Lymphocyte Count Derived Polygenic Score and Interindividual Variability in CD4 T-cell Recovery in Response to Antiretroviral Therapy

Kathleen M. Cardone1, Scott Dudek1, Karl Kear1, Yuki Bradford1, Zinhle Cindi1,2, Eric S. Daar3, Roy Gulick4, Sharon A. Riddler5, Jeffrey L. Lennox6, Phumla Sinxadi7, David W. Haas8,9, Marylyn D. Ritchie1,10
1Department of Genetics, 2Genomics and Computational Biology Graduate Program University of Pennsylvania, Philadelphia, PA, USA
3Lundquist Institute at Harbor-UCLA Medical Center, Torrance, CA, USA
4Weill Cornell Medicine New York, New York, NY, USA
5University of Pittsburgh, Pittsburgh, PA, USA
6Emory University School of Medicine, Atlanta, GA, USA
7Division of Clinical Pharmacology, Department of Medicine University of Cape Town, Cape Town, South Africa
8Vanderbilt University Medical Center, Nashville, TN, USA
9Meharry Medical College, Nashville, TN, USA
10Institute for Biomedical Informatics University of Pennsylvania, Philadelphia, PA, USA
*Email: marylyn@pennmedicine.upenn.edu

Access to safe and effective antiretroviral therapy (ART) is a cornerstone in the global response to the HIV pandemic. Among people living with HIV, there is considerable interindividual variability in absolute CD4 T-cell recovery following initiation of virally suppressive ART. The contribution of host genetics to this variability is not well understood. We explored the contribution of a polygenic score which was derived from large, publicly available summary statistics for absolute lymphocyte count from individuals in the general population (PGSlymph) due to a lack of publicly available summary statistics for CD4 T-cell count. We explored associations with baseline CD4 T-cell count prior to ART initiation (n=4959) and change from baseline to week 48 on ART (n=3274) among treatment-naïve participants in prospective, randomized ART studies of the AIDS Clinical Trials Group. We separately examined an African-ancestry-derived and a European-ancestry-derived PGSlymph, and evaluated their performance across all participants, and also in the African and European ancestral groups separately. Multivariate models that included PGSlymph, baseline plasma HIV-1 RNA, age, sex, and 15 principal components (PCs) of genetic similarity explained ~26-27% of variability in baseline CD4 T-cell count, but PGSlymph accounted for <1% of this variability. Models that also included baseline CD4 T-cell count explained ~7-9% of variability in CD4 T-cell count increase on ART, but PGSlymph accounted for <1% of this variability. In univariate analyses, PGSlymph was not significantly associated with baseline or change in CD4 T-cell count. Among individuals of African ancestry, the African PGSlymph term in the multivariate model was significantly associated with change in CD4 T-cell count while not significant in the univariate model. When applied to lymphocyte count in a general medical biobank population (Penn Medicine BioBank), PGSlymph explained ~6-10% of variability in multivariate models (including age, sex, and PCs) but only ~1% in univariate models. In summary, a lymphocyte count PGS derived from the general population was not consistently associated with CD4 T-cell recovery on ART. Nonetheless, adjusting for clinical covariates is quite important when estimating such polygenic effects.

Keywords: HIV; Polygenic Scores; Lymphocyte Count; CD4 T-Cell Count; Pharmacogenomics

© 2023 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.
1. Introduction

1.1. Incomplete CD4 T-Cell Recovery in Response to Antiretroviral Therapy

Human immunodeficiency virus type 1 (HIV-1) is a global health challenge, with 38.4 million individuals worldwide living with HIV\(^1\), including nearly 1.2 million in the United States\(^2\). This virus depletes CD4 T lymphocytes (hereafter referred to as CD4 cells), a critical component of the immune system\(^3\). Effective antiretroviral therapy (ART) controls viral replication, improves health and prevents transmission\(^4\). With viral load reduction, CD4 cell counts may return to normal levels, but in many individuals this is not achieved\(^5\—^7\). Understanding the etiology of CD4 cell recovery is important because individuals with lower CD4 cell counts may be at increased risk for non-AIDS conditions such as hepatic cirrhosis, cardiovascular disease, kidney disease, and cancer\(^8\).

The etiology of incomplete CD4 cell recovery has not been fully elucidated, but many biological, demographic, treatment, and genetic factors have been associated\(^9\). Individuals who begin ART with CD4 cell counts <200 cells/mm\(^3\) are less likely to achieve normal CD4 cell counts >500 cells/mm\(^3\)^5—^7. Other biological factors associated with this treatment response include higher body mass index (BMI), lower naïve/memory CD4\(^+\) cell ratio, lower CD4/CD8 cell ratios, and other immunological factors\(^9\). Demographic factors have also been associated with poor CD4 cell recovery including older age, male sex, and Eastern African ancestry, as well as specific ART regimens\(^9, 10\). Additionally, variants that influence the absorption, distribution, metabolism, and elimination of ART may also play a role\(^1\). Genes with single nucleotide polymorphisms (SNPs) reported to be associated with CD4 cell recovery on ART have included IL-2, IL-2Rβ, IL-2Rγ, IL-15, IL-15Rα, TRAIL, Bim, TNF-α, and IFN-γ\(^12\). One particular SNP (rs6897932) in IL7RA was associated with a faster CD4 cell count increase in individuals of both European and African ancestry, but another SNP in this gene (rs3194051) was only associated with this response in individuals of African ancestry\(^13, 14\). Another study suggested that differences in CCR5 genotype and CCL3L1 dosage were associated with the extent and rate of CD4 cell recovery\(^15\). Additionally, HLA-Bw4 homozygosity was associated with impaired CD4 cell recovery\(^16\). Particular mitochondrial DNA haplogroups were associated with CD4 cell recovery in individuals of European and African ancestry\(^17, 18\). More recently, whole exome sequencing associated 41 genes with CD4 cell response in females\(^19\).

Although multiple genes and SNPs have been associated with poor CD4 cell count recovery on ART, these explain a small fraction of the variance. Previous studies considered effects of SNPs individually, which fails to consider whether combinations of many SNPs may explain a larger portion of the variance. Many conditions are polygenic (e.g., coronary artery disease), meaning that many genes and variants have impact\(^20\). It is conceivable that CD4 cell recovery on ART is also polygenic, so it is worth exploring whether polygenic scores may explain a larger portion of the genetic variance, which has never been investigated for this treatment response. Furthermore, understanding the
pharmacogenomic underpinnings of treatment response has the potential to better individualize therapy.

1.2 Polygenic Scores May Predict Complex Treatment Responses

One way to assess the contribution of many variants in combination is by applying Polygenic Scores (PGS), which are the mathematical, cumulative aggregation of risk derived from the total contribution of numerous variants in the genome. PGS effectively predict phenotypes such as schizophrenia, bipolar disorder, breast cancer, type 2 diabetes, coronary artery disease, and atrial fibrillation. Given their success in other disease areas, it is plausible that PGS could predict poor CD4 cell recovery in response to ART.

When using PGS, it is important to consider the potential for ancestral health disparity. Across many phenotypes, PGS is more predictive for individuals of European ancestry because this population has more readily available summary statistics from large genome-wide association studies (GWAS). An ultimate goal of PGS is clinical implementation so that patients can be informed of their genetic risk for disease. However, clinical implementation could create a larger health disparity whereby individuals of European ancestry may more readily benefit from these risk prediction models. Thus, it is important to improve risk prediction for global populations. This is particularly important for HIV given its global distribution of prevalence, particularly in Africa. We hope to better predict genetic risk in individuals of African ancestry by generating a PGS based on summary statistics generated in a dataset of individuals largely of African ancestry, in addition to a PGS generated in a dataset of individuals largely of European ancestry. Additionally, we plan to use PRScsx, a method that more effectively predicts polygenic risk in global populations.

In this study, we assess whether the PGS generated from a general population is predictive of CD4 cell recovery in persons living with HIV (PWH). A similar approach used a body mass index PGS generated from a general population to study ART-associated weight gain. As there are no large GWAS studies of CD4 cell count, either in the general population or in PWH, we generate statistical power by using summary statistics on total lymphocyte count from a general population, for which large sample sizes are publicly available. Finally, the principle of predicting phenotypic effects in a population affected by a health condition by using genetics from the general population was effective in one study that found that variants associated with cardiac QRS duration in individuals without cardiac diseases were also associated with arrhythmia and atrial fibrillation. We assess whether this same principle applies to treatment response by testing whether the genetic underpinnings of lymphocyte count in a general population predicts CD4 cell recovery in PWH. We hypothesize that cumulative genetic variants that affect total lymphocyte count also affect recovery of the CD4 T cell subset in response to ART (i.e., that a lymphocyte count PGS generated from the general population will be associated with CD4 cell recovery on ART). We also hypothesize that PGS will be associated with CD4 cell counts prior to initiating ART.
2. Methods

Figure 1: Study Overview: EUR and AFR PGSlymph were trained using lymphocyte count GWAS summary statistics. Both PGSlymph were applied to individuals in the AIDS Clinical Trials Group (ACTG) to assess its predictability of CD4 cell response to ART.

2.1 Data and Study Participants

2.1.1 Lymphocyte Count Meta Analysis

We used publicly available summary statistics from a published meta-analysis of existing GWAS for lymphocyte count in populations of European and African ancestry in the general population\(^1\). The meta-analysis included 524,923 individuals of European ancestry with 47,264,266 SNPs, and 13,477 individuals of African ancestry with 34,121,887 SNPs\(^1\). The European ancestry summary statistics were subset to 1,120,498 SNPs that were present on the European linkage disequilibrium (LD) panels and the African ancestry summary statistics were subset to 1,225,091 SNPs that were present on the African LD reference panels.

2.1.2 AIDS Clinical Trials Group

Participants were ART-naïve individuals who had initiated ART in prospective, randomized clinical trials of the AIDS Clinical Trials Group (ACTG), and had consented to genetic research and provided DNA under ACTG protocol A5128\(^3\). Data were generated by conducting a retrospective analysis of
this cohort. Individuals had initiated ART in the United States in studies ACTG384, A5095 (NCT00013520), A5142 (NCT00050895), A5202 (NCT00118898), and A5257 (NCT25285539)44–47. All participants provided written, informed consent for genetic testing. Drug class components of regimens were randomly assigned except for nucleoside reverse transcriptase inhibitor (NRTI) choice in A5142. Included individuals had the following data: imputed genotype, sex, genetically inferred ancestry (GIA), lymphocyte count or CD4 cell count data. Additional eligibility criteria included HIV-1 RNA <400 copies/mL at week 48 on ART.

2.1.3 Penn Medicine BioBank

The Penn Medicine BioBank (PMBB) is an electronic health record (EHR)-linked biobank research program at the University of Pennsylvania48. PMBB participants included in this study provided consent for research including access to their medical records, blood sample collection, and generation of genetic data48. Individuals with both imputed genotype data from PMBB v2.0 and with lymphocyte count data were included in PGS analysis as a positive control. Included individuals had the following data: imputed genotype, lymphocyte count, sex, and GIA.

2.2 Genotyping and Quality Control

2.2.1 AIDS Clinical Trials Group

DNA extracted from whole blood was labeled with coded identifiers and genotyped in seven phases. Phases 1-3 were genotyped at the Broad Institute (Phases 1 and 2 with HumanHap650Yv3_A, and Phase 3 with Human1M-Duov3_B). Phases 4-7 were genotyped at the Vanderbilt Technologies for Advanced Genomics (VANTAGE) facility (Phase 4 using the Human Core Exome chip, phase 5 with the HumanOmmi2.5Exome-8-v1.1_A1 chip, Phase 6 with the HumanOmmi25-8v1-2_A1 chip, and phase 7 with the Illumina Infinium Multi-Ethnic Global BeadChip (MEGAEX).

Post-genotype quality control procedures utilizing PLINK v1.949 were conducted by Vanderbilt Technologies for Advanced Genomics Analysis and Research Design (VANGARD). Prior to imputation, samples with genotyping efficiency < 99% or with discordance between genotype sex and reported sex were removed. After completing these quality control procedures, each genotyping phase was imputed separately utilizing the TOPMed reference panel, which was parallelized by chromosome to increase computational efficiency50. During the imputation process, liftOver was used to transform genotype data to genome build 3850. After imputation, PLINK was used to merge the seven imputed datasets, and variants with imputation R² scores < 0.3, genotyping call rates < 95%, or minor allele frequency (MAF) < 0.05 were dropped50. GIA was determined using principal component analysis (PCA) with 1000 Genomes as the reference, subsequently assigning each participant to one of six
superpopulations: African (AFR), Admixed American (AMR), East Asian (EAS), European (EUR), South Asian (SAS), and Other.

2.2.2 Penn Medicine BioBank

DNA was extracted from blood samples. Approximately 80% of samples were genotyped by the Regeneron Genomics Center (RGC) using an Illumina Global Screening Array v.2.0 (GSAv2)\(^4\), while the remaining 20% were genotyped by the Center for Applied Genomics (CAG) at the Children’s Hospital of Philadelphia using the GSAv1 and GSAv2 genotyping array\(^4\).

Prior to imputation, sample level quality control was conducted\(^4\). Using PLINK v1.9, variants with genotyping call rates < 95%, individuals with sample call rates < 90%, and individuals with discordance between reported sex and genotype sex were dropped\(^4\). Autosomes were imputed utilizing a TOPMed version R2 genome build 38 reference panel\(^4,5\). After imputation, variants with imputation R\(^2\) scores < 0.3, genotype call rate < 99%, MAF < 1%, and/or were multi-allelic were dropped using PLINK v1.9\(^4\). Individuals with sample call rate < 99% or discordant sex information were also dropped\(^4\). PCA was done to identify GIA using 1000 Genomes as the reference and subsequently separated individuals into six superpopulations: African (AFR), Admixed American (AMR), East Asian (EAS), European (EUR), South Asian (SAS), Other\(^4\).

2.3 Polygenic Score Calculation

The PGS\(_{\text{lymph}}\) was constructed using PRScsx (version released on July 29 2021), which integrates summary statistics and LD panels across genetically diverse populations to better predict polygenic risk in global populations\(^3\). 1000 Genomes phase 3 LD reference panels were used in the calculation\(^5\). Summary statistics from the lymphocyte count meta-analysis were used to train the PGS\(_{\text{lymph}}\)\(^4,6\). The PGS\(_{\text{lymph}}\) was applied to ACTG study participants with CD4 cell count data using PLINK2 “--score” function\(^4\). As positive controls, the PGS\(_{\text{lymph}}\) was also applied to individuals with lymphocyte count data in ACTG as well as individuals with lymphocyte count data in PMBB.

2.4 Statistical Analysis

The results were analyzed to assess model predictability across all ancestries combined, and in European and African ancestries separately. Linear regressions were calculated, and performance was assessed with an R\(^2\) value generated from a multivariate linear regression between the phenotype of interest and the PGS\(_{\text{lymph}}\). Additionally, performance of individual covariates was assessed with effect sizes generated from these regressions. We used a p-value threshold of 0.05 to assess significance. Regressions were calculated in individuals of European and African ancestry only, as well as individuals of all superpopulations combined. PGS\(_{\text{lymph}}\) was applied to two different cohorts, ACTG...
and PMBB. In ACTG, the predictability of the PGS_{lymph} for three different phenotypes was assessed: the square root (SQRT) of CD4 cell count at study entry prior to ART (baseline), change in CD4 cell count from study entry to 48 weeks of ART (a measure of treatment response), and inverse normal lymphocyte count prior to ART (a control variable). We performed two regressions for each phenotype, one without correcting for any covariates, and one correcting for age, sex, principal components (PC) of genetic similarity 1-15, as well as log_{10} HIV-1 RNA (a measure of viral load). Additionally, we adjusted for SQRT of baseline CD4 cell count in regression models between PGS_{lymph} and change in CD4 cell count on ART. In addition to these regressions, we also evaluated interactions between the PGS_{lymph} and age, sex, viral load, and baseline CD4 cell count to identify whether PGS_{lymph} interacts with any covariate. In PMBB, the predictability of PGS_{lymph} for inverse normal lymphocyte count was assessed as a positive control and to understand predictability in a general medical biobank population. Similarly, two regressions were performed, one without correcting for covariates, and one correcting for age, sex, and PC1-15. These results were visualized using SynthesisView\textsuperscript{52}.

3. Results

<table>
<thead>
<tr>
<th>Table 1: ACTG Participant Demographics at Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lymphocyte Count Data</strong></td>
</tr>
<tr>
<td>Total, N</td>
</tr>
<tr>
<td>European ancestry, n (%)</td>
</tr>
<tr>
<td>African ancestry, n (%)</td>
</tr>
<tr>
<td>Male/Female, n (%)</td>
</tr>
<tr>
<td>Age, mean (range)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2: PMBB Demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lymphocyte Count Data</strong></td>
</tr>
<tr>
<td>Total, N</td>
</tr>
<tr>
<td>European ancestry, n (%)</td>
</tr>
<tr>
<td>African ancestry, n (%)</td>
</tr>
<tr>
<td>Male/Female, n (%)</td>
</tr>
<tr>
<td>Mean Age (Range)</td>
</tr>
</tbody>
</table>
4. Discussion

A lymphocyte count PGS trained in the general population did not effectively predict baseline CD4 cell count or change in CD4 cell count in response to ART, leading to rejection of our hypothesis that poor CD4 cell recovery in response to ART is dependent on each individual’s overall genetic predisposition to this outcome. When running regressions without correcting for covariates, $R^2$ values were low across all ancestry groups and most regressions were not statistically significant (Figure 2, Supplementary Table 1. In contrast, clinical covariates were predictive of these phenotypes. When correcting for covariates, performance of the model improved markedly. Baseline regressions performed modestly ($R^2 = 0.278$) while on-treatment regressions were not very predictive ($R^2 = 0.073$), although all values were statistically significant (Figure 3, Supplementary Table 2). However, because the PGSlymph itself was not highly predictive, the success of this model was mostly due to the contribution of covariates. Additionally, when including covariates in the model, the model including the African PGSlymph better predicted change in CD4 cell count on-treatment in individuals of African ancestry than the model including the European PGSlymph ($R^2$ was greater by 0.003) (Figure 3, Supplementary Table 2). This is the only case where we see improved performance by an AFR PGSlymph compared to a EUR PGSlymph. Interestingly, when considering effects of individual covariates in this model, the influence of the AFR
PGS\text{lymph} is significant \((p = 0.044)\) in individuals of African ancestry with an effect size of -2.062 (Supplementary Table 3). In comparison to other covariates, this effect size is minimal, but suggests that the AFR PGS\text{lymph} is playing a role. Furthermore, this shows that our methods improved PGS\text{lymph} performance in individuals of African ancestry, which was likely because of a combination of a PGS\text{lymph} based on African ancestry summary statistics and utilizing PRScsx for calculation.

In univariate analyses, lymphocyte count PGS did not effectively predict baseline lymphocyte count in ACTG participants. \(R^2\) values were also low and insignificant (Figure 2, Supplementary Table 5). Performance improved when including covariates in this model, as \(R^2\) values rose to ~0.10 and regressions became statistically significant (Figure 3, Supplementary Table 6). Within the covariate models, the influence of the EUR PGS\text{lymph} is significant in individuals of European ancestry \((p = 0.018)\) with a minimal effect size of 0.025 (Supplementary Table 7). However, as the effect size is small, though significant, the EUR PGS\text{lymph} is not adding much to this model. Still, this significant effect is exhibited as the \(R^2\) value of the EUR PGS\text{lymph} covariate model in individuals of European ancestry \((0.103)\) is slightly higher than the \(R^2\) value of the AFR PGS\text{lymph} covariate model in individuals of European ancestry \((0.101)\) (Figure 3, Supplementary Table 6). Additionally, in the multivariable model, the influence of the AFR PGS\text{lymph} is significant in the multi-ancestry group \((p=8.7e-3)\) with an effect size of 8.3e-3 (Supplementary Table 9). Although this evidently did not have a large impact on the model, the effects of this are still present as the \(R^2\) value of the AFR PGS\text{lymph} covariate model in the multi-ancestry group \((0.098)\) is slightly higher than the \(R^2\) value of the EUR PGS\text{lymph} covariate model in the multi-ancestry group \((0.097)\) (Figure 3, Supplementary Table 6). Also, it is interesting that the \(R^2\) value did not increase as high as in CD4 cell count regressions, perhaps because viral load was the greatest contributing covariate (viral load had the lowest \(p\)-value of all variables in all CD4 cell count regressions), and total lymphocyte counts are not greatly affected by viral load, in contrast to CD4 cell counts\(^3\) (Supplementary Table 3).

Although this model did not perform well in PWH, it performed slightly better when applied to a general medical biobank population. The PGS\text{lymph} best predicted lymphocyte count in a general medical biobank population. Regressions were highly statistically significant, likely due to a large sample size (~37,000 individuals). In the univariate model, the African PGS\text{lymph} applied to the multi-ancestry group and the European PGS\text{lymph} applied to the European population had the highest \(R^2\) values (~0.01) (Figure 2, Supplementary Table 11). It is interesting that these regressions had the highest \(R^2\) values, as these are the only ACTG lymphocyte count regressions that had a significant contribution from PGS\text{lymph} in the multivariable model. Seeing these patterns across the general population and PWH shows that the AFR PGS\text{lymph} performs best in a multi-ancestry group and the EUR PGS\text{lymph} performs best in individuals of European ancestry. When controlling for covariates, performance of the model increased. \(R^2\) values rose to ~0.06-0.10 and \(p\)-values dropped even lower (Figure 3, Supplementary Table 12). This mirrors the impact of covariates seen in PWH. The effect size of the EUR PGS\text{lymph} was ~0.01 in all ancestry groups (Supplementary Table 13). It is interesting that without covariates, the EUR PGS\text{lymph} in individuals of European ancestry was the only regression mirroring this effect size (Figure
2, Supplementary Table 11). The effect size of the AFR PGSlymph was much lower, \(~5e-3\) (Supplementary Table 14). This effect size was mirrored in the AFR PGSlymph regressions without covariates in European and African ancestry, as the \(R^2\) values were also low (\(~3e-4\) or \(8e-4\)), but interestingly the \(R^2\) value was higher when the AFR PGSlymph was applied to the multi-ancestry group (\(~0.01\)) (Figure 2, Supplementary Table 11).

Although these results showed that PGSlymph itself is not predictive of this treatment response, some results show that in combination with covariates, the impact of PGSlymph can become significant, suggesting a possible synergistic effect between PGSlymph and clinical covariates in the model. In the regressions between AFR PGSlymph and change in CD4 cell count in individuals of African ancestry, the impact of the PGSlymph was insignificant, but when including clinical covariates in the regression, the impact of the PGSlymph became significant (Supplementary Table 3). However, the AFR PGSlymph did not significantly interact with any covariates, eliminating the possibility of a synergistic effect (Supplementary Table 4). Additionally, in the regressions between the AFR PGSlymph and baseline lymphocyte count in PWH of all ancestry groups, as well as in the regressions between the EUR PGSlymph and baseline lymphocyte count in individuals of European ancestry, the same patterns were observed (Supplementary Table 7, Supplementary Table 9). Similarly, the AFR PGSlymph did not significantly interact with any covariates, but the EUR PGSlymph significantly interacted with age (Supplementary Table 8, Supplementary Table 10). Thus, it is possible that in PWH, there are synergistic effects between the EUR PGSlymph and covariates, thus leading the PGSlymph to become significant. These findings highlight the importance of including clinical covariates in PGS analyses, not only because the covariates themselves are predictive of treatment response, but also because they seem to interact with the PGSlymph in some way. Another explanation for this observation is that covariates with strong effects overshadow the effects of PGSlymph when not controlled for. Covariates such as viral load have such high significance and large effect sizes, that the effects of smaller impact variables such as PGSlymph are not seen unless these covariates were controlled for. Thus, it is important to consider clinical covariates when implementing PGS in a clinical setting.

This study had several limitations. First, the sample size of the African ancestry summary statistics that were used to generate the African PGSlymph were small (~13,000 individuals), which is due to the lack of availability of lymphocyte count summary statistics for individuals of African ancestry. To improve these results, more lymphocyte count GWAS data are needed in future studies, as it is possible that the AFR PGSlymph could have performed better with a larger base sample size. Additionally, the ACTG sample size was modest (~4600 individuals) which was subset to even smaller groups when stratified by ancestry. It is possible that associations with PGSlymph may have become statistically significant with a larger sample size. Subsequent work in this area could investigate whether this model is predictive of other drug response traits, specifically other ART treatment responses.

Polygenic scores have the potential to leverage large, publicly available datasets to find novel genetic discoveries in pharmacogenomic cohorts. This study utilized a novel method to predict CD4 cell recovery in response to ART and illustrated the importance of including clinical covariates in a
PGS model. As more associations or lack thereof are found, we continue to narrow down the biological underpinnings of responses to ART including suboptimal CD4 cell recovery.

5. Acknowledgements

The authors are grateful to the many persons living with HIV who volunteered for ACTG protocols ACTG384, A5095, A5142, A5202 and A5257. 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 Bristol-Myers Squibb, 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 A1068636 and UM1 A1106701. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Research reported in this publication was also 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.); A1110527, A1077505, TR000445, and A1069439 (to D.W.H.); A1077505 (to K.M.S. and M.D.R). This work was supported by the Tennessee Center for AIDS Research (P30) A1110527.


We acknowledge the Penn Medicine BioBank (PMBB) for providing data and thank the patient-participants of Penn Medicine who consented to participate in this research program. We would also like to thank the Penn Medicine BioBank team and Regeneron Genetics Center for providing genetic variant data for analysis. The PMBB is approved under IRB protocol# 813913 and supported by Perelman School of Medicine at University of Pennsylvania, a gift from the Smilow family, and the
National Center for Advancing Translational Sciences of the National Institutes of Health under CTSA award number UL1TR001878.

6. Supplementary Tables
All supplemental data can be found at:
https://ritchielab.org/files/PSB_supplemental_data/PSB_LymphocytePGSHIV_Cardone_2023_Supplementary.pdf

References
1. HIV. https://www.who.int/data/gho/data/themes/hiv-aids#cms.


