Leveraging Multi-Ancestry Polygenic Risk Scores for Body Mass Index to Predict Antiretroviral Therapy-Induced Weight Gain*

Karl Keat[1], Daniel Hui[1], Brenda Xiao[1], Yuki Bradford[2], Zinhle Cindi[3], Eric S. Daar[4], Roy Gulick[5], Sharon A. Riddler[6], Phumla Sinxadi[3], David W. Haas[7,8], Marylyn D. Ritchie[2,9*]
[1]*Genomics and Computational Biology Graduate Program,* [2]*Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA*
[3]*Division of Clinical Pharmacology, Department of Medicine University of Cape Town, Cape Town, South Africa*
[4]*Lundquist Institute at Harbor-UCLA Medical Center Torrance, CA 90502, USA*
[5]*Weill Cornell Medicine, New York, New York, NY 10065, USA*
[6]*University of Pittsburgh, Pittsburgh, PA 15260, USA*
[7]*Vanderbilt University Medical Center, Nashville, TN, USA*
[8]*Meharry Medical College, Nashville, TN, USA*
[9]*Institute for Biomedical Informatics University of Pennsylvania, Philadelphia, PA 19104, USA*
**Email: marylyn@pennmedicine.upenn.edu*

Widespread availability of antiretroviral therapies (ART) for HIV-1 have generated considerable interest in understanding the pharmacogenomics of ART. In some individuals, ART has been associated with excessive weight gain, which disproportionately affects women of African ancestry. The underlying biology of ART-associated weight gain is poorly understood, but some genetic markers which modify weight gain risk have been suggested, with more genetic factors likely remaining undiscovered. To overcome limitations in available sample sizes for genome-wide association studies (GWAS) in people with HIV, we explored whether a multi-ancestry polygenic risk score (PRS) derived from large, publicly available non-HIV GWAS for body mass index (BMI) can achieve high cross-ancestry performance for predicting baseline BMI in diverse, prospective ART clinical trials datasets, and whether that PRS_{BMI} is also associated with change in BMI over 48 weeks on ART. We show that PRS_{BMI} explained ~5-7% of variability in baseline (pre-ART) BMI, with high performance in both European and African genetic ancestry groups, but that PRS_{BMI} was not associated with change in BMI on ART. This study argues against a shared genetic predisposition for baseline (pre-ART) BMI and ART-associated weight gain.

Keywords: HIV; AIDS; Polygenic Risk Scores; BMI; Pharmacogenomics.

1. Introduction

1.1. *Many antiretroviral therapies for HIV are associated with weight gain*

There are ~1.2 million individuals in the United States and ~38 million worldwide living with HIV-1.[1] With >30 FDA-approved antiviral agents for treating HIV-1, many available in combination co-formulated tablets, and with long-acting injectable agents now available, HIV is now a chronic treatable infection in most patients with access to contemporary antiretroviral therapy (ART). However, there remains considerable interindividual variability in HIV treatment responses including drug toxicity, immune recovery, and drug-drug interactions. Variable responses may be influenced by polymorphisms in drug absorption, distribution, metabolism, and

elimination (ADME) genes and/or off-target genes. Beyond the need to develop novel therapies and optimize current therapies are newer priorities which include achieving functional or sterilizing cure of HIV and reducing HIV-associated inflammation and immune activation so as to prevent end-organ complications.

Weight gain following ART initiation is common with most modern ART regimens.[2] The greatest weight gain has been observed in individuals of African ancestry, especially among women of African ancestry. While environmental and social factors likely play a role, there is also the potential for an underlying genetic predisposition.[3] As a few examples among many, it has been shown that, among patients who switched from efavirenz- to integrase strand transfer inhibitor (INSTI)-based ART, *CYP2B6* genotype was associated with weight gain, possibly reflecting withdrawal of inhibitory effects of higher efavirenz levels.[4] Analyses using Phase 1 clinical trials data showed that *CYP2B6* slow metabolizers who switch from efavirenz to dolutegravir will have more prolonged subtherapeutic dolutegravir levels.[5] In ART-naïve AIDS Clinical Trial Group (ACTG) studies, *CYP2B6* slow metabolizers had less weight gain at week 48 in participants receiving efavirenz with tenofovir disoproxil fumarate (TDF) but not those receiving efavirenz with abacavir.[4] We previously discovered and replicated an association between *CYP2B6* 15582C→T (rs4803419) and efavirenz C_{min} in self-identified Black, Hispanic, and white individuals, showed that this single nucleotide polymorphism (SNP) improved prediction of efavirenz plasma exposure in individuals living with HIV in South Africa, and showed that this polymorphism is associated with decreased plasma nevirapine clearance in Asians.[6,7] While we and others have identified potential genetic associations with weight gain, a large proportion of variation remains unexplained. Given this discrepancy, it is plausible that susceptibility to ART-associated excessive weight gain will be affected by each individual's overall genetic predisposition at many genetic loci.

1.2. *Polygenic risk scores allow for prediction of complex traits such as body mass index*

Polygenic risk scores (PRS) are the cumulative, mathematical aggregation of risk derived from the contributions of many DNA variants across the genome. PRS are a powerful technology in the field of disease risk prediction and have been shown to be correlated with disease incidence in coronary artery disease, type 2 diabetes, atrial fibrillation, breast cancer, schizophrenia, and many other traits.[8–15] In recent years there have been advances in PRS methodology that incorporate diverse ancestry groups, quantitative and qualitative phenotypes, and consider different linkage disequilibrium (LD) reference panels.[16–19] In addition, PRS and SNP-based heritability estimation have been applied to body mass index (BMI) in large biobank populations and genome-wide significant SNPs have been shown to explain ~6% of trait interindividual variation in BMI (while considering all common SNPs, the estimate is greater than 20%).[20,21] When considering the underlying genetic predisposition to weight gain in response to ART, is it possible that the underlying genetic background for BMI in populations without HIV will also be predictive of weight gain in response to ART? In this paper, we explore whether susceptibility to ART-associated weight gain is influenced by each individual's overall genetic predisposition to higher BMI as reflected by PRS for BMI (PRS_{BMI}) derived from large datasets from populations without

HIV. Figure 1 shows an overview of our study design, which is described in more details in *Methods*.

2. Methods

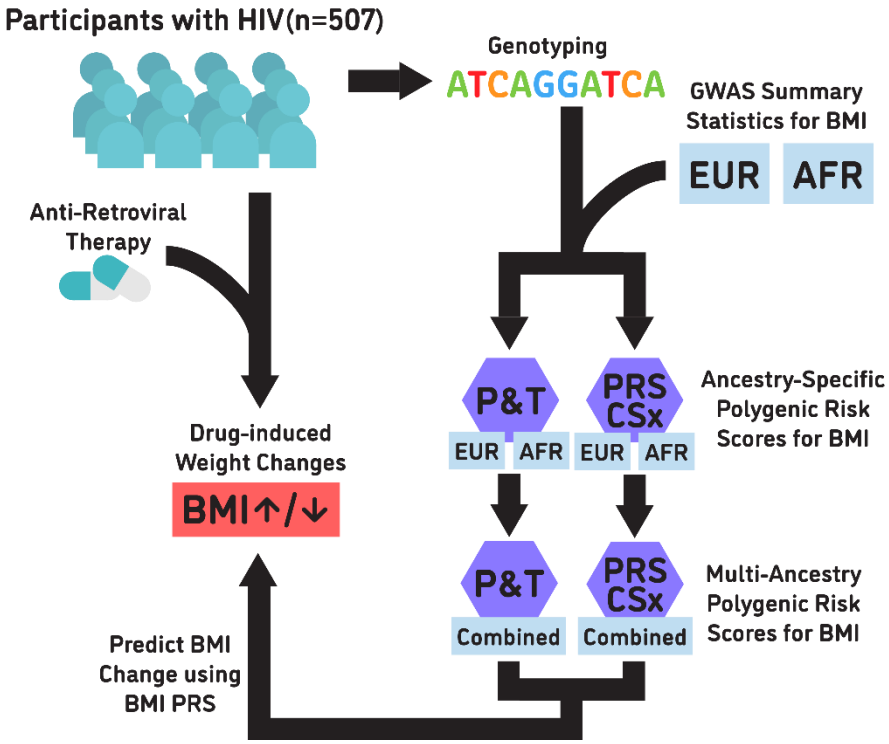

Fig. 1. Study Overview

2.1. *Data and Study Participants*

2.1.1. *GWAS Summary Statistics*

We used publicly available summary statistics from existing genome-wide association studies (GWAS) for BMI in European and African ancestry populations. The European ancestry summary statistics come from the GIANT consortium's meta-analysis of ~700,000 individuals of European ancestry which contained 2,336,269 SNPs.[21] The African ancestry summary statistics come from the African American Anthropometry Genetics Consortium's GWAS of 42,752 individuals which included ~18,000,000 variants.[22] Both sets of summary statistics were subset to the ~1.1 million HapMap3 SNPs included in the PRS-CSx LD reference for the PRS-CSx analysis.[18]

2.1.2. *AIDS Clinical Trials Group Data*

These study data are from a retrospective analysis of a clinical trials cohort from efavirenz-containing arms of prospective, randomized ACTG protocols. Data were from ART-naïve individuals who initiated efavirenz-containing regimens in ACTG studies A5095 (NCT00013520), A5142 (NCT00050895), and A5202 (NCT00118898) in the United States and consented to genetic testing.[23–27] All participants provided written informed consent for genetic research and

provided DNA for analysis. Drug class components of regimens were randomly assigned (efavirenz-based versus comparator) except for nucleoside reverse transcriptase inhibitor (NRTI) choice in A5142. Eligible individuals met the following criteria: initial efavirenz-containing regimens included TDF or abacavir; available weight data at entry and week 48 (± 4 weeks); >100 CD4 T-cells/mm3 at baseline and week 48; HIV-1 RNA <400 copies/mL at week 48; and available *CYP2B6* genotypes. This cohort did not receive INSTIs. The participants' sex was 78.4% male (n = 413) and 21.6% female (n = 114). Data on participants' gender was not available.

2.2. *Quality Control*

2.2.1. *Genotypic Data*

DNA was extracted from whole blood collected from consenting participants, and DNA extracted. Samples were labelled with coded identifiers. Stored DNA was genotyped in seven different phases using different genotyping arrays. Phases 1, 2, and 3 were genotyped at the Broad Institute with HumanHap650Yv3_A for phases 1 and 2, and Human1M-Duov3_B for phase 3. For phases 4-7, genotyping was performed at the Vanderbilt Technologies for Advanced Genomics (VANTAGE) facility using the Human Core Exome chip for phase 4, HumanOmni2.5Exome-8-v1.1_A1 chip for phase 5, the HumanOmni25-8v1-2_A1 chip for phase 6, and the Illumina Infinium Multi-Ethnic Global BeadChip (MEGAEX) for phase 7.

Post-genotype quality control was performed by Vanderbilt Technologies for Advanced Genomics Analysis and Research Design (VANGARD). All quality control steps were performed using PLINK version 1.9.[28] Genotyping efficiency per participant was > 99% for all samples, and discordant samples between genotype sex and reported sex were removed from the datasets prior to imputation. After quality control steps, each genotyping phase was imputed separately using the TOPMed reference panel after transforming to genome build 38 using liftOver and stratification by chromosome to parallelize the imputation process.[29] The seven imputed datasets were merged using PLINK, and we excluded imputed polymorphisms with imputation R^2 scores < 0.3, genotyping call rates < 95%, or minor allele frequency (MAF) < 0.05.[28] Genotype data were transformed back to genome build 37 using liftOver to allow compatibility with the PRS-CSx LD reference panels. Genetic ancestry was inferred using principal component analysis with 1000 Genomes as the reference, to assign each participant to a superpopulation of African (AFR), Admixed American (AMR), East Asian (EAS), European (EUR), South Asian (SAS), or Other.

2.3. *Polygenic Risk Score Construction*

2.3.1. *Pruning and thresholding*

A PRS for baseline BMI (PRS_{BMI}) was created using PRSice 2.3.5 (2021-09-20) for LD clumping and p-value thresholding with default optimization parameters.[17] A multi-ancestry LD reference was generated using data from the 1000 Genomes Project.[30] Optimal p-value thresholds were estimated in a subset of the target data comprising 20% of the total target set (n=105/527) for both the European and African ancestry summary statistics. This threshold was then used to calculate an EUR-derived PRS_{BMI} and AFR-derived PRS_{BMI} for the remaining 80% of individuals. This approach was also used to separately optimize p-value thresholds for predicting BMI change on ART.

2.3.2. *PRS-CSx*

PRS-CSx (version July 29, 2021) was used to construct a multi-ancestry PRS_{BMI}, where both the European and African ancestry summary statistics were jointly adjusted by the model using default optimization parameters to learn the shrinkage factor.[18] The output was then converted to risk scores using the PLINK '--score' function as described in the PRS-CSx documentation.[28] The resulting PRSs were analyzed independently for their performance in each ancestry group and were also linearly combined to create a multi-ancestry PRS_{BMI}. A mixing parameter for the combined PRS_{BMI} was optimized in a subset of the target data comprising 20% of the total target set (n=105/527) and was optimized to minimize the difference in mean PRS_{BMI} between the AFR and EUR ancestry groups. The resulting PRS_{COMB} took the form of $PRS_{COMB} = PRS_{EUR} + \alpha*PRS_{AFR}$ where α is the mixing parameter.

2.4. *Computational and statistical analysis*

All data analyses were performed using python3, scipy, and pandas in a jupyter notebook.[31–33] The distribution of PRS_{BMI} scores was compared between ancestry groups to evaluate systematic ancestry-dependent trends and biases. Performance of each PRS_{BMI} was evaluated as the R^2 value of the PRS_{BMI} in the test set against the phenotype of interest (baseline BMI or change in BMI). Linear regression was used to calculate a p-value for each PRS_{BMI}. For the pre-ART BMI phenotype, we also adjusted for the first 10 principal components, age, sex, and baseline weight in our regression and calculated the incremental performance of our PRS_{BMI} by comparing the PRS_{BMI} + covariates R^2 to the covariates-only model and recorded the p-value for the PRS_{BMI} parameter in the PRS_{BMI} + covariates model. For BMI change, we also adjusted for the first 10 principal components, age and sex, as well as baseline BMI.

3. Results

3.1. *PRS-CSx produces a high-performing multi-ancestry PRS for baseline BMI*

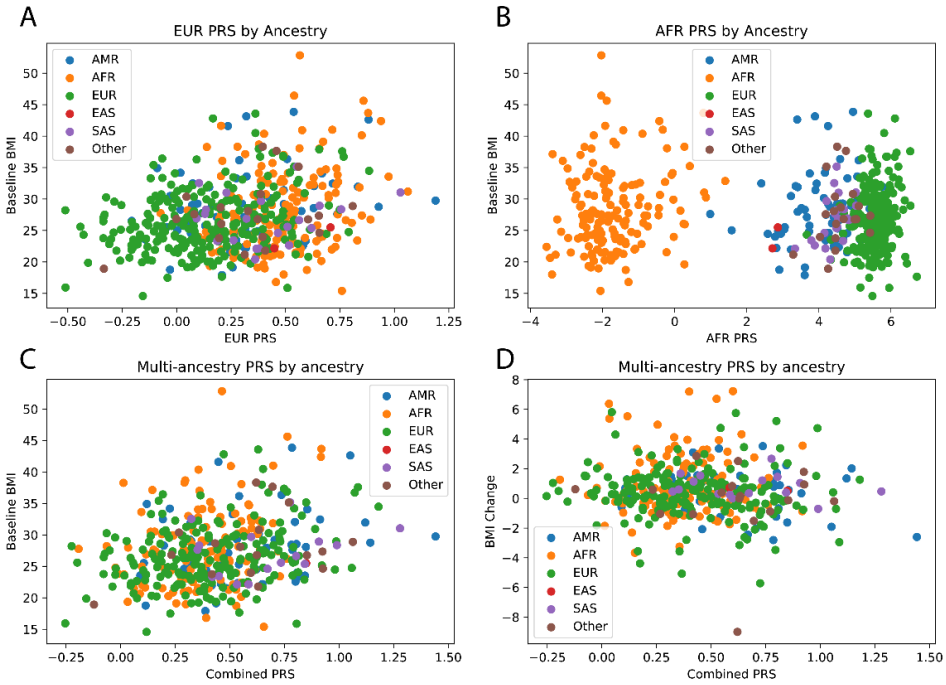

Fig. 2. Distribution of PRS_{BMI} from PRS-CSx in each ancestry group. (A) European-derived PRS_{BMI} vs baseline BMI. (B) African-derived PRS_{AFR} vs baseline BMI. (C) Combined PRS_{AFR} + PRS_{EUR} vs baseline BMI. (D) Combined PRS_{AFR} + PRS_{EUR} vs BMI change from baseline to week 48 on ART.

3.1.1. *PRS_{BMI} generated from European summary statistics systematically overestimate BMI in African ancestry individuals*

Consistent with other work applying PRS across ancestry groups, the EUR-derived PRS_{BMI} (PRS_{EUR}) from PRSice and PRS-CSx both perform best in the EUR ancestry subset of our data and have significant performance decreases in other ancestry groups. Before covariate adjustment, PRS_{EUR} from PRSice performs better at predicting baseline BMI in EUR than the PRS_{EUR} from PRS-CSx, with an R^2 of 0.080 versus 0.070. However, the PRSice PRS_{EUR} performs very poorly in AFR compared to the PRS-CSx PRS_{EUR}, with R^2 in AFR of 0.0032 and 0.055 respectively. Scatterplots of the PRS vs BMI show that the discrepancy in performance is accompanied by a systematic overestimation of AFR BMI in the PRSice PRS_{EUR} (Supplementary Figure 1). This trend is also present in the PRS-CSx results (Figure 1A). Full PRS performance results are provided in Supplementary Table 1. Interestingly, the performance of the PRSice PRS_{EUR} in AMR was high, with an R^2 of 0.110.

3.1.2 *PRS generated from African summary statistics produces a bimodal distribution*

Similar to the trend in PRS_{EUR}, the AFR-derived BMI PRS (PRS_{AFR}) performs better in the AFR ancestry subset of our data, with R^2 in AFR of 0.052 and 0.062 for PRSice and PRS-CSx

respectively. However, the PRS_{AFR} from PRSice performs much worse in EUR than the PRS-CSx one does, with R^2 of 0.0063 and 0.034 respectively. In both the PRSice and PRS-CSx results, the distribution of PRS varies by ancestry, but the difference is particularly pronounced between AFR and EUR, where scores in the AFR population and EUR population from both PRSice and PRS-CSx are entirely disjoint, with the highest AFR score being lower than the lowest EUR score (Figure 1B, Supplementary Figure 1).

3.1.2. *Linear combination of the European and African PRS_{BMI} improves performance in both European and African ancestry populations*

Table 1. Multi-ancestry PRS-CSx PRS_{COMB} performance for BMI prediction in each ancestry group

Target Ancestry	R^2	p-value
EUR (n=206)	0.0725	9.1e-5
AFR (n=128)	0.0795	1.3e-3
AMR (n=43)	0.0674	0.060
Multi-ancestry (n=422)	0.0663	8.1e-8

Given that PRS_{EUR} overestimates BMI in AFR compared to EUR and that PRS_{AFR} underestimates BMI in EUR compared to AFR, we combined the two PRS additively, tuning a mixing parameter such that we minimized the difference in mean combined PRS (PRS_{COMB}) between the AFR and EUR test sets (Table 1). Beyond outperforming both PRS_{AFR} and PRS_{EUR} in AFR test set, the PRS_{COMB} also improves performance in the EUR set. The PRS_{COMB} also improves performance for admixed individuals (AMR) over the PRS-CSx PRS_{EUR} which achieved an R^2 0.056. For comparison purposes, we explored a similar linear combination of the PRSice scores, but to avoid further reducing the sample size, we opted to optimize the combination in the entire test set by also minimizing the difference in mean PRS. Despite the possibility of overfitting to the test data, we found that this approach resulted in drastically diminished performance in the AFR test set, with an R^2 of 0.0016. This seems to indicate that linear combination of PRS_{BMI} from pruning and thresholding is not as effective for creating an unbiased multi-ancestry PRS_{BMI}. Full PRS_{BMI} performance results for predicting BMI in each ancestry group are provided in Supplementary Table 1. Additionally, when we adjust our PRS_{BMI} for the first 10 principal components, age, sex, and height, the incremental performance of PRS-CSx PRS_{COMB} on the entire population is greater than the incremental performance of the PRSice PRS_{COMB} with R^2 increases of 0.053 and 0.038 respectively over the covariates alone. Furthermore, we see that the incremental performance of the PRS-CSx PRS_{COMB} is greater than the incremental performance of the single-ancestry PRS-CSx PRSs (Supplementary Table 2).

3.2. *PRS_{BMI} is not correlated with weight change on antiretroviral therapy*

Table 2. Multi-ancestry PRS_{BMI} performance for weight change prediction in each target ancestry group

Target Ancestry	R^2	p-value
EUR (n=206)	0.0085	0.186
AFR (n=128)	8.97e-07	0.992
AMR (n=43)	0.020	0.305
Multi-ancestry (n=422)	0.0073	0.080

With our high-performing multi-ancestry PRS_{BMI} from PRS-CSx, we then measured its performance in predicting BMI change from baseline to week 48 following initiation of ART. Across all ancestry groups, the PRS_{BMI} was not a significant predictor of weight change and had small R^2 values in all analyses (Table 2). The performance of the other PRSs for BMI change prediction can be found in Supplementary Table 3 with concurrent results. When we subsequently adjust for the first 10 principal components, age, sex, and baseline BMI, we see negligible change in prediction performance or statistical significance (Supplementary Table 4). This evidence further supports the conclusion that weight gain following ART shares little to no underlying genetic predisposition with baseline BMI.

4. Discussion

Our work carries interesting implications for the underlying biology of ART-associated weight gain and for the application of PRS derived from large population GWAS for predicting potentially related traits. First, we were able to successfully construct PRS for BMI (PRS_{BMI}) using large, publicly-available GWAS summary statistics for BMI in different ancestry groups. We showed that while pruning and thresholding produced higher performance in EUR using the EUR summary statistics, PRS-CSx produced a better multi-ancestry PRS, with the exception of the AMR population subset, where pruning and thresholding-based combined PRS performed higher than any other ancestry or PRS. A larger validation set of AMR individuals will be needed to see whether this performance holds, but this could be a consequence of the use of a multi-ancestry subset of the dataset to tune the p-value threshold. Notably, we also demonstrated that our PRS_{BMI} derived from summary statistics from a population without HIV is highly predictive of BMI pre-treatment in individuals with HIV. Through the use of PRS-CSx, we were subsequently able to create a multi-ancestry PRS_{BMI} that performed very well in both EUR and AFR populations. This followed from the peculiar observation that the PRS_{AFR} from both PRSice and PRS-CSx showed a disjoint bimodal distribution where PRS_{AFR} is drastically lower in the AFR subset of the population. Since PRS_{EUR} tends to overestimate BMI in the AFR subset, the PRS_{AFR} can be seen as a "correction factor" for the PRS_{EUR}, increasing scores for EUR and decreasing scores for AFR to mitigate the bias. Despite this trend appearing from both PRSice and PRS-CSx, PRSice did not produce a very effective multi-ancestry PRS.

Despite the strong correlation between our PRS_{BMI} and baseline BMI, the PRS_{BMI} was not well correlated with BMI change in response to ART, and we did not find statistically significant evidence that PRS_{BMI} is associated with BMI change in response to efavirenz-based therapy, even when adjusting for covariates including baseline BMI. Our results provide compelling evidence

that an individual's genetic predisposition based on a common variant PRS for higher BMI may not contribute to greater ART-associated weight gain. It is still possible that other genetic models and/or low frequency variants not captured by PRS may play a role in ART-associated weight gain. Future research on the causes of ART-associated weight gain should explore distinct mechanisms beyond our canonical understanding of the genetics of obesity and BMI.

There are limitations to this work which may have influenced our results. First, our PRS_{BMI} testing sample size was limited to approximately 500 individuals, and when subdivided by ancestry the sample sizes become smaller, limiting our power to find associations between our PRS and target traits. As such, it remains a possibility that PRS_{BMI} could be associated with ART-associated weight gain, but at a smaller effect size than we could detect given our statistical power. Additionally, due to particularly small sample sizes of East Asian and South Asian individuals, we mostly focused on cross-ancestry performance in EUR, AFR, and AMR populations, as well as in the entire population. Finally, it is also worth noting that integrase inhibitor-associated weight gain is greater than efavirenz-associated weight gain and that integrase inhibitors are currently the preferred initial therapy for most people. The ACTG cohorts included in this study did not receive INSTIs; thus the effect sizes may be larger if this investigation was repeated in a cohort of individuals who experienced weight gain after receiving INSTIs.

Subsequent work in this area could investigate how other covariates may influence BMI change. In further exploration of the use of large sample-size GWAS to construct PRS for drug response traits, one could study other phenotypes, such as how GWAS for liver function tests (such as alanine transaminase (ALT) and aspartate transaminase (AST)) may be predictive of adverse liver events, or whether a PRS derived from GWAS for major depressive disorder is predictive of neurological effects of ART. These approaches have the potential to leverage large, publicly available datasets to generate new discoveries in smaller pharmacogenetic cohorts. As more associations or lack thereof are found, we continue to narrow down the likely biological causes of adverse drug reactions such as excessive weight gain, bringing us closer to the true etiology.

5. Acknowledgments

The authors are grateful to the many persons living with HIV who volunteered for ACTG protocols A5095, A5142 and A5202. In addition, they acknowledge the contributions of study teams and site staff for these protocols. We thank Paul J. McLaren, PhD (Public Health Agency of Canada, Winnipeg, Canada) for prior involvement and collaborations that used these genome-wide genotype data. Study drugs were provided by DuPont Pharmaceutical Company, Bristol-Myers Squibb, Inc., GlaxoWellcome, Inc., Gilead Sciences, Inc., GlaxoSmithKline, Inc.. The clinical trials were A5095 (NCT00013520), A5142 (NCT00050895), and A5202 (NCT00118898).

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number UM1 AI068634, UM1 AI068636 and UM1 AI106701. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of

Health, Supported in part by grants funded by the National Center for Research Resources and the National Center for Advancing Translational Sciences.

Grant support included TR000124 (to E.S.D.); AI110527, AI077505, TR000445, and AI069439 (to D.W.H.). This work was supported by the Tennessee Center for AIDS Research (P30) AI110527.

Clinical research sites that participated in ACTG protocols A5095, A5142 and/or A5202, and collected DNA under protocol A5128 were supported by the following grants from the National Institutes of Health (NIH): A1069412, A1069423, A1069424, A1069503, AI025859, AI025868, AI027658, AI027661, AI027666, AI027675, AI032782, AI034853, AI038858, AI045008, AI046370, AI046376, AI050409, AI050410, AI050410, AI058740, AI060354, AI068636, AI069412, AI069415, AI069418, AI069419, AI069423, AI069424, AI069428, AI069432, AI069432, AI069434, AI069439, AI069447, AI069450, AI069452, AI069465, AI069467, AI069470, AI069471, AI069472, AI069474, AI069477, AI069481, AI069484, AI069494, AI069495, AI069496, AI069501, AI069501, AI069502, AI069503, AI069511, AI069513, AI069532, AI069534, AI069556, AI072626, AI073961, RR000046, RR000425, RR023561, RR024156, RR024160, RR024996, RR025008, RR025747, RR025777, RR025780, TR000004, TR000058, TR000124, TR000170, TR000439, TR000445, TR000457, TR001079, TR001082, TR001111, and TR024160.

Supplementary Figures/Tables

Supplementary Figure 1. PRSice PRS for BMI plotted against baseline BMI
Supplementary Table 1. Performance of each PRS for predicting baseline BMI
Supplementary Table 2. Incremental performance of each PRS for predicting baseline BMI
Supplementary Table 3. Performance of each PRS for predicting BMI change
Supplementary Table 4. Incremental performance of each PRS for predicting BMI change

All supplemental data can be found at the links above or at:
https://ritchielab.org/publications/supplementary-data/psb-2023/actg-bmi-prs.

References

1. HIV/AIDS. https://www.who.int/data/gho/data/themes/hiv-aids.
2. Lake, J. E. *et al.* Risk Factors for Weight Gain Following Switch to Integrase Inhibitor–Based Antiretroviral Therapy. *Clin. Infect. Dis.* **71**, e471–e477 (2020).
3. Erlandson, K. M. *et al.* Mitochondrial DNA haplogroups and weight gain following switch to integrase strand transfer inhibitor-based antiretroviral therapy. *AIDS* **35**, 439–445 (2021).
4. Leonard, M. A. *et al.* Efavirenz Pharmacogenetics and Weight Gain Following Switch to Integrase Inhibitor-Containing Regimens. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **73**, e2153–e2163 (2021).
5. Haas, D. W. *et al.* Pharmacogenetic interactions of rifapentine plus isoniazid with efavirenz or nevirapine. *Pharmacogenet. Genomics* **31**, 17–27 (2021).
6. Holzinger, E. R. *et al.* Genome-wide association study of plasma efavirenz pharmacokinetics in AIDS Clinical Trials Group protocols implicates several CYP2B6 variants. *Pharmacogenet. Genomics* **22**, 858–867 (2012).
7. Bertrand, J. *et al.* Multiple genetic variants predict steady-state nevirapine clearance in HIV-infected Cambodians. *Pharmacogenet. Genomics* **22**, 868–876 (2012).

8. Rao, A. S. & Knowles, J. W. Polygenic risk scores in coronary artery disease. *Curr. Opin. Cardiol.* **34**, 435–440 (2019).
9. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
10. McCarthy, M. I. & Mahajan, A. The value of genetic risk scores in precision medicine for diabetes. *Expert Rev. Precis. Med. Drug Dev.* **3**, 279–281 (2018).
11. Pulit, S. L. *et al.* Atrial fibrillation genetic risk differentiates cardioembolic stroke from other stroke subtypes. *Neurol. Genet.* **4**, (2018).
12. Mavaddat, N. *et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am. J. Hum. Genet.* **104**, 21–34 (2019).
13. Li, H. *et al.* Breast cancer risk prediction using a polygenic risk score in the familial setting: a prospective study from the Breast Cancer Family Registry and kConFab. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **19**, 30–35 (2017).
14. Wimberley, T. *et al.* Polygenic Risk Score for Schizophrenia and Treatment-Resistant Schizophrenia. *Schizophr. Bull.* **43**, 1064–1069 (2017).
15. Miller, A. P. *et al.* Polygenic liability for schizophrenia predicts shifting-specific executive function deficits and tobacco use in a moderate drinking community sample. *Psychiatry Res.* **279**, 47–54 (2019).
16. Ni, G. *et al.* A Comparison of Ten Polygenic Score Methods for Psychiatric Disorders Applied Across Multiple Cohorts. *Biol. Psychiatry* **90**, 611–620 (2021).
17. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience* **8**, giz082 (2019).
18. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* **54**, 573–580 (2022).
19. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2020).
20. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
21. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
22. Ng, M. C. Y. *et al.* Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium. *PLOS Genet.* **13**, e1006719 (2017).
23. Gulick, R. M. *et al.* Triple-Nucleoside Regimens versus Efavirenz-Containing Regimens for the Initial Treatment of HIV-1 Infection. *N. Engl. J. Med.* **350**, 1850–1861 (2004).
24. Gulick, R. M. *et al.* Three- vs Four-Drug Antiretroviral Regimens for the Initial Treatment of HIV-1 InfectionA Randomized Controlled Trial. *JAMA* **296**, 769–781 (2006).
25. Riddler, S. A. *et al.* Class-Sparing Regimens for Initial Treatment of HIV-1 Infection. *N. Engl. J. Med.* **358**, 2095–2106 (2008).
26. Daar, E. S. *et al.* Atazanavir plus ritonavir or efavirenz as part of a 3-drug regimen for initial treatment of HIV-1. *Ann. Intern. Med.* **154**, 445–456 (2011).
27. Haas, D. W. *et al.* A multi-investigator/institutional DNA bank for AIDS-related human genetic studies: AACTG Protocol A5128. *HIV Clin. Trials* **4**, 287–300 (2003).
28. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).

29. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590-598 (2006).
30. Xiao, B. *et al.* Inference of causal relationships based on the genetics of cardiometabolic traits and conditions unique to females in >50,000 participants. 2022.02.02.22269844 Preprint at https://doi.org/10.1101/2022.02.02.22269844 (2022).
31. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
32. Reback, J. *et al.* pandas-dev/pandas: Pandas 1.4.3. (2022) doi:10.5281/zenodo.6702671.
33. McKinney, W. Data Structures for Statistical Computing in Python. *Proc. 9th Python Sci. Conf.* 56–61 (2010) doi:10.25080/Majora-92bf1922-00a.