

A Comprehensive Genome-Wide and Phenome-Wide Examination of BMI and Obesity in a Northern Nevadan Cohort

Karen A. Schlauch,^{*,†} Robert W. Read,^{*,†} Vincent C. Lombardi,[‡] Gai Elhanan,^{*,†} William J. Metcalf,^{*,†} Anthony D. Slonim,^{*,§} the 23andMe Research Team,^{**} and Joseph J. Grzymalski^{*,†,1}

^{*}Institute for Health Innovation, Renown Health, Reno, NV [†]Desert Research Institute, Reno, NV, [‡]Department of Microbiology and Immunology, University of Nevada, School of Medicine, Reno, NV, [§]Renown Health, Reno, NV, and ^{**}23andMe, Inc. Mountain View, CA

ORCID IDs: 0000-0001-6916-8571 (K.A.S.); 0000-0002-6967-5634 (R.W.R.); 0000-0003-4474-8662 (V.C.L.); 0000-0003-2646-8958 (J.J.G.)

ABSTRACT The aggregation of Electronic Health Records (EHR) and personalized genetics leads to powerful discoveries relevant to population health. Here we perform genome-wide association studies (GWAS) and accompanying phenome-wide association studies (PheWAS) to validate phenotype-genotype associations of BMI, and to a greater extent, severe Class 2 obesity, using comprehensive diagnostic and clinical data from the EHR database of our cohort. Three GWASs of 500,000 variants on the Illumina platform of 6,645 Healthy Nevada participants identified several published and novel variants that affect BMI and obesity. Each GWAS was followed with two independent PheWASs to examine associations between extensive phenotypes (incidence of diagnoses, condition, or disease), significant SNPs, BMI, and incidence of extreme obesity. The first GWAS examines associations with BMI in a cohort with no type 2 diabetics, focusing exclusively on BMI. The second GWAS examines associations with BMI in a cohort that includes type 2 diabetics. In the second GWAS, type 2 diabetes is a comorbidity, and thus becomes a covariate in the statistical model. The intersection of significant variants of these two studies is surprising. The third GWAS is a case vs. control study, with cases defined as extremely obese (Class 2 or 3 obesity), and controls defined as participants with BMI between 18.5 and 25. This last GWAS identifies strong associations with extreme obesity, including established variants in the *FTO* and *NEGR1* genes, as well as loci not yet linked to obesity. The PheWASs validate published associations between BMI and extreme obesity and incidence of specific diagnoses and conditions, yet also highlight novel links. This study emphasizes the importance of our extensive longitudinal EHR database to validate known associations and identify putative novel links with BMI and obesity.

KEYWORDS

GWAS
PheWAS
BMI
Obesity

The rate of obesity is growing at an alarming rate worldwide – fast enough to call it an epidemic. As obesity is a risk factor for developing typically related diseases such as type 2 diabetes mellitus (DM2),

cardiovascular disease and some cancers (Wang *et al.* 2011), the situation is becoming a public health concern. The percentage of obesity is rising nationwide, with current adult obesity rates at close to 40%, up from 32% in 2004 (Ogden *et al.* 2006; Warren *et al.* 2018). In Nevada, the current adult obesity rate (BMI \geq 30) is 27%, an increase from 21% in 2005 (Warren *et al.* 2018). Additionally, since 2016, Nevada witnessed a significant increase in the percentage of adults who are overweight (the current rate is 66%) (Warren *et al.* 2018). Studies identified several genetic factors that influence the development of obesity with estimates on the heritability of the disease (40%-75%) (Stunkard *et al.* 1986; 1990; Maes *et al.* 1997; Herrera and Lindgren 2010) and 65-80% (Malis *et al.* 2005).

High body mass index (BMI) and DM2 are known from many sources to be strongly related both epidemiologically and genetically

Copyright © 2020 Schlauch *et al.*

doi: <https://doi.org/10.1534/g3.119.400910>

Manuscript received August 2, 2019; accepted for publication December 10, 2019; published Early Online December 26, 2019.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at figshare: <https://doi.org/10.25387/g3.8266430>.

¹Corresponding author: 2215 Raggio Pkwy, Reno, NV 89512. E-mail: joeg@dri.edu

(Kopelman 2007; Bays *et al.* 2007; Grarup *et al.* 2014; Cronin *et al.* 2014); however, these two conditions share very few known causative variants (Grarup *et al.* 2014; Karaderi *et al.* 2015). This study presents first a GWAS of BMI in a cohort without DM2, followed by a GWAS of a cohort with DM2, to identify the differences in the genetic mechanisms of obesity (BMI ≥ 30) without DM2 and with DM2, and show that DM2 is indeed, an important predictor of high BMI when included as comorbidity. Although a number of large meta-analyses of multiple genome-wide association studies (GWASs) have detected possible causative single nucleotide polymorphisms (SNPs) of obesity and increased BMI (Scuteri *et al.* 2007; Frayling *et al.* 2007; Dina *et al.* 2007; Zeggini *et al.* 2007; Yanagiya *et al.* 2007; Hinney *et al.* 2007; Hunt *et al.* 2008; Price *et al.* 2008; Grant *et al.* 2008; Hotta *et al.* 2008; Loos *et al.* 2008; Tan *et al.* 2008; Villalobos-Comparán *et al.* 2008; Thorleifsson *et al.* 2008; Willer *et al.* 2009; Meyre *et al.* 2009; Wing *et al.* 2009; Liu *et al.* 2010; Shimaoka *et al.* 2010; Fawcett and Barroso 2010; Speliotes *et al.* 2010; Wang *et al.* 2011; Prakash *et al.* 2011; Okada *et al.* 2012; Cha *et al.* 2012; Berndt *et al.* 2013; Wheeler *et al.* 2013; Graff *et al.* 2013; Olza *et al.* 2013; Boender *et al.* 2014; Qureshi *et al.* 2017; Huđek *et al.* 2018; González-Herrera *et al.* 2019), none, to the best of our knowledge, have included comprehensive GWASs on the quantitative BMI metric and on extreme obesity case-control simultaneously, as well as investigated phenotypic associations with BMI, obesity, and significant loci identified by the GWAS.

Our study begins with the Healthy Nevada Project (HNP), a project centered around a Northern Nevada cohort formed in 2016 and 2017 by Renown Health and the Desert Research Institute in Reno, NV to investigate factors that may contribute to health outcomes in Northern Nevada. Its first phase provided 10,000 individuals in Northern Nevada with genotyping using the 23andMe platform at no cost. Renown Health is the only tertiary care health system in the area, and 75% of these 10,000 individuals are cross-referenced in its extensive electronic health records (EHR) database. The Renown EHR database contains 86,610 BMI measurements for these 10,000 individuals over twelve years, along with comprehensive disease diagnoses, (*e.g.* diabetes or eating disorders) and other general conditions such as pregnancy, allowing for precise individual phenotypic classifications and thereby leading to more robust and meaningful phenotype-genotype associations.

The focus of the comprehensive GWAS-PheWAS examinations of the Healthy Nevada Project (HNP) cohort and its EHR database is two-fold: the first is to establish infrastructure to perform large-scale genome-wide and phenome-wide association investigations in alliance with complex electronic health care records; the second is to validate well-known published variants and associations with BMI and obesity in this cohort, as well as to identify possibly novel genotypic and phenotypic associations with BMI and extreme obesity.

The three GWASs identified several of the "usual suspects" for both BMI and obesity, such as *FTO* and *NEGR1*, that were shown to have a role in weight regulation (Scuteri *et al.* 2007; Frayling *et al.* 2007; Dina *et al.* 2007; Zeggini *et al.* 2007; Hinney *et al.* 2007; Hunt *et al.* 2008; Price *et al.* 2008; Grant *et al.* 2008; Hotta *et al.* 2008; Loos *et al.* 2008; Tan *et al.* 2008; Villalobos-Comparán *et al.* 2008; Thorleifsson *et al.* 2008; Willer *et al.* 2009; Meyre *et al.* 2009; Wing *et al.* 2009; Shimaoka *et al.* 2010; Fawcett and Barroso 2010; Speliotes *et al.* 2010; Herrera and Lindgren 2010; Wang *et al.* 2011; Prakash *et al.* 2011; Okada *et al.* 2012; Berndt *et al.* 2013; Wheeler *et al.* 2013; Graff *et al.* 2013; Olza *et al.* 2013; Boender *et al.* 2014; Qureshi *et al.* 2017; González-Herrera *et al.* 2019). However, this study also identified a number of novel BMI and obesity associations to genes which are differentially expressed in obese patients (Jiao *et al.* 2008; Pietiläinen *et al.* 2008; Nakajima *et al.* 2016).

Additionally, using linked EHR, the PheWASs examined the pleiotropy of HNP BMI and obesity associated SNPs: whether these variants are linked with other endocrine or metabolic diagnoses or conditions of this nature. A second PheWAS identified many known phenotypes related to BMI and obesity, especially to DM2, abnormal glucose levels, hypertension, hyperlipidemia, sleep apnea, asthma and other less-studied BMI-related diagnoses.

MATERIALS AND METHODS

The Renown EHR database

The Renown Health EHR system was instated in 2007 on the EPIC system (EPIC System Corporation, Verona, Wisconsin, USA), and currently contains lab results, diagnosis codes (ICD9 and ICD10) and demographics of more than one million patients seen in the hospital system since 2005.

Sample collection

Saliva as a source of DNA was collected from 10,000 adults in Northern Nevada as the first phase of the Healthy Nevada Project to contribute to comprehensive population health studies in Nevada. Via a donation from the Renown Health Foundation, the study offered the DNA testing for free to any and all participants willing to sign up for the study. The personal genetics company 23andMe, Inc. was used to genotype these individuals using the Oragene DX OGD-500.001 saliva kit [DNA Genotek, Ontario, Canada]. Genotypes are based on the Illumina Human OmniExpress-24 BeadChip platform [San Diego, CA, USA], that include approximately 570,000 SNPs.

IRB and ethics statement

This study was reviewed and approved by the University of Nevada, Reno Institutional Review Board (IRB, project 956068-12). Participants in the Healthy Nevada Project undergo written and informed consent to having genetic information associated with electronic health information in a deidentified manner. All participants were 18 years of age or older. Neither researchers nor participants have access to the complete EHR data and cannot map participants to patient identifiers. Patient identifiers are not incorporated into the EHR; rather, EHR and genetic data are linked in a separate environment via a unique identifier as approved by the IRB.

Processing of EHR data

Most participants had multiple BMI recordings across the thirteen years of EHR; the mean number of BMI records across the individuals was 12.2 records, with 215 the maximum number of records for the cohort. For the 5,811 individuals with more than one recorded BMI measure, a simple quality control step was first performed before computing the average BMI value. More specifically, if a participant had multiple BMI records, the coefficient of variation (CV) of the BMI values was computed; if the CV was greater than 0.096, any outlying BMI values greater than 1.5 standard deviations from the mean BMI for that individual was excluded for that individual. As 10% of the multiple BMI records presented a CV of 0.096 or greater, this threshold was chosen. Note that the relatively small percentage of participants presenting a CV of approximately 10% or greater indicates that variation across multiple records is unremarkable in most individuals. Indeed, the additional quality control step excluded one or more outlying BMI records in only 106 individuals; these 701 BMI records included values such as "2823.42" and values less than 10. Examples of outliers include 158.38 in an individual's set of values with mean 25.3 and "2874" in an individual with mean BMI measure of 22.4. Additionally, this quality

control step allowed the study to include pregnant women: of the 464 pregnant women with BMI recorded for pregnant and non-pregnant phases, outlying pregnancy-related BMI records were easily identified and removed. The raw BMI values and quality-controlled average BMI values are presented in Supplementary Figure S1.

Genotyping and quality control

Genotyping was performed by 23andMe using the Illumina Infinum DNA Human OmniExpress-24 BeadChip V4 (Illumina, San Diego, CA). This genotyping platform consists of approximately 570,000 SNPs. DNA extraction and genotyping were performed on saliva samples by the National Genetics Institute (NG1), a CLIA licensed clinical laboratory and a subsidiary of the Laboratory Corporation of America.

Raw genotype data were processed through a standard quality control process (Anderson *et al.* 2010; Verma *et al.* 2016; Schlauch *et al.* 2016; Verma *et al.* 2018; Schlauch *et al.* 2018). SNPs with a minor allele frequency (MAF) less than 0.005 were removed. SNPs that were out of Hardy Weinberg equilibrium (p -value $< 1 \times 10^{-6}$) were also excluded. Any SNP with a call rate less than 95% was removed; any individual with a call rate less than 95% was also excluded from further study. There was an observable bias within the African American sub-cohort, thus 89 African American participants were excluded from this study. Specifically, upon plotting the first two principal components there was a clear population stratification between the African Americans and all other individuals who showed no other distinct groupings. The first principal component explained 14% of the total population stratification. Additionally, 107 patients with type I diabetes were removed, as were 29 participants with eating disorders recorded into their health record. After quality control, this left 500,508 high-quality SNPs and 6,645 participants in the BMI cohort with mean autosomal heterozygosity of 0.318. The same process yielded 5,994 participants when all individuals with DM2 diagnoses were removed. Within the extreme obesity study, participants with BMI values between 18.5 and 25 were considered as controls, while any participant with BMI at least 35 kg/m² was considered a case subject. Again, any individual with type I diabetes and recorded eating disorders was removed. This resulted in a cohort size of 2,994 participants with 984 extreme obese cases and 2,012 lean controls with a mean autosomal heterozygosity of 0.316.

A standard principal component analysis (PCA) was performed on the genotype data to identify principal components to correct for population substructure. Genotype data were pruned to exclude SNPs with high linkage disequilibrium using *PLINK* v1.9 (Purcell *et al.* 2007) and standard pruning parameters of 50 SNPs per sliding window; window size of five SNPs; $r^2=0.5$ (Anderson *et al.* 2010). The remaining 270,160 SNPs were used to calculate the principal components. Regression models were adjusted by the first four principal components, decreasing the genomic inflation factor of all obesity and BMI traits to $\lambda \leq 1.06$.

Genome-wide association studies

Using *PLINK*, we first performed a simple linear regression of BMI *vs.* genotype using the additive model (number of copies of the minor allele) including age, gender and the first four principal components as covariates to correct for any bias generated by the population substructure. In the first BMI study, participants with DM2 were excluded. The second BMI study included DM2-diagnosed participants and included DM2 as a covariate in the statistical model. To test associations between obesity and genotype, a standard case-control logistic regression was applied, adjusting for age, gender and the first four principal components. Total phenotypic variance explained by all 500,508 SNPs

was calculated by first producing a genetic relationship matrix of all SNPs on autosomal chromosomes in *PLINK*. Subsequently, a restricted maximum likelihood analysis was conducted using GTCA (Yang *et al.* 2011) on the relationship matrix to estimate the variance explained by the SNPs.

Analysis of variance

The mean BMI values across genotypes presented in Supplementary Tables S1 and S2 correlate with negative and positive effect sizes: SNPs showing a negative effect size have a decrease in mean BMI values across the genotypes from left to right (homozygous in major allele, heterozygous, homozygous in minor allele). The 6,645 log-transformed quality-controlled and averaged BMI measures were nearly normally distributed. As one-way ANOVA computations are robust against even moderate deviations of normality (Blanca *et al.* 2017), parametric ANOVA methods were used to make comparisons across the genotypes. All ANOVA F-test p -values of the significant SNPs identified in the two BMI studies are statistically significant at the $\alpha=0.05$ level, even after a simple Bonferroni correction ($.05/27=0.0019$, and $.05/20=0.0025$, respectively). Supplementary Table S3 presents the proportion of obese cases across each genotype. A simple test of equal proportions (Pearson's chi-square test) is performed across these proportions. All p -values associated with the test of equal proportions in Supplementary Table S3 are also statistically significant at the 0.05 significance level upon a conservative Bonferroni multiple testing adjustment ($.05/34=.0015$).

Power of GWAS

The software program QUANTO (Gauderman 2002) was used to calculate sample sizes to detect effect sizes in the range [0.5,1] and odds ratios in [1,1.5] with at least 80% power under the additive model, at a two-sided Type I error level of 5%. Using the rate of extreme obesity (BMI ≥ 35 kg/m²) as 14.5% from Ogden *et al.* (Ogden *et al.* 2006), the case-control GWAS study of approximately 1,000 cases and 2,000 controls has sufficient statistical power ($\geq 80\%$) with MAFs of 16% or greater to detect odds ratios of 1.225 or greater. As the MAF increases, the power to detect smaller odds ratios increases: for example, with a MAF of 25%, our sample size was adequate to detect odds ratios of size 1.18 or higher. With a small MAF of 8%, the power was also at least 80% to detect effect sizes as small as 0.58 in the BMI GWAS cohort of 6,645. With MAF of 17%, power was at least 80% to detect effect sizes as small as 0.425. Larger MAFs clearly can detect larger effect sizes with the same sample size. Specific effect sizes and MAFs can be seen in Table 3.

Phenome-wide association study

The R package PheWAS (Carroll *et al.* 2014) was used to perform two independent PheWAS analyses for each of our studies. The first examined associations between statistically significant SNPs identified in the respective GWASs and EHR phenotypes based on ICD codes. The second PheWAS identified associations between BMI levels or incidence of obesity, respectively, and ICD-based diagnoses. ICD9 and ICD10 codes for each individual in the cohort recorded in the Renown EHR were aggregated via a mapping from the Center for Medicare and Medicaid services (<https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-CM-and-GEMs.html>). A total of 22,693 individual diagnoses mapped to 4,769 documented ICD9 codes. ICD9 codes were aggregated and converted into 1,814 individual phenotype groups ("phecodes") using the PheWAS package as described in Carroll and Denny (Denny *et al.* 2013; Carroll *et al.* 2014). Of these, only the phecodes that included at least 20 cases were used for downstream

analyses, following Carroll's protocol (Carroll *et al.* 2014). Age and gender were standard covariates included in the PheWAS models. The first type of PheWAS detected associations between statistically significant SNPs ($p < 1 \times 10^{-5}$) identified in each of the three GWASs above and case/control status of EHR phenotypes represented by ICD codes. Specifically, a logistic regression between the incidence (number of cases) of each phenotype group (phecode) and the additive genotypes of each statistically significant SNP was performed, including age and gender as covariates. Possible associations of the phecodes with at least 20 individuals with each previously detected SNP were assessed. Two levels of significance were computed: the first, on which the reported results are based, was generated by first calculating the adjusted p -values for the multiple hypothesis tests using the Benjamini-Hochberg false discovery rate (FDR) (Benjamini and Hochberg 1995) and selecting the raw p -value corresponding to the FDR = 0.1 significance level, following Denny's protocol (Denny *et al.* 2013). This level is represented by a red line in PheWAS images. For visual purposes only, a threshold based on a Bonferroni correction for all possible associations made in this analysis ($p = 0.05 / N_{ps}$, where N_{ps} is the sum of the number of phecodes tested for each individual SNP, across all identified SNPs), is represented by a blue line in PheWAS images.

A second PheWAS, as outlined in Carroll *et al.* (2014) (Carroll *et al.* 2014), was performed to examine associations between BMI, and secondarily, obesity, and the phecodes. Specifically, a linear regression between the BMI measures and the case/control status of a phecode was performed (with age and gender as covariates) for each phecode including at least 20 individuals. Significance levels corresponded to the FDR value of 0.1 and are not shown in either figure due to space constraints. The Bonferroni corrections for the BMI study and obesity study were 3.3×10^{-5} and 3.7×10^{-5} , respectively. In Figures 4 and 6, only phenotypes above the red line representing $p = 1 \times 10^{-15}$ are annotated for ease of viewing.

DATA AVAILABILITY

EHR data

EHR data for the Healthy Nevada cohort are subject to HIPAA and other privacy and compliance restrictions. Mean quality-controlled BMI values for the 6,645 individuals are available in Supplementary Table S6. Supplemental material available at figshare: <https://doi.org/10.25387/fig3.8266430>.

Genotype data

The data that support the findings of this study are available from 23andMe but restrictions apply to the availability of these data, which were used under license for the current study, and thus are not publicly available. The data are, however, available for qualified researchers upon reasonable request and with permission of the Institute for Health Innovation and 23andMe. Researchers who would like to obtain the raw genotype data related to this study will be presented with a data user agreement, which requires that the participants will not be re-identified and no data will be shared between researchers or uploaded onto public domains. Due to the public nature of this article, and genetic privacy requirements, the Institute for Health Innovation and 23andMe require that the statistics for only 10,000 SNPs be made publicly available. This is the amount of data considered to be insufficient to enable a reidentification attack. The statistical summary results of the top 10,000 SNPs for the 23andMe data are available here: www.dri.edu/HealthyNVProjectGenetics. All column definitions are listed in Table 1.

■ **Table 1 Column Identifiers for GWAS Results**

Column name	Definition
CHR	Chromosome
SNP	Individual SNP identifier
BP	Location of SNP on relative chromosome
A1	Alternative Allele
TEST	Selected statistical test – ADD represents the additive effect
NMISS	Indicates the number of observations – non-missing genotypes
BETA	The effect size for this variant, defined per copy of the A1 allele
SE	The standard error of the effect size
LE	Lower end of the 95% confidence interval for the effect size
UE	Upper end of the 95% confidence interval for the effect size
STAT	The value of the test statistic
P	The p-value for the association test

Table describing the column headers for the results file of our genome-wide associations. This summary results file only lists the top 10,000 SNPs in order to prevent a re-identification attack.

The IHI collaborates with scientific researchers on an individual basis. Examples of restrictions that will be considered in requests to data access include but are not limited to:

1. Whether the request comes from an academic institution in good standing and will collaborate with the Institute for Health Innovation to protect the privacy of the participants and the security of the data requested
2. Type and amount of data requested
3. Feasibility of the research suggested
4. Amount of resource allocation for the IHI and Renown Hospital required to support the collaboration

Any correspondence and data availability requests should be addressed to Joe Grzymiski at (joeg@dri.edu) or Craig Kugler (craig.kugler@dri.edu).

PheWAS results

Summarized counts of each ICD classification (ICD-9 and ICD-10) and phenotype group (phecode) are presented in Supplementary Table S7.

Institutional review board

This study was reviewed and approved by the University of Nevada, Reno Institutional Review Board (IRB, project 956068-12).

RESULTS

Characteristics of cohort

10,000 participants were genotyped of which 6,870 participants were linked to Renown Hospital electronic health records with the covariates required to conduct this study: age at consent, gender, ethnicity and BMI. We removed 225 participants based on the exclusion criteria as stated in our methods. Our final cohort characteristics of the 6,645 individuals are described in Table 2, which illustrates the makeup of the cohort with respect to gender, age, ethnic origin, and standardized value of BMI after removal of outliers using our custom algorithm. For the extreme obesity (Class 2 and 3) case vs. control study, the normal (healthy) control range consisted of BMI values between 18.5 and 25 kg/m², while the case obese values were any BMI ≥ 35 kg/m² (Hruby and

■ **Table 2 Cohort Characteristics**

	Association with BMI Measures	Association with Obesity
Cohort Size	6,645	2,996
Age (years)	50.91 ± 15.97	48.91 ± 15.96
Male (%)	2145 (32.28)	747 (24.93)
African American (%)	0 (0)	0 (0)
Asian (%)	157 (2.36)	84 (2.80)
Caucasian (%)	5,945 (89.47)	2653 (88.55)
Latino (%)	173 (2.60)	81 (2.70)
Native American (%)	40 (0.60)	18 (0.60)
Pacific Islander (%)	13 (0.20)	4 (0.13)
Unknown (%)	317 (4.77)	156 (5.21)
DM2	651 (9.8)	246 (8.2)
Quality-Controlled BMI	28.58 ± 6.41	22.57 ± 1.64 (Control), N=2,012 40.15 ± 4.99 (Cases), N=984

Table of cohort characteristics. Continuous variables are presented as mean ± SD; categorical variables are presented as counts and percentages. All values were standardized by using a custom algorithm to remove outliers. BMI has the units of kg/m².

Hu 2014). The number of participants in each range is displayed in Table 2.

GWAS of BMI in the healthy Nevada cohort

Using quality-controlled BMI values, two separate GWASs were performed to find genotypic associations with BMI using *PLINK*. In the first association study, all individuals diagnosed with DM2 were excluded to focus on the association between the genotype and BMI under the additive model with adjustments for gender, age and the first four principal components (PC1-PC4). The second association analysis included all DM2-diagnosed individuals and added DM2 as a covariate to the model, again using the additive model with adjustments for gender, age, diabetes status, and PC1-PC4. Genomic inflation coefficients (lambda) were computed for the two separate cohorts: 1.06 for the association without DM as a covariate, and 1.06 for the association where DM is a covariate. Any SNP with association *p*-value of $p < 1 \times 10^{-5}$ was considered a statistically significant association, based on the standard of the NHGRI-EBI Catalog of published genome-wide association studies [<https://www.ebi.ac.uk/gwas/docs/methods/criteria>], as well as obesity studies performed by Frayling *et al.* (Frayling *et al.* 2007). Genetic variance in the BMI study with DM2 cases removed was 15.78%; genetic variance was 17.49% in the BMI study with DM2 cases included.

The first GWAS was performed on 5,994 total participants without DM2 and identified 20 SNPs across seven chromosomes at statistical significance defined by $p < 1 \times 10^{-5}$ (Table 3). The majority of these mapped to the *FTO* gene on chromosome 16, while two SNPs mapped to *TDH* on chromosome 8 (Supplementary Figure S2). Of the 20 SNPs, 15 were shown to be associated with BMI in previous publications (Scuteri *et al.* 2007; Frayling *et al.* 2007; Dina *et al.* 2007; Zeggini *et al.* 2007; Yanagiya *et al.* 2007; Hinney *et al.* 2007; Hunt *et al.* 2008; Price *et al.* 2008; Grant *et al.* 2008; Hotta *et al.* 2008; Loos *et al.* 2008; Tan *et al.* 2008; Villalobos-Comparán *et al.* 2008; Thorleifsson *et al.* 2008; Willer *et al.* 2009; Meyre *et al.* 2009; Wing *et al.* 2009; Liu *et al.* 2010; Shimaoka *et al.* 2010; Fawcett and Barroso 2010; Speliotes *et al.* 2010; Wang *et al.* 2011; Prakash *et al.* 2011; Okada *et al.* 2012; Cha *et al.* 2012; Berndt *et al.* 2013; Wheeler *et al.* 2013; Graff *et al.* 2013; Olza *et al.* 2013; Boender *et al.* 2014; Qureshi *et al.* 2017; Huđek *et al.* 2018; González-Herrera *et al.* 2019). A large majority of the SNPs (17/20) lie within noncoding regions of genes and are intronic in nature. It is interesting to note that our strongest associations lie within the *FTO* gene ($p < 3.5 \times 10^{-6}$). Results are presented in Table 3: **BMI without DM2** lists the significant associations of our cohort that exclude

all DM2 diagnoses. **BMI with DM2** presents significant associations with BMI in the cohort that includes participants with a DM2 diagnosis. Effect sizes (and their standard deviations) are presented as change in BMI per each copy of the minor allele. Raw per-SNP *p*-values are presented.

Statistical analysis with *PLINK* demonstrated that DM2 is a significant predictor of BMI, with the *p*-value of its coefficient consistently less than $p < 2 \times 10^{-16}$ in each per-SNP linear regression. The entire cohort includes 6,645 participants: of those, 651 have a diagnosis of DM2 in their twelve-year medical history. A GWAS applied to this larger cohort identified 27 statistically significant SNPs across seven chromosomes associated with BMI at $p < 1 \times 10^{-5}$ (Figure 1). Furthermore, 77% of the SNPs in this second GWAS (21/27) were previously associated with BMI in previous research studies (Scuteri *et al.* 2007; Frayling *et al.* 2007; Dina *et al.* 2007; Zeggini *et al.* 2007; Yanagiya *et al.* 2007; Hinney *et al.* 2007; Hunt *et al.* 2008; Price *et al.* 2008; Grant *et al.* 2008; Hotta *et al.* 2008; Loos *et al.* 2008; Tan *et al.* 2008; Villalobos-Comparán *et al.* 2008; Thorleifsson *et al.* 2008; Willer *et al.* 2009; Meyre *et al.* 2009; Wing *et al.* 2009; Liu *et al.* 2010; Shimaoka *et al.* 2010; Fawcett and Barroso 2010; Speliotes *et al.* 2010; Wang *et al.* 2011; Prakash *et al.* 2011; Okada *et al.* 2012; Cha *et al.* 2012; Berndt *et al.* 2013; Wheeler *et al.* 2013; Graff *et al.* 2013; Olza *et al.* 2013; Boender *et al.* 2014; Christensen *et al.* 2015; Nakajima *et al.* 2016; Thomsen *et al.* 2016; Qureshi *et al.* 2017; Huđek *et al.* 2018; González-Herrera *et al.* 2019). With the addition of DM2 as a covariate, the GWAS identified several additional SNPs on chromosome 8, as well as SNPs on chromosomes 17 and 19. These additional SNPs were previously linked to BMI and obesity in other studies (Christensen *et al.* 2015; Nakajima *et al.* 2016; Thomsen *et al.* 2016). Manhattan plots for the two BMI GWAS studies are presented in Figure 1 and Supplementary Figure S2, with the linear associations results presented in Table 3.

The SNP on chromosome 17 is of particular interest, as it has the largest effect of any SNP identified in our study ($\beta=0.90$). It is also the rarest SNP tested in our cohort with minor allele frequency (MAF) 8.17%. The median MAF across the strongest associative SNPs in both studies is 40%, which demonstrates that most of the SNPs are common and thus result in relatively moderate individual effect sizes. Most of the SNPs lie within noncoding intronic regions; this is of interest as although noncoding SNPs do not alter the amino acid sequence of the translated protein, they may be linked to disease, as several previous studies show that polymorphisms within introns can affect intron splicing

■ **Table 3 Statistically Significant BMI GWAS SNPs**

rsID	Chrom	Cyto Region	Associated Gene	Minor Allele	MAF	β	(SE)	GWAS p -value	Mutation Classification
BMI without DM2									
rs1620977	chr1	p31.1	NEGR1	A	27.29	0.5819	0.125	3.30x10 ⁻⁶	intron
rs871122	chr5	p15.32	ADAMTS16	T	16.61	0.6837	0.1528	7.78x10 ⁻⁶	intron
rs4839813	chr6	q16.1	FUT9	T	9.69	0.8755	0.1876	3.13x10 ⁻⁶	intron
rs11774673	chr8	p23.1	NA	C	48.21	0.5083	0.1119	5.69x10 ⁻⁶	unknown
rs2060457	chr8	p23.1	TDH	T	48.58	0.5584	0.1119	6.24x10 ⁻⁷	intron
rs2293859	chr8	p23.1	TDH	G	48.73	0.5506	0.1119	8.80x10 ⁻⁷	ncRNA
rs10733990	chr10	p12.1	NA	A	30.53	0.5857	0.1199	1.06x10 ⁻⁶	unknown
rs10875969	chr12	q13.12	NCKAP5L	A	42.38	0.505	0.1133	8.41x10 ⁻⁶	intron
rs9937053	chr16	q12.2	FTO	A	40.76	0.5193	0.1133	4.69x10 ⁻⁶	intron
rs9930333	chr16	q12.2	FTO	G	40.78	0.5161	0.113	5.08x10 ⁻⁶	intron
rs9940128	chr16	q12.2	FTO	A	40.75	0.5272	0.1131	3.19x10 ⁻⁶	intron
rs1421085	chr16	q12.2	FTO	C	38.43	0.5999	0.1147	1.75x10 ⁻⁷	intron
rs1558902	chr16	q12.2	FTO	A	38.46	0.6007	0.1147	1.67x10 ⁻⁷	intron
rs1121980	chr16	q12.2	FTO	A	40.85	0.5272	0.1128	3.03x10 ⁻⁶	intron
rs17817449	chr16	q12.2	FTO	G	37.93	0.5562	0.1143	1.17x10 ⁻⁶	intron
rs8043757	chr16	q12.2	FTO	T	37.98	0.5612	0.1143	9.29x10 ⁻⁷	intron
rs8050136	chr16	q12.2	FTO	A	37.94	0.5636	0.1143	8.34x10 ⁻⁷	intron
rs3751812	chr16	q12.2	FTO	T	37.52	0.548	0.1149	1.90x10 ⁻⁶	intron
rs9939609	chr16	q12.2	FTO	A	38.06	0.5504	0.1143	1.50x10 ⁻⁶	intron
rs12149832	chr16	q12.2	FTO	A	39.1	0.534	0.1142	3.01x10 ⁻⁶	intron
BMI with DM2									
rs1776012	chr1	p31.1	NEGR1	G	47.74	-0.4848	0.1092	9.13x10 ⁻⁶	intron
rs11774673	chr8	p23.1	NA	C	48.21	0.5149	0.1086	2.14x10 ⁻⁶	unknown
rs1435277	chr8	p23.1	NA	C	44.39	-0.4944	0.1112	8.90x10 ⁻⁶	near-gene-5
rs11250129	chr8	p23.1	TDH	A	48.12	0.5115	0.1089	2.67x10 ⁻⁶	intron
rs2060457	chr8	p23.1	TDH	T	48.58	0.5802	0.1085	9.30x10 ⁻⁸	intron
rs2293859	chr8	p23.1	TDH	G	48.73	0.5722	0.1085	1.37x10 ⁻⁷	ncRNA
rs2246606	chr8	p23.1	TDH	G	42.95	-0.539	0.1115	1.37x10 ⁻⁶	intron
rs2736280	chr8	p23.1	TDH	C	48.22	-0.5308	0.1098	1.36x10 ⁻⁶	intron
rs2572386	chr8	p23.1	FAM167A	G	42.27	-0.5308	0.1115	1.96x10 ⁻⁶	intron
rs2948300	chr8	p23.1	NA	T	49.03	0.5034	0.1099	4.70x10 ⁻⁶	unknown
rs12412241	chr10	p14	NA	A	29.62	-0.5474	0.1171	3.01x10 ⁻⁶	unknown
rs11041833	chr11	p15.4	LMO1	A	41.01	-0.4841	0.1082	7.77x10 ⁻⁶	intron
rs9937053	chr16	q12.2	FTO	A	40.76	0.5002	0.1105	6.04x10 ⁻⁶	intron
rs9930333	chr16	q12.2	FTO	G	40.78	0.4989	0.1102	6.10x10 ⁻⁶	intron
rs9940128	chr16	q12.2	FTO	A	40.75	0.5078	0.1102	4.18x10 ⁻⁶	intron
rs1421085	chr16	q12.2	FTO	C	38.43	0.5616	0.1117	5.04x10 ⁻⁷	intron
rs1558902	chr16	q12.2	FTO	A	38.46	0.5625	0.1116	4.80x10 ⁻⁷	intron
rs1121980	chr16	q12.2	FTO	A	40.85	0.5063	0.11	4.27x10 ⁻⁶	intron
rs17817449	chr16	q12.2	FTO	G	37.93	0.5206	0.1113	2.94x10 ⁻⁶	intron
rs8043757	chr16	q12.2	FTO	T	37.98	0.525	0.1112	2.41x10 ⁻⁶	intron
rs8050136	chr16	q12.2	FTO	A	37.94	0.5244	0.1112	2.47x10 ⁻⁶	intron
rs3751812	chr16	q12.2	FTO	T	37.52	0.5119	0.1118	4.74x10 ⁻⁶	intron
rs9939609	chr16	q12.2	FTO	A	38.06	0.513	0.1112	4.06x10 ⁻⁶	intron
rs12149832	chr16	q12.2	FTO	A	39.1	0.5034	0.1111	5.98x10 ⁻⁶	intron
rs11651343	chr17	p13.3	RTN4RL1	T	8.17	0.9039	0.1974	4.75x10 ⁻⁶	intron
rs750456	chr19	q13.33	CABP5	C	26.19	-0.5486	0.1221	7.15x10 ⁻⁶	intron
rs8105198	chr19	q13.33	CABP5	G	17.18	-0.6466	0.1417	5.14x10 ⁻⁶	coding-synon

This table represents statistically significant associations with BMI in our cohort.

as well as transcriptional and translational efficiency (Lalonde *et al.* 2011).

Case-control GWAS of extreme obesity in the healthy Nevada cohort

A complementary GWAS was performed to identify genotype-phenotype links in extreme obesity (BMI ≥ 35) versus non-obese (BMI between 18.5 and 25 kg/m²) in our cohort. This study incorporated 2996 participants (984 extreme obese cases, 2012 non-obese controls), and under the log-additive model with adjustments for gender, age and the first four principal components, identified 26 SNPs across six

chromosomes that were associated with obesity at $p < 1 \times 10^{-5}$, with approximately 70% associated with obesity and BMI in prior studies (Figure 2). The percentage of phenotypic variance attributed to genetic variation based on all the SNPs was 15.7%. The genomic inflation coefficient (λ) for the obesity cohort was computed as 1.05. We also include eight SNPs found slightly above the significance threshold in the *FTO* gene that are reported in several studies as obesity-related (Ehrlich and Friedenberg 2016; West *et al.* 2018). In comparison to the two quantitative-trait BMI GWASs, this study identified several more associations around the *NEGR1* gene on chromosome 1. We also identified SNPs in two genes, *PFKFB3* and



Figure 1 Manhattan plot of GWAS results of BMI including DM2-diagnosed individuals. This study includes DM2-diagnosed individuals and the statistical model includes DM2 as a bimodal covariate. The x-axis represents the genomic position of 500,508 SNPs. The y-axis represents $-\log_{10}$ -transformed raw p -values of each genotypic association. The red horizontal line indicates the significance level 1×10^{-5} .

CABP5, which are associated with obesity in other studies (Scuteri *et al.* 2007; Jiao *et al.* 2008; Nakajima *et al.* 2016). Note that all the mutations in the *FTO* gene increase the odds of obesity risk. Table 4 lists the strongest SNPs associated in our extreme obese vs. non-obese GWAS. Effect sizes and their standard deviations are presented as odds ratios. Raw *p*-values generated by the GWAS are also presented.

From our three separate GWASs, we identified fifteen different chromosomal cytoband regions across ten chromosomes associated with at least one BMI or obesity-related trait. All but three of those cytoband regions contained a gene, while the remaining cytoband hits were in noncoding regions of the genome. Approximately 70% of the SNPs identified in this study were linked to BMI and obesity in prior studies (Tables 3 and 4), validating our methods (Scuteri *et al.* 2007; Frayling *et al.* 2007; Dina *et al.* 2007; Yanagiya *et al.* 2007; Hinney *et al.* 2007; Jiao *et al.* 2008; Pietiläinen *et al.* 2008; Grant *et al.* 2008; Hotta *et al.* 2008; Thorleifsson *et al.* 2008; Joe *et al.* 2009; Nakajima *et al.* 2016; Thomsen *et al.* 2016; Justice *et al.* 2017). The functions of the genes which lie within the cytoband regions are outlined in Table 5.

Analysis of variance

The mean BMI values across genotypes presented in Supplementary Tables S1 and S2 correlate with negative and positive effect sizes: SNPs showing a negative effect size have a decrease in mean BMI values across the genotypes from left to right (homozygous in major allele, heterozygous, homozygous in minor allele). Note that BMI levels increase with the increase of the number of minor alleles, which is typical of variants in *FTO* (Frayling *et al.* 2007). All ANOVA F-test *p*-values of the significant SNPs identified in the two BMI studies are statistically significant at the $\alpha=0.05$ level, even after a simple Bonferroni correction ($.05/27=0.0019$, and $.05/20=0.0025$, respectively). Supplementary Table S3 presents the proportion of extremely obese cases across each genotype. A box and whisker figure of ANOVA results for one of the strongest associations (rs9939609) is shown in Supplementary Figure S3.

PheWAS of BMI and obesity

Beyond the GWASs, we present here two comprehensive PheWAS studies that follow each GWAS. The first examines pleiotropy, *i.e.*, whether additional phenotypic associations exist between the statistically significant SNPs associated with BMI or obesity in our cohort. The second investigates which EHR phenotype groups are associated with BMI; more specifically, the analysis identifies whether the number of individuals in an EHR phenotype group is a predictor of BMI and/or extreme obesity.

A PheWAS tested the 20 statistically significant SNPs identified in the first BMI GWAS (the sub-cohort with no DM2-diagnoses) for association with 562 EHR phenotypes and resulted in no statistically significant associations at the false discovery rate of 0.1 (data not shown). The top two associations showed that a locus on *FUT9* (rs4839813) associated with obesity, and rs1620977 on *NEGR1* associated with morbid obesity with raw *p*-values $p=2.3 \times 10^{-5}$ and 2.8×10^{-5} , respectively. The second PheWAS identified 179 EHR (phenotypic) associations of BMI (DM2-diagnosed participants excluded) with $p < 2 \times 10^{-2}$, that associates to an adjusted *p*-value of 0.1 (see Materials and Methods). Included in the strongest phenotypic associations are obesity, morbid obesity, and overweight ($p < 1 \times 10^{-80}$), sleep apnea ($p < 1 \times 10^{-45}$), hypertension ($p < 1 \times 10^{-40}$), abnormal glucose ($p < 1 \times 10^{-25}$), hyperlipidemia, asthma, GERD, osteoporosis, and others. (Data not shown).

The PheWAS of the second BMI GWAS (DM2-diagnosed individuals included) examined whether 633 EHR phenotype groups containing at least 20 participants are dependent on the genotypes of the

27 statistically significant SNPs associated with BMI in our cohort (Figure 3). Results of this PheWAS indicate that *TDH* and *FAMA167-AS1* are strongly associated with DM2. Variants in the *FTO* gene associate with obesity and overweight phenotypes. An association with hypertension and essential hypertension and the locus rs12412241 on chromosome 10 was detected ($p < 1 \times 10^{-4}$). Several strong associations with *FTO* loci and the prostate-specific antigen (PSA) were also found. Variants in the *FTO* gene also associated with obesity at the $p=3 \times 10^{-4}$ level, and with hypercholesterolemia with $p=1 \times 10^{-3}$ level. Significant associations ($p < 1.02 \times 10^{-4}$, associated to an adjusted *p*-value of 0.1, see Materials and Methods) are included in Table 6 and illustrated in Figure 3, where the blue line represents the Bonferroni correction of $p=3 \times 10^{-6}$. The second PheWAS examined links between BMI and the 1,523 EHR phenotype groups containing at least 20 individuals in this cohort and showed that 301 such clinical phenotypes groups associated with significance $p < 1.96 \times 10^{-2}$. (This significance level is associated to an adjusted *p*-value of 0.1, as described in the Materials and Methods) These are shown in Figure 4. Included in the highest associations are obesity, morbid obesity, and overweight ($p < 1 \times 10^{-100}$), DM2 ($p < 1 \times 10^{-87}$), hypertension ($p < 1 \times 10^{-82}$), sleep apnea ($p < 1 \times 10^{-80}$), abnormal glucose ($p < 1 \times 10^{-53}$), hyperlipidemia, asthma and other respiratory disorders, GERD, edema, liver disease, mood disorders, polycystic ovaries, and others. Significant associations are presented in Supplementary Table S4 and Figure 4. Only associations at $p < 1 \times 10^{-15}$ are annotated in the image for ease of viewing. Note that a single-SNP Bonferroni correction results in a significance level of 3.3×10^{-5} .

We performed the same two PheWASs on the case-control GWAS. The first PheWAS identified possible associations between 34 SNPs and 372 phenotype groups with at least 20 individuals (Figure 5). The significance level corresponding to an FDR of 0.1 was $p=3.85 \times 10^{-4}$, resulting in 50 significant associations. The Bonferroni correction *p*-value is 4×10^{-6} and shown in blue (see Materials and Methods). The strongest associations occurred between the *NEGR1* and *KLRB1* gene and obesity and morbid obesity and the *NEGR1* gene and abnormal glucose. The *NEGR1* gene, along with *TDH* and *FAMA167A-AS1*, also associated with impaired fasting glucose and diabetes, respectively, at a slightly lower, yet still significant *p*-values. The locus rs2948300 on chromosome 8 associated with essential hypertension. The SNP rs1620977 on *NEGR1* is linked with incidence of bronchitis. Additionally, several strong associations of irritable bowel syndrome (IBS) and digestive disorders with loci in *CABP5* are shown. Results are pictured in Figure 5 and included in Table 7.

The second PheWAS in this case-control study identified possible links between 1,362 EHR phenotype groups with at least 20 individuals and the incidence of extreme obesity (Figure 6). The significant threshold of $p=1.4 \times 10^{-2}$ enabled an FDR of 0.1. At this significance level, 191 significant associations were identified, including obesity ($p < 1 \times 10^{-134}$), hypertension ($p < 1 \times 10^{-65}$), sleep apnea ($p < 1 \times 10^{-43}$), abnormal glucose ($p < 1 \times 10^{-40}$), hyperlipidemia, asthma, and GERD. The high level of association with obesity validates our methods. These associations are shown in Figure 6, with phenotypes annotated above a significance level of 1×10^{-15} for ease of viewing. A line is drawn at the significance level 1×10^{-15} as guidance. Results are included in Supplementary Table S5. Note that a single-SNP Bonferroni correction results in a significance level of 3.7×10^{-5} .

DISCUSSION

GWAS of healthy Nevada BMI and obesity

Here we present three GWASs on participants in the Healthy Nevada Project. The first investigates associations with BMI on subjects that do

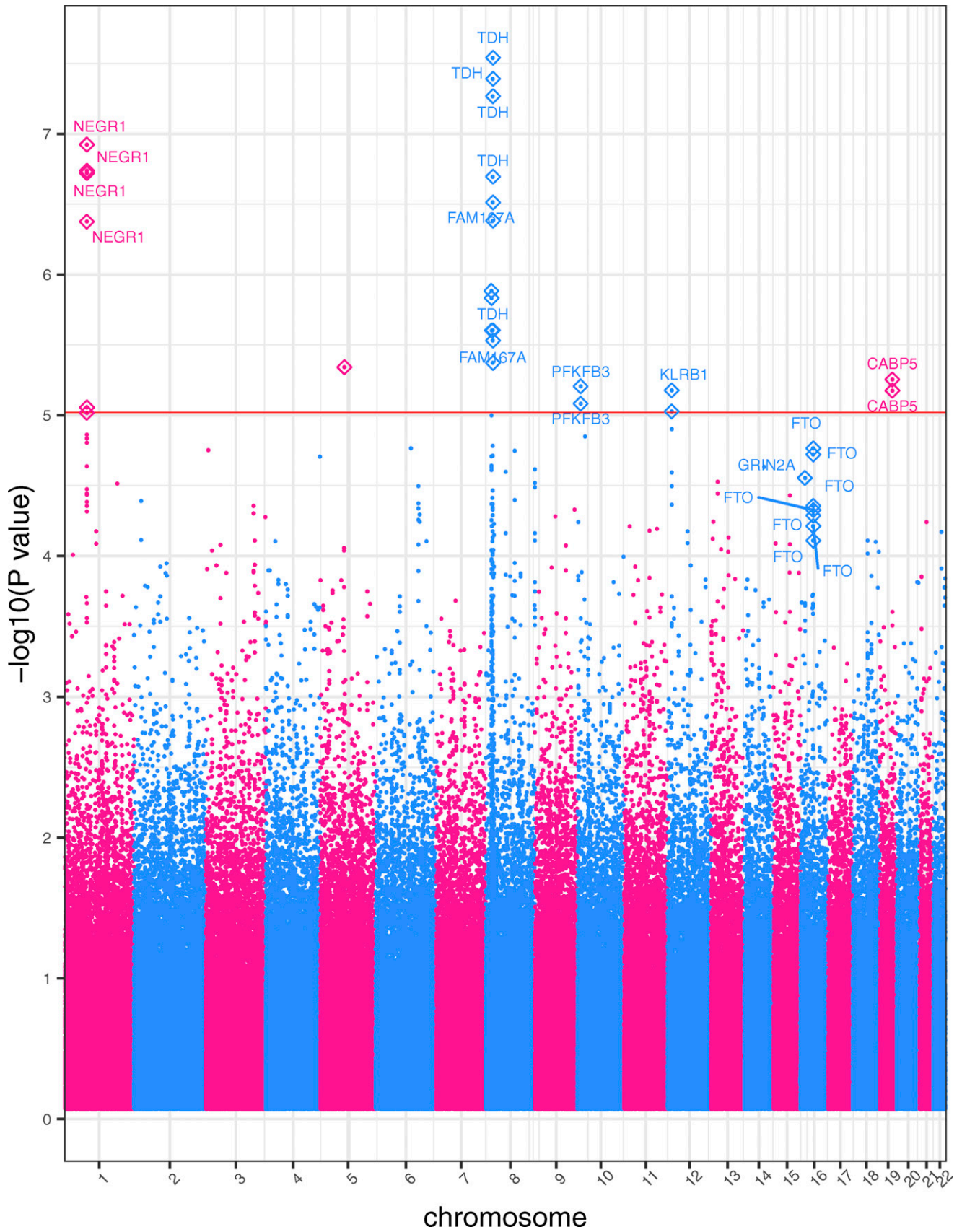


Figure 2 Obesity Case-Control GWAS Manhattan Plot. This cohort includes DM2-diagnosed individuals. The x-axis represents the genomic position of 500,508 SNPs. The y-axis represents $-\log_{10}$ -transformed raw p -values of each genotypic association. The red horizontal line indicates the significance level 1×10^{-5} .

■ **Table 4 Statistically Significant Obesity GWAS SNPs**

rsID	Chr	Cyto Region	Associated Gene	Minor Allele	MAF	Odds Ratio	(SE)	GWAS <i>p</i> -value	Mutation Classification
rs1776012	chr1	p31.1	NEGR1	G	47.36	0.7365	0.05775	1.19x10 ⁻⁷	intron
rs9424977	chr1	p31.1	NEGR1	C	47.09	0.7411	0.05745	1.83x10 ⁻⁷	intron
rs1620977	chr1	p31.1	NEGR1	A	27.31	1.369	0.06213	4.21x10 ⁻⁷	intron
rs1870676	chr1	p31.1	NEGR1	T	46.97	0.7403	0.05773	1.90x10 ⁻⁷	intron
rs3101336	chr1	p31.1	NA	T	35.93	0.7678	0.0597	9.60x10 ⁻⁶	unknown
rs2568958	chr1	p31.1	NA	G	35.94	0.767	0.05967	8.79x10 ⁻⁶	unknown
rs2815752	chr1	p31.1	NA	G	35.94	0.767	0.05967	8.79x10 ⁻⁶	unknown
rs2173676	chr5	q14.1	NA	C	26.52	0.7433	0.0647	4.56x10 ⁻⁶	unknown
rs11774673	chr8	p23.1	NA	C	48.25	1.304	0.05778	4.25x10 ⁻⁶	unknown
rs1435277	chr8	p23.1	NA	C	44.28	0.7419	0.05898	4.15x10 ⁻⁷	near-gene-5
rs11250129	chr8	p23.1	TDH	A	48.26	1.314	0.05796	2.49x10 ⁻⁶	intron
rs2060457	chr8	p23.1	TDH	T	48.76	1.379	0.05797	2.88x10 ⁻⁸	intron
rs2293859	chr8	p23.1	TDH	G	48.81	1.375	0.05798	4.06x10 ⁻⁸	ncRNA
rs2246606	chr8	p23.1	TDH	G	42.76	0.7357	0.05905	2.01x10 ⁻⁷	intron
rs2736280	chr8	p23.1	TDH	C	48.23	0.727	0.05863	5.41x10 ⁻⁸	intron
rs2572386	chr8	p23.1	FAM167A	G	41.91	0.7378	0.05941	3.07x10 ⁻⁷	intron
rs1435282	chr8	p23.1	FAM167A	A	45.57	0.7631	0.05786	2.95x10 ⁻⁶	intron
rs2948300	chr8	p23.1	NA	T	49.06	1.325	0.05809	1.30x10 ⁻⁶	unknown
rs2953802	chr8	p23.1	NA	A	42	1.318	0.05734	1.47x10 ⁻⁶	unknown
rs435581	chr8	p23.1	NA	A	45.19	0.7609	0.05802	2.50x10 ⁻⁶	unknown
rs680951	chr10	p15.1	PFKFB3	G	33.83	0.7614	0.06116	8.26x10 ⁻⁶	intron
rs666595	chr10	p15.1	PFKFB3	A	31.47	0.7537	0.06256	6.22x10 ⁻⁶	intron
rs2058426	chr12	p13.31	NA	A	49.67	0.7798	0.05614	9.40x10 ⁻⁶	near-gene-3
rs2241005	chr12	p13.31	KLRB1	C	49.98	0.7768	0.05608	6.65x10 ⁻⁶	intron
rs1421085	chr16	q12.2	FTO	C	38.18	1.28	0.05766	1.90x10 ⁻⁵	intron
rs1558902	chr16	q12.2	FTO	A	38.21	1.281	0.0576	1.72x10 ⁻⁵	intron
rs17817449	chr16	q12.2	FTO	G	37.63	1.261	0.05732	5.15x10 ⁻⁵	intron
rs8043757	chr16	q12.2	FTO	T	37.66	1.264	0.05733	4.42x10 ⁻⁵	intron
rs8050136	chr16	q12.2	FTO	A	37.66	1.263	0.05733	4.72x10 ⁻⁵	intron
rs3751812	chr16	q12.2	FTO	T	37.27	1.26	0.05764	6.13x10 ⁻⁵	intron
rs9939609	chr16	q12.2	FTO	A	37.83	1.254	0.05737	7.79x10 ⁻⁵	intron
rs2267770	chr16	p13.2	GRIN2A	T	41.01	0.7859	0.0575	2.79x10 ⁻⁵	intron
rs750456	chr19	q13.33	CABP5	C	26.1	0.7405	0.06615	5.58x10 ⁻⁶	intron
rs8105198	chr19	q13.33	CABP5	G	16.77	0.7026	0.07836	6.67x10 ⁻⁶	coding-synon

This table presents statistically significant associations with extreme obesity in the case-control study.

not have DM2 diagnoses. The second identifies associations with BMI in which participants with DM2 are included as a comorbidity. The third GWAS is a case-control study of extreme obesity that complements outcomes of the first two quantitative trait studies. Each GWAS is followed with two independent PheWASs to examine pleiotropy and additional phenotypic associations with quantitative BMI levels and incidence of obesity.

The first GWAS tested the association between genotype and BMI without DM2, to remove DM2 effects on BMI. As expected, the majority of the resulting associations were found in SNPs that lie within the *FTO* gene. This gene has been associated to BMI and obesity in several studies and is a major focal point in obesity-related research (Scuteri *et al.* 2007; Frayling *et al.* 2007; Dina *et al.* 2007; Zeggini *et al.* 2007; Hinney *et al.* 2007; Hunt *et al.* 2008; Price *et al.* 2008; Grant *et al.* 2008; Hotta *et al.* 2008; Loos *et al.* 2008; Tan *et al.* 2008; Villalobos-Comparán *et al.* 2008; Thorleifsson *et al.* 2008; Willer *et al.* 2009; Meyre *et al.* 2009; Wing *et al.* 2009; Shimaoka *et al.* 2010; Fawcett and Barroso 2010; Wang *et al.* 2011; Prakash *et al.* 2011; Okada *et al.* 2012; Berndt *et al.* 2013; Wheeler *et al.* 2013; Graff *et al.* 2013; Olza *et al.* 2013; Qureshi *et al.* 2017; González-Herrera *et al.* 2019). Moreover, the two strongest associations in the GWAS were from SNPs in *FTO* (rs1558902/ rs1421085, $p = 1.67 \times 10^{-7} / 1.75 \times 10^{-7}$), highlighting the overall importance of this gene in relation to obesity. Frayling suggests that the association of *FTO* SNPs with DM2 is mediated through BMI (Frayling *et al.* 2007). The exact mechanism by which the *FTO* gene affects BMI is not

understood; however, it has been discovered that the gene product of *FTO* mediates oxidative demethylation of several different RNA species, such as mRNA, snRNA and tRNA (Jia *et al.* 2011; Wei *et al.* 2018). This indicates that protein produced from *FTO* likely operates as a RNA regulatory molecule, which can affect both gene expression as well as translation initiation and elongation (Wei *et al.* 2018).

Two SNPs within *TDH* were found to be strongly associated to BMI. This gene codes for a nonfunctional L-threonine dehydrogenase, lacking most of the C-terminus found in other species, and is thus characterized as a putative pseudogene. Previous research has identified this gene as a possible susceptibility gene for obesity (Yanagiya *et al.* 2007); however, relatively little is known about any functional consequences of SNPs within this pseudogene. We also observed a strongly associated locus in *NEGR1*, one of the first genes shown to have variants associated to BMI (Thorleifsson *et al.* 2008; Willer *et al.* 2009; Speliotes *et al.* 2010; Boender *et al.* 2014). This gene codes for a cell adhesion molecule, although its function in relation to BMI is still unknown. Previous research in mice determined that deletions of *NEGR1* cause a decrease in weight and a change in the regulation of energy balance, implying that *NEGR1* most likely functions to control the regulation of energy balance (Lee *et al.* 2012; Boender *et al.* 2014).

We hypothesized that including DM2 participants (and thus DM2 as a covariate in the genetic model) would produce a more parsimonious fit, as many studies show a relationship between diabetes and BMI. We discovered that diabetes was indeed an important predictor of BMI for all

■ **Table 5 Table Presenting Gene Functions**

Gene	Gene Description	Region	Trait	Function	Reference
NEGR1	Neuronal Growth Regulator 1	p31.1	BMI No DM2 BMI w DM2 obesity	Cell-adhesion molecule and regulator of cellular processes as neurite outgrowth and synapse formation	(Hashimoto et al. 2008; Boender et al. 2014)
ADAMTS16	ADAM Metalloproteinase with Thrombospondin Type 1 Motif 16	p15.32	BMI No DM2	May be a secreted proteinase	(SurrIDGE et al. 2009)
FUT9	Fucosyltransferase 9	q16.1	BMI No DM2	Plays a role in the biosynthesis of Lewis X (LeX) antigen precursor polysaccharides	(Gouveia et al. 2012)
TDH	L-Threonine Dehydrogenase	p23.1	BMI No DM2 BMI w DM2 obesity	Likely pseudogene that cannot produce functional protein	(Yanagiya et al. 2007)
NCKAP5L	NCK Associated Protein 5 Like	q13.12	BMI No DM2	Regulates microtubule organization and stabilization	(Mori et al. 2015)
FTO	Fat Mass and Obesity Associated	q12.2	BMI No DM2 BMI w DM2 obesity	Mediates oxidative demethylation of different RNA species. Acts as a regulator of fat and energy homeostasis	(Frayling et al. 2007; Jia et al. 2011; Wei et al. 2018)
FAM167A	Family with Sequence Similarity 167 Member A	p23.1	BMI w DM2 obesity	Unknown function. May have a role in autoimmune diseases	(Chen et al. 2015)
LMO1	LIM domain only 1	p15.4	BMI w DM2	Transcriptional regulator through binding of transcription factors	(Wang et al. 2010)
RTN4RL1	Reticulon 4 Receptor Like 1	p13.3	BMI w DM2	Cell surface receptor	(Pignot et al. 2003)
CABP5	Calcium Binding Protein 5	q13.33	BMI w DM2 obesity	Plays a role in calcium mediated cellular signal transduction	(Haeseleer et al. 2002)
PFKFB3	6-Phosphofructo-2-kinase/Fructose-2,6-biphosphatase 3	p15.1	obesity	Potent regulator of glycolysis	(Nelson and Cox 2005; De Bock et al. 2013)
KLRB1	Killer Cell Lectin-like Receptor Subfamily B, Member 1	p13.31	obesity	Possible regular of natural killer cells	(Rother et al. 2015)
GRIN2A	Glutamate Ionotropic Receptor NMDA Type Subunit 2A	p13.2	obesity	Possibly has a role in learning and long-term memory	(Micu et al. 2006)

This table presents functions of genes associated to all SNPs found significantly associated to one or more BMI and obesity traits in the GWASs.

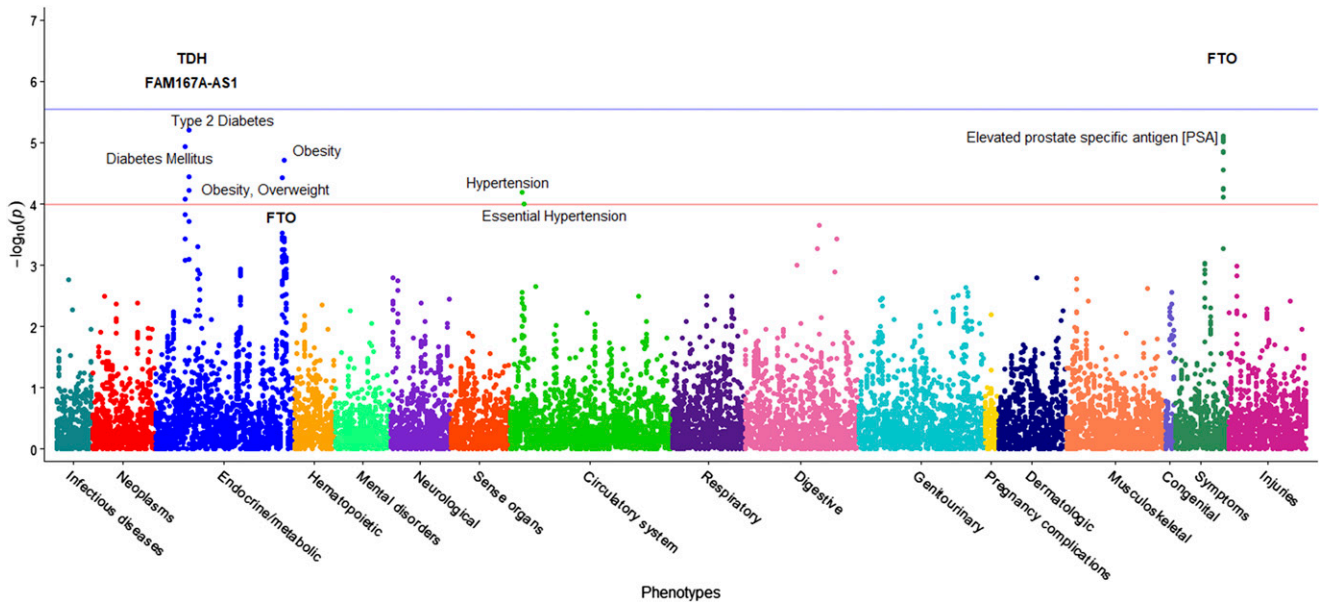


Figure 3 PheWAS results between BMI-significant SNPs and EHR Phenotypes. This figure shows the results of individual logistic regressions between incidence of 633 phenotype groups (phecodes) and the genotypes of 27 SNPs found to have statistically significant associations with BMI in a cohort with DM2 patients. Each point represents the p -value of one SNP and one of 633 phecodes with at least 20 cases assigned to it. The horizontal red line represents the significance level $p=1.02 \times 10^{-4}$, and the blue line represents the Bonferroni correction of $p=3 \times 10^{-6}$.

500,000 regressions performed ($p < 2 \times 10^{-16}$ for every SNP), regardless of age, gender or genotype. Furthermore, adding DM2 as a predictor in the additive model increases the significance of associations between SNPs in the *TDH* gene and BMI; five out the top eight most significant associations fall within this gene (Table 3). It is clear that incidence of DM2 in our cohort affects the genetic association of BMI. Specifically, when DM2 patients are excluded in our cohort, there are two associations within *TDH*. When DM2 participants are included, we observe five associations with the *TDH* gene. This indicates, along with the BMI PheWAS results, that not only does *TDH* influence BMI measurements, it also has an association with DM2.

It is rare for SNPs to be effectors of two separate diseases, even those as intertwined as BMI and DM2 (Grarup *et al.* 2014). A possible explanation why *TDH* has not been previously associated with DM2 is due to a lack of statistical power to observe the small risk increases *TDH* may impose on DM2 (Grarup *et al.* 2014). Nonetheless, the increased rate of DM2 diagnosis worldwide makes this an interesting candidate gene. How the SNPs in the *TDH* pseudogene may influence either BMI or DM2 is unknown, as evidence supporting the association between *TDH* and BMI/DM2 is scant; however, previous research has discovered that not all pseudogenes are "junk" DNA. Some of these genes can be actively transcribed to produce short interfering RNAs (siRNAs), which can regulate gene expression (Pink *et al.* 2011). In certain cases, they can even competitively bind micro-RNAs (miRNAs), which can attenuate repression of cellular mRNA (An *et al.* 2016). Additionally, the expression of pseudogene transcripts tends to be tissue-specific. Given that the greatest expression of *TDH* transcripts is found in the pancreas (Fagerberg *et al.* 2014), one might speculate that *TDH* affects the production of insulin and/or digestive enzymes. If true, this may account for our observation where *TDH* influences BMI measurements, and is associated with DM2. Given the strong associations between *TDH* SNPs and DM2, as well as potential regulatory functions of pseudogenes, we believe it is essential that future studies focus on determining the function of the *TDH* pseudogene in a tissue-specific context. Although studies using genetically modified mice with a *TDH*

polymorphism and proteomics analysis of their pancreatic tissue would be straight-forward, to the best of our knowledge, no such studies have been reported. We see this as a possible future direction.

Adding DM2 as a covariate into the statistical model also identified two additional genes that may influence BMI: *RTN4RL1* and *CABP5*. *RTN4RL1* is a gene that codes for a cell surface receptor and was previously found to be upregulated approximately 2-fold when exposed to bone morphogenetic protein 4 (BMP4), a protein that is increased in diabetic animals and may reduce insulin secretion (Christensen *et al.* 2015). This implies that the effects of *RTN4RL1* on BMI may be secondary to its main effect on diabetes. Moreover, this gene has also been listed as a potential candidate gene for DM2 in previous GWAS (Thomsen *et al.* 2016). The gene *CABP5*, which codes for a calcium binding protein that has role in calcium mediated cellular signal transduction (Haeseleer *et al.* 2002) may have a more direct effect on BMI. It was previously discovered as part of a group of several genes that were upregulated in obese individuals, although its exact function relative to obesity is still unknown (Nakajima *et al.* 2016).

A case-control association study examining the effects between genetics and the risk of extreme obesity ($\text{BMI} \geq 35 \text{ kg/m}^2$) was the final GWAS we conducted. It has been determined by the World Health Organization (WHO) that more than 1.9 billion adults are overweight and over 650 million are obese. Moreover, obesity is associated with several other chronic diseases, such as cardiovascular disease, DM2 and cancer, all of which could lead to premature death (Kopelman 2007). Overall, our obesity results consist of many of the same SNPs and genes found to be associated with BMI. However, the obesity results did demonstrate an increase in the genetic associations at the significance level of $p < 1 \times 10^{-5}$ and very close to the *NEGR1* gene compared to previous BMI associations. Previous studies using genetically modified mice with *NEGR1* deficiency or *NEGR1*-loss-of-function support a role for *NEGR1* in the control of body weight; however, the mechanism of its involvement is not clear (Lee *et al.* 2012). Contradictory with anticipated results, these mutant mice display a small but steady reduction of body mass. Notwithstanding, these studies do

■ **Table 6 Statistically Significant BMI PheWAS SNPs including DM2 as a comorbidity**

Phenotype	Description	SNP	Gene	β	(SE)	Odds Ratio	PheWAS p -value	N	Cases	Controls
250.2	Type 2 Diabetes	rs2246606	TDH	-0.331	0.073	0.718	6.32×10^{-6}	5566	459	5107
796	Elevated prostate specific antigen [PSA]	rs9937053	FTO	0.882	0.197	2.416	7.76×10^{-6}	6293	71	6222
796	Elevated prostate specific antigen [PSA]	rs9930333	FTO	0.881	0.197	2.413	7.88×10^{-6}	6297	71	6226
796	Elevated prostate specific antigen [PSA]	rs9940128	FTO	0.875	0.197	2.398	8.66×10^{-6}	6297	71	6226
796	Elevated prostate specific antigen [PSA]	rs1121980	FTO	0.872	0.197	2.391	9.60×10^{-6}	6298	71	6227
250	Diabetes Mellitus	rs2246606	TDH	-0.318	0.072	0.728	1.14×10^{-5}	5577	470	5107
796	Elevated prostate specific antigen [PSA]	rs1421085	FTO	0.855	0.197	2.351	1.41×10^{-5}	6298	71	6227
796	Elevated prostate specific antigen [PSA]	rs1558902	FTO	0.854	0.197	2.350	1.43×10^{-5}	6298	71	6227
278.1	Obesity	rs2948300	N/A	0.199	0.047	1.221	1.90×10^{-5}	5805	1172	4633
796	Elevated prostate specific antigen [PSA]	rs17817449	FTO	0.823	0.196	2.277	2.75×10^{-5}	6298	71	6227
796	Elevated prostate specific antigen [PSA]	rs8043757	FTO	0.822	0.196	2.275	2.82×10^{-5}	6298	71	6227
250.2	Type 2 Diabetes	rs2572386	FAM167A	-0.303	0.073	0.739	3.57×10^{-5}	5564	459	5105
278	Obesity, Overweight	rs2948300	N/A	0.182	0.044	1.200	3.71×10^{-5}	5980	1347	4633
796	Elevated prostate specific antigen [PSA]	rs8050136	FTO	0.788	0.195	2.198	5.55×10^{-5}	6298	71	6227
796	Elevated prostate specific antigen [PSA]	rs9939609	FTO	0.787	0.196	2.197	5.71×10^{-5}	6298	71	6227
250.2	Type 2 Diabetes	rs1435277	N/A	-0.292	0.073	0.747	5.94×10^{-5}	5554	458	5096
401	Hypertension	rs12412241	N/A	-0.221	0.055	0.802	6.49×10^{-5}	6057	1385	4672
796	Elevated prostate specific antigen [PSA]	rs3751812	FTO	0.772	0.195	2.163	7.87×10^{-5}	6297	71	6226
250	Diabetes Mellitus	rs2572386	FAM167	-0.284	0.072	0.753	8.44×10^{-5}	5575	470	5105
401.1	Essential Hypertension	rs12412241	N/A	-0.216	0.056	0.806	1.02×10^{-4}	6037	1365	4672

suggest that loss of *NEGR1* function in the mouse models has a negative effect on body mass as well as lean mass, supporting the possibility that *NEGR1* may contribute to a change body mass. It is also important to note that animal models are a representation of human physiology but not necessarily a precise depiction.

Our extreme obesity vs. non-obese study also identified two new genes, *PFKFB3* and *KLRB1*, that are not yet found to be significantly associated to BMI. The odds ratios associated with all SNPs found in these genes were less than one, indicating that they decreased the odds of extreme obesity risk. These results are potentially supported by work conducted by Huo *et al.*, who reported that mice transgenically modified to selectively overexpress *PFKFB3* in adipocytes show increases fat deposition in their adipose tissue (Huo *et al.* 2012). In contrast, an earlier study reported that transgenic mice with reduced *PFKFB3* expression show exacerbated diet-induced insulin resistance (Huo *et al.* 2010). Kerr *et al.* showed that inhibition of *PFKFB3* mRNA impairs basal and insulin stimulated lipogenesis and also proposed that decreasing the expression of *PFKFB3* may inhibit adipocyte lipid storage (Kerr *et al.* 2019). This would support a hypothesis that polymorphisms leading to a decrease in *PFKFB3* expression may be protective from the development of obesity; however, tissue-specific transcriptional studies in humans would be required to fully support this assertion.

We also observed polymorphisms in *KLRB1* correlate with a decrease of odds in extreme obesity risk. *KLRB1* expression produces a type II transmembrane glycoprotein also known as CD161; a member of the C-type lectin superfamily. CD161 is expressed on the surface of most natural killer (NK) cells and natural killer T (NKT) but also on subsets of peripheral T cells and CD3⁺ thymocytes. While the biological

function of CD161 is not firmly established, it was suggested that it serves either as a stimulatory receptor or to inhibit NK cell-mediated cytotoxicity and cytokine production (Lanier *et al.* 1994). Indeed, NK cells were shown to be upregulated in the fat of obese twins (Pietiläinen *et al.* 2008); moreover, BMI and *KLRB1* expression may be correlated in that *KLRB1* transcription has been reported to increase as BMI increases (Rai *et al.* 2014). Additionally, CD161^{bright} CD8⁺ mucosal associated invariant T (MAIT) cells play a central role in maintaining mucosal immunity and therefore, changes in CD161 expression on these cells may lead to alterations in mucosal immunity and gut microbiota homeostasis. These changes may in turn manifest as alterations of dietary metabolism. It is noteworthy that increases in MAIT cells are associated with Juvenile Type 1 Diabetes and polymorphisms in *KLRB1* have been associated with ischemic heart disease (Makeeva *et al.* 2015), and differential transcription of *KLRB1* has been reported in DM2 and coronary artery disease (Gong *et al.* 2017). Furthermore, another gene, *GRIN2A*, that is part of the family of genes {*GRIN1*, *GRIN2A*, *GRIN2B*, *GRIN2C*, *GRIN2D*, *GRIN3A*, and *GRIN3B*}, which encode proteins that form a receptor in charge of sending chemical messages between neurons in the brain, was found to be associated to obesity in adult women defined as metabolically healthy in Schlauch *et al.* ($p=1.7 \times 10^{-5}$). (Schlauch *et al.* 2018).

Associations between *FTO* and obesity were just under genome-wide significance levels. This is a possible indication that *FTO* polymorphisms cause small changes in BMI, rather than the wide range differences observed between extreme obese cases and controls. Nonetheless, previous research has demonstrated that a combination of several *FTO* mutations will increase the likelihood of a participant being

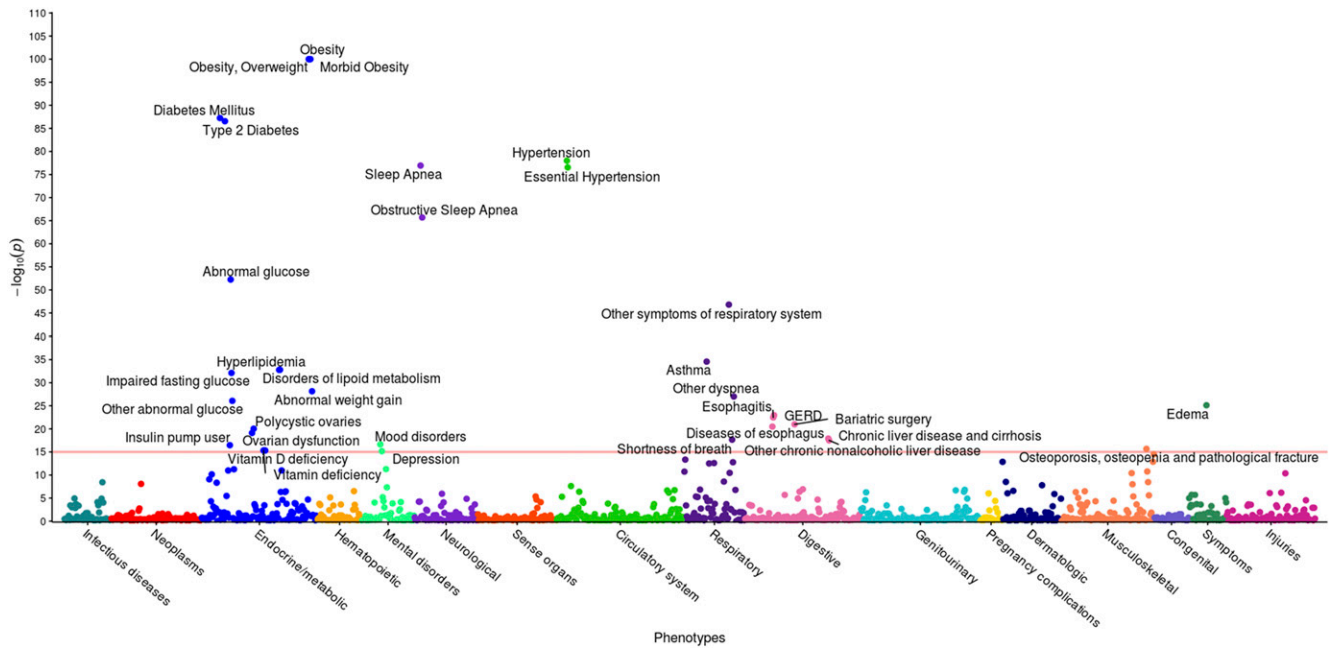


Figure 4 PheWAS results between BMI and EHR Phenotypes. This figure illustrates the results of individual linear regression between incidence of phenotype groups (phecodes) and the continuous BMI metric of all 6,645 individuals. Each of the 301 points represents the p -value of the association between one of 1,523 phecodes with at least 20 cases assigned to it, and BMI. Statistical significance was assessed by using the False Discovery Rate of 0.1, corresponding to a raw p -value of 1.96×10^{-2} . Only associations with $p < 1 \times 10^{-15}$ are annotated for ease of viewing, represented by a horizontal line at 15 on the y -axis.

classified as obese (Li *et al.* 2009). Speakman *et al.* further stated that Frayling showed the *FTO* was significantly associated with diabetes only through its association with BMI (Speakman *et al.* 2018).

The comprehensive series of GWASs presented here validates associations of obesity and BMI found in previous studies, such as the *FTO* and *NEGR1* loci (Willer *et al.* 2009; Speliotes *et al.* 2010; Okada

et al. 2012; Locke *et al.* 2015). Many larger studies identify associative loci in *MC4R* (Willer *et al.* 2009; Speliotes *et al.* 2010; Okada *et al.* 2012). While our studies did not detect SNPs in *MC4R* with genome-wide significance, they did identify associations at $p=1 \times 10^{-4}$ and $p=1.7 \times 10^{-4}$ of SNPs rs17782313 and rs571312 (Willer *et al.* 2009; Speliotes *et al.* 2010). A number of obesity case-control studies have found variations

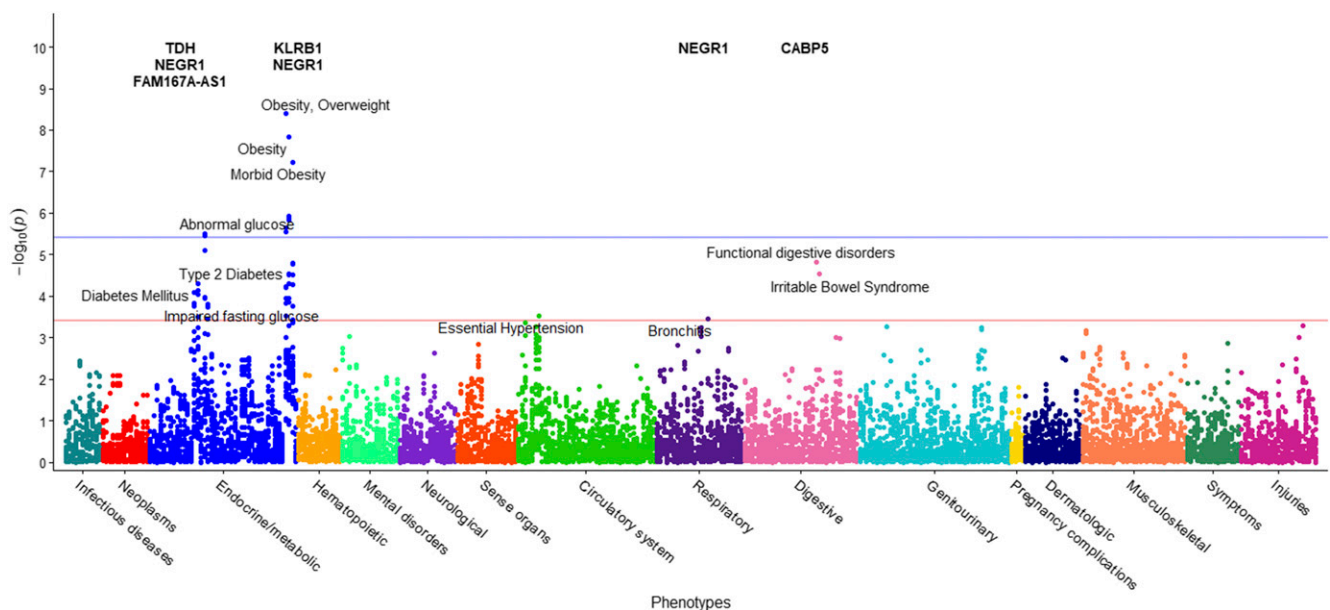


Figure 5 PheWAS results between obesity-significant SNPs and EHR Phenotypes. This figure presents results of individual logistic regressions between incidence of 372 phenotype groups (phecodes) and the genotypes of 34 SNPs found to be associated with extreme obesity. Each point represents the p -value of one SNP and one of 372 phecodes with at least 20 cases assigned to it. The horizontal red line represents the significance level $p=3.85 \times 10^{-4}$, and the blue line represents the Bonferroni correction of $p=4 \times 10^{-6}$.

■ Table 7 Statistically Significant Obesity PheWAS SNPs

Phenotype	Description	SNP	Gene	β	(SE)	Odds Ratio	PheWAS p -value	N	Cases	Controls
278	Overweight, Obesity	rs1620977	NEGR1	0.409	0.070	1.506	4.09x10 ⁻⁹	2819	640	2179
278.1	Obesity	rs1620977	NEGR1	0.405	0.071	1.499	1.45x10 ⁻⁸	2779	600	2179
278.11	Morbid Obesity	rs1620977	NEGR1	0.432	0.080	1.541	5.96x10 ⁻⁸	2624	445	2179
278.1	Obesity	rs1776012	NEGR1	-0.326	0.067	0.722	1.18x10 ⁻⁶	2762	594	2168
278.1	Obesity	rs1870676	NEGR1	-0.324	0.067	0.723	1.40x10 ⁻⁶	2757	593	2164
278.1	Obesity	rs9424977	NEGR1	-0.321	0.067	0.725	1.46x10 ⁻⁶	2776	599	2177
278	Overweight, Obesity	rs1776012	NEGR1	-0.308	0.065	0.735	2.23x10 ⁻⁶	2802	634	2168
278	Overweight, Obesity	rs1870676	NEGR1	-0.305	0.065	0.737	2.84x10 ⁻⁶	2797	633	2164
278	Overweight, Obesity	rs9424977	NEGR1	-0.303	0.065	0.739	2.87x10 ⁻⁶	2816	639	2177
250.4	Abnormal glucose	rs1776012	NEGR1	-0.444	0.095	0.642	3.20x10 ⁻⁶	2664	272	2392
250.4	Abnormal glucose	rs1870676	NEGR1	-0.444	0.096	0.642	3.58x10 ⁻⁶	2655	270	2385
250.4	Abnormal glucose	rs9424977	NEGR1	-0.424	0.095	0.655	7.81x10 ⁻⁶	2674	273	2401
564	Functional digestive disorders	rs750456	CABP5	0.989	0.229	2.689	1.55x10 ⁻⁵	2518	40	2478
278.11	Morbid Obesity	rs2241005	KLRB1	-0.321	0.075	0.725	1.64x10 ⁻⁵	2625	445	2180
278.11	Morbid Obesity	rs2058426	N/A	-0.322	0.075	0.725	1.66x10 ⁻⁵	2621	445	2176
278.1	Obesity	rs2568958	N/A	-0.294	0.070	0.745	2.89x10 ⁻⁵	2781	601	2180
278.1	Obesity	rs2815752	N/A	-0.294	0.070	0.745	2.89x10 ⁻⁵	2781	601	2180
564.1	Irritable Bowel Syndrome	rs750456	CABP5	0.992	0.238	2.697	2.98x10 ⁻⁵	2515	37	2478
278.11	Morbid Obesity	rs1870676	NEGR1	-0.315	0.075	0.730	3.06x10 ⁻⁵	2605	441	2164
278.1	Obesity	rs3101336	N/A	-0.293	0.070	0.746	3.08x10 ⁻⁵	2781	601	2180
278.11	Morbid Obesity	rs1776012	NEGR1	-0.315	0.076	0.730	3.14x10 ⁻⁵	2606	438	2168
250.2	Type 2 Diabetes	rs2246606	TDH	-0.468	0.115	0.626	4.98x10 ⁻⁵	2591	187	2404
278.1	Obesity	rs2953802	N/A	0.266	0.066	1.305	5.13x10 ⁻⁵	2781	601	2180
278.11	Morbid Obesity	rs9424977	NEGR1	-0.304	0.075	0.738	5.21x10 ⁻⁵	2620	443	2177
278	Overweight, Obesity	rs2568958	N/A	-0.274	0.068	0.760	5.88x10 ⁻⁵	2821	641	2180
278	Overweight, Obesity	rs2815752	N/A	-0.274	0.068	0.760	5.88x10 ⁻⁵	2821	641	2180
278	Overweight, Obesity	rs3101336	N/A	-0.273	0.068	0.761	6.27x10 ⁻⁵	2821	641	2180
250.2	Type 2 Diabetes	rs2736280	TDH	-0.452	0.114	0.636	7.48x10 ⁻⁵	2588	187	2401
250	Diabetes Mellitus	rs2246606	TDH	-0.447	0.114	0.639	8.11x10 ⁻⁵	2596	192	2404
250.2	Type 2 Diabetes	rs2953802	N/A	0.431	0.110	1.539	8.90x10 ⁻⁵	2591	187	2404
250.4	Abnormal glucose	rs2568958	N/A	-0.389	0.100	0.678	1.08x10 ⁻⁴	2678	274	2404
250.4	Abnormal glucose	rs2815752	N/A	-0.389	0.100	0.678	1.08x10 ⁻⁴	2678	274	2404
278	Overweight, Obesity	rs2241005	KLRB1	-0.247	0.064	0.781	1.11x10 ⁻⁴	2821	641	2180
250.4	Abnormal glucose	rs3101336	N/A	-0.388	0.101	0.678	1.12x10 ⁻⁴	2678	274	2404
278.1	Obesity	rs2241005	KLRB1	-0.253	0.066	0.776	1.14x10 ⁻⁴	2781	601	2180
278.1	Obesity	rs2058426	N/A	-0.250	0.066	0.779	1.43x10 ⁻⁴	2777	601	2176
278	Overweight, Obesity	rs2058426	N/A	-0.243	0.064	0.784	1.46x10 ⁻⁴	2817	641	2176
250	Diabetes Mellitus	rs2736280	TDH	-0.426	0.112	0.653	1.48x10 ⁻⁴	2593	192	2401
250.41	Impaired fasting glucose	rs1870676	NEGR1	-0.471	0.125	0.625	1.59x10 ⁻⁴	2535	150	2385
278.11	Morbid Obesity	rs2953802	N/A	0.278	0.074	1.321	1.76x10 ⁻⁴	2625	445	2180
250	Diabetes Mellitus	rs2953802	N/A	0.407	0.109	1.502	1.77x10 ⁻⁴	2596	192	2404
250.41	Impaired fasting glucose	rs1776012	NEGR1	-0.464	0.124	0.629	1.80x10 ⁻⁴	2543	151	2392
401.1	Essential Hypertension	rs2948300	N/A	0.282	0.078	1.326	3.02x10 ⁻⁴	2737	521	2216
278	Overweight, Obesity	rs2953802	N/A	0.231	0.064	1.260	3.06x10 ⁻⁴	2821	641	2180
250.2	Type 2 Diabetes	rs2572386	FAM167A	-0.415	0.115	0.660	3.15x10 ⁻⁴	2590	187	2403
497	Bronchitis	rs1620977	NEGR1	0.673	0.188	1.960	3.47x10 ⁻⁴	2118	60	2058
250.41	Impaired fasting glucose	rs9424977	NEGR1	-0.441	0.124	0.644	3.60x10 ⁻⁴	2552	151	2401
278.11	Morbid Obesity	rs2568958	N/A	-0.283	0.079	0.753	3.66x10 ⁻⁴	2625	445	2180
278.11	Morbid Obesity	rs2815752	N/A	-0.283	0.079	0.753	3.66x10 ⁻⁴	2625	445	2180
278.11	Morbid Obesity	rs3101336	N/A	-0.282	0.080	0.754	3.85x10 ⁻⁴	2625	445	2180

in the *MC4R* gene (Xi *et al.* 2012; Evans *et al.* 2014). Our case-control study does reveal that rs17782313 in *MC4R* associates with obesity at $p=3 \times 10^{-4}$. Our cohort is a controlled, regional population. The next two stages of the Healthy Nevada Project will add between 40,000 (2019) and 150,000 (late 2020) more Nevadans to the current cohort. With these much larger cohort sizes, it is our hope that a stronger associate link with *MC4R* will be identified.

PheWAS of healthy Nevada BMI and obesity

To the best of our knowledge, this is the first dual-PheWAS targeted at BMI and obesity. Cronin *et al.* present a comprehensive PheWAS

targeted at *FTO* variants, which also show strong associations with overweight and obesity phenotypes, hypertension and hyperlipidemia (Cronin *et al.* 2014). Milliard *et al.* perform a large PheWAS study to examine phenotypic associations with BMI that focus on the nervousness phenotypes: the study identified known associations such as diabetes and hypertension (Milliard *et al.* 2019).

The PheWAS performed on the SNP associations in this study's BMI cohort identified strong associations of elevated PSA levels with variants in *FTO*, and indicates that the number of minor alleles of these variants is predictive of elevated PSA. This finding is in contradiction to the reports (Bañez *et al.* 2007; Oh *et al.* 2013; Zhang *et al.* 2016; Bonn

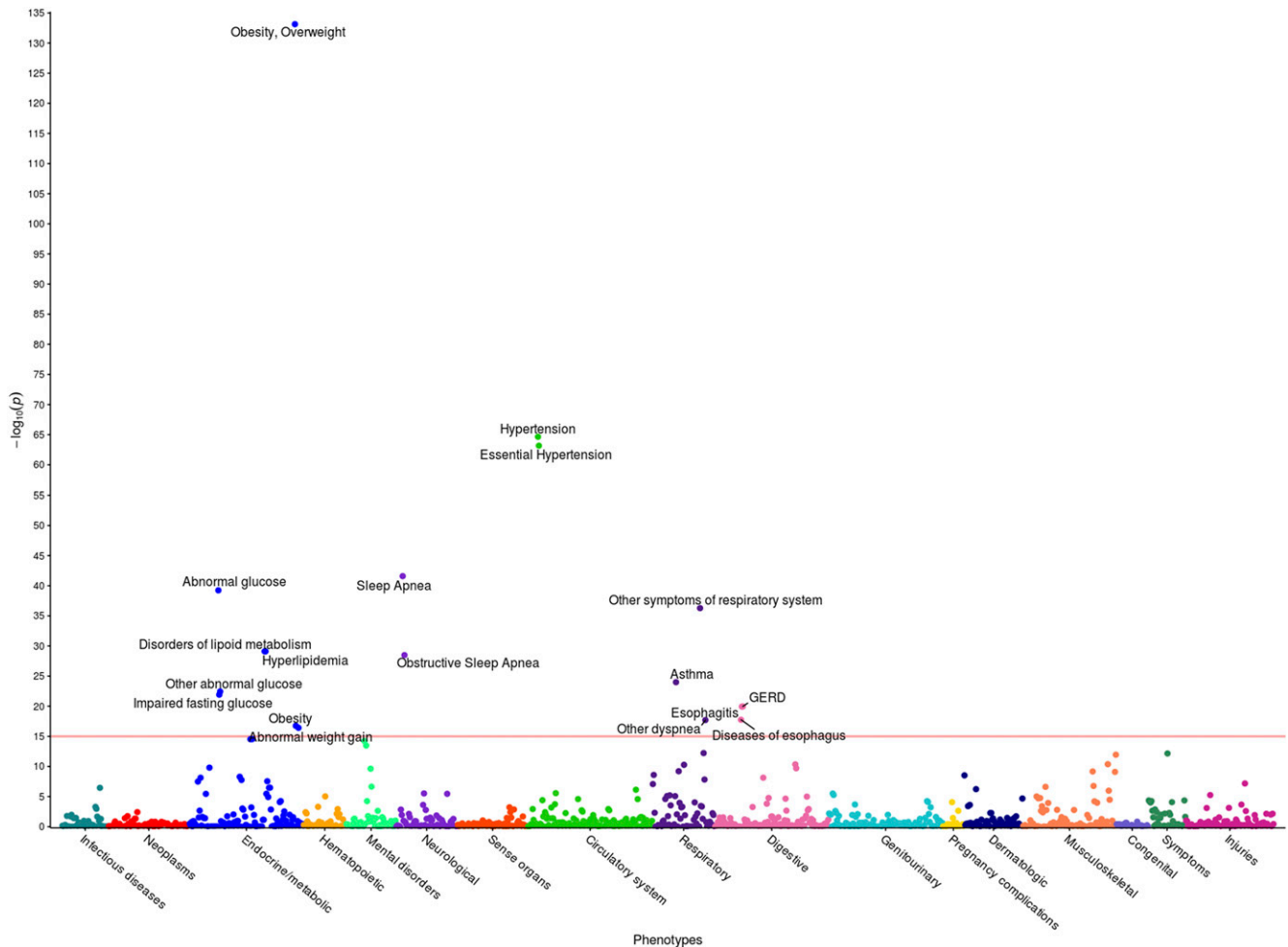


Figure 6 PheWAS results between extreme obesity and EHR Phenotypes. This figure illustrates the results of individual linear regression between incidence of phenotype groups (phecodes) and the incidence of extreme obesity in 2,996 individuals. Each of the 191 points represents the p -value of the association between one of 1,362 phecodes with at least 20 cases assigned to it, and extreme obesity. Statistical significance was assessed by using the False Discovery Rate of 0.1, corresponding to a raw p -value of 1.4×10^{-2} . Only associations with $p < 1 \times 10^{-15}$ are annotated for ease of viewing, represented by a horizontal line at 15 on the y -axis.

et al. 2016) indicating an inverse relationship between PSA levels and BMI. However, serum levels of PSA may be elevated due to reasons other than prostatic malignancy. Benign prostatic hyperplasia (BPH), prostatitis (Nadler *et al.* 1995), ejaculation (Herschman *et al.* 1997), or manipulation of the prostate gland (Chybowski *et al.* 1992; Crawford *et al.* 1992; Tarhan *et al.* 2005) may cause elevated levels of serum PSA. Our study did not control for such parameters. Our sample includes only 71 individuals with ICD codes indicating high PSA levels, of which a number are morbidly obese. Increased BMI is often associated with increased age and our study population was significantly older than the general median age of the U.S. population. Thus, it is possible that older age contributed to increased likelihood of BPH and, subsequently, elevated serum PSA, negating the reported inverse effect of BMI on PSA levels and possibly exposing a novel association with variants in *FTO*.

Many of the clinical associations observed in the PheWASs of the HNP in relation to various degrees of increased BMI and the presence of obesity are part of the cluster of clinical conditions associated with metabolic syndrome (Alberti *et al.* 2009). Obesity is a risk factor for respiratory conditions such as chronic obstructive pulmonary disease (COPD), asthma, obstructive sleep apnea and obesity hypoventilation

syndrome, and may influence the development and presentation of these diseases (Poulain *et al.* 2006). Accumulation of fat tissue impairs ventilatory function in adults (Lazarus *et al.* 1997) and increased BMI is associated with a reduction in forced expiratory volume in one second (FEV1), forced vital capacity (FVC), total lung capacity, functional residual capacity and expiratory reserve volume (Rubinstein *et al.* 1990; Chinn *et al.* 1996; Lazarus *et al.* 1997; Biring *et al.* 1999). Peripheral edema has long been recognized as associated with extreme obesity (Alexander *et al.* 1962). In the U.K. Community Nursing Services study, obesity was found as an independent risk factor for chronic edema (Moffatt *et al.* 2019).

To the best of our knowledge, this work was the first to conduct simultaneous and comprehensive genome-wide and phenome-wide association studies to validate genetic associations of BMI and severe class 2 obesity. While GWAS/PheWAS studies have been conducted in the past, the present study leveraged a unique EHR database to identify 26 SNPs across six chromosomes associated with extreme obesity. Moreover, while approximately 70% were associated with obesity and BMI in prior studies, 30% are novel with respect to these metrics, suggesting that they may be of special significance. For example, our unique approach allowed us to identify a novel polymorphism associated

with extreme obesity in a previously uncharacterized pseudogene that is primarily expressed in the pancreas. Given that the pancreas serves an essential role in converting food into fuel and regulating blood sugar, future studies may support a critical role for these SNPs in extreme obesity cases and may suggest targeted treatment strategies.

ACKNOWLEDGMENTS

We thank Michele Henderson, Toni Curreri and all the ambassadors of the Healthy Nevada Project. We thank Renown Health, DRI marketing and all the folks at 23andMe who helped launch the project. Research support was provided by the Governor's Office of Economic Development Knowledge Fund. Support for the Healthy Nevada Project and personal genetics was provided by the Renown Health Foundation.

Members of the 23andMe Research Team are: Michelle Agee, Adam Auton, Robert K. Bell, Robert Borkowski, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, David A. Hinds, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Matthew H. McIntyre, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, and Catherine H. Wilson.

LITERATURE CITED

- Alberti, K. G. M. M., R. H. Eckel, S. M. Grundy, P. Z. Zimmet, J. I. Cleeman *et al.*, 2009 Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* 1640–1645. <https://doi.org/10.1161/CIRCULATIONAHA.109.192644>
- Alexander, J. K., K. H. Amad, and V. W. Cole, 1962 Observations on some clinical features of extreme obesity, with particular reference to cardio-respiratory effects. *Am. J. Med.* 32: 512–524. [https://doi.org/10.1016/0002-9343\(62\)90052-9](https://doi.org/10.1016/0002-9343(62)90052-9)
- An, Y., K. L. Furber, and S. Ji, 2016 Pseudogenes regulate parental gene expression via RNA network. *J. Cell. Mol. Med.* 21: 185–192. <https://doi.org/10.1111/jcmm.12952>
- Anderson, C. A., F. H. Pettersson, G. M. Clarke, L. R. Cardon, A. P. Morris *et al.*, 2010 Data quality control in genetic case-control association studies. *Nat. Protoc.* 5: 1564–1573. <https://doi.org/10.1038/nprot.2010.116>
- Bañez, L. L., R. J. Hamilton, A. W. Partin, R. T. Vollmer, L. Sun *et al.*, 2007 Obesity-Related Plasma Hemodilution and PSA Concentration Among Men With Prostate Cancer. *JAMA* 298: 2275–2280. <https://doi.org/10.1001/jama.298.19.2275>
- Bays, H. E., R. H. Chapman, and S. Grandy, SHIELD Investigators' Group, 2007 The relationship of body mass index to diabetes mellitus, hypertension and dyslipidaemia: comparison of data from two national surveys. *Int. J. Clin. Pract.* 61: 737–747. <https://doi.org/10.1111/j.1742-1241.2007.01336.x>
- Benjamini, Y., and Y. Hochberg, 1995 Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57: 289–300.
- Berndt, S. I., S. Gustafsson, R. Mägi, A. Ganna, E. Wheeler *et al.*, 2013 Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat. Genet.* 45: 501–512. <https://doi.org/10.1038/ng.2606>
- Biring, M. S., M. I. Lewis, J. T. Liu, and Z. Mohsenifar, 1999 Pulmonary physiologic changes of morbid obesity. *Am. J. Med. Sci.* 318: 293–297. [https://doi.org/10.1016/S0002-9629\(15\)40641-X](https://doi.org/10.1016/S0002-9629(15)40641-X)
- Blanca, M. J., R. Alarcón, J. Arnau, R. Bono, and R. Bendayan, 2017 Non-normal data: Is ANOVA still a valid option? *Psicothema* 29: 552–557.
- Boender, A. J., M. A. van Gestel, K. M. Garner, M. C. M. Luijendijk, and R. A. H. Adan, 2014 The obesity-associated gene *Negr1* regulates aspects of energy balance in rat hypothalamic areas. *Physiol. Rep.* 2: pii: e12083. <https://doi.org/10.14814/phy2.12083>
- Bonn, S. E., A. Sjölander, A. Tillander, F. Wiklund, H. Grönberg *et al.*, 2016 Body mass index in relation to serum prostate-specific antigen levels and prostate cancer risk. *Int. J. Cancer* 139: 50–57. <https://doi.org/10.1002/ijc.30052>
- Carroll, R. J., L. Bastarache, and J. C. Denny, 2014 R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* 30: 2375–2376. <https://doi.org/10.1093/bioinformatics/btu197>
- Cha, S. W., S. M. Choi, K. S. Kim, B. L. Park, J. R. Kim *et al.*, 2012 Replication of Genetic Effects of FTOPolymorphisms on BMI in a Korean Population. *Obesity (Silver Spring)* 16: 2187–2189. <https://doi.org/10.1038/oby.2008.314>
- Chen, S., W. Wu, J. Li, Q. Wang, Y. Li *et al.*, 2015 Single nucleotide polymorphisms in the FAM167A-BLK gene are associated with polymyositis/dermatomyositis in the Han Chinese population. *Immunol. Res.* 62: 153–162. <https://doi.org/10.1007/s12026-015-8646-0>
- Chinn, D. J., J. E. Cotes, and J. W. Reed, 1996 Longitudinal effects of change in body mass on measurements of ventilatory capacity. *Thorax* 51: 699–704. <https://doi.org/10.1136/thx.51.7.699>
- Christensen, G. L., M. L. B. Jacobsen, A. Wendt, I. G. Mollet, J. Friberg *et al.*, 2015 Bone morphogenetic protein 4 inhibits insulin secretion from rodent beta cells through regulation of calbindin1 expression and reduced voltage-dependent calcium currents. *Diabetologia* 58: 1282–1290. <https://doi.org/10.1007/s00125-015-3568-x>
- Chybowski, F. M., E. J. Bergstralh, and J. E. Oesterling, 1992 The Effect of Digital Rectal Examination on the Serum Prostate Specific Antigen Concentration: Results of a Randomized Study. *J. Urol.* 148: 83–86. [https://doi.org/10.1016/S0022-5347\(17\)36517-5](https://doi.org/10.1016/S0022-5347(17)36517-5)
- Crawford, E. D., M. J. Schutz, S. Clejan, J. Drago, M. I. Resnick *et al.*, 1992 The Effect of Digital Rectal Examination on Prostate-Specific Antigen Levels. *JAMA* 267: 2227–2228. <https://doi.org/10.1001/jama.1992.03480160085039>
- Cronin, R. M., J. R. Field, Y. Bradford, C. M. Shaffer, R. J. Carroll *et al.*, 2014 Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. *Front. Genet.* 5: 1061. <https://doi.org/10.3389/fgene.2014.00250>
- De Bock, K., M. Geogiadous, S. Schoors, A. Kuchnio, B. W. Wong *et al.*, 2013 Role of PFKFB3-Driven Glycolysis in Vessel Sprouting. *Cell* 154: 651–663. <https://doi.org/10.1016/j.cell.2013.06.037>
- Denny, J. C., L. Bastarache, M. D. Ritchie, R. J. Carroll, R. Zink *et al.*, 2013 Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* 31: 1102–1111. <https://doi.org/10.1038/nbt.2749>
- Dina, C., D. Meyre, S. Gallina, E. Durand, A. Körner *et al.*, 2007 Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat. Genet.* 39: 724–726. <https://doi.org/10.1038/ng2048>
- Ehrlich, A. C., and F. K. Friedenberg, 2016 Genetic Associations of Obesity: The Fat-Mass and Obesity-Associated (FTO) Gene. *Clin. Transl. Gastroenterol.* 7: e140. <https://doi.org/10.1038/ctg.2016.1>
- Evans, D. S., M. A. Calton, M. J. Kim, P.-Y. Kwok, I. Miljkovic *et al.*, 2014 Genetic association study of adiposity and melanocortin-4 receptor (MC4R) common variants: replication and functional characterization of non-coding regions. *PLoS One* 9: e96805. <https://doi.org/10.1371/journal.pone.0096805>
- Fagerberg, L., B. M. Hallström, P. Oksvold, C. Kampf, D. Djureinovic *et al.*, 2014 Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* 13: 397–406. <https://doi.org/10.1074/mcp.M113.035600>
- Fawcett, K. A., and I. Barroso, 2010 The genetics of obesity: FTO leads the way. *Trends Genet.* 26: 266–274. <https://doi.org/10.1016/j.tig.2010.02.006>
- Frayling, T. M., N. J. Timpson, M. N. Weedon, E. Zeggini, R. M. Freathy *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. <https://doi.org/10.1126/science.1141634>
- Gauderman, W. J., 2002 Sample size requirements for association studies of gene-gene interaction. *Am. J. Epidemiol.* 155: 478–484. <https://doi.org/10.1093/aje/155.5.478>
- Gong, R., M. Chen, C. Zhang, M. Chen, and H. Li, 2017 A comparison of gene expression profiles in patients with coronary artery disease, type 2 diabetes, and their coexisting conditions. *Diagn. Pathol.* 12: 44. <https://doi.org/10.1186/s13000-017-0630-7>
- González-Herrera, L., J. Zavala-Castro, C. Ayala-Caceres, G. Pérez-Mendoza, M. J. López-González *et al.*, 2019 Genetic variation of FTO: rs1421085 T>C, rs8057044 G>A, rs9939609 T>A, and copy number (CNV) in Mexican Mayan school-aged children with obesity/overweight and with normal weight. *Am. J. Hum. Biol.* 31: e23192. <https://doi.org/10.1002/ajhb.23192>
- Gouveia, R., L. Schaffer, S. Papp, N. Grammel, S. Kandzia *et al.*, 2012 Expression of glycogenes in differentiating human NT2N neurons. Downregulation of fucosyltransferase 9 leads to decreased Lewis(x) levels and impaired neurite outgrowth. *Biochim. Biophys. Acta* 1820: 2007–2019. <https://doi.org/10.1016/j.bbagen.2012.09.004>
- Graff, M., J. S. Ngwa, T. Workalemahu, G. Homuth, S. Schipf *et al.*, 2013 Genome-wide analysis of BMI in adolescents and young adults reveals additional insight into the effects of genetic loci over the life course. *Hum. Mol. Genet.* 22: 3597–3607. <https://doi.org/10.1093/hmg/ddt205>
- Grant, S. F., M. Li, J. P. Bradfield, C. E. Kim, K. Annaiah *et al.*, 2008 Association Analysis of the FTO Gene with Obesity in Children of Caucasian and African Ancestry Reveals a Common Tagging SNP. *PLoS One* 3: e1746. <https://doi.org/10.1371/journal.pone.0001746>
- Grarup, N., C. H. Sandholt, T. Hansen, and O. Pedersen, 2014 Genetic susceptibility to type 2 diabetes and obesity: from genome-wide association studies to rare variants and beyond. *Diabetologia* 57: 1528–1541. <https://doi.org/10.1007/s00125-014-3270-4>
- Haeseleer, F., Y. Imanishi, I. Sokal, S. Filipek, and K. Palczewski, 2002 Calcium-Binding Proteins: Intracellular Sensors from the Calmodulin Superfamily. *Biochem. Biophys. Res. Commun.* 290: 615–623. <https://doi.org/10.1006/bbrc.2001.6228>
- Hashimoto, T., M. Yamada, S. Maekawa, T. Nakashima, and S. Miyata, 2008 IgLON cell adhesion molecule Kilon is a crucial modulator for synapse number in hippocampal neurons. *Brain Res.* 1224: 1–11. <https://doi.org/10.1016/j.brainres.2008.05.069>
- Herrera, B. M., and C. M. Lindgren, 2010 The genetics of obesity. *Curr. Diab. Rep.* 10: 498–505. <https://doi.org/10.1007/s11892-010-0153-z>
- Herschman, J. D., D. S. Smith, and W. J. Catalona, 1997 Effect of ejaculation on serum total and free prostate-specific antigen concentrations. *Urology* 50: 239–243. [https://doi.org/10.1016/S0090-4295\(97\)00209-4](https://doi.org/10.1016/S0090-4295(97)00209-4)
- Hinney, A., T. T. Nguyen, A. Scherag, S. Friedel, G. Brönner *et al.*, 2007 Genome Wide Association (GWA) Study for Early Onset Extreme Obesity Supports the Role of Fat Mass and Obesity Associated Gene (FTO) Variants. *PLoS One* 2: e1361. <https://doi.org/10.1371/journal.pone.0001361>
- Hotta, K., Y. Nakata, T. Matsuo, S. Kamohara, K. Kotani *et al.*, 2008 Variations in the FTO gene are associated with severe obesity in the Japanese. *J. Hum. Genet.* 53: 546–553. <https://doi.org/10.1007/s10038-008-0283-1>
- Hruby, A., and F. B. Hu, 2014 The Epidemiology of Obesity: A Big Picture. *Pharmacoeconomics* 33: 673–689. <https://doi.org/10.1007/s40273-014-0243-x>
- Hunt, S. C., S. Stone, Y. Xin, C. A. Scherer, C. L. Magness *et al.*, 2008 Association of the FTO Gene With BMI. *Obesity (Silver Spring)* 16: 902–904. <https://doi.org/10.1038/oby.2007.126>
- Huo, Y., X. Guo, H. Li, H. Wang, W. Zhang *et al.*, 2010 Disruption of Inducible 6-Phosphofructo-2-kinase Ameliorates Diet-induced Adiposity but Exacerbates Systemic Insulin Resistance and Adipose Tissue Inflammatory Response. *J. Biol. Chem.* 285: 3713–3721. <https://doi.org/10.1074/jbc.M109.058446>
- Huo, Y., X. Guo, H. Li, H. Xu, V. Halim *et al.*, 2012 Targeted overexpression of inducible 6-phosphofructo-2-kinase in adipose tissue increases fat deposition but protects against diet-induced insulin resistance and inflammatory responses. *J. Biol. Chem.* 287: 21492–21500. <https://doi.org/10.1074/jbc.M112.370379>
- Huđek, A., L. Škara, B. Smolkovič, S. Kazazić, S. Ravlić *et al.*, 2018 Higher prevalence of FTO gene risk genotypes AA rs9939609, CC rs1421085, and GG rs17817449 and saliva containing *Staphylococcus aureus* in obese women in Croatia. *Nutr. Res.* 50: 94–103. <https://doi.org/10.1016/j.nutres.2017.12.005>
- Jia, G., Y. Fu, X. Zhao, Q. Dai, G. Zheng *et al.*, 2011 N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* 7: 885–887. Erratum: 8: 1008. <https://doi.org/10.1038/nchembio.687>
- Jiao, H., M. Kaaman, E. Dungner, J. Kere, P. Arner *et al.*, 2008 Association analysis of positional obesity candidate genes based on integrated data from transcriptomics and linkage analysis. *Int. J. Obes.* 32: 816–825. <https://doi.org/10.1038/sj.ijo.0803789>
- Joe, B., Y. Saad, N. H. Lee, B. C. Frank, O. H. Achinike *et al.*, 2009 Positional identification of variants of *Adams16* linked to inherited hypertension. *Hum. Mol. Genet.* 18: 2825–2838. <https://doi.org/10.1093/hmg/ddp218>
- Justice, A. E., T. W. Winkler, M. F. Feitosa, M. Graff, V. A. Fisher *et al.*, 2017 Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat. Commun.* 8: 14977. <https://doi.org/10.1038/ncomms14977>
- Karaderi, T., A. W. Drong, and C. M. Lindgren, 2015 Insights into the Genetic Susceptibility to Type 2 Diabetes from Genome-Wide Association Studies of Obesity-Related Traits. *Curr. Diab. Rep.* 15: 83. <https://doi.org/10.1007/s11892-015-0648-8>
- Kerr, A. G., I. Sinha, S. Dadvar, P. Arner, and I. Dahlman, 2019 Epigenetic regulation of diabetogenic adipose morphology. *Mol. Metab.* 25: 159–167. <https://doi.org/10.1016/j.molmet.2019.04.009>
- Kopelman, P., 2007 Health risks associated with overweight and obesity. *Obes. Rev.* 8: 13–17. <https://doi.org/10.1111/j.1467-789X.2007.00311.x>
- Lalonde, E., K. C. H. Ha, Z. Wang, A. Bemmo, C. L. Kleinman *et al.*, 2011 RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. *Genome Res.* 21: 545–554. <https://doi.org/10.1101/gr.111211.110>
- Lanier, L. L., C. Chang, and J. H. Phillips, 1994 Human NKR-PIA. A disulfide-linked homodimer of the C-type lectin superfamily expressed by a subset of NK and T lymphocytes. *J. Immunol.* 153: 2417–2428.
- Lazarus, R., D. Sparrow, and S. T. Weiss, 1997 Effects of Obesity and Fat Distribution on Ventilatory Function: The Normative Aging Study. *Chest* 111: 891–898. <https://doi.org/10.1378/chest.111.4.891>
- Lee, A. W. S., H. Hengstler, K. Schwald, M. Berriel-Diaz, D. Loreth *et al.*, 2012 Functional inactivation of the genome-wide association study obesity gene neuronal growth regulator 1 in mice causes a body mass phenotype. *PLoS One* 7: e41537. <https://doi.org/10.1371/journal.pone.0041537>
- Li, S., J. H. Zhao, J. Luan, R. N. Luben, S. A. Rodwell *et al.*, 2009 Cumulative effects and predictive value of common obesity-susceptibility variants identified by genome-wide association studies. *Am. J. Clin. Nutr.* 91: 184–190. <https://doi.org/10.3945/ajcn.2009.28403>
- Liu, Y., Z. Liu, Y. Song, D. Zhou, D. Zhang *et al.*, 2010 Meta-analysis Added Power to Identify Variants in FTO Associated With Type 2 Diabetes and Obesity in the Asian Population. *Obesity (Silver Spring)* 18: 1619–1624. <https://doi.org/10.1038/oby.2009.469>
- Locke, A. E., B. Kahali, S. I. Berndt, A. E. Justice, T. H. Pers *et al.*, 2015 Genetic studies of body mass index yield new insights for obesity biology. *Nat. Biotechnol.* 518: 197–206.
- Loos, R. J. F., C. M. Lindgren, S. Li, E. Wheeler, J. H. Zhao *et al.*, 2008 Common variants near *MC4R* are associated with fat mass, weight and risk of obesity. *Nat. Genet.* 40: 768–775. <https://doi.org/10.1038/ng.140>
- Maes, H. H., M. C. Neale, and L. J. Eaves, 1997 Genetic and environmental factors in relative body weight and human adiposity. *Behav. Genet.* 27: 325–351. <https://doi.org/10.1023/A:1025635913927>
- Makeeva, O. A., A. A. Sleptsov, E. V. Kulish, O. L. Barbarash, A. M. Mazur *et al.*, 2015 Genomic Study of Cardiovascular Continuum Comorbidity.

- Acta Naturae 7: 89–99. <https://doi.org/10.32607/20758251-2015-7-3-89-99>
- Malis, C., E. L. Rasmussen, P. Poulsen, I. Petersen, K. Christensen *et al.*, 2005 Total and regional fat distribution is strongly influenced by genetic factors in young and elderly twins. *Obes. Res.* 13: 2139–2145. <https://doi.org/10.1038/oby.2005.265>
- Meyre, D., J. Delplanque, J.-C. Chèvre, C. Lecocq, S. Lobbens *et al.*, 2009 Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat. Genet.* 41: 157–159. <https://doi.org/10.1038/ng.301>
- Micu, I., Q. Jiang, E. Coderre, A. Ridsdale, L. Zhang *et al.*, 2006 NMDA receptors mediate calcium accumulation in myelin during chemical ischaemia. *Nature* 439: 988–992. <https://doi.org/10.1038/nature04474>
- Millard, L. A. C., N. M. Davies, K. Tilling, T. R. Gaunt, and G. Davey Smith, 2019 Searching for the causal effects of body mass index in over 300 000 participants in UK Biobank, using Mendelian randomization. *PLoS Genet.* 15: e1007951. <https://doi.org/10.1371/journal.pgen.1007951>
- Moffatt, C. J., R. Gaskin, M. Sykora, E. Dring, A. Aubeeluck *et al.*, 2019 Prevalence and Risk Factors for Chronic Edema in U.K. Community Nursing Services. *Lymphat. Res. Biol.* 17: 147–154. <https://doi.org/10.1089/lrb.2018.0086>
- Mori, Y., Y. Inoue, S. Tanaka, S. Doda, S. Yamanaka *et al.*, 2015 Cep169, a Novel Microtubule Plus-End-Tracking Centrosomal Protein, Binds to CDK5RAP2 and Regulates Microtubule Stability. *PLoS One* 10: e0140968. <https://doi.org/10.1371/journal.pone.0140968>
- Nadler, R. B., P. A. Humphrey, D. S. Smith, W. J. Catalona, and T. L. Ratliff, 1995 Effect of inflammation and benign prostatic hyperplasia on elevated serum prostate specific antigen levels. *J. Urol.* 154: 407–413. [https://doi.org/10.1016/S0022-5347\(01\)67064-2](https://doi.org/10.1016/S0022-5347(01)67064-2)
- Nakajima, S., V. Koh, L.-F. Kua, J. So, L. Davide *et al.*, 2016 Accumulation of CD11c+CD163+ Adipose Tissue Macrophages through Upregulation of Intracellular 11 β -HSD1 in Human Obesity. *J. Immunol.* 197: 3735–3745. <https://doi.org/10.4049/jimmunol.1600895>
- Nelson, D. L., and M. M. Cox, 2005 *Lehninger Principles of Biochemistry*, W. H. Freeman and Company, NY.
- Ogden, C. L., M. D. Carroll, L. R. Curtin, M. A. McDowell, C. J. Tabak *et al.*, 2006 Prevalence of overweight and obesity in the United States, 1999–2004. *JAMA* 295: 1549–1555. <https://doi.org/10.1001/jama.295.13.1549>
- Oh, J. J., S. J. Jeong, B. K. Lee, C. W. Jeong, S.-S. Byun *et al.*, 2013 Does obesity affect the accuracy of prostate-specific antigen (PSA) for predicting prostate cancer among men undergoing prostate biopsy. *BJU Int.* 112: E265–E271. <https://doi.org/10.1111/j.1464-410X.2012.11766.x>
- Okada, Y., M. Kubo, H. Ohmiya, A. Takahashi, N. Kumasaka *et al.*, 2012 Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations. *Nat. Genet.* 44: 302–306. <https://doi.org/10.1038/ng.1086>
- Olza, J., A. I. Ruperez, M. Gil-Campos, R. Leis, D. Fernandez-Orth *et al.*, 2013 Influence of FTO variants on obesity, inflammation and cardiovascular disease risk biomarkers in Spanish children: a case-control multicentre study. *BMC Med. Genet.* 14: 123. <https://doi.org/10.1186/1471-2350-14-123>
- Pietiläinen, K. H., J. Naukkarinen, A. Rissanen, J. Saharinen, P. Ellonen *et al.*, 2008 Global transcript profiles of fat in monozygotic twins discordant for BMI: pathways behind acquired obesity. *PLoS Med.* 5: e51. <https://doi.org/10.1371/journal.pmed.0050051>
- Pignot, V., A. E. Hein, C. Barske, C. Wiessner, A. R. Walmsley *et al.*, 2003 Characterization of two novel proteins, NgrH1 and NgrH2, structurally and biochemically homologous to the Nogo-66 receptor. *J. Neurochem.* 85: 717–728. <https://doi.org/10.1046/j.1471-4159.2003.01710.x>
- Pink, R. C., K. Wicks, D. P. Caley, E. K. Punch, L. Jacobs *et al.*, 2011 Pseudogenes: Pseudo-functional or key regulators in health and disease? *RNA* 17: 792–798. <https://doi.org/10.1261/rna.2658311>
- Poulain, M., M. Doucet, G. C. Major, V. Drapeau, F. Sériès *et al.*, 2006 The effect of obesity on chronic respiratory diseases: pathophysiology and therapeutic strategies. *CMAJ* 174: 1293–1299. <https://doi.org/10.1503/cmaj.051299>
- Prakash, J., N. Srivastava, S. Awasthi, C. G. Agarwal, S. M. Natu *et al.*, 2011 Association of FTO rs17817449 SNP with obesity and associated physiological parameters in a north Indian population. *Ann. Hum. Biol.* 38: 760–763. <https://doi.org/10.3109/03014460.2011.614278>
- Price, R. A., W.-D. Li, and H. Zhao, 2008 FTO gene SNPs associated with extreme obesity in cases, controls and extremely discordant sister pairs. *BMC Med. Genet.* 9: 4. <https://doi.org/10.1186/1471-2350-9-4>
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira *et al.*, 2007 PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 81: 559–575. <https://doi.org/10.1086/519795>
- Qureshi, S. A., A. Mumtaz, S. U. Shahid, and N. A. Shabana, 2017 rs3751812, a common variant in fat mass and obesity-associated (FTO) gene, is associated with serum high- and low-density lipoprotein cholesterol in Pakistani individuals. *Nutrition* 39-40: 92–95. <https://doi.org/10.1016/j.nut.2016.04.008>
- Rai, M. F., D. Patra, L. J. Sandell, and R. H. Brophy, 2014 Relationship of Gene Expression in the Injured Human Meniscus to Body Mass Index: A Biologic Connection Between Obesity and Osteoarthritis. *Arthritis Rheumatol.* 66: 2152–2164. <https://doi.org/10.1002/art.38643>
- Rother, S., J. Hundrieser, C. Pokoyski, S. Kollrich, K. Borns *et al.*, 2015 The c.503T>C Polymorphism in the Human KLRB1 Gene Alters Ligand Binding and Inhibitory Potential of CD161 Molecules. *PLoS One* 10: e0135682. <https://doi.org/10.1371/journal.pone.0135682>
- Rubinstein, I., N. Zamel, L. DuBarry, and V. Hoffstein, 1990 Airflow Limitation in Morbidly Obese, Nonsmoking Men. *AIM* 112: 828–832.
- Schlauch, K. A., S. F. Khaiboullina, K. L. De Meirleir, S. Rawat, J. Petereit *et al.*, 2016 Genome-wide association analysis identifies genetic variations in subjects with myalgic encephalomyelitis/chronic fatigue syndrome. *Transl. Psychiatry* 6: e730. <https://doi.org/10.1038/tp.2015.208>
- Schlauch, K. A., D. Kulick, K. Subramanian, K. L. De Meirleir, A. Palotás *et al.*, 2018 Single-nucleotide polymorphisms in a cohort of significantly obese women without cardiometabolic diseases. *Int. J. Obes.* 106: 1656.
- Scuteri, A., S. Sanna, W.-M. Chen, M. Uda, G. Albai *et al.*, 2007 Genome-Wide Association Scan Shows Genetic Variants in the FTO Gene Are Associated with Obesity-Related Traits. *PLoS Genet.* 3: e115. <https://doi.org/10.1371/journal.pgen.0030115>
- Shimaoka, I., K. Kamide, M. Ohishi, T. Katsuya, H. Akasaka *et al.*, 2010 Association of gene polymorphism of the fat-mass and obesity-associated gene with insulin resistance in Japanese. *Hypertens. Res.* 33: 214–218. <https://doi.org/10.1038/hr.2009.215>
- Speakman, J. R., R. J. F. Loos, S. O’Rahilly, J. N. Hirschhorn, and D. B. Allison, 2018 GWAS for BMI: a treasure trove of fundamental insights into the genetic basis of obesity. *Int. J. Obes.* 42: 1524–1531. <https://doi.org/10.1038/s41366-018-0147-5>
- Splietes, E. K., C. J. Willer, S. I. Berndt, K. L. Monda, G. Thorleifsson *et al.*, 2010 Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* 42: 937–948. <https://doi.org/10.1038/ng.686>
- Stunkard, A. J., T. T. Foch, and Z. Hrubec, 1986 A Twin Study of Human Obesity. *JAMA* 256: 51–54. <https://doi.org/10.1001/jama.1986.03380010055024>
- Stunkard, A. J., J. R. Harris, N. L. Pedersen, and G. E. McClearn, 1990 The body-mass index of twins who have been reared apart. *N. Engl. J. Med.* 322: 1483–1487. <https://doi.org/10.1056/NEJM199005243222102>
- Surridge, A. K., U. R. Rodgers, T. E. Swingler, R. K. Davidson, L. Kevorkian *et al.*, 2009 Characterization and regulation of ADAMTS-16. *Matrix Biol.* 28: 416–424. <https://doi.org/10.1016/j.matbio.2009.07.001>
- Tan, J. T., R. Dorajoo, M. Seielstad, X. L. Sim, R. T.-H. Ong *et al.*, 2008 FTO variants are associated with obesity in the Chinese and Malay populations in Singapore. *Diabetes* 57: 2851–2857. <https://doi.org/10.2337/db08-0214>
- Tarhan, F., A. Orçun, İ. Küçükercan, N. Çamursoy, and U. Kuyumcuoğlu, 2005 Effect of prostatic massage on serum complexed prostate-specific antigen levels. *Urology* 66: 1234–1238. <https://doi.org/10.1016/j.urology.2005.06.077>
- Thomsen, S. K., A. Ceroni, M. van de Bunt, C. Burrows, A. Barrett *et al.*, 2016 Systematic Functional Characterization of Candidate Causal

- Genes for Type 2 Diabetes Risk Variants. *Diabetes* 65: 3805–3811. <https://doi.org/10.2337/db16-0361>
- Thorleifsson, G., G. B. Walters, D. F. Gudbjartsson, V. Steinthorsdottir, P. Sulem *et al.*, 2008 Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* 41: 18–24. <https://doi.org/10.1038/ng.274>
- Verma, A., A. O. Basile, Y. Bradford, H. Kuivaniemi, G. Tromp *et al.*, 2016 Phenome-Wide Association Study to Explore Relationships between Immune System Related Genetic Loci and Complex Traits and Diseases. *PLoS One* 11: e0160573. <https://doi.org/10.1371/journal.pone.0160573>
- Verma, A., A. Lucas, S. S. Verma, Y. Zhang, N. Josyula *et al.*, 2018 PheWAS and Beyond: The Landscape of Associations with Medical Diagnoses and Clinical Measures across 38,662 Individuals from Geisinger. *Am. J. Hum. Genet.* 102: 592–608. <https://doi.org/10.1016/j.ajhg.2018.02.017>
- Villalobos-Comparán, M., M. Teresa Flores-Dorantes, M. Teresa Villarreal-Molina, M. Rodríguez-Cruz, A. C. García-Ulloa *et al.*, 2008 The FTO gene is associated with adulthood obesity in the Mexican population. *Obesity (Silver Spring)* 16: 2296–2301. <https://doi.org/10.1038/oby.2008.367>
- Wang, K., S. J. Diskin, H. Zhang, E. F. Attiyeh, C. Winter *et al.*, 2010 Integrative genomics identifies LMO1 as a neuroblastoma oncogene. *Nature* 469: 216–220. <https://doi.org/10.1038/nature09609>
- Wang, K., W.-D. Li, C. K. Zhang, Z. Wang, J. T. Glessner *et al.*, 2011 A Genome-Wide Association Study on Obesity and Obesity-Related Traits. *PLoS One* 6: e18939. Erratum: 2012 7:2. <https://doi.org/10.1371/journal.pone.0018939>
- Warren, M., S. Beck, and J. Rayburn, 2018 The State of Obesity: Better Policies for a Healthier America. 1–68, Trust for America's Health and Princeton, NJ: Robert Wood Johnson Foundation, Washington, DC. <https://www.tfah.org/report-details/the-state-of-obesity-2018/>
- Wei, J., F. Liu, Z. Lu, Q. Fei, Y. Ai *et al.*, 2018 Differential m6A, m6Am, and m1A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. *Mol. Cell* 71: 973–985.e5. <https://doi.org/10.1016/j.molcel.2018.08.011>
- West, N. R., J. Dorling, A. E. Thackray, N. C. Hanson, S. E. Decombel *et al.*, 2018 Effect of Obesity-Linked FTO rs9939609 Variant on Physical Activity and Dietary Patterns in Physically Active Men and Women. *J. Obes.* 2018: 7560707. <https://doi.org/10.1155/2018/7560707>
- Wheeler, E., N. Huang, E. G. Bochukova, J. M. Keogh, S. Lindsay *et al.*, 2013 Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. *Nat. Genet.* 45: 513–517. <https://doi.org/10.1038/ng.2607>
- Willer, C. J., E. K. Speliotes, R. J. F. Loos, S. Li, C. M. Lindgren *et al.*, 2009 Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* 41: 25–34. <https://doi.org/10.1038/ng.287>
- Wing, M. R., J. Ziegler, C. D. Langefeld, M. C. Y. Ng, S. M. Haffner *et al.*, 2009 Analysis of FTO gene variants with measures of obesity and glucose homeostasis in the IRAS Family Study. *Hum. Genet.* 125: 615–626. <https://doi.org/10.1007/s00439-009-0656-3>
- Xi, B., G. R. Chandak, Y. Shen, Q. Wang, and D. Zhou, 2012 Association between common polymorphism near the MC4R gene and obesity risk: a systematic review and meta-analysis. *PLoS One* 7: e45731. <https://doi.org/10.1371/journal.pone.0045731>
- Yanagiya, T., A. Tanabe, A. Iida, S. Saito, A. Sekine *et al.*, 2007 Association of single-nucleotide polymorphisms in MTMR9 gene with obesity. *Hum. Mol. Genet.* 16: 3017–3026. <https://doi.org/10.1093/hmg/ddm260>
- Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher, 2011 GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88: 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>
- Zeggini, E., M. N. Weedon, C. M. Lindgren, T. M. Frayling, K. S. Elliott *et al.*, 2007 Replication of Genome-Wide Association Signals in UK Samples Reveals Risk Loci for Type 2 Diabetes. *Science* 316: 1336–1341. Erratum: 317: 1035–1036. <https://doi.org/10.1126/science.1142364>
- Zhang, J., M. Ma, X. Nan, and B. Sheng, 2016 Obesity inversely correlates with prostate-specific antigen levels in a population with normal screening results of prostate cancer in northwestern China. *Braz. J. Med. Biol. Res.* 49: 2893–2898. <https://doi.org/10.1590/1414-431x20165272>

Communicating editor: D. Gresham