Genome-wide association studies and polygenic risk score phenome-wide association studies across complex phenotypes in the human phenotype project

Genome Wide Association Studies and Polygenic Risk Score Phenome Wide Association Studies across complex phenotypes in the Human Phenotype Project

Zachary Levine[1,2,3], Iris Kalka[1,2,3], Dmitry Kolobkov[1], Hagai Rossman[1,4], Anastasia Godneva[1], Smadar Shilo[1], Ayya Keshet[1,4], Daphna Weissglas-Volkov[1,4], Tal Shor[1,4], Alon Diament[1,4], Yeela Talmor-Barkan[1,5,6] Yaron Aviv[5,6], Tom Sharon[1], Adina Weinberger[1,4], Eran Segal[1,*]

Summary

Background

Genome-Wide-Association Studies (GWAS) associate phenotypes and genetic variants across a study cohort. GWAS require large-scale cohorts with both phenotype and genetic sequencing data, limiting studied phenotypes. The Human Phenotype Project is a longitudinal study which has measured a wide range of clinical and biomolecular features from a self-assignment cohort over 5 years. The phenotypes collected are quantitative traits, providing higher resolution insights into the genetics of complex phenotypes.

Methods

We present the results of GWAS and Polygenic Risk Score-Phenome-Wide Association Studies with 729 clinical phenotypes and 4,043 molecular features from the Human Phenotype Project. This includes clinical traits that have not been previously associated with genetics, including measures from continuous sleep monitoring, continuous glucose monitoring, liver ultrasound, hormonal status and fundus imaging.

Findings

In GWAS of 8,706 individuals, we found significant associations between 169 clinical traits and 1,184 Single Nucleotide Polymorphisms. We found genes associated with both glycemic control and mental disorders, and we quantify the strength of genetic signals in serum metabolites. In Polygenic Risk Score Phenome-Wide-Association Studies for clinical traits, we found 16,047 significant associations.

Conclusions

The entire set of findings, which we disseminate publicly, provides newfound resolution into the genetic architecture of complex human phenotypes. Our findings are available through an interactive online dashboard at https://zacharylevine.shinyapps.io/GWASDashboard/.

Funding

E.S. is supported by the Minerva foundation with funding from the Federal German Ministry for Education and Research and by the European Research Council and the Israel Science Foundation.

Keywords: Genome-Wide-Association-Study, Polygenic Risk Score, Biobank, UKBB, Genetics, SNPs

[1] Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, 76100, Israel

[2] Senior author

[3] These authors contributed equally

[4] Pheno.AI, Tel-Aviv, Israel

[5] Sackler Faculty of Medicine, Tel Aviv University, Tel-Aviv, 6997801 Israel

[6] Department of Cardiology, Rabin Medical Center, Petah-Tikva, 49100 Israel

* Correspondence and lead contact: eran.segal@weizmann.ac.il

Introduction

The Human Phenotype Project (HPP)

The present study aimed to study the genetic associations for novel complex phenotypes from the HPP. HPP is a large-scale longitudinal study focused on the deep profiling of healthy participants aged between 40 and 70 years from the Israeli general population.The project is centered around the cohort based on voluntary self-assignment and a screening survey and aims to identify novel diagnostic biomarkers and targets for disease. Over the course of five years, the study has collected a wide range of clinical data and molecular data, and aims to uncover markers for disease and treatment targets[1]. Table 1 displays the basic demographic characteristics of the study population. The baseline characteristics of the cohort vary significantly between sexes. As assessed by standard health markers such as waist circumference and Body-Mass-Index (BMI), female HPP participants are healthier than their male counterparts. The HPP cohort is additionally largely composed of educated European (Ashkenazi) Jews.

Why Study Quantitative Traits?

The vast majority of our traits are quantitative, and here we focused on the continuous-scale traits. This not only increases statistical power in a GWAS setting[2], but also allows for detection of genetic signals before disease diagnoses thresholds are reached. Overall, our phenotypes provide unique viewpoints to both identify new genetic associations across the many domains of human health we studied, and understand existing genetic associations in a new light.

What is PRS-PheWAS?

Phenome-wide association studies (PheWAS) test for associations between many phenotypes and a single genetic variant or signal. For most human traits, the proportion of variance explained by individual genetic variants is small[3]. Thus, one approach to study complex phenotypes is to aggregate the contributions of multiple SNPs (weighted by effect sizes) to a phenotype using Polygenic Risk Scores (PRS). PRS-PheWAS is an analogue of standard PheWAS, in which a single genetic variant is replaced by a PRS[4], which is tested for associations with phenotypes. PheWAS has been shown previously to yield insights on the genetic underpinnings of disease[5,6], with follow-up applications to causal inference using Mendelian Randomization[7]. PheWAS normally includes genotypes and phenotypes from the same population. In our PheWAS, we associated each of our traits with PRS from the UK Biobank (UKBB)[8], which we chose because it is the largest biobank with comparable phenotypes to the The Human Phenotype Project (HPP). The UKBB also features pre-existing GWAS results for its traits. Compared to standard PheWAS with SNPs from the HPP, testing for associations to PRS from the UKBB allows for detection of more robust signals, leveraging insights from the larger UKBB study cohort.

Main Findings

We found 1,184 and 16,047 significant associations in GWAS and PRS-PheWAS, respectively. In GWAS we identified genes associated with both glycemic control and mental disorders, and present unexpected genetic associations for glycemic variance. We also found strong genetic signals in serum metabolites.

Results and Discussion

Trait and Variant Filtering

The core components of our datasets are quantitative high-resolution traits extracted from continuous time measurements of sleep monitoring and blood glucose. For blood glucose, these include eA1c, an estimate of Glycated Hemoglobin A1c[9] (HbA1c) levels in the blood, which is useful for diagnosing diabetes, and CoGV, the coefficient of glucose variation over time. These

features are based on continuous glucose monitors (CGMs) worn by participants for two weeks. For sleep, our feature set is based on at-home sleep monitoring data from three nights, and includes clinical tools used to diagnose sleep apnea such as the respiratory disturbance index (RDI) and Apnea-Hypopnea Index (AHI). We additionally included fundus microvasculature features, computed via a deep-learning pipeline for analysis of retinal vascular morphology[10]. These include ocular biomarkers for systemic disease, such as vein vessel density and fractal dimension, which is a measure of the complexity and density of retinal vasculature. We also performed measurements related to liver health (liver ultrasound and two-dimensional shear wave elastography as a surrogate for fibrosis), the ankle brachial index (ABI), pulse wave velocity and blood pressure, body composition via Dual-energy X-ray absorptiometry (DEXA) scan, and lifestyle and hormonal status questionnaires from the HPP. Rather than standard disease diagnoses, the aforementioned measured traits allow for higher resolution insights into the development of disease, thus enabling us to detect signals before a disease diagnosis threshold is crossed. Genetics data is based on shallow sequencing performed by the external third-party Gencove [11,12] on 8,706 subjects, which has been shown to be viable and accurate[13,14].
In our study, we first applied filters for Hardy-Weinberg Equilibrium, minor allele frequency, and first-degree relatives to arrive at a final genetics set consisting of genotypes for 8,932,865 SNPs and 8,706 individuals. We analyzed the clinical phenotypes separately from the molecular phenotypes. After excluding phenotypes with missing data and trait imbalance, we included a total of 729 clinical phenotypes in our analysis, of which 701 were quantitative and 28 were binary (as shown in Table 2). We included observations from the beginning of the study, January 2019, to September 17th 2023, the cut off point.

Considering Clinical Traits, Only SNPs Associated to Glycemic Control Reach Study Wide Significance

To investigate the relationship between individual SNPs and all 729 clinical traits in the HPP, we conducted GWAS at two different thresholds (single trait and study-wide). Using the standard genome-wide significant threshold (P < 5e-8), we identified 160 traits with at least one significant hit, and 14 of these traits remained significant after strict correction for multiple testing (P < 5e-8/729) (Figure 1). We denoted the results that were significant after correction for multiple testing as “multi-trait” hits, and studied these in more detail. The only multi-trait hits were related to glycemic control (e.g., e1Ac, COGI, and quartiles of glucose distribution). Thus, while we provide full results for all data domains in the online dashboard, the present discussion only considers glycemic control traits.

In our insulin genome-wide association study (GWAS) after clumping (originally 8,932,865tests), we observed that the first quartile of each individual’s glucose distribution (a trait computed by the iglu package) had the largest number (4) of multi-trait significant SNPs. Genes containing or located near these SNPs were also implicated in other GWAS studies for childhood obesity[15] (*PCDH15,* P<2E-19), smoking[16] (*PCDH15,* P<4E-21), and mental disorders, predominantly schizophrenia[17–23] (*GPM6A*, *PCDH15*) (P<3E-11, P<2E-19) (Figure 2). Our results are inline with the previously reported shared genetic associations between glycemic control and mental disorders[24].

Features are Related to the Genetic Signals (PRS) of Similar Static Measurements

Due to the polygenic nature of most commonly-studied human traits [25–27], and limitations of GWAS in terms of small effect sizes for most SNPs, we generated 4,116 PRS based on GWAS results from the UKBB for our HPP cohort. The entire set includes PRS for disease diagnoses, quantitative blood test measurements, behavior patterns, and family medical history. We associated each PRS from the UKBB with all of our phenotypes in a PRS-PheWAS analysis (Methods). The strongest PRS associations were found for HPP glycemic traits with the PRS for

glucose, as shown in Table S1. Consistent results were also observed across all data domains (see Supplementary Figures). Overall, our findings suggest that the novel traits are strongly associated with the genetic signals of similar but lower-resolution traits. Based on this validation, any other associations found may be clinically significant.

Within-Day and Between-Day Glycemic Variance contain different Genetic Signals

In PRS-PheWAS, the standard deviation of the mean glucose across days at each time point in the day[28] (SdHHMM) had the highest number of significant associations to PRS, among the entire set of computed PRS. SdHHMM is a measure of within-day glycemic variance, the extent to which blood glucose changes over the course of a day[28]. As our association between this trait and genetics is novel, it provides another angle into previous associations for glycemic control.

The iglu package characterizes glycemic variability using the set of measures assembled by Rodbard[28], including MAGE[29,30], CONGA_n[31], SdW[28], MODD[30,32], SdDM[28], SdHHMM. These measures distinguish between within-day variability and between-day variability. Within-day variability includes MAGE, SdW, and SdHHMM, while between-day variability includes MODD, SdDM, and CONGA. In our study, we used CONGA_24, defined on an interval for the entire day. Rodbard observed weak correlations between the measures of within-day variability and between-day variability[28]. We found this similar dichotomy in our results. We found two different groups of PRS, one that associated significantly with measures of between-day variability, and a different set associated with the within-day measures. For SdHHMM, the significantly associated PRS all predicted body fat mass, including BMI (P<2E-05) and arm (P <1E-3) and leg (P<8E-5) fat mass, and the direction of parameter estimates was negative. This suggests that in people who are genetically predisposed to have higher fat mass, SdHHMM, or within day variability, was significantly lower. This association is not limited to the genetic space: we saw significant negative pearson correlations between SdHHMM and the BMI (P<3E-39), DEXA-measured total fat mass (P<2E-17) and waist circumference (P<2E-18) of HPP participants (see Methods). We found different behavior in between-day variability for the Mean Of Daily paired Differences during consecutive 24h periods (MODD). We found significant associations for MODD with PRS for whether one takes the blood sugar medication metformin (P = 0.0002), diabetes diagnosis (P < 2E-2), and diabetes diagnosis of the mother (P<3E-7), all linked to genetic risk of diabetes. These parameter estimates were all positive, so that increased genetic risk of diabetes is associated with larger between-day glycemic variance. These findings suggest that the relationship between glycemic variability and Type 2 diabetes may involve more complex interactions than initially thought[33], warranting further research to elucidate the underlying mechanisms.

Our Association of Continuous Time Monitors with Genetics is Novel

iglu CGM features have never been associated with genetics before. In addition, while sleep traits have been the subject of GWAS before, the majority of these studies are based on self-reported data, which has been shown to disagree with objective sleep measures [34,35], or were based on data from a single night or with lower resolution[36]. GWAS have been conducted previously on quantitative sleep data from the UK Biobank. These analyses on sleep quality, quantity, and timing estimates are based on actigraphy, and focus only on a limited number of traits[37]. In contrast, we measure more precise sleep indices, reflecting both previously measured sleep patterns, but also sleep apnea related phenomena, from both actigraphy and photoplethysmography.

Molecular Traits: Negative Results for Gut Microbiome but Hope for Metabolites

The gut microbiome is an important determinant of health[38–46]. We therefore tested for associations between gut microbiome composition, specifically relative abundance of different species, and genetics. Compared to most studies, our GWAS is on a larger scale.

To quantify gut microbiome composition, we measured the relative abundance of different species in the gut from a self collected stool sample by HPP participants. We passed these feature sets through the same filtering as other traits (see Methods). Of these, 3645 metabolites and 398 species of gut microbiota were selected based on a balanced distribution, which we defined by no more than 95% of the data having the same value, and a low proportion of missing values, in which we set the minimum allowable sample size to be 2000. These are quantitative molecular traits, as opposed to the clinical traits discussed above, thus, we performed GWAS and PRS analysis separately on them.

Our results suggest that there is untapped potential in genetic analyses of metabolomic features. We found 70 metabolites with at least one multi-trait (study-wide) hit, adjusting for 3,645 metabolites. After a much stricter multiple hypothesis correction than for our clinical traits (differing by a factor of 5), 1.9% of the metabolites had at least one multi-trait significant hit. The metabolite with the most hits had 48 significant SNPs after clumping. These mapped to 32 unique genes, many of which have previous associations to metabolomic[47] (rs62529934, *ARHGAP39*, P<3E-12*)* , mental[19,48–50] (rs184086599, *HCN1*, P<3E-15*)*, and lipidomic[51] (rs61873077, *OLMALINC*, P<4E-15*)* traits in the literature. For each metabolite, we ran over representation analysis on the set of genes containing its significant multi-trait hits after clumping. After correction for the multiple pathways tested for each metabolite, we found 38 pathways with significant overlap, for 9 metabolites. Table 3 describes the pathways we found, and the metabolites that are linked to them.

Among all gut microbiome species cataloged, we found only one gut microbiota species whose relative abundance had multi-trait hits after clumping, Flavionifractor (rs117217109, P<1E-10).

The Connection between Host Genetics and Gut Microbiome Composition is Poorly Understood

With respect to the gut microbiome, we found largely negative results. However, the role of host genetics in determining gut microbiome composition is currently poorly understood[52]. Previous work has found that host genetics perhaps plays only a minor role in determining gut microbiome composition[53]. According to a systematic review published in 2022, while published studies including more than 1,000 subjects have reported dozens of hits for microbiome composition at the genome wide significance level, SNPS in only two genes (LCT, ABO) have been consistently replicated in at least three studies or more. Beyond these two loci, another 546 were reported at genome wide-significance levels in at least one study. While for the LCT gene, genetic variants are consistently associated with the genus Bifidobacterium and its related species, taxa associations for the ABO gene are not consistent across studies[54]. There is a large variance in gut microbiome composition between individuals. In one study, less than 3% of genera were detected in more than 95% of the samples. However, this study pooled multiple cohorts that were heterogenous both in ancestry and methodological processing of samples [55]. Still, this diversity, together with statistical power considerations, could potentially explain the poor replicability of genetic analyses of the gut microbiome[54].

Conclusion

In this study, we performed comprehensive genetic analyses on novel phenotypes from the HPP cohort. Using GWAS we showed that clinical traits measured share genetic associations with related traits in the literature. We compared computed PRS to our traits through PRS-Phenome-Wide Association Studies (PRS-PheWAS). We validated these traits by showing that many of them were associated with the expected PRS from the UK Biobank. We continued to discuss novel findings, including negative linear associations between within-day glycemic variance and

genetic risk for obesity, and positive correlations between genetic risk for diabetes and between-day glycemic variance. Finally, we provided evidence for the presence of strong genetic signals in serum metabolomics. We believe that our findings for glycemic variance could serve as a foundation for further studies on the role played by glycemic control mechanisms in human health. Our results showed novel genotype-phenotype relationships which may lead to improved understanding of human health and disease, and are available for download and interactive exploration (Data Availability).

Limitations of Study

In this study, we transferred PRS from the UK Biobank to the HPP cohort. While prediction performance of the PRS may be reduced as a result of potential ancestry differences between the two cohorts[56], as the 10K cohort is largely composed of European Jews, these effects are assumed to be minimal. Our study is limited by sample size: in any cohort study, there is an inherent tradeoff between phenotyping depth and viable cohort size[57]. As well, as many of our phenotypes have never been connected to genetic data before on a large scale, there is no other data that can be used to replicate our GWAS results on CGM phenotypes, for instance. This lack of replicability is another drawback of our study, though our results are also novel for this reason. As well, our population is a single sample of the general healthy Israeli population, and thus is not fully representative of all people everywhere. Future work should replicate our analysis on a separate cohort that is also more representative of populations. Exclusion criteria such as age restrictions and specific medical conditions, while theoretically limit generalizability, were elected to build a healthy cohort in which diseases may be detectable a few years through follow-up. Still, it is our hope that the public availability of our results can facilitate further investigations into the genetic determinants of human health and disease.

Acknowledgements

We thank members of the Segal lab for useful discussions. E.S. is supported by the Crown Human Genome Center; Larson Charitable Foundation New Scientist Fund; Else Kroener Fresenius Foundation; White Rose International Foundation; Ben B. and Joyce E. Eisenberg Foundation; Nissenbaum Family; Marcos Pinheiro de Andrade and Vanessa Buchheim; Lady Michelle Michels; Aliza Moussaieff; and grants funded by the Minerva foundation with funding from the Federal German Ministry for Education and Research and by the European Research Council and the Israel Science Foundation.

Author Contributions

Z.L. conceived the project, designed and conducted the analyses, interpreted the results and wrote the manuscript. I.K. conceived the project, coordinated and designed data collection, designed the analyses, and interpreted the results. D.K. coordinated and designed data collection, and designed the analyses. H.R. and A.K. provided data and interpreted the results, and A.K. coordinated and designed data collection. A.G. coordinated and designed data collection and provided data. S.S. interpreted the results. D.W.V., A.D., and T.S. provided data, designed the analyses, and interpreted the results. Y.T.B. and Y.A. conceived the project and designed data collection, T.S. conducted the analyses, A.W. coordinated and designed data collection and provided data. E.S. conceived, directed, and supervised the project and analyses

Data Oversight: Z.L. and I.K. had unrestricted access to all data. Z.L. prepared the first draft of the manuscript, and I.K., D.K., H.R., A.K., S.S, A.G., D.W.V., A.D., T.S., Y.T.B. Y.A., and E.S. reviewed and edited it. All authors agreed to submit the manuscript, read and approved the final draft, and take full responsibility for its content, including the accuracy of the data and fidelity of

the trial to the registered protocol and its statistical analyses. All authors read and approved the final article and take responsibility for its content.

Declaration of Interests

H.R., D.W.V, T.S, and A.D. are employees of Pheno.AI, Ltd, a biomedical data science company from Tel-Aviv, Israel. A.W., A.K. and E.S. are paid consultants to Pheno.AI, Ltd. The rest of the authors declare no competing interests.

Main figure titles and legends

Figure 1: There are few multi-trait significant hits after clumping, but many single-GWAS significant SNPs.

Study summary and distribution of significant GWAS associations for all 729 clinical traits both before (P < 5e-8) and after (P < 5e-8/727) multi-trait correction and clumping.

Figure 2: Glycemic variance traits have the highest number of multi-SNP (polygenic risk score) associations. The strongest associations are to the polygenic risk score for blood glucose.

a) GWAS results for all CGM in related traits b) Top multi-trait significant hits (after clumping) for first quartile of glucose c) Manhattan plot for first quartile of glucose d) PRS associations heatmap for glycemic-related traits. “(WD)” denotes measures of within-day glycemic variability, and “(BD)” denotes measures of between-day variability.

Main tables and legends

Feature	Males (46.6%)	Females (53.4%)	All
	Mean (SD)	Mean (SD)	Mean (SD)
Age	51.05 (7.84)	51.91 (7.92)	51.5 (7.90)
Waist Circumference (cm)	94.23 (10.77)	84.80 (11.12)	89.19 (11.93)
BMI	26.54 (3.75)	25.67 (4.40)	26.08 (4.13)
Sitting BP Systolic	126.1 (14.61)	114.53 (15.45)	119.92 (16.13)
Sitting BP Diastolic	80.00 (9.76)	76.61 (9.47)	78.20 (9.76)

Table 1: Baseline features of the HPP cohort

Baseline features of the HPP cohort differed between the sexes.

Domain (Average N)	# Included features
Sleep (N = 7291)	70
Glycemic control(N = 7466)	47
Liver (N = 3672)	97
Body Composition (N = 8440)	436
Retina Scans (N = 7018)	50
Lifestyle Questionnaires (N = 6014)	29
Total Clinical Features	729
Gut Microbiome (N = 4432)	398
Serum Metabolomics (N = 4463)	3645
Total	4772

Table 2: Number of included features for each trait domain

In our analysis, we included 729 clinical and 4,043 molecular features from the HPP, the majority of which measured body composition, and serum metabolite levels, respectively.

Pathway	P. Adj
equilibrioception (GO:0050957)	0.0036
neuromuscular process controlling balance (GO:0050885)	0.0042
negative regulation of amyloid fibril formation (GO:1905907)	0.0045
regulation of amyloid fibril formation (GO:1905906)	0.0045
sensory perception of mechanical stimulus (GO:0050954)	0.0095
sensory perception (GO:0007600)	0.0095
sensory perception of sound (GO:0007605)	0.0095
sensory perception of light stimulus (GO:0050953)	0.0095
negative regulation of protein metabolic process (GO:0051248)	0.0104
negative regulation of supramolecular fiber organization (GO:1902904)	0.0108
dephosphorylation (GO:0016311)	0.0114

protein dephosphorylation (GO:0006470)	0.0114
protein localization to endosome (GO:0036010)	0.0149
cell part morphogenesis (GO:0032990)	0.01496
plasma membrane tubulation (GO:0097320)	0.0149
plasma membrane raft assembly (GO:0044854)	0.0141
caveola assembly (GO:0070836)	0.0141
caveolin-mediated endocytosis (GO:0072584)	0.0149
dosage compensation by inactivation of X chromosome (GO:0009048)	0.0246
microtubule-based transport (GO:0099111)	0.0244
brain morphogenesis (GO:0048854)	0.0245
extracellular transport (GO:0006858)	0.0246
cell projection organization (GO:0030030)	0.0271
epithelial cilium movement involved in extracellular fluid movement (GO:0003351)	0.0291
regulation of sodium ion transport (GO:0002028)	0.038
regulation of ion transport (GO:0043269)	0.0420
regulation of endocytosis (GO:0030100)	0.0422
plasma membrane organization (GO:0007009)	0.0422
spermatid development (GO:0007286)	0.0451
cilium movement (GO:0003341)	0.0451
respiratory system development (GO:0060541)	0.04512
skeletal system morphogenesis (GO:0048705)	0.0451
regulation of vesicle-mediated transport (GO:0060627)	0.0472
plasma membrane bounded cell projection morphogenesis (GO:0120039)	0.0474
neuron migration (GO:0001764)	0.0474
positive regulation of filopodium assembly (GO:0051491)	0.0474
stem cell differentiation (GO:0048863)	0.0474
regulation of filopodium assembly (GO:0051489)	0.0474

335
336

Table 3: Pathways Significantly Genetically Associated to Serum Metabolites

Over representation analysis was performed on sets of genes containing multi-trait significant hits for each metabolite tested. Significant pathways were implicated in pathways including those for sensory perception and cellular mechanisms.

STAR METHODS

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Eran Segal (eran.segal@weizmann.ac.il).

Materials Availability

This study did not generate new unique reagents.

Data and code availability

- GWAS and PR-PheWAS results are publicly accessible at https://zacharylevine.shinyapps.io/GWASDashboard/. Data in this paper is part of the Human Phenotype Project. Individual-level data cannot be released to the public due to ethical considerations. Individual level data is accessible to researchers from universities and other research institutions at https://humanphenotypeproject.org/home, by submitting an application, and signing a Material Transfer Agreement (MTA). Restrictions on the utilization of individual-level data include access through a Trusted Research Environment (TRE), with computational resources available within secure collaborative workspaces.
- All original code is publicly available at Zenodo as of the date of publication. The link is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The age and gender of the participants in the study is described in Table 1. The study population is Israeli subjects, and our cohort is largely composed of Ashkenazi (European-ancestry) Jewish people, of above average socioeconomic status. We do not believe this limits the generalizability of our work, as we adjusted our models for population-level stratification using principal component analysis (PCA). This data was self-reported based on questionnaires. We did not perform genetic analyses that were stratified by sex and gender, opting instead to include and therefore adjust for sex as a covariate in all of the models we used. We do not believe this limits the generalizability of our work. Information on the sample size used for each analysis is available both in Table 2 (stratified by each data modality), and in the raw downloadable results files and visualization dashboards we provide to the public. The Weizmann Institute of Science review board (IRB) approved the study and its protocols. All identifying details of the participants were erased prior to statistical analysis, so informed consent was waived by the IRB. All participants had full knowledge of data handling, storage, and sharing methods. This information was given to all participants, and is in agreement with the data privacy and protection policy of the Weizmann Institute of Science (https://www.weizmann.ac.il/pages/ privacy-policy).

METHOD DETAILS

Study Cohort

HPP is a large-scale longitudinal study focused on the deep physiological and molecular profiling of a diverse cohort of healthy participants from the general population aged between 40 and 70 years in Israel. The HPP cohort is largely composed of European (Ashkenazi) Jews. Over the course of 5 years, the study has collected a wide range of clinical data including vital signs, anthropometry and body composition, blood tests, liver ultrasound and two-dimensional shear wave elastography, continuous glucose monitoring (CGM), continuous food intake logging, sleep monitoring, measures of vascular health, and fundus imaging. Molecular data collected as part of the study includes serum metabolomics and gut microbiome composition. The project is centered around the cohort based on voluntary self-assignment and a screening survey and aims to identify novel diagnostic biomarkers and targets for diseases[1].

Phenotypes

We set the data collection period to be from the beginning of the study, January 2019 to September 2023. Participants in our study wear a continuous glucose monitor (CGM), measuring interstitial fluid glucose every 5 minutes for two weeks using subcutaneous sensors[58]. A set of 47 phenotypes was computed from CGMs using the iglu R package[59], which is a standard method for quantifying glucose control and variability from CGM data[60]. Sleep quality data was collected using a clinically validated sleep monitor for three nights, with quantitative sleep features computed from the continuous time data [61,62].

To measure the relative abundance of different species in the gut microbiome, we analyzed a self collected stool sample by HPP participants. The process of estimating bacterial relative abundance estimation has been detailed previously[63,64]. Denoted as serum metabolites, thousands of small molecules circulating in blood may provide newfound understandings of biological processes[65]. We therefore tested serum metabolites for associations with genetics. Within this data, single features may represent the levels of single metabolites or clusters of related metabolites in blood collected from participants as part of the HPP. As the serum metabolites are untargeted, the identity of individual metabolites is unknown.

QUANTIFICATION AND STATISTICAL ANALYSIS

Outlier Detection and Exclusion

For each of the features included in the HP study, we kept only the latest of multiple entries; removing outliers from the data by clipping it to five standard deviations of the mean. Additionally, we excluded microbiome species with an abundance below 10^{-4} as this was our detection limit. Lastly, we excluded features with less than 2000 entries, and those that were highly unbalanced (>95% frequency for any individual value).

Genotypes

Genomes of participants were sequences at an average depth of 0.6x coverage before being imputed based on HapMap 3[66]. PLINK version 1.9[67] was used for variant filtering. Subjects and SNPs with a minor allele frequency (MAF) below 0.01 were excluded from the analysis. We additionally also removed SNPS that did not meet the Hardy-Weinberg equilibrium threshold of 1e-6. To account for small-scale family structure, we used PLINK's built-in King-cutoff to estimate pairs of first degree relatives, and removed them based on a threshold of 0.22, which accounts for the higher baseline relatedness in our population of primarily Ashkenazi Jews. SNPs with a minimum allele count below 20 were also removed at this stage. We ran Plink's LD-based clumping procedure on each GWAS result afterwards for traits with the trait had at least one

significant (multi-trait) hit, with a significance threshold of 0.0001, a secondary significance threshold for clumped SNPS of 0.01, a LD threshold of 0.5, and a 250 kb physical distance for clumping.

GWAS Analysis

We used PLINK version 2[67] to perform GWAS analyses. To correct for demography and population stratification, for each GWAS we included 12 covariates in addition to the variant term under an additive model. These twelve covariates were composed of the top 10 eigenvectors estimated from principal component analysis as well as the reported age and sex of each participant. We used PLINK's built-in linear transform to normalize each covariate and phenotype to zero mean and unit variance.

PRS Computation

We projected PRS using PRSice-2[68] using the latest GWAS results from the Neale Lab[69] as base data and the HPP genotypes data as target data. We included all traits for which summary statistics were available, using the "both sexes" and inverse rank normalized, "irnt" version for each trait. The summary statistics from the Neale Lab are based on the UK Biobank. We used LD clumping cutpoints of 0.5 (r2), and 250 (kb) and a P value threshold of 0.0001. For each PRS, we excluded points above the 95% percentile or below the 5% percentile then calculated the mean and standard deviation of the distribution, and normalized the data by these values using z-score normalization.

PRS-PheWAS

For PRS involving taking medications, family history of medical diagnoses, or medical diagnoses, we excluded from the association test the HPP participants who met these criteria based on International Classification of Diseases 10th Revision (ICD10) codes[70]. For every trait, we tested its association with every PRS projected from the UK Biobank under a multiple linear model, including as covariates the top 10 principal components as well as age and sex. We then tested for the significance of the difference between the parameter estimate for the PRS and zero using the Wald test, and obtained a P value. We used Bonferroni correction to adjust P values for every test we did. This corresponds to correction for each of 4,116 PRS and 4,772 traits, or a total of 19,641,552 tests.

Overrepresentation Analysis

We mapped the significant multi-trait SNPs from each metabolite to genes based on dbSNP[71]. To perform overrepresentation analysis, we used GSEAPY[72] to match gene sets to the GO Gene Ontology Resource[73–75]. We adjusted original P values using Bonferroni test correction for all gene sets with any overlap for all metabolites, for a total of 2,387 tests.

Phenotype Correlations

We calculated the Pearson correlation coefficients between the SdHHMM and each of BMI, total fat DEXA-measured fat mass, and the waist circumference of individuals. We adjusted P values for multiple tests using Bonferroni correction for each of the two response traits (SdHHMM and BMI) and for each the two fat mass indicators, for a total of four tests.

ADDITIONAL RESOURCES

Our publicly accessible interface for download and visualization of results:

https://zacharylevine.shinyapps.io/GWASDashboard/

Sources

1. Shilo, S., Bar, N., Keshet, A., Talmor-Barkan, Y., Rossman, H., Godneva, A., Aviv, Y., Edlitz, Y., Reicher, L., Kolobkov, D., et al. (2021). 10 K: a large-scale prospective longitudinal study in Israel. Eur. J. Epidemiol. *36*, 1187–1194. 10.1007/s10654-021-00753-5.
2. Yang, J., Wray, N.R., and Visscher, P.M. (2010). Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. Genet. Epidemiol. *34*, 254–257. 10.1002/gepi.20456.
3. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. *101*, 5–22. 10.1016/j.ajhg.2017.06.005.
4. Leppert, B., Millard, L.A.C., Riglin, L., Smith, G.D., Thapar, A., Tilling, K., Walton, E., and Stergiakouli, E. (2020). A cross-disorder PRS-pheWAS of 5 major psychiatric disorders in UK Biobank. PLOS Genet. *16*, e1008185. 10.1371/journal.pgen.1008185.
5. Denny, J.C., Bastarache, L., Ritchie, M.D., Carroll, R.J., Zink, R., Mosley, J.D., Field, J.R., Pulley, J.M., Ramirez, A.H., Bowton, E., et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat. Biotechnol. *31*, 1102–1110. 10.1038/nbt.2749.
6. Fritsche, L.G., Gruber, S.B., Wu, Z., Schmidt, E.M., Zawistowski, M., Moser, S.E., Blanc, V.M., Brummett, C.M., Kheterpal, S., Abecasis, G.R., et al. (2018). Association of Polygenic Risk Scores for Multiple Cancers in a Phenome-wide Study: Results from The Michigan Genomics Initiative. Am. J. Hum. Genet. *102*, 1048–1061. 10.1016/j.ajhg.2018.04.001.
7. Richardson, T.G., Harrison, S., Hemani, G., and Davey Smith, G. (2019). An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. eLife *8*, e43657. 10.7554/eLife.43657.
8. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLOS Med. *12*, e1001779. 10.1371/journal.pmed.1001779.
9. Nathan, D.M., Kuenen, J., Borg, R., Zheng, H., Schoenfeld, D., Heine, R.J., and for the A1c-Derived Average Glucose (ADAG) Study Group (2008). Translating the A1C Assay Into Estimated Average Glucose Values. Diabetes Care *31*, 1473–1478. 10.2337/dc08-0545.
10. Zhou, Y., Wagner, S.K., Chia, M.A., Zhao, A., Woodward-Court, P., Xu, M., Struyven, R., Alexander, D.C., and Keane, P.A. (2022). AutoMorph: Automated Retinal Vascular Morphology Quantification Via a Deep Learning Pipeline. Transl. Vis. Sci. Technol. *11*, 12. 10.1167/tvst.11.7.12.
11. Li, J.H., Mazur, C.A., Berisa, T., and Pickrell, J.K. (2021). Low-pass sequencing increases the power of GWAS and decreases measurement error of polygenic risk scores compared to genotyping arrays. Genome Res. *31*, 529–537. 10.1101/gr.266486.120.
12. The Technical Guide to Low-Pass Sequencing Gencove. https://gencove.com/lp/low-pass-tech-guide.
13. Snelling, W.M., Hoff, J.L., Li, J.H., Kuehn, L.A., Keel, B.N., Lindholm-Perry, A.K., and Pickrell, J.K. (2020). Assessment of Imputation from Low-Pass Sequencing to Predict Merit of Beef Steers. Genes *11*, 1312. 10.3390/genes11111312.
14. Pasaniuc, B., Rohland, N., McLaren, P.J., Garimella, K., Zaitlen, N., Li, H., Gupta, N., Neale, B.M., Daly, M.J., Sklar, P., et al. (2012). Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. Nat. Genet. *44*, 631–635. 10.1038/ng.2283.
15. Costa-Urrutia, P., Colistro, V., Jiménez-Osorio, A.S., Cárdenas-Hernández, H., Solares-Tlapechco, J., Ramirez-Alcántara, M., Granados, J., Ascencio-Montiel, I. de J., and

Rodríguez-Arellano, M.E. (2019). Genome-Wide Association Study of Body Mass Index and Body Fat in Mexican-Mestizo Children. Genes *10*, 945. 10.3390/genes10110945.
16. Saunders, G.R.B., Wang, X., Chen, F., Jang, S.-K., Liu, M., Wang, C., Gao, S., Jiang, Y., Khunsriraksakul, C., Otto, J.M., et al. (2022). Genetic diversity fuels gene discovery for tobacco and alcohol use. Nature *612*, 720–724. 10.1038/s41586-022-05477-4.
17. Sherva, R., Tripodis, Y., Bennett, D.A., Chibnik, L.B., Crane, P.K., de Jager, P.L., Farrer, L.A., Saykin, A.J., Shulman, J.M., Naj, A., et al. (2014). Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. Alzheimers Dement. J. Alzheimers Assoc. *10*, 45–52. 10.1016/j.jalz.2013.01.008.
18. Nazarian, A., Yashin, A.I., and Kulminski, A.M. (2019). Genome-wide analysis of genetic predisposition to Alzheimer's disease and related sex disparities. Alzheimers Res. Ther. *11*, 5. 10.1186/s13195-018-0458-8.
19. Li, Z., Chen, J., Yu, H., He, L., Xu, Y., Zhang, D., Yi, Q., Li, C., Li, X., Shen, J., et al. (2017). Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. Nat. Genet. *49*, 1576–1583. 10.1038/ng.3973.
20. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427. 10.1038/nature13595.
21. Goes, F.S., McGrath, J., Avramopoulos, D., Wolyniec, P., Pirooznia, M., Ruczinski, I., Nestadt, G., Kenny, E.E., Vacic, V., Peters, I., et al. (2015). Genome-wide association study of schizophrenia in Ashkenazi Jews. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet. *168*, 649–659. 10.1002/ajmg.b.32349.
22. Bigdeli, T.B., Fanous, A.H., Li, Y., Rajeevan, N., Sayward, F., Genovese, G., Gupta, R., Radhakrishnan, K., Malhotra, A.K., Sun, N., et al. (2021). Genome-Wide Association Studies of Schizophrenia and Bipolar Disorder in a Diverse Cohort of US Veterans. Schizophr. Bull. *47*, 517–529. 10.1093/schbul/sbaa133.
23. Trubetskoy, V., Pardiñas, A.F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T.B., Bryois, J., Chen, C.-Y., Dennison, C.A., Hall, L.S., et al. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature *604*, 502–508. 10.1038/s41586-022-04434-5.
24. Fanelli, G., Franke, B., De Witte, W., Ruisch, I.H., Haavik, J., van Gils, V., Jansen, W.J., Vos, S.J.B., Lind, L., Buitelaar, J.K., et al. (2022). Insulinopathies of the brain? Genetic overlap between somatic insulin-related and neuropsychiatric disorders. Transl. Psychiatry *12*, 59. 10.1038/s41398-022-01817-0.
25. Shi, H., Kichaev, G., and Pasaniuc, B. (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. Am. J. Hum. Genet. *99*, 139–153. 10.1016/j.ajhg.2016.05.013.
26. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An expanded view of complex traits: from polygenic to omnigenic. Cell *169*, 1177–1186. 10.1016/j.cell.2017.05.038.
27. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. *46*, 1173–1186. 10.1038/ng.3097.
28. Rodbard, D. (2009). New and Improved Methods to Characterize Glycemic Variability Using Continuous Glucose Monitoring. Diabetes Technol. Ther. *11*, 551–565. 10.1089/dia.2009.0015.
29. Service, F.J., Molnar, G.D., Rosevear, J.W., Ackerman, E., Gatewood, L.C., and Taylor, W.F. (1970). Mean amplitude of glycemic excursions, a measure of diabetic instability. Diabetes *19*, 644–655. 10.2337/diab.19.9.644.
30. Service, F.J., O'Brien, P.C., and Rizza, R.A. (1987). Measurements of Glucose Control. Diabetes Care *10*, 225–237. 10.2337/diacare.10.2.225.

31. McDonnell, C.M., Donath, S.M., Vidmar, S.I., Werther, G.A., and Cameron, F.J. (2005). A novel approach to continuous glucose analysis utilizing glycemic variation. Diabetes Technol. Ther. *7*, 253–263. 10.1089/dia.2005.7.253.
32. Service, F.J., and Nelson, R.L. (1980). Characteristics of glycemic stability. Diabetes Care *3*, 58–62. 10.2337/diacare.3.1.58.
33. Salkind, S.J., Huizenga, R., Fonda, S.J., Walker, M.S., and Vigersky, R.A. (2014). Glycemic Variability in Nondiabetic Morbidly Obese Persons. J. Diabetes Sci. Technol. *8*, 1042–1047. 10.1177/1932296814537039.
34. Girschik, J., Fritschi, L., Heyworth, J., and Waters, F. (2012). Validation of Self-Reported Sleep Against Actigraphy. J. Epidemiol. *22*, 462–468. 10.2188/jea.JE20120012.
35. Hughes, J.M., Song, Y., Fung, C.H., Dzierzewski, J.M., Mitchell, M.N., Jouldjian, S., Josephson, K.R., Alessi, C.A., and Martin, J.L. (2018). Measuring Sleep in Vulnerable Older Adults: A Comparison of Subjective and Objective Sleep Measures. Clin. Gerontol. *41*, 145–157. 10.1080/07317115.2017.1408734.
36. Cade, B.E., Chen, H., Stilp, A.M., Gleason, K.J., Sofer, T., Ancoli-Israel, S., Arens, R., Bell, G.I., Below, J.E., Bjonnes, A.C., et al. (2016). Genetic Associations with Obstructive Sleep Apnea Traits in Hispanic/Latino Americans. Am. J. Respir. Crit. Care Med. *194*, 886–897. 10.1164/rccm.201512-2431OC.
37. Jones, S.E., van Hees, V.T., Mazzotti, D.R., Marques-Vidal, P., Sabia, S., van der Spek, A., Dashti, H.S., Engmann, J., Kocevska, D., Tyrrell, J., et al. (2019). Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. Nat. Commun. *10*, 1585. 10.1038/s41467-019-09576-1.
38. Pronovost, G.N., and Hsiao, E.Y. (2019). Perinatal Interactions between the Microbiome, Immunity, and Neurodevelopment. Immunity *50*, 18–36. 10.1016/j.immuni.2018.11.016.
39. Leshem, A., Liwinski, T., and Elinav, E. (2020). Immune-Microbiota Interplay and Colonization Resistance in Infection. Mol. Cell *78*, 597–613. 10.1016/j.molcel.2020.03.001.
40. Foster, J.A., and McVey Neufeld, K.-A. (2013). Gut–brain axis: how the microbiome influences anxiety and depression. Trends Neurosci. *36*, 305–312. 10.1016/j.tins.2013.01.005.
41. Koeth, R.A., Wang, Z., Levison, B.S., Buffa, J.A., Org, E., Sheehy, B.T., Britt, E.B., Fu, X., Wu, Y., Li, L., et al. (2013). Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat. Med. *19*, 576–585. 10.1038/nm.3145.
42. Larsen, N., Vogensen, F.K., Berg, F.W.J. van den, Nielsen, D.S., Andreasen, A.S., Pedersen, B.K., Al-Soud, W.A., Sørensen, S.J., Hansen, L.H., and Jakobsen, M. (2010). Gut Microbiota in Human Adults with Type 2 Diabetes Differs from Non-Diabetic Adults. PLoS One *5*, e9085. 10.1371/journal.pone.0009085.
43. Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. Nature *444*, 1027–1031. 10.1038/nature05414.
44. Arrieta, M.-C., Stiemsma, L.T., Dimitriu, P.A., Thorson, L., Russell, S., Yurist-Doutsch, S., Kuzeljevic, B., Gold, M.J., Britton, H.M., Lefebvre, D.L., et al. (2015). Early infancy microbial and metabolic alterations affect risk of childhood asthma. Sci. Transl. Med. *7*, 307ra152-307ra152. 10.1126/scitranslmed.aab2271.
45. Hills, R.D., Pontefract, B.A., Mishcon, H.R., Black, C.A., Sutton, S.C., and Theberge, C.R. (2019). Gut Microbiome: Profound Implications for Diet and Disease. Nutrients *11*, 1613. 10.3390/nu11071613.
46. Ghosh, T.S., Shanahan, F., and O'Toole, P.W. (2022). Toward an improved definition of a healthy microbiome for healthy aging. Nat. Aging *2*, 1054–1069. 10.1038/s43587-022-00306-9.
47. Genome-wide association studies of metabolites in Finnish men identify disease-relevant loci | Nature Communications https://www.nature.com/articles/s41467-022-29143-5.

48. Periyasamy, S., John, S., Padmavati, R., Rajendren, P., Thirunavukkarasu, P., Gratten, J., Vinkhuyzen, A., McRae, A., Holliday, E.G., Nyholt, D.R., et al. (2019). Association of Schizophrenia Risk With Disordered Niacin Metabolism in an Indian Genome-wide Association Study. JAMA Psychiatry *76*, 1026–1034. 10.1001/jamapsychiatry.2019.1335.
49. Hysi, P.G., Choquet, H., Khawaja, A.P., Wojciechowski, R., Tedja, M.S., Yin, J., Simcoe, M.J., Patasova, K., Mahroo, O.A., Thai, K.K., et al. (2020). Meta-analysis of 542,934 subjects of European ancestry identifies new genes and mechanisms predisposing to refractive error and myopia. Nat. Genet. *52*, 401–407. 10.1038/s41588-020-0599-0.
50. Gouveia, C., Gibbons, E., Dehghani, N., Eapen, J., Guerreiro, R., and Bras, J. (2022). Genome-wide association of polygenic risk extremes for Alzheimer's disease in the UK Biobank. Sci. Rep. *12*, 8404. 10.1038/s41598-022-12391-2.
51. Cadby, G., Giles, C., Melton, P.E., Huynh, K., Mellett, N.A., Duong, T., Nguyen, A., Cinel, M., Smith, A., Olshansky, G., et al. (2022). Comprehensive genetic analysis of the human lipidome identifies loci associated with lipid homeostasis with links to coronary artery disease. Nat. Commun. *13*, 3124. 10.1038/s41467-022-30875-7.
52. Lopera-Maya, E.A., Kurilshikov, A., van der Graaf, A., Hu, S., Andreu-Sánchez, S., Chen, L., Vila, A.V., Gacesa, R., Sinha, T., Collij, V., et al. (2022). Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. Nat. Genet. *54*, 143–151. 10.1038/s41588-021-00992-y.
53. Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., et al. (2018). Environment dominates over host genetics in shaping human gut microbiota. Nature *555*, 210–215. 10.1038/nature25973.
54. Sanna, S., Kurilshikov, A., van der Graaf, A., Fu, J., and Zhernakova, A. (2022). Challenges and future directions for studying effects of host genetics on the gut microbiome. Nat. Genet. *54*, 100–106. 10.1038/s41588-021-00983-z.
55. Kurilshikov, A., Medina-Gomez, C., Bacigalupe, R., Radjabzadeh, D., Wang, J., Demirkan, A., Le Roy, C.I., Raygoza Garay, J.A., Finnicum, C.T., Liu, X., et al. (2021). Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat. Genet. *53*, 156–165. 10.1038/s41588-020-00763-1.
56. Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., Peterson, R., and Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. Nat. Commun. *10*, 3328. 10.1038/s41467-019-11112-0.
57. Shilo, S., Rossman, H., and Segal, E. (2020). Axes of a revolution: challenges and promises of big data in healthcare. Nat. Med. *26*, 29–38. 10.1038/s41591-019-0727-5.
58. Cohort profile: 10K - a large-scale prospective longitudinal study in Israel | medRxiv https://www.medrxiv.org/content/10.1101/2021.02.19.21251487v1.
59. Broll, S., Buchanan, D., Chun, E., Muschelli, J., Fernandes, N., Seo, J.H., Shih, J., Urbanek, J., Schwenck, J., Hicban, M., et al. (2022). iglu: Interpreting Glucose Data from Continuous Glucose Monitors.
60. New and Improved Methods to Characterize Glycemic Variability Using Continuous Glucose Monitoring | Diabetes Technology & Therapeutics https://www.liebertpub.com/doi/10.1089/dia.2009.0015.
61. Pillar, G., Berall, M., Berry, R.B., Etzioni, T., Henkin, Y., Hwang, D., Marai, I., Shehadeh, F., Manthena, P., Rama, A., et al. (2022). Detection of Common Arrhythmias by the Watch-PAT: Expression of Electrical Arrhythmias by Pulse Recording. Nat. Sci. Sleep *14*, 751–763. 10.2147/NSS.S359468.
62. Choi, J.H., Lee, B., Lee, J.Y., and Kim, H.J. (2018). Validating the Watch-PAT for Diagnosing Obstructive Sleep Apnea in Adolescents. J. Clin. Sleep Med. JCSM Off. Publ. Am. Acad. Sleep Med. *14*, 1741–1747. 10.5664/jcsm.7386.
63. Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D., Weinberger, A., Ben-Yacov, O., Lador, D., Avnit-Sagi, T., Lotan-Pompan, M., et al. (2015). Personalized Nutrition by

Prediction of Glycemic Responses. Cell *163*, 1079–1094. 10.1016/j.cell.2015.11.001.
64. Leviatan, S., Shoer, S., Rothschild, D., Gorodetski, M., and Segal, E. (2022). An expanded reference map of the human gut microbiome reveals hundreds of previously unknown species. Nat. Commun. *13*, 3863. 10.1038/s41467-022-31502-1.
65. Bar, N., Korem, T., Weissbrod, O., Zeevi, D., Rothschild, D., Leviatan, S., Kosower, N., Lotan-Pompan, M., Weinberger, A., Le Roy, C.I., et al. (2020). A reference map of potential determinants for the human serum metabolome. Nature *588*, 135–140. 10.1038/s41586-020-2896-2.
66. HapMap 3 - Wellcome Sanger Institute https://www.sanger.ac.uk/resources/downloads/human/hapmap3.html.
67. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7. 10.1186/s13742-015-0047-8.
68. Choi, S.W., and O'Reilly, P.F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. GigaScience *8*, giz082. 10.1093/gigascience/giz082.
69. Abbot, L., Bryant, S., Churchhouse, C., Ganna, A., Howrigan, D., Neale, B., Walters, R., Carey, C., and Hail Team UK Biobank. Neale Lab. https://zenodo.org/record/7186871.
70. World Health Organization (2019). International statistical classification of diseases and related health problems (11th ed.). https://icd.who.int/en.
71. Home - SNP - NCBI https://www.ncbi.nlm.nih.gov/snp/.
72. Fang, Z., Liu, X., and Peltz, G. (2023). GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. Bioinformatics *39*, btac757. 10.1093/bioinformatics/btac757.
73. Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. *47*, D419–D426. 10.1093/nar/gky1038.
74. The Gene Ontology Consortium, Carbon, S., Douglass, E., Good, B.M., Unni, D.R., Harris, N.L., Mungall, C.J., Basu, S., Chisholm, R.L., Dodson, R.J., et al. (2021). The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. *49*, D325–D334. 10.1093/nar/gkaa1113.
75. Gene Ontology: tool for the unification of biology https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3037419/.