up to millions of SNPs. The method evaluates each SNP individually fitting a simplified model comprising only that SNP (with no other SNPs involved) and calculating a weighted mean squared error. This is then used to rank the SNPs. In our study, the top 10,000 SNPs were selected with this feature screening step. See Supplementary Note S1 for additional details about FDA feature selection.

After feature screening, we used FLAME (*Functional Linear Adaptive Mixed Estimation*)[41], a method that simultaneously selects important predictors and produces smooth estimates for function-on-scalar linear models. This method further downselects from the pool of the top 10,000 SNPs as ranked within our screening step. In addition, it provides smooth estimates of the effects of the selected SNPs on the growth curves. To tune the penalty involved in FLAME, we split our observations into training (75%) and test (25%) sets. This procedure resulted in 24 SNPs and their corresponding estimated effect curves. FLAME was also used to assess the statistical robustness of SNP selection in a 20-fold sub-sampling scheme (selection was repeated 20 times, each time on 19/20 of the data); for this exercise we fixed the penalty level, to be consistent across folds. See Supplementary Note S2 for additional details about FLAME.

Next, we used the estimated effect curves produced by FLAME for each of the 24 selected SNPs to construct our *FDA-based Polygenic risk score*. This was done choosing SNP-specific weights that maximize the covariance between weighted SNP counts and growth curves fitted through the FLAME[41] estimates—thus incorporating both the dynamic nature of the SNP effects and linkage disequilibrium between the SNPs themselves. We applied the weights to the allele counts of each child, and computed his/her FDA PRS as the weighted sum of counts across the selected SNPs. Thus, FDA allows us to exploit the longitudinal structure of our data to not only screen and select SNPs, but also weigh them using estimates of how their effects change over time.

We assessed the association between growth curves and the FDA PRS fitting function-on-scalar linear models[69]. The significance of the FDA PRS was determined based on three tests[70] employing different types of weighted quadratic forms. One employs a simple L2 norm of the parameter estimate (L2), another uses principal components to reduce dimension prior to a Wald-type test (PCA), and the last blends the two through the addition of a weighted scheme in the PCA (Choi). We reported the more conservative of the three values.

Polygenic Risk Scores constructed by other studies

To run the calculations of Belsky PRS[26], Elks PRS[25], den Hoed PRS[24], and Li PRS[27], we employed the Allelic Scoring function in PLINK v1.9[62,63]. In some cases proxy alleles had to be used in place of SNPs that were not assessed on the PMRA. Such proxies were determined using linkage disequilibrium with LDlink[72]. Tables describing the composition of each PRS can be found in the Supplemental Materials (Tables S2-S5).

Validation datasets

We used two validation datasets downloaded from dbGaP. The first dataset was obtained from Neurodevelopmental Genomics: Trajectories of Complex Phenotypes (dbGaP dataset accession number phs000607.v3.p2[46–48]). We considered 525 adolescents between the ages of 12 to 15 years who self-reported as being of European descent. Using their height and weight measurements, we calculated BMI and then categorized the BMI based on the Centers for Disease Control and Prevention BMI-for-age (and gender) recommendations. The second dataset was obtained from the eMERGE Network Imputed for 41 Phenotypes (dbGaP dataset accession number phs000888.v1.p1, variable number phv00225989.v1.p1). We considered 3,486 adults with either a case or control diagnosis of extreme obesity. For both datasets the FDA GRS was calculated using the score function in PLINK v1.9[62,63]. Proxies for SNPs were determined using LDLink[72] and are summarized in Table S6.

Analysis of environmental and behavioral covariates

Using the Bayesian Information Criterion option of the `bestglm` function in R[73], we applied best subset selection to the regression of conditional weight gain scores[34] on 11 potentially confounding covariates (described in the Results section). We did not consider interaction terms for this analysis, but we included (separately) Belsky PRS or FDA PRS as a 12th predictor in the regression. Once the best subset of predictors was selected, we fitted a linear model on it using the lm function in the R stats package.

We also applied LASSO and Group LASSO procedures to the same regressions for conditional weight gain, using the `glmnet`[74] and `gglasso`[75] packages respectively. We considered again the 11 potentially confounding covariates along with a PRS, and all possible two-way interactions. The LASSO methods were tuned via 10 fold cross-validation, fixing the penalty parameter to be within 1 standard error of the overall minimum cross-validation error. This is considered the more parsimonious approach, usually favoring a sparser final model[71].

Data Availability

Phenotypic and Genetic data are/will be available under dbGaP study number: phs001498.v2.p1. Code for carrying out the statistical methods (screening, applying FLAME, PRS construction and evaluation) can be found at https://github.com/makovalab-psu/InsightPRSConstruction.

Ethics Statement

This project has been approved by Penn State IRB (PRAMS 34493).

14

Acknowledgements

Funding

Author contributions

SJCC, KDM, IMP, LLB, FC, and MR conceived the project and devised the project study design. AK, SJCC, JL, MR, FC, and KDM were involved in the data analysis. SJCC, AK, KDM, FC, and MR contributed to the writing of the manuscript with comments from co-authors. MP, LLB, JS, and MM provided resources such as access to the study population and the associated data.

Competing Interests

The authors declare no competing interests.

References

1. Hales, C. M., Fryar, C. D., Carroll, M. D., Freedman, D. S. & Ogden, C. L. Trends in Obesity and Severe Obesity Prevalence in US Youth and Adults by Sex and Age, 2007-2008 to 2015-2016. *JAMA* **319**, 1723–1725 (2018).

2. Ogden, C. L. *et al.* Trends in Obesity Prevalence Among Children and Adolescents in the United States, 1988-1994 Through 2013-2014. *JAMA* **315**, 2292–2299 (2016).

3. Cunningham, S. A., Kramer, M. R. & Narayan, K. M. V. Incidence of childhood obesity in the United States. *N. Engl. J. Med.* **370**, 1660–1661 (2014).

4. Ang, Y. N., Wee, B. S., Poh, B. K. & Ismail, M. N. Multifactorial Influences of Childhood Obesity. *Curr. Obes. Rep.* **2**, 10–22 (2012).

5. Sahoo, K. *et al.* Childhood obesity: causes and consequences. *J Family Med Prim Care* **4**, 187–192 (2015).

6. Barriuso, L. *et al.* Socioeconomic position and childhood-adolescent weight status in rich countries: a systematic review, 1990-2013. *BMC Pediatr.* **15**, 129 (2015).

7. Boney, C. M., Verma, A., Tucker, R. & Vohr, B. R. Metabolic syndrome in childhood: association with birth weight, maternal obesity, and gestational diabetes mellitus. *Pediatrics* **115**, e290–6 (2005).

8. Kries, R. von & von Kries, R. Maternal Smoking during Pregnancy and Childhood Obesity. *American Journal of Epidemiology* **156**, 954–961 (2002).

9. Carnell, S. & Wardle, J. Measuring behavioural susceptibility to obesity: Validation of the child eating behaviour questionnaire. *Appetite* **48**, 104–113 (2007).

10. Wardle, J. *et al.* Obesity Associated Genetic Variation inFTOIs Associated with Diminished Satiety. *The Journal of Clinical Endocrinology & Metabolism* **93**, 3640–3643 (2008).

11. Maes, H. H., Neale, M. C. & Eaves, L. J. Genetic and environmental factors in relative body weight and human adiposity. *Behav. Genet.* **27**, 325–351 (1997).

12. Pigeyre, M., Yazdi, F. T., Kaur, Y. & Meyre, D. Recent progress in genetics, epigenetics and metagenomics unveils the pathophysiology of human obesity. *Clin. Sci.* **130**, 943–986 (2016).

13. Llewellyn, C. H., Trzaskowski, M., Plomin, R. & Wardle, J. From modeling to measurement: developmental trends in genetic influence on adiposity in childhood. *Obesity* **22**, 1756–1761 (2014).

14. Saeed, S. *et al.* Loss-of-function mutations in ADCY3 cause monogenic severe obesity. *Yearbook of Paediatric Endocrinology* (2018). doi:10.1530/ey.15.11.5

15. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

16. Meyre, D. *et al.* Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat. Genet.* **41**, 157–159 (2009).

17. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).

18. Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **41**, 18–24 (2009).

19. the Early Growth Genetics (EGG) Consortium. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat. Genet.* **44**, 526–531 (2012).

20. Warrington, N. M. *et al.* A genome-wide association study of body mass index across early life and childhood. *Int. J. Epidemiol.* **44**, 700–712 (2015).

21. Felix, J. F. *et al.* Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Hum. Mol. Genet.* **25**, 389–403 (2016).

22. Goodarzi, M. O. Genetics of obesity: what genetic association studies have taught us about the biology of obesity and its complications. *Lancet Diabetes Endocrinol* **6**, 223–236 (2018).

23. Sugrue, L. P. & Desikan, R. S. What Are Polygenic Scores and Why Are They Important? *JAMA* (2019). doi:10.1001/jama.2019.3893

24. den Hoed, M. *et al.* Genetic susceptibility to obesity and related traits in childhood and adolescence: influence of loci identified by genome-wide association studies. *Diabetes* **59**, 2980–2988 (2010).

25. Elks, C. E. *et al.* Genetic markers of adult obesity risk are associated with greater early infancy weight gain and growth. *PLoS Med.* **7**, e1000284 (2010).

26. Belsky, D. W. *et al.* Polygenic risk, rapid childhood growth, and the development of obesity: evidence from a 4-decade longitudinal study. *Arch. Pediatr. Adolesc. Med.* **166**, 515–521 (2012).

27. Li, A. *et al.* Parental and child genetic contributions to obesity traits in early life based on 83 loci validated in adults: the FAMILY study. *Pediatr. Obes.* **13**, 133–140 (2018).

28. Andersson, E. A. *et al.* Do gene variants influencing adult adiposity affect birth weight? A population-based study of 24 loci in 4,744 Danish individuals. *PLoS One* **5**, e14190 (2010).

29. Sovio, U. *et al.* 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 (2011).

30. Graff, M. *et al.* Genome-wide analysis of BMI in adolescents and young adults reveals additional

    insight into the effects of genetic loci over the life course. *Human Molecular Genetics* **22**,

    3597–3607 (2013).

31. Llewellyn, C. H. & Fildes, A. Behavioural Susceptibility Theory: Professor Jane Wardle and the

    Role of Appetite in Genetic Risk of Obesity. *Curr. Obes. Rep.* **6**, 38–45 (2017).

32. Frayling, T. M. *et al.* A Common Variant in the FTO Gene Is Associated with Body Mass Index and

    Predisposes to Childhood and Adult Obesity. *Science* **316**, 889–894 (2007).

33. Paul, I. M. *et al.* The Intervention Nurses Start Infants Growing on Healthy Trajectories (INSIGHT)

    study. *BMC Pediatr.* **14**, 184 (2014).

34. Savage, J. S., Birch, L. L., Marini, M., Anzman-Frasca, S. & Paul, I. M. Effect of the INSIGHT

    Responsive Parenting Intervention on Rapid Infant Weight Gain and Overweight Status at Age 1

    Year: A Randomized Clinical Trial. *JAMA Pediatr.* **170**, 742–749 (2016).

35. Paul, I. M. *et al.* Effect of a Responsive Parenting Educational Intervention on Childhood Weight

    Outcomes at 3 Years of Age: The INSIGHT Randomized Clinical Trial. *JAMA* **320**, 461–468 (2018).

36. Griffiths, L. J., Smeeth, L., Hawkins, S. S., Cole, T. J. & Dezateux, C. Effects of infant feeding

    practice on weight gain from birth to 3 years. *Arch. Dis. Child.* **94**, 577–582 (2009).

37. Zhou, J. *et al.* Rapid Infancy Weight Gain and 7- to 9-year Childhood Obesity Risk: A Prospective

    Cohort Study in Rural Western China. *Medicine* **95**, e3425 (2016).

38. Daniels, S. R., Hassink, S. G. & COMMITTEE ON NUTRITION. The Role of the Pediatrician in

    Primary Prevention of Obesity. *Pediatrics* **136**, e275–92 (2015).

39. Taveras, E. M. *et al.* Weight Status in the First 6 Months of Life and Obesity at 3 Years of Age.

    PEDIATRICS **123**, 1177–1183 (2009).

40. Chu, W., Li, R. & Reimherr, M. FEATURE SCREENING FOR TIME-VARYING COEFFICIENT MODELS WITH ULTRAHIGH DIMENSIONAL LONGITUDINAL DATA. *Ann. Appl. Stat.* **10**, 596–617 (2016).

41. Parodi, A. & Reimherr, M. *FLAME: Simultaneous variable selection and smoothing for functionon-scalar regression*. (Technical report, Pennsylvania State University, 2017).

42. Fall, T. & Ingelsson, E. Genome-wide association studies of obesity and metabolic syndrome. *Mol. Cell. Endocrinol.* **382**, 740–757 (2014).

43. McQueen, M. B. *et al.* The National Longitudinal Study of Adolescent to Adult Health (Add Health) sibling pairs genome-wide data. *Behav. Genet.* **45**, 12–23 (2015).

44. Wheeler, E. *et al.* Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. *Nat. Genet.* **45**, 513–517 (2013).

45. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Series B Stat. Methodol.* **58**, 267–288 (1996).

46. Glessner, J. T. *et al.* Strong synaptic transmission impact by copy number variations in schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10584–10589 (2010).

47. Calkins, M. E. *et al.* The psychosis spectrum in a young U.S. community sample: findings from the Philadelphia Neurodevelopmental Cohort. *World Psychiatry* **13**, 296–305 (2014).

48. Calkins, M. E. *et al.* The Philadelphia Neurodevelopmental Cohort: constructing a deep phenotyping collaborative. *J. Child Psychol. Psychiatry* **56**, 1356–1369 (2015).

49. Li, W. *et al.* Association between Obesity and Puberty Timing: A Systematic Review and Meta-Analysis. *Int. J. Environ. Res. Public Health* **14**, (2017).

50. Martinez-Herrera, M., Silvestre-Rangil, J. & Silvestre, F.-J. Association between obesity and periodontal disease. A systematic review of epidemiological studies and controlled clinical trials. *Med. Oral Patol. Oral Cir. Bucal* **22**, e708–e715 (2017).

51. Reimherr, M. & Nicolae, D. A functional data analysis approach for genetic association studies. *The Annals of Applied Statistics* **8**, 406–429 (2014).

52. Llewellyn, C. H., Trzaskowski, M., van Jaarsveld, C. H. M., Plomin, R. & Wardle, J. Satiety mechanisms in genetic risk of obesity. *JAMA Pediatr.* **168**, 338–344 (2014).

53. Johnson, L., Llewellyn, C. H., van Jaarsveld, C. H. M., Cole, T. J. & Wardle, J. Genetic and environmental influences on infant growth: prospective analysis of the Gemini twin birth cohort. *PLoS One* **6**, e19918 (2011).

54. McLoone, P. & Morrison, D. S. Risk of child obesity from parental obesity: analysis of repeat national cross-sectional surveys. *Eur. J. Public Health* **24**, 186–190 (2014).

55. Bahreynian, M. *et al.* Association between Obesity and Parental Weight Status in Children and Adolescents. *J. Clin. Res. Pediatr. Endocrinol.* **9**, 111–117 (2017).

56. Zalbahar, N., Najman, J., McIntyre, H. D. & Mamun, A. Parental pre-pregnancy obesity and the risk of offspring weight and body mass index change from childhood to adulthood. *Clin. Obes.* **7**, 206–215 (2017).

57. Craig, S. J. C. *et al.* Child Weight Gain Trajectories Linked To Oral Microbiota Composition. *Sci. Rep.* **8**, 14030 (2018).

58. Llewellyn, C. H., van Jaarsveld, C. H. M., Plomin, R., Fisher, A. & Wardle, J. Inherited behavioral susceptibility to adiposity in infancy: a multivariate genetic analysis of appetite and weight in the Gemini birth cohort. *Am. J. Clin. Nutr.* **95**, 633–639 (2012).

59. Yao, F., Müller, H.-G. & Wang, J.-L. Functional Data Analysis for Sparse Longitudinal Data. *J. Am. Stat. Assoc.* **100**, 577–590 (2005).

60. Baird, J. *et al.* Being big or growing fast: systematic review of size and growth in infancy and later obesity. *BMJ* **331**, 929 (2005).

61. Ong, K. K. & Loos, R. J. F. Rapid infancy weight gain and subsequent obesity: systematic reviews

and hopeful suggestions. *Acta Paediatr.* **95**, 904–908 (2006).

62. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

63. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

64. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011).

65. O'Connell, J. *et al.* A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLoS Genet.* **10**, e1004234 (2014).

66. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

67. Howie, B. N., Donnelly, P. & Marchini, J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet.* **5**, e1000529 (2009).

68. Fan, J. & Lv, J. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**, 849–911 (2008).

69. Kokoszka, P. & Reimherr, M. *Introduction to Functional Data Analysis*. (CRC Press, 2017).

70. Choi, H. & Reimherr, M. A geometric approach to confidence regions and bands for functional parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**, 239–260 (2018).

71. Hastie, T., Tibshirani, R. & Friedman, J. The elements of statistical learning: data mining, inference, and prediction, Springer Series in Statistics. (2009).

72. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555–3557 (2015).

73. McLeod, A. I. & Xu, C. *bestglm: Best Subset GLM*. (2014).

74. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1 (2010).

75. Yang, Y. & Zou, H. A fast unified algorithm for solving group-lasso penalize learning problems. *Stat. Comput.* **25**, 1129–1141 (2015).

Table 1. SNPs identified as significant predictors of children's weight gain patterns by functional data analysis

SNP	Associated gene[1]	Chr	Position (bp)	Effect Allele	Weight	GWAS association[2]	Folds[3]
rs72679478	DNAJC6	1	65826847	C	0.210	Age at menarche	12
rs12039940	ZNF648	1	181952711	C	0.231	HDL cholesterol, cardiovascular heart disease in diabetics	17
rs10494802	NR5A2	1	199663441	A	0.164	Chronic inflammatory disease, age at onset of menarche	9
rs4915535	NAV1	1	201500537	T	-0.215	BMI, waist circumference, obesity related traits	12
rs638348	RHOU	1	229100364	T	-0.260	Type 2 Diabetes, dental caries	20
rs113822101	NOL10	2	10763339	-	0.234		19
rs9837708	FOXP1	3	71487582	T	-0.222	LDL cholesterol, subcutaneous adipose tissue	14
rs921551	TNIP3	4	122110975	C	-0.141		11
rs17057519	FABP6	5	159637815	G	0.129		8
rs16889349	RNF144B	6	36948528	G	0.227	Chronic periodontitis, age at voice drop	18
rs17626544	MTCH1	6	18404450	T	0.144	HDL cholesterol	10
rs4716760	chr7_54678150	7	157376644	T	0.225	VSTM2A (close gene): obesity related traits, waist-to-hip ratio adjusted for BMI	19
rs1701822	POR	7	112947322	G	0.128		11
rs62475261	chr7_112947322	7	75533388	C	0.225		20
rs10227226	PTPRN2	7	54678150	T	-0.209	Obesity-related traits, BMI	14
rs58307428	MIR147A	9	138038493	C	0.197		13
rs9409226	OLFM1	9	122850776	A	0.247	Obesity related traits (estradiol measurement), pediatric bone mineral density	19
rs2389157	LINC00557	13	95433968	A	0.170		11
rs17648524	RBFOX1	16	7459683	C	0.173		10
rs72815409	DNAH9	17	11751186	A	-0.277		20
rs4969367	BAIAP2	17	79032015	A	0.214		18
rs141177192	TGIF1	18	3451388	AA	-0.201	Periodontitis	13
rs1539759	TIAM1	21	32895851	C	0.212	Hypertension	17
rs133709	ELFN2	22	37827248	A	-0.144		7

[1]As determined by Affymetrix

[2]Association determined by the NHGRI-EBI GWAS catalog of published genome-wide association studies (https://www.ebi.ac.uk/gwas/)

[3]The number of times the SNP is selected in 20-fold cross validation

Table 2. Description of the study participants

	Rapid infant weight gain (n=104)	Non-rapid infant weight gain (n=122)
Children (n=226)		
Gender: # males/ # females	53/51	67/55
Birth weight (g): mean(S.D.)	3486 (449)	3420 (412)
Delivery Mode: Vaginal (%)	36 (35)	33 (27)
Weight-for-Length (kg/m) at 1 year: mean (S.D.)	13.67 (0.98)	12.23 (0.93)
Weight-for-Height (kg/m) at 2 years: mean (S.D.)	15.13 (1.18)[1]	13.79 (1.21)
Weight-for-Height (kg/m) at 3 years: mean (S.D.)	16.06 (1.29)[2]	14.70 (1.14)[3]
Appetite score: mean (S.D.)	4.011 (0.92)[4]	3.38 (1.00)[5]
Mothers (n=226)		
Pre-pregnancy BMI: mean (S.D.)	25.90 (5.75)	25.14 (4.97)
Gestational weight gain status: did not gain enough/met guidelines/gained excess	16/29/59	25/43/54
Gestational diabetes: had gestational diabetes (%)	6 (5.8)	7 (5.7)
Smoked during pregnancy: smoked (%)	8 (7.7)	5 (4.1)
Fathers (n=209)		
BMI: mean (S.D.)	29.53 (6.28)[6]	27.72 (4.79)[7]
Ethnicity of the children (as reported by their mothers)		
Black	6 (5.8)	5 (4.1)
White	94 (90.4)	109 (89.3)
Native Hawaiian or Pacific Islander	1 (0.4)	0 (0)
Asian	2 (0.9)	5 (4.1)
Other	1 (0.4)	3 (1.3)
Annual Household Income		
< $10,000	2	4
$10,000 - $24,999	6	9
$25,000 - $49,999	13	8
$50,000 - $74,999	30	34
$75,000 - $99,999	25	23
$100,000 or more	24	38

1. Missing 3 measurements
2. Missing 9 measurements
3. Missing 7 measurements
4. Missing 16 measurements
5. Missing 20 measurements
6. Missing 9 measurements
7. Missing 8 measurements