nature portfolio

https://doi.org/10.1038/s41586-026-10330-z

Supplementary information

Genetic predictors of GLP1 receptor agonist weight loss and side effects

In the format provided by the authors and unedited

Genetic predictors of GLP-1 receptor agonist weight loss and side effects

Supplementary Materials

1. Supplementary Methods

1.1. Imputation Panel

Imputation was performed using an approach that combined two independent reference panels: the publicly available Human Reference Consortium (HRC) panel [1] and a 23andMe reference panel, which was built by 23andMe using internal and external cohorts. The publicly available HRC data were downloaded from the European Genome-Phenome Archive at the European Bioinformatics Institute (accession EGAD00001002729). The HRC data included 27,165 samples. Variants were lifted to hg38 and excluded if their new positions were on a different chromosome. Variants were then re-phased using SHAPEIT4 [2]. Finally, singletons were excluded. The final HRC reference panel included 39,057,040 SNPs (no indels). The 23andMe reference included 12,217 samples from multiple internal and external WGS datasets (Table below), with sequencing reads having been remapped to the build 38 reference genome, and variant calls obtained via DeepVariant [3]. The final 23andMe reference panel included 82,078,539 variants (73,852,355 SNPs + 8,226,184 indels).

Sequencing Project	**Sample size**	**Notes**
NYGC 1000G 30X genomes	2,493	1000 Genomes samples sequenced to 30X [4]
CHM syndip	1	CHM benchmark dataset [5]
GTEx	838	GTEx v8 samples [6]
HGDP	770	HGDP samples from global ancestral groups [7]
SGDP	188	SGDP samples from diverse populations [8]
Parkinson's cohort	2,721	Parkinson's cases of predominantly European ancestry
Ashkenazi Jewish cohort	553	Individuals of Ashkenazi ancestry
LRRK2 carrier cohort	303	LRRK2 Parkinson's risk allele carriers
African American cohort	2,389	Individuals of African American Ancestry [3]
Latino cohort	1,961	Individuals of Latino ancestry
Total	12,217	

Imputation was performed against the combined HRC and custom panel. More than 36M variants were found in both the HRC and 23andMe reference panels, 46M in the 23andMe panel only, and 3M in the HRC panel only. Using Beagle 5 [9], variant imputation was performed separately for the three sets of variants (HRC only, 23andMe only, and intersect), using the relevant reference samples as appropriate.

Because participants were genotyped on one of the five genotyping platforms (v1 to v5), the imputation was performed independently for each genotyped platform. For the non-pseudoautosomal region of the X chromosome, males and females were phased together in segments, treating males as already phased; pseudoautosomal regions were phased

separately. We then imputed males and females together, treating males as homozygous pseudodiploids for the non-pseudoautosomal region.

We assessed the above approach by measuring aggregated r-squared of the imputed genotypes against true genotypes ascertained in held out genomes, and established that it yielded meaningful improvements in imputation performance across ancestries relative to the HRC panel alone. In the table below, we show the imputation performance as a function of allele frequency for the original HRC panel compared to the combined 23andMe custom panel (referred to as the 'r6' panel) in 3 populations. Imputation performance was assessed using held-out samples for which whole-genome sequences were available.

	HRC			**23andMe r6 panel**		
Allele Frequency	**European**	**African American**	**Latino**	**European**	**African American**	**Latino**
0.001	0.54	0.34	0.2	0.66	0.67	0.5
0.01	0.84	0.67	0.55	0.88	0.86	0.8
0.1	0.97	0.86	0.92	0.98	0.94	0.95

1.2. Additional predictors of efficacy

Previous literature has indicated that ancestry and T2D status may be relevant indicators in determining GLP-1 medication efficacy [10,11]. To investigate these factors in our dataset, we first used the baseline model described in equation 1, and added additional terms to account for European / non-European ancestry and T2D diagnosis status. Specifically, we tested tested the model:

$$\Delta BMI_{\%} \sim age + sex + BMI_1 + drugType + dose + days_{treat} + drugType{:}\,dose + drugType{:}\,days_{treat} + dose{:}\,days_{treat} + dose{:}\,days_{treat}{:}\,drugType + T2D + european\_ancestry$$

where $T2D$ indicates if an individual has self-reported a previous diagnosis of type 2 diabetes, and $european\_ancestry$ is equal to 1 for individuals of predominantly European ancestry and 0 otherwise.

To expand on the ancestry analysis, we fit an additional model that included terms for Latino, African American, and 'other' ancestry, as follows:

$$\Delta BMI_{\%} \sim age + sex + BMI_1 + drugType + dose + days_{treat} + drugType{:}\,dose + drugType{:}\,days_{treat} + dose{:}\,days_{treat} + dose{:}\,days_{treat}{:}\,drugType +$$

$$T2D + latino\_ancestry + african\_american\_ancestry + other\_ancestry$$

where $latino\_ancestry$, $african\_american\_ancestry$, and $other\_ancestry$ define membership in the ancestral groups. Note that this model pivots on European ancestry.

The results of these models are given in Supplementary Table 4. Given these models, we find that T2D status is a significant predictor of GLP-1 medication efficacy, with individuals having a T2D diagnosis experiencing less weight loss than those without a diagnosis, as expected from previous literature [12]. Furthermore, our data suggests that GLP-1 medications to be most effective in individuals of European ancestry, less effective in individuals of Latino ancestry, and least effective in individuals of African American ancestry (Supplementary Table 4).

1.3. Additional genetic association tests

We performed additional association analyses of the index variant, which are detailed here:

1.3.1. Dominance testing

To ascertain if rs10305420 T allele has a non-additive effect on efficacy, we performed a dominance test. Specifically, the SNP was re-coded in terms of additive and dominant effects. Given an imputed dosage, g, that encodes the T allele dosage for a given individual such that $0 \leq g \leq 2$, the additive component was encoded as $a = g - 1$. The dominance component was defined as $d = 1 - abs(1 - g)$. This encoding represents a continuous equivalent to the encoding that would be used for discrete genotype calls; i.e.: an additive encoding of -1, 0, 1 for AA, AB, BB, and a dominance encoding 0, 1, 0 for AA, AB, BB. Given this encoding, we tested the for dominance using the model:

$$\Delta BMI_{\%} \sim age + sex + BMI_1 + drugType + dose + days_{treat} + drugType{:}\,dose + drugType{:}\,days_{treat} + dose{:}\,days_{treat} + dose{:}\,days_{treat}{:}\,drugType + pc0 + \ldots + pc4 + I(platform) + a + d$$

where $pc0$ and $pc4$ represent the first and fifth principal components respectively, and $I(platform)$ represents indicator functions to account for genotyping platforms, as used within the main GWAS analysis. Applying this model to rs10305420, we find the additive component to be highly significant (p = 1e-11, effect = -0.66%), whereas the dominant component is not (p = 0.80, effect = 0.03%). Given this, we conclude that there is no evidence for deviation from additivity in the contribution of rs10305420 to GLP-1 medication efficacy.

1.3.2. Robustness to type 2 diabetes status

GLP-1 medications were initially indicated to treat type 2 diabetes. In addition, the rs10305420 T allele has been associated with a lower risk of type 2 diabetes (T2D) in prior studies [13,14], as well as in our own data (Supplementary Table 17). Indeed, of the individuals in the GWAS sample with a self-reported T2D status, approximately 22% reported having a T2D diagnosis (3300 /

15043), which is higher than the expected population prevalence of 14.7% in the adult US population [15], and consistent with the dataset being enriched for people taking GLP-1 medication in relation to T2D diagnosis.

It is therefore plausible that the observed association between rs10305420 and greater weight loss could be partially explained by the enrichment of individuals with T2D in the sample. To address this question, we used an interaction model to determine if the SNP association is altered by T2D status. Specifically, we extended the GWAS regression model to include T2D status, as follows:

$$\begin{aligned}\Delta BMI_{\%} \sim\ & age + sex + BMI_1 + drugType + dose + days_{treat} + drugType{:}\,dose + \\ & drugType{:}\,days_{treat} + dose{:}\,days_{treat} + dose{:}\,days_{treat}{:}\,drugType + \\ & pc0 + \ldots + pc4 + I(platform) + \\ & T2D + SNP + SNP{:}\,T2D\end{aligned}$$

where $T2D$ represents T2D diagnosis status, as self-reported by 23andMe research participants (self-reported doctor T2D diagnosis = 1, no diagnosis = 0).

Applying this model to rs10305420 in the European population (Supplementary Table 7), we find that individuals with T2D report significantly less weight loss while on GLP-1 medication, after accounting for other factors (BMI% effect = 2.78%, p = 1.0e-52). However, rs10305420 remains significantly associated with a comparable effect size (effect = -0.59%, p = 4.28e-8), while the interaction is not significant (effect = -0.29%, p = 0.24). We therefore conclude that T2D status does not explain the relationship between rs10305420 and GLP-1 medication efficacy.

To confirm the above result, we repeated the analysis using a stratified approach; retesting rs10305420 separately in individuals with or without T2D. As with the interaction analysis, we see no evidence of a differential effect of rs10305420 depending on T2D status (Supplementary Table 7).

1.3.3. Robustness to smoking status

In a similar vein to the T2D analysis above, rs10305420 has also been observed as associated with tobacco use. It is therefore reasonable to ask if smoking status could influence the observed association with GLP-1 medication efficacy. We again employed an interaction model to test this hypothesis:

$$\begin{aligned}\Delta BMI_{\%} \sim\ & age + sex + BMI_1 + drugType + dose + days_{treat} + drugType{:}\,dose + \\ & drugType{:}\,days_{treat} + dose{:}\,days_{treat} + dose{:}\,days_{treat}{:}\,drugType + \\ & pc0 + \ldots + pc4 + I(platform) + \\ & smoking\_status + SNP + SNP{:}\,smoking\_status\end{aligned}$$

where $smoking\_status$ represents the self-reported smoking status of the participants. In this case, we tested two definitions of smoking status; 'current tobacco user', or 'ever tobacco user'. 'Current tobacco users' were defined as individuals who self-reported smoking cigarettes, using tobacco, or typically smoking cigarettes either every day or on some days; controls reported never smoking or never smoking at least 100 cigarettes. Approximately 8.4% of the participants with complete data classified as current tobacco users (1,239 / 14,676). 'Ever tobacco users' were defined as individuals who self-reported either regular smoking or having smoked at least 100 cigarettes in their lifetime; controls both did not report having smoked at least 100 cigarettes nor reported being a current regular smoker. Approximately 45% of the participants with complete data were classified as ever tobacco users (6,698 / 14,965).

We fitted the regression model separately for each smoking definition. Neither current smoking status nor ever smoker status appears to be significantly associated with GLP-1 medication efficacy (p = 0.24 and p = 0.25, respectively; Supplementary Table 8). Likewise, neither definition shows any evidence for an interaction with rs10305420 genotype (p = 0.99 and = 0.68 respectively). As such, we conclude that smoking status is unlikely to be significantly mediating the rs10305420 association.

To confirm the above result, we repeated the analysis using a stratified approach; retesting rs10305420 separately in individuals who were ever smokers, or never smokers. As with the interaction analysis, we see no evidence of a differential effect of rs10305420 depending on 'ever smoker' status (Supplementary Table 8).

1.3.4. Conditional analysis

To identify if multiple independent associations could be identified within the GLP1R locus, we utilized conditional step-down analysis. For each phenotype with a GLP1R GWAS association, we performed a stepwise conditional analysis by including the index variant in the regression model, and testing for any additional significantly associated variants within the locus (+/- 200kb of the index SNP in each case). The forward selection procedure began with the index variant, adding one additionally significant lead variant at each step. The procedure concluded when the last added variant had a p-value > 10^{-5}. For all three phenotypes ($\Delta BMI_{\%}$, nausea, and vomiting), we did not detect compelling evidence of secondary associations (Extended Data Figure 2).

1.3.5. Interaction analysis

To test for interaction between the GLP1R variant (rs10305420) and the GIPR variant (rs1800437) in the vomiting side effect for the tirzepatide population, we applied the model:

$$vomiting_{tirzepatide} \sim age + sex + BMI_1 + dose + days_{treat} + pc0 + \ldots + pc4 + I(platform) + dose{:}\,days_{treat} + SNP1 + SNP2 + SNP1{:}\,SNP2$$

where $vomiting_{tirzepatide}$ represents the vomiting phenotype restricted to the tirzepatide treated population, SNP1 represents rs10305420, and SNP2 represents rs1800437. For the purposes of interaction testing, we mean centered the SNP data. The model was fitted in the European population, and results are shown in Supplementary Table 15.

To provide a more interpretable assessment of the risk associated with a potential interaction between these loci, we performed a separate analysis in which we encoded the carriers of homozygous risk alleles at both loci as cases, and homozygous carriers of non-risk alleles at both loci as controls. Heterozygous carriers were set to missing. Performing association testing between these two groups gave an estimated odds ratio of 14.8 (95% CI 6.2 - 35.8) for tirzepatide-mediated vomiting.

1.4. Fine mapping and functional annotation

1.4.1. GLP1R locus

To better understand which variants in the efficacy locus were likely to be causal, we considered the 42 variants within the credible set of the efficacy association (Figure 1b). We annotated these variants using the Ensembl Variant Effect Predictor (VEP) [16]. Of the 42 variants, 37 were intronic within the DNAH8 gene, one was downstream of DNAH8, and 3 were upstream or downstream of a processed ribosomal protein pseudogene (ENSG00000220556). Only rs10305420 is a missense protein coding variant (Supplementary Table 9).

We further characterized the locus by performing colocalization analysis with eQTL datasets. We did not find any eQTLs from the EBI eQTL catalogue release 7 [17] that had index SNPs within 500 kb of rs10305420 and having $r^2 > 0.5$, nor any that colocalized with H4 > 0.5 using susie-coloc.

Based on the above analysis, we conclude that rs10305420 represents the most likely causal variant from a functional perspective. This variant represents a proline to leucine change on residue 7 (p.Pro7Leu), which lies in the N-terminal signal peptide. As a common variant (with allele frequency between 7% and 40%), pathogenicity scoring algorithms judge the variant as likely benign, with SIFT scoring the variant as 0.18 (“tolerated low confidence”), while AlphaMissense scores the variant as 0.1411 (“likely benign”).

Given the variant is contained within the signal peptide domain, the associated residue will be cleaved prior to expression of the receptor on the cell surface, and as such, the genetic variant is unlikely to alter the final protein structure directly, and it is more likely that the variant influences trafficking within the cell on its journey to the cell surface.

To better understand the influence of p.Pro7Leu on the signal peptide, we utilized the SignalP protein language model [18] to analyse how the variant may be altering the peptide structure. Using SignalP 6.0, we performed SignalP 6.0 analysis using a 40-amino acid window, capturing the full signal peptide and the N-terminal domain of the mature receptor, and repeated the

analysis for both the proline and leucine variants.

The results of this analysis are shown in Extended Data Figure 9. The analysis indicates that the variant influences the boundary of the N-terminal region (n-region) and the hydrophobic region (h-region) of the signal peptide. Specifically, the probability of the 7th amino acid being in the h-region is ~70% for the wild-type proline variant, but is ~100% for the leucine variant. This suggests that the Pro7 variant may be inducing structural instability in the secondary structure that propagates downstream, with residues 8 and 9 also failing to achieve the maximal helical probability (<100%). This observation is consistent with leucine being strongly hydrophobic and helix-promoting, which are properties that stabilize the h-region and create a canonical signal sequence.

Given the above results, it is likely that the leucine variant (associated with the rs10305420 T allele) results in more efficient translocation to the endoplasmic reticulum, and thus may result in more efficient expression of GLP1R on the cell surface. However, direct *in vitro* data on p.Pro7Leu is limited, and published studies have not detected differences in surface expression or signaling [19,20]. As such, any effect of p.Pro7Leu on receptor biosynthesis/trafficking appears to be too small to manifest as a robust change in surface expression or classical signalling in these overexpression systems. Alternatively, it may be that the effect is cell type or otherwise context dependent.

We repeated the above analysis for the GLP1R-locus nausea and vomiting associations, which had 13 and 21 variants in the 99% credible set respectively (Supplementary Table 9). Of the 13 variants in the nausea association, 9 were intronic to DNAH8, one was downstream of DNAH8, and 3 were upstream or downstream of the processed ribosomal protein pseudogene (ENSG00000220556). Of the 21 variants in the vomiting credible set, 1 was intronic to GLP1R, 16 were intronic to DNAH8, 1 was downstream of DNAH8, and 3 were upstream or downstream of the same processed ribosomal protein pseudogene (ENSG00000220556). Notably, none of the variants in either credible set were missense variants, and neither credible set included rs10305420, despite the fact that rs10305420 is in moderate LD with the index variant of both associations ($r^2 = 0.75$ and 0.57 for nausea and vomiting respectively). We also characterized these loci against the EBI eQTL catalogue in the same manner as above, and observed no eQTLs that colocalized with either signal (i.e. within 500 kb of the index SNP and having $r^2 > 0.5$, nor any that colocalized with H4 > 0.5 using coloc). Given this lack of functional evidence, it is much harder to independently speculate which variant is likely causal for these associations without relying on colocalization with the efficacy signal.

1.4.2. GIPR locus

For the GIPR locus, the 99% credible set is much smaller, with just 6 variants (Supplementary Table 13), and all variants in very tight LD. Of these variants, only rs1800437 is a protein-altering variant, and we therefore conclude that this is the most plausible causal variant.

1.5. Signal colocalization analysis

For between-trait colocalization analysis, we initially applied the method of Giambartolomei *et al.* [21]. Specifically, we considered the window around rs10305420 +/- 500kb, and used default prior parameters (p1 = 1e-04, p2 = 1e-04, p12 = 1e-05). We applied the method to all pairs of the three traits; $\Delta BMI_{\%}$, nausea, and vomiting. As detailed in the main text, we found good colocalization probabilities between the three traits ($\Delta BMI_{\%}$ vs nausea: H4 = 96.6%, $\Delta BMI_{\%}$ vs vomiting: H4 = 88.5%, nausea vs vomiting: H4 = 92.1%).

We performed a further colocalization analysis using HyPrColoc [22], which allows simultaneous colocalization analysis of multiple traits. We selected all variants with non-missing effect sizes and standard errors that lie within the genomic window spanning the minimum to maximum positions across the efficacy, nausea, and vomiting loci, extended by ±500 kb on each side. As input parameters we included the estimated GWAS effects and standard errors, an LD matrix derived from 200k samples derived from the same population, and the sample overlap matrix. This method concluded that shared colocalization between the 3 traits is the most probable model, with a colocalization posterior probability of 72.6%.

1.6. Comparison of self-report data to EHR data obtained via Apple Healthkit

We processed the vital signs and medication records obtained from 23andMe research participants via Apple HealthKit. All participants included in this analysis had provided additional consent to allow research of data derived from mobile devices. We extracted BMI measurements and GLP-1 prescriptions for the comparison of self-reported data to EHRs.

1.6.1. BMI calculation

BMI records were extracted from vital signs data using specific LOINC codes (39156-5, 89270-3) as the primary identification method. For entries with missing LOINC codes, text matching searches for keywords including "BMI", "Body Mass Index" in the display field were applied. Manual examination excluded inaccurate measurements containing terms such as "percentile", "Z-Score", "body surface area", and "Weight in (lb) to have BMI = 25" to ensure only direct BMI measurements were retained.

To maximize sample size, height and weight data were extracted similarly using established LOINC code (weight: 34372-9, 29463-7, 3141-9, 3142-7, 75292-3, 79348-9, 8350-1, 8351-9; height: 34373-7, 3137-7, 3138-5, 8302-2, 8306-3, 8308-9). Additional keyword matching identified records containing "weight", "height", or "length" when LOINC codes were missing. Manual review excluded inappropriate measurements including percentiles, Z-scores, birth measurements, pre-pregnancy weights, ideal weights for quality control. All measurements for weight were standardized to grams (from grams, kg, and pounds), and height measurements were converted to meters (from centimeters, inches, and feet). BMI was calculated using the standard formula (weight_kg / height_m²) for height-weight pairs. Since height measurements were less frequent than weight measurements, we paired each weight measurement with the closest available height measurement within one year.

For data consolidation, BMI measurements directly recorded in the EHR took priority over calculated values when both existed for the same individual on the same date. Multiple BMI records on the same day were averaged. Using this approach, we extracted 378,933 BMI records from 16,325 individuals.

1.6.2. GLP-1 medication extraction

GLP-1 medication entries were identified through text matching of medication records with specific brand names (Ozempic, Wegovy, Mounjaro, Zepbound) and generic compounds (semaglutide, tirzepatide) if no brand name was identified. Dosages were extracted using regular expression parsing to identify numeric values preceding "dose" or "mg" in free-text medication descriptions. We further applied manual review of more complicated dosing regiments. For example, medication strings containing "0.25 mg or 0.5 mg" patterns were systematically split into separate records with 0.25 mg assigned to the first four weeks and 0.5 mg to the remaining treatment course.

Dosages within clinically typical ranges (0.25, 0.5, 1.0, 1.7, 2.0, 2.4, 2.5, 5.0, 7.5, 10.0, 12.5, 15.0 mg, depending on medication) were retained. Furthermore, all GLP-1 records were filtered against FDA approval dates for each specific drug: Ozempic (2017-12-05), Wegovy (2021-06-04), Mounjaro (2022-05-13), Zepbound (2023-11-08), with compounded formulations using their respective drug approval dates (compounded semaglutide: 2017-12-05 and compounded tirzepatide: 2022-05-13).

Prescriptions with status "active", "stopped", or "completed" were retained, while prescriptions labelled "canceled", "draft", "intended", or "on-hold" were removed to focus on actual medication exposures. A final consolidation process merged prescriptions when gaps between treatment periods were 30 days or less with identical patient-drug-dose combinations in order to reflect real-world medication adherence patterns. The final dataset provided 8,975 entries covering start/end dates, drug name, drug type, and dosages from 1,348 individuals.

1.6.3. Cohort construction

With the extracted BMI and GLP-1 records from previous steps, we aligned the begin and end BMI with the medication data. For each GLP-1 record, the first BMI measurement (BMI_1) was defined as the closest measurement within a window starting 60 days before to 30 days after the record date. The second BMI measurement (BMI_2) was taken as the lowest recorded value during the extended treatment window (from BMI_1 measurement date to BMI_2 measurement date). For records with missing prescription end dates, we imputed the end dates using the last recorded BMI measurement within 3 years of treatment initiation, adjusting the treatment duration accordingly.

For individuals with multiple GLP-1 medication records, the drug with the longest treatment duration, and the most recent dose for the corresponding drug was used. The final step filtered records based on baseline BMI ranges (25-70 kg / m²), end BMI values (14-70 kg / m²), and

reasonable BMI change boundaries (-45% to +20%). The final cohort consisted of 909 individuals with GLP-1 medication records and BMI measurements (Supplementary Table 5).

1.7. Replication

1.7.1. All of Us replication

We performed replication of the identified efficacy lead SNP (rs10305420) in the All of Us cohort [23], using Controlled Tier Dataset v8. We defined the cohort as 9,579 individuals who have drug exposure to semaglutide or tirzepatide, and have genomic data and EHR data. The sequential filtering steps and exclusion criteria used to derive the final study cohorts are detailed in the table below.

Selection step	Description of inclusion / exclusion criteria	Remaining sample size (N)
1	Individuals with available genomic data.	447,281
2	Individuals with available EHRs.	340,582
3	Individuals with drug exposure information including drug codes for Semaglutide or Tirzepatide.	9,579
4	Individuals with valid pre- and post-treatment BMIs + passing genotype QC.	4,889
5 (Analysis A)	Final Cohort (Complete Case Analysis): Individuals with complete covariates.	3,948
5 (Analysis B)	Final Cohort (Imputed drug dose): Individuals included after drug dose imputation.	4,885

Specific details of the steps are as follows:

1.7.1.1. Drug dataframe

We extracted drug type and dosage information from the prescription in drug exposure standard_concept_name. We next collapsed the rows for the same [person_id, drug type] to keep the minimum drug_exposure_start_datetime as drug_start_date, the maximum drug_exposure_end_datetime as drug_end_date, and the maximum weekly dose as weekly_dose. With these steps, we generated a drug dataframe with person_id, drugType, weekly_dose, drug_start_date, and drug_end_date.

1.7.1.2. Drug-BMI dataframe

We then merged the above drug dataframe with BMI measurement dataframe (removing rows where BMI values are less than 15 or larger than 60), and defined BMI_1 as the latest BMI measurement before or on the drug_start_date within 30 days of drug_start_date, and defined BMI_2 as the earliest BMI measurement on or after drug_end_date within 30 days of drug_end_date. If BMI_1 was unavailable, we instead used the earliest BMI measurement after the drug_start_date within a 30-day window, requiring the BMI measurement date to be earlier

than the drug_end_date, and adjusted the drug_start_date to this BMI measurement date. If BMI_2 was unavailable, we instead used as BMI_2 the latest BMI measurement before the drug_end_date within a 30-day window, requiring the BMI measurement date to be later than the drug_start_date, and adjusted the drug_end_date to this BMI measurement date. We defined the treatment days as the difference between drug_end_date and drug_start_date, with the adjustment if applicable. We calculated $\Delta BMI_{\%}$ according to the formula in the GWAS methods, and excluded rows with $\Delta BMI_{\%}$ above 20% or below -45%.

Using the above definitions, we obtained 4,231 rows, among which 206 person_ids contributed 2 rows for both semaglutide and tirzepatide. We selected the one with longer treatment days to keep, and the resulted drug-BMI dataframe contained 4,025 rows. Finally, we merged sex_at_birth and date_of_birth data into the drug-BMI dataframe and calculated age_at_start.

1.7.1.3. Drug-calculated BMI dataframe

Separately, we merged the drug dataframe with body weight measurements (removing rows where weight values are less than 36 kg and larger than 362 kg) and height measurements (removing rows where height values are less than 1.13 meters and larger than 2.286 meters). We selected the median height value for each person_id, and calculated_BMI as:

*calculated_BMI = value_kg / (value_meter * value_meter)*

We then applied the same procedure as described in 2) Drug-BMI dataframe to define treatment days, BMI_1, BMI_2, and $\Delta BMI_{\%}$ with the boundary of [-45%, 20%]. We obtained 4,541 rows, among which 229 person_ids contributed 2 rows and the one with longer treatment days was kept. The final drug-calculated_BMI dataframe contained 4,312 rows with the additional columns including sex_at_birth, date_of_birth, and age_at_start.

1.7.1.4. Final phenotype-genotype dataframe

For the lead efficacy variant rs10305420, we extracted genotype data from WGS_ACAF_THRESHOLD_SPLIT_HAIL_PATH with interval chr6:39,048,860-39,048,861, and defined GT as the count of the alternative allele (T). 3,758 person_ids from Drug-BMI dataframe have genotyping data for the lead efficacy variant rs10305420, and 3,992 person_ids from Drug-calculated_BMI dataframe have data for rs10305420.

We added to the drug-BMI dataframe new person_id rows from drug-calculated_BMI dataframe and obtained a union phenotype-genotype dataframe. We retrieved the WGS QC file and excluded person_ids present in the QC flagged samples. The final phenotype-genotype dataframe contained 4,889 rows.

1.7.1.5. Association of lead efficacy variant with $\Delta BMI_{\%}$

We retrieved genetic principal components (PCs) and performed a standardization of the raw PC features before merging with the final phenotype-genotype dataframe. We also centered

age_at_start by subtracting the mean value. We performed linear regression with python3 statsmodels (v0.14.1) OLS fit with the following formula:

$$\Delta BMI_{\%} \sim rs10305420\_T + age + sex + BMI_1 + drugType + dose + days_{treat} + drugType{:}\, dose + drugType{:}\, days_{treat} + dose{:}\, days_{treat} + dose{:}\, days_{treat}{:}\, drugType + pc0 + \ldots + pc9$$

There were 3,948 rows with all non-NA covariates. In order to boost sample size, we repeated the association analysis, having imputed missing data in the weekly_dose covariate, specifically by filling in NAs with the mean weekly_dose for the matching drugType, which boosted the sample count to 4,855. We present results both with and without imputation (Supplementary Table 10a and b respectively).

For the replication analyses described above, we chose to maximize statistical power by relaxing the ancestral restriction to maximize sample size. However, to ensure that the replication result is not driven by population structure, repeated the All of Us replication analysis having restricted to individuals of genetically-inferred European ancestry. While the reduced sample size resulted in less significant p-values compared to the full multi-ancestry cohort, the association remains statistically significant ($p < 0.05$) in this restricted European analysis, both with and without drug dose imputation (Supplementary Table 10c).

1.8. UK Biobank replication

To replicate the efficacy association in the UK Biobank (UKBB), we first assembled a GLP-1 user cohort by extracting dated GLP-1 prescriptions from the gp_scripts table. Prescriptions were identified by searching the *drug_name* field for any of the following GLP-1 receptor agonist terms: exenatide, bydureon, byetta, liraglutide, victoza, saxenda, lixisenatide, lyxumia, dulaglutide, trulicity, albiglutide, and eperzan. This yielded 1,004 distinct participants classified as GLP-1 users. We note that the data in UKBB represent prescription information prior to 2017, so pre-dates the availability of semaglutide and tirzepatide.

For these individuals, we retrieved BMI and height measurements and their associated assessment dates from the participant table. In parallel, we extracted BMI and weight measurements recorded during primary care encounters from the gp_clinical table. We applied quality-control filters to remove biologically implausible observations, excluding BMI values <15 or >60 kg/m², weight values <36 kg or >362 kg, and height values <1.13 m or >2.286 m. For records with valid weight and height measurements, BMI was calculated as weight (kg) divided by height (m) squared and incorporated as additional dated BMI observations. These steps yielded the BMI measurement dataset used in downstream analyses.

For each GLP-1 drug with at least 100 users (liraglutide, exenatide, and victoza), we selected individuals with at least two prescription records and defined the treatment start and treatment end dates as the minimum and maximum prescription dates, respectively. This resulted in 472,

324, and 117 individuals with defined treatment intervals for liraglutide, exenatide, and victoza, respectively.

We merged the drug-specific treatment intervals with the BMI measurement dataset to determine baseline and post-treatment BMI values. BMI_1 was defined as the latest BMI measurement on or before the treatment start date within a 365-day window. If no such measurement existed, we used instead the earliest BMI obtained after the treatment start date (within 365 days), provided that it preceded the treatment end date. In such cases, the treatment start date was redefined to the date of this BMI measurement. BMI_2 was defined as the earliest BMI measurement on or after the treatment end date within a 365-day window. If unavailable, we used the latest BMI measurement preceding the treatment end date (within 365 days), provided that it occurred after the treatment start date; when this occurred, the treatment end date was set to that BMI measurement date. Treatment duration was defined as the difference between the adjusted treatment end and treatment start dates. BMI percent change ($\Delta BMI_{\%}$) was calculated as described in the GWAS analysis, and we excluded observations with $\Delta BMI_{\%}$ >20% or <−45%.

Sex information was obtained from the participant table, and age at treatment initiation was calculated from birth year and month relative to the treatment start date. Genotype at rs10305420 was extracted from the chromosome 6 PLINK files in the *UKB imputation from genotype* dataset, and genetic principal components (PCs) were retrieved from the participant table.

After requiring non-missing BMI_1, BMI_2, genotype, sex, age, and PCs, we retained 108, 79, and 34 individuals for liraglutide, exenatide, and victoza, respectively.

We performed linear regression using the Python3 `statsmodels` (v0.14.1) ordinary least squares model with the following formula:

$$\Delta BMI_{\%} \sim age + sex + BMI_1 + days_{treat} + \text{rs10305420\_T} + pc0 + \ldots + pc9$$

Because individual drug-specific sample sizes were small, we additionally conducted an analysis pooling all GLP-1 drugs. In the combined GLP-1 dataset, 933 individuals had defined treatment intervals, and 233 had complete covariates for association testing.

The results of the association testing are shown in Supplementary Table 11. Given the available UKBB sample sizes, we estimate the power to detect an association of a similar magnitude to that observed with rs10305420 in our study to be approximately 10%, even if we assume liraglutide, exenatide, and victoza all have similar efficacy to semaglutide or tirzepatide. However, these medications are generally less effective than semaglutide for weight loss [24], so it may be that power is even lower. In addition, virtually all individuals taking one of these drugs in UKBB were type 2 diabetic which would further be expected to moderate any signal regarding weight loss efficacy. We therefore conclude that there is insufficient power to replicate the efficacy association in UKBB.

1.9. Multivariable modeling of genetic and non-genetic risk factors

1.9.1. Efficacy outcome and Predictor Variables

The primary objective was to construct a multivariable linear model to predict weight loss efficacy in response to GLP-1 receptor agonists. The primary outcome of medication efficacy was defined as the percent change in BMI from baseline ($\Delta BMI_{\%}$). The model included the following treatment variables: drug type (semaglutide = 1, or tirzepatide = 0), maximum dose administered (in mg), treatment duration (in days), and a drug-by-dose interaction term. Clinical and demographic variables were: baseline BMI, age at treatment initiation, genetic sex, years of education as a proxy for socioeconomic status (SES), and binary indicators for type 2 diabetes, hypertension, and non-alcoholic fatty liver disease (NAFLD), following previous studies [25]. Genetic variables were: allele dosage for the index variant(s) identified in the primary genome-wide association study (GWAS). To account for potential population stratification in our ancestrally diverse cohort, the first ten global principal components (PCs) of genetic ancestry were included as covariates. All continuous predictor variables were standardized (z-scored) prior to modeling to allow for the comparison of effect sizes. To allow for the comparison of effect sizes, all continuous predictor variables were standardized using the mean and standard deviation estimated from the training set only, which were then applied to scale the held-out test set (detailed below). The outcome variable was not standardized in order to maintain interpretability in original units (i.e. percent BMI change, with negative values representing *more loss*).

1.9.2. Modeling the Likelihood of Side Effects

In addition to efficacy, we modeled the risk for each specific side effect. We developed models for all reported side effects in our data, using the same predictor variables as the efficacy model. For side effect phenotypes with a significant GWAS association (i.e. nausea and vomiting), the SNPs identified from the GWAS were included, specifically we selected the likely causal variants at each locus for inclusion in the models (i.e. *GLP1R* rs10305420, *GIPR* rs1800437). An interaction between rs1800437 and drug type was included to account for the GWAS observation of this variant having an effect in the tirzepatide-treated population, but not in the semaglutide-treated population. Given the binary nature of the side effect phenotype outcomes, we fit multivariable logistic regression models (equivalent to a generalized linear model with a binomial family and logit link function). The resulting models were used to estimate the odds ratios (ORs) for each predictor's association with the likelihood of experiencing the specific side effect.

1.9.3. Cohort Definition and Data Preprocessing

The analysis was restricted to participants with complete data for all model variables. To isolate the active weight loss phase and exclude periods of weight maintenance, we established a specific treatment duration window for inclusion. The rationale for the upper bound was based on evidence from the SELECT trial[10], which demonstrated that weight loss with semaglutide typically plateaus after two to three years, which was confirmed via observation in our own data. Therefore, the analysis cohort was limited to individuals with at least 1 day and no more than 30 months of treatment.

1.9.4. Model Evaluation within self-report data

The dataset was randomly partitioned into a training set (70% of the sample), used for model fitting, and a held-out test set (30%), used for independent model evaluation. Participants were required to have complete data to be included in a model. The resulting sample sizes for training and testing are given in the table below:

Phenotype	Training N	Testing N
$\Delta BMI_{\%}$	11,151	4,780
Nausea	10,823	4,613
Vomiting	10,821	4,613

A multivariate linear or logistic model (for continuous or binary outcomes) was fitted to the training data using the statsmodels library (v0.14.1) in Python3. Model performance was scored by R^2 for linear models and area under the receiver operator curve (AUROC) for binary outcomes. For AUROC plots (Figures 3b and c), 95% confidence intervals were determined via 1000 bootstrap samples. Model performance was assessed and reported by evaluation of calibration (the agreement between predicted and observed outcomes) and overall fit in the held-out test set.

1.9.5. Evaluation of self-report derived risk models in HealthKit EHRs

Using the efficacy model derived from the self-reported data (as described above), we aimed to assess the performance of the model in a set of 642 individuals for whom we had EHR data and had not been used within the training of the model. All individuals in this validation cohort had provided HealthKit EHR data with documented GLP-1 medication usage, but had not participated in the GLP-1 survey, ensuring no overlap with the model's training and testing sets.

For each individual, the GLP1 medication start date in the EHR cohort was established as the baseline date. Preliminary inspection revealed occasional outlier BMI measurements that differed substantially from preceding and succeeding measurements. These implausible BMI values were removed by calculating a rolling median and median absolute deviation (MAD) within 5-measurement windows, and excluding measurements with modified Z-scores exceeding 3 standard deviations from the local median. Given the longitudinal nature of EHR BMI measurements, we then interpolated BMI measurements to a consistent 90-day interval for each individual using piecewise linear interpolation.

Inputs to the model were derived as follows. Pre-treatment BMI at baseline date were derived from the EHR. Clinical factors were also derived from EHR condition records, specifically whether an individual had an EHR diagnosis of type 2 diabetes, hypertension, or non-alcoholic fatty liver disease (NAFLD). Age at the baseline date was obtained from the self-reported birth date. To mimic a realistic pre-treatment scenario in which treatment time, drug type, and final dosage are unknown, we fixed the input parameters for the model for these variables across all individuals. Specifically, the days of treatment input to the model was set to be 180 days, and all

individuals were modeled as if they were receiving a standard dose of semaglutide (2mg). Allele dosage for the genetic variants were obtained from the non-EHR data sources (i.e. from the 23andMe participant's genetic data), as were the first ten global principal components (PCs) of genetic ancestry. All continuous predictor variables were standardized using the mean and standard deviation estimated from the self-report training set.

Having obtained predictions from the model, individuals were stratified into three groups based on their expected weight-loss response from the model: top 25%, middle 25-75%, and bottom 25%. Within each group, we averaged the longitudinal BMI trajectories relative to the baseline and plotted the longitudinal BMI estimates up to 24 months (Figure 3a).

2. References

1. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
2. Delaneau, O., Zagury, J.-F., Robinson, M. R., Marchini, J. L. & Dermitzakis, E. T. Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019).
3. O'Connell, J. *et al.* A population-specific reference panel for improved genotype imputation in African Americans. *Commun. Biol.* **4**, 1269 (2021).
4. Byrska-Bishop, M. *et al.* High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* **185**, 3426-3440.e19 (2022).
5. Li, H. *et al.* A synthetic-diploid benchmark for accurate variant-calling evaluation. *Nat Methods* **15**, 595–597 (2018).
6. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
7. Bergström, A. *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020).
8. Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
9. Browning, B. L., Zhou, Y. & Browning, S. R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genetics* **103**, 338–348 (2018).
10. German, J. *et al.* Association between plausible genetic factors and weight loss from GLP1-RA and bariatric surgery. *Nat. Med.* **31**, 2269–2276 (2025).
11. Yang, Y. *et al.* Sex Differences in the Efficacy of Glucagon‐Like Peptide‐1 Receptor Agonists for Weight Reduction: A Systematic Review and Meta‐Analysis. *J. Diabetes* **17**, e70063 (2025).
12. Vosoughi, K., Roghani, R. S. & Camilleri, M. Effects of GLP-1 agonists on proportion of weight loss in obesity with or without diabetes: Systematic review and meta-analysis. *Obes. Med.* **35**, 100456 (2022).
13. Vujkovic, M. *et al.* Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat Genet* **52**, 680–691 (2020).
14. Verma, A. *et al.* Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science* **385**, eadj1182 (2024).

15. Stierman, B. *et al.* NHSR 158. National Health and Nutrition Examination Survey 2017–March 2020 Pre-pandemic Data Files. *Natl. Heal. Stat. Rep.* (2021) doi:10.15620/cdc:106273.
16. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
17. Kerimov, N. *et al.* A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat. Genet.* **53**, 1290–1299 (2021).
18. Teufel, F. *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **40**, 1023–1025 (2022).
19. Fortin, J.-P., Schroeder, J. C., Zhu, Y., Beinborn, M. & Kopin, A. S. Pharmacological Characterization of Human Incretin Receptor Missense Variants. *J. Pharmacol. Exp. Ther.* **332**, 274–280 (2010).
20. Koole, C. *et al.* Polymorphism and Ligand Dependent Changes in Human Glucagon-Like Peptide-1 Receptor (GLP-1R) Function: Allosteric Rescue of Loss of Function Mutation. *Mol. Pharmacol.* **80**, 486–497 (2011).
21. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10**, e1004383 (2014).
22. Foley, C. N. *et al.* A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat Commun* **12**, 764 (2021).
23. The All of Us Research Program Genomics Investigators *et al.* Genomic data in the All of Us Research Program. *Nature* **627**, 340–346 (2024).
24. Xie, Z., Yang, S., Deng, W., Li, J. & Chen, J. Efficacy and Safety of Liraglutide and Semaglutide on Weight Loss in People with Obesity or Overweight: A Systematic Review. *Clin. Epidemiology* **14**, 1463–1476 (2022).
25. Levy, M. E. *et al.* Influence of BMI-associated genetic variants and metabolic risk factors on weight loss with semaglutide: a longitudinal clinico-genomic cohort study. *medRxiv* 2024.10.31.24316494 (2024) doi:10.1101/2024.10.31.24316494.