Genetics and Infectious Diseases: From Host Transcriptome of HIV Infection to Bioethical Considerations in International Cohorts

Yunqi Li

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Jairam R. Lingappa, Chair

Romel D. Mackelprang

Kathleen F. Kerr

Joon-Ho Yu

Program Authorized to Offer Degree:

Public Health Genetics

University of Washington

Abstract

Genetics and Infectious Diseases: From Host Transcriptome of HIV Infection to Bioethical Considerations in International Cohorts

Yunqi Li

Chair of the Supervisory Committee:
Professor Jairam R. Lingappa
Departments of Global Health and Medicine

Host genetic research has fueled infectious disease discoveries, including for HIV-1, from early candidate gene studies to recent multi-omics efforts. It offers opportunities for a more comprehensive understanding of infection and pathogenesis, which may elucidate targets for new preventive and therapeutic interventions. The first two aims of this dissertation analyze the host transcriptome to understand the HIV acquisition and treatment response: Chapter 2 aims to identify genes and pathways associated with extreme HIV-1 susceptibility phenotypes; Chapter 3 aims to identify a signature of interferon-stimulated gene expression that differentiates people living with HIV-1 who are virally suppressed on antiretroviral treatment from HIV-uninfected individuals. Many of the host genetic research on infectious diseases are multinational collaborations. Participants are recruited from culturally diverse environments, e.g., Africa, where bioethics issues need to be considered to better safeguard participants' autonomy and voluntary participation. The third aim of this dissertation, described in Chapter 4, explores ethical, legal, and social implication issues raised in the consent process of genetic research conducted in international settings.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The past few years of pursuing my PhD have been a challenging but rewarding journey. I could not have completed this adventure without the support from so many different people.

First, I would like to acknowledge my committee members, whom I have trained under, engaged with, and learned from both within and outside the academic arena. I am appreciative of the exceptional mentorship, support, and respect given to me by them, especially during this unprecedented period of pandemic. My chair Prof. Jairam R. Lingappa, for devoting his enthusiasm and energy to direct me in research. Jai is the warmest and most patient mentor, allowing me to explore intriguing problems in genetics and biomedicine. Prof. Romel D. Mackelprang, who afforded me all the support and independence possible. Romel's deep understanding of the area has inspired me to work out the puzzles, and his exceptionally high standard in research is something I always want to pursue. Prof. Katie F. Kerr, who taught the Applied Biostatistics class and provided guidance, helped me to navigate the twists and turns of the rigorous scientific process. Prof. Joon-Ho Yu, for stepping up and providing the most thorough overview of the field of bioethics and translation of genomics. And Prof. Li Hsu, for performing as my Graduate School Representative (GSR).

Additionally, I would like to acknowledge colleagues and scientist who I have collaborated with. Xuanlin (Sharon) Hou from International Clinical Research Center (ICRC) for data management assistance. Profs. Cheryl M. and Mark J. Cameron from Case Western Research University for their experimental and computational insight. I also would like to thank couples registered in Partners in Prevention HSV/HIV Transmission Study, the Couples Observational Study, and the Partners PrEP Study for their participation and the research staff at the study sites and the UW Coordinating Center for their steadfast efforts recruiting and following the cohort.

Next, I would like to acknowledge the Institute for Public Health Genetics. Prof. Bruce S. Weir, who was my first-year academic advisor and offered me great tutorial in Statistical & Population Genetics as well as research assistant position in his lab when I had just joined the university. Prof. Anna Mastroianni, who taught the Genetics and Law class and for her instrumental in developing the ELSI chapter. Prof. Deborah J. Bowen, in memoriam, who taught the Social Science Research Methods class and initially helped me with the ELSI chapter from early conceptualization to later literature screening. Her kind, caring, passion for her work, and featured direct & humorous communication style will always be remembered. Last but not least, I am incredibly grateful to Ms. Annique Atwater, for her warm-hearted administrative support, lavish care, and for supporting all of us students, step-by-step, as we completed our educational journey.

I would like to acknowledge my friends and classmates, who I was fortunate to interact with during the program: Nandana Rao, Hanley Kingston, Alyna Khan, Priyanka Murali,

Hang Yin, Diane Xue, Sanne Aalbers, Taylor Johnson, EJ Dusic, Jingheng Chen. They are brilliant young researchers, and their varied knowledge and fruitful perspectives always provide me inspiration to continue exploration in genetics and public health.

Finally, I would like to acknowledge my family: my parents, Shue and Mengqiang, for providing the most pleasant harbor for me; and my husband Hanxin, for his patience and attentiveness, for understanding my difficulties, and for exploring science with me as a comrade.

Chapter 1. Overview

With the goal of pursuing research to translate genetics/genomics for public health and population-level benefit, the Institute for Public Health Genetics (IPHG) at the University of Washington is a pioneer in designing its interdisciplinary training program. IPHG graduates receive training in two core knowledge areas: *Genomics in Public Health*[†] and *Implications of Genetics for Society*[‡]. Graduates usually choose one of these two fundamental areas to focus on, while also having some flexibility to integrate elements of both areas into their research and training. In my dissertation study, I conducted three major projects: 1) Identifying genes and pathways associated with extreme HIV-1 susceptibility phenotypes, 2) identifying interferon-stimulated genes (ISGs) differentially expressed between people living with HIV-1 (PLWH) who are virally suppressed on antiretroviral treatment (ART) compared to HIV-1 uninfected individuals, and 3) exploring ethical, legal, social implication (ELSI) issues raised in the informed consent process of genetic/genomic research in the context of international settings, primarily Africa.

1.1 Human Genetics in HIV

Since the first clinical AIDS case reported in 1980s, HIV-1infection has been one of the most profound global public health challenges[1]. Important progress in human genetics research

[†] *Genomics in Public Health* focuses on applications of genetic/genomic data and research in population health, such as identifying genetic and environmental risk factors on disease in the diverse human population. Core coursework for PhD program includes genetic epidemiology, statistical genetics, human genetics, pharmacogenetics and toxicogenomics, bioinformatics, biostatistics, and epidemiology.

[‡] *Implication of Genetics for Society* focuses on the ethical, legal, social, and financial issues arising when genetic/genomic knowledge and technologies are developed and made available to individuals and populations. Core coursework for PhD program includes public health, bioethics, social sciences, law and policy, health economics and outcomes research.

has been made in the past decades to combat the HIV epidemic (reviewed in references[2; 3]). In early candidate gene studies, efforts were made to understand how host genetic variations influence infection susceptibility. These studies resulted in the identification of the *CCR5* gene[4-6] and the HLA class I locus[7] as the major genetic factors regulating the natural history of HIV infection. More recently, advances in sequencing technology have enabled innovative approaches in HIV host genetics research, including genome-wide[8-13], multi-omics, and system biology, such as transcriptome[14-16] and proteome[17; 18], analysis. Topics of study also have expanded broadly, ranging from various measures of HIV-1 control and disease progression phenotypes, to pharmacogenetics of long-term therapy[19-21], to comorbidities in PLWH[22; 23], etc.

As another step in that research tradition, in **Chapter 2** of this dissertation I analyzed whole blood transcriptional signatures from Africans with quantified HIV exposure in order to identify genes and pathways that identify individuals at higher risk of HIV infection. In that Chapter, I identified several genes, most notably including *PTPRC* (which encodes CD45 antigen), as well as interferon-response pathways, both markers of immune activation, that had significantly higher expression among individuals who went on to acquire HIV.

In **Chapter 3** of this dissertation, I compared the gene expression signatures in peripheral blood lymphocytes from PLWH who are virally suppressed on ART compared to HIV-uninfected individuals. Genes involved in CD8+ T cell responses and multiple immune-related pathways were upregulated, demonstrating ongoing inflammation among PLWH.

Most of the enrichment of this inflammatory response was contributed by ISGs, such as *IFI27*, *CXCL10*, *OASL*, *IFI35*, and *MX1.*

1.2 AFRICAN STUDY COHORTS

Sub-Saharan Africa is the most heavily affected region, with about two-thirds of HIV-1 cases, followed by the Asia and the Pacific region[24]. Since its inception in 2007, the International Clinical Research Center (ICRC) in the Department of Global Health at the University of Washington has coordinated recruitment and follow-up of HIV-1 serodifferent heterosexual couple cohorts in Sub-Sahara Africa. In this dissertation, study subjects, samples, and data are drawn from three cohorts coordinated by the ICRC: the Partners in Prevention HSV/HIV-1 Transmission study (Partners cohort)[25], the Couples Observational Study (COS cohort)[11], and the Partners PrEP study (PrEP cohort)[26].

The three cohorts together recruited about 8,500 couples from about 20 study sites in East and Southern Africa (**Figure 1.1**). At the time of recruitment, heterosexual couples were identified with one partner already living with HIV-1, while the other partner was HIV-1 negative (**Figure 1.2**). Couples were followed-up at intervals of at least once every three-months for between one to three years. All partners were provided safe-sex information and condoms, and offered ART according to national guidelines if they became HIV-1-infected. Pre-exposure prophylaxis had not been demonstrated effective during the time when these studies had been conducted. Despite these prevention efforts, during the study, the HIV-uninfected partner was potentially exposed to the virus through heterosexual sex with their HIV-1-infected partner. HIV-1 serostatus was determined along with collection of

epidemiological, clinical, and behavioral data and biological samples throughout the study. Data and samples are now archived in the ICRC repository.

1.3 Informed Consent for Genetic/Genomic Research

Participants from these HIV-1 serodifferent heterosexual cohorts who provided written informed consent for host genetic studies were included in the dissertation analyses. The relevant consent documents were approved by the University of Washington, and at local and affiliated institutional review boards at the study sites where the individuals were enrolled.

A practical ethical issue raised by international genomics research is the validity of informed consent. Researchers may question whether their participants could comprehend information about the study, and whether they could make an informed decision on whether or not to participate. For example. Researchers from UW ICRC, who have field work experiences in Kenya, once observed that many of their research participants did not finish high-school education, and their local language, Kiswahili, is traditionally spoken rather than written. It raises concerns about whether the traditional text-based informed consent could effectively communicate complex genetic/genomic concepts to study participants. In **Chapter 4**, I performed a systematic review of the empirical studies that examined the informed consent process involving genetic/genomic data collection in international cohorts, primarily in sub-Saharan Africa. Major issues proposed by stakeholders (including participants, local community, research team, and institutional ethics review boards, etc)

include (1) recall and comprehension, (2) voluntary participation, (3) consent elements, (4) consent model, and (5) community engagement.

1.4 SUMMARY

The three chapters of this dissertation that constitute my original analyses are linked by the thread of the interdisciplinary nature of training in IPHG. This hybrid approach could help enhance the potential for innovation and creativity in the application of genomic research to population health. **Chapters 2** and **3** constitute the science-side of genetics. Data and Results presented in these two chapters have been submitted or prepared for publication. The identified transcriptional signatures associated with HIV-1 acquisition and ART-treated virally suppressed individuals could be utilized to provide novel biological insights in HIV-1 infection and pathogenesis. The IPHG program adds the valuable opportunity to approach the field through a novel lens of ethical, legal, financial, and social implications (**Chapter 4**). Researchers' awareness and attention to the identified bioethical issues and managing these concerns would further ensure the rights, safety, and well-being of genetic/genomic research participants, and support the successful conduct of research locally.

1.5 TABLES & FIGURES

Figure 1.1 Overview of ICRC HIV-1 Cohorts.

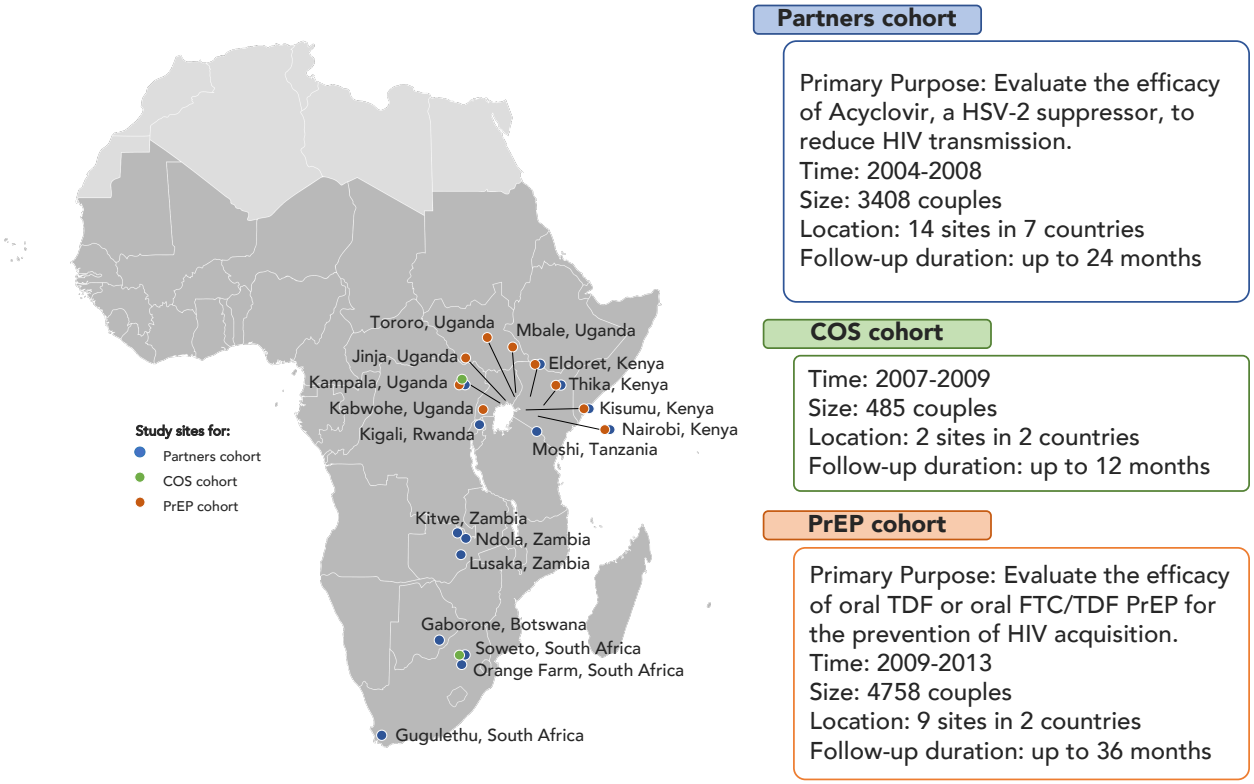
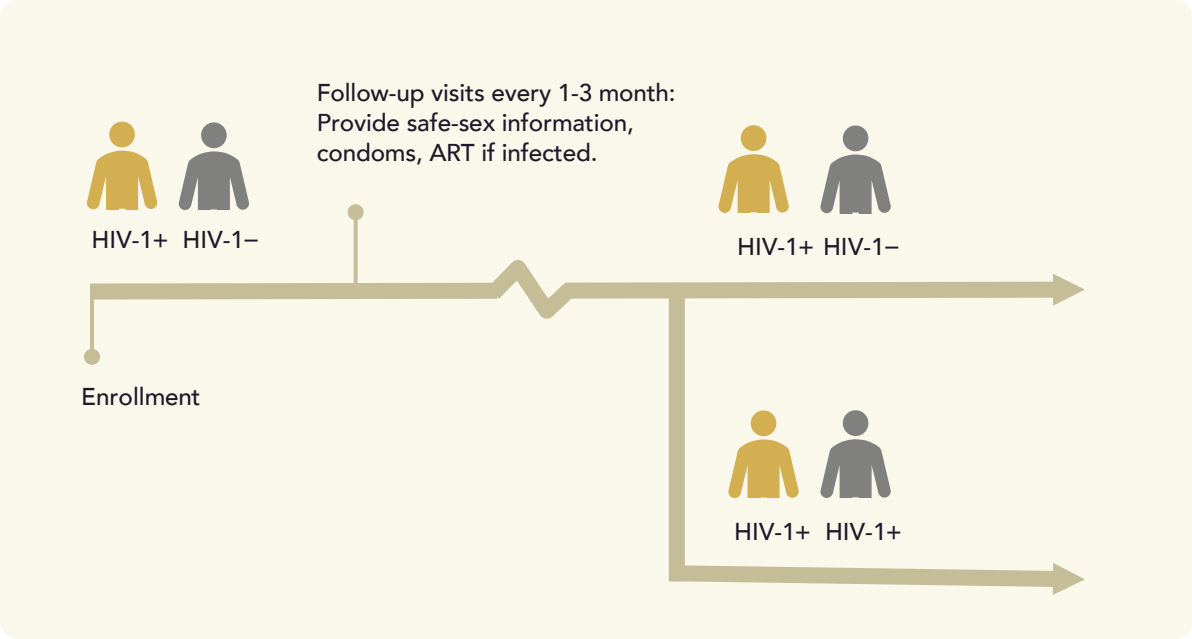


Figure 1.2 HIV-1 Serodifferent Couple Cohort Design



1.6 REFERENCES

1. Deeks, S.G., Overbaugh, J., Phillips, A., and Buchbinder, S. (2015). HIV infection. Nat Rev Dis Primers 1, 15035.

2. McLaren, P.J., and Fellay, J. (2021). HIV-1 and human genetic variation. Nat Rev Genet 22, 645-657.

3. Fellay, J., Shianna, K.V., Telenti, A., and Goldstein, D.B. (2010). Host genetics and HIV-1: the final phase? PLoS Pathog 6, e1001033.

4. Liu, R., Paxton, W.A., Choe, S., Ceradini, D., Martin, S.R., Horuk, R., MacDonald, M.E., Stuhlmann, H., Koup, R.A., and Landau, N.R. (1996). Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell 86, 367-377.

5. Huang, Y., Paxton, W.A., Wolinsky, S.M., Neumann, A.U., Zhang, L., He, T., Kang, S., Ceradini, D., Jin, Z., Yazdanbakhsh, K., et al. (1996). The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. Nat Med 2, 1240-1243.

6. Dean, M., Carrington, M., Winkler, C., Huttley, G.A., Smith, M.W., Allikmets, R., Goedert, J.J., Buchbinder, S.P., Vittinghoff, E., Gomperts, E., et al. (1996). Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. Science 273, 1856-1862.

7. Kaslow, R.A., Carrington, M., Apple, R., Park, L., Munoz, A., Saah, A.J., Goedert, J.J., Winkler, C., O'Brien, S.J., Rinaldo, C., et al. (1996). Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. Nat Med 2, 405-411.

8. Fellay, J., Shianna, K.V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A., et al. (2007). A whole-genome association study of major determinants for host control of HIV-1. Science 317, 944-947.

9. Petrovski, S., Fellay, J., Shianna, K.V., Carpenetti, N., Kumwenda, J., Kamanga, G., Kamwendo, D.D., Letvin, N.L., McMichael, A.J., Haynes, B.F., et al. (2011). Common human genetic variants and HIV-1 susceptibility: a genome-wide survey in a homogeneous African population. AIDS 25, 513-518.

10. Mackelprang, R.D., Bamshad, M.J., Chong, J.X., Hou, X., Buckingham, K.J., Shively, K., deBruyn, G., Mugo, N.R., Mullins, J.I., McElrath, M.J., et al. (2017). Whole genome sequencing of extreme phenotypes identifies variants in CD101 and UBE2V1 associated with increased risk of sexually acquired HIV-1. PLoS Pathog 13, e1006703.

11. Lingappa, J.R., Petrovski, S., Kahle, E., Fellay, J., Shianna, K., McElrath, M.J., Thomas, K.K., Baeten, J.M., Celum, C., Wald, A., et al. (2011). Genomewide association study for determinants of HIV-1 acquisition and viral set point in HIV-1 serodiscordant couples with quantified virus exposure. PLoS One 6, e28632.

12. Luo, M., Sainsbury, J., Tuff, J., Lacap, P.A., Yuan, X.Y., Hirbod, T., Kimani, J., Wachihi, C., Ramdahin, S., Bielawny, T., et al. (2012). A genetic polymorphism of FREM1 is associated with resistance against HIV infection in the Pumwani sex worker cohort. J Virol 86, 11899-11905.

13. Pelak, K., Goldstein, D.B., Walley, N.M., Fellay, J., Ge, D., Shianna, K.V., Gumbs, C., Gao, X., Maia, J.M., Cronin, K.D., et al. (2010). Host determinants of HIV-1 control in African Americans. J Infect Dis 201, 1141-1149.

14. Lederer, S., Favre, D., Walters, K.A., Proll, S., Kanwar, B., Kasakow, Z., Baskin, C.R., Palermo, R., McCune, J.M., and Katze, M.G. (2009). Transcriptional profiling in pathogenic and non-pathogenic SIV infections reveals significant distinctions in kinetics and tissue compartmentalization. PLoS Pathog 5, e1000296.

15. Bosinger, S.E., Li, Q., Gordon, S.N., Klatt, N.R., Duan, L., Xu, L., Francella, N., Sidahmed, A., Smith, A.J., Cramer, E.M., et al. (2009). Global genomic analysis reveals rapid control of a robust innate response in SIV-infected sooty mangabeys. J Clin Invest 119, 3556-3572.

16. Jacquelin, B., Mayau, V., Targat, B., Liovat, A.S., Kunkel, D., Petitjean, G., Dillies, M.A., Roques, P., Butor, C., Silvestri, G., et al. (2009). Nonpathogenic SIV infection of African green monkeys induces a strong but rapidly controlled type I IFN response. J Clin Invest 119, 3544-3555.

17. Fu, W., Sanders-Beer, B.E., Katz, K.S., Maglott, D.R., Pruitt, K.D., and Ptak, R.G. (2009). Human immunodeficiency virus type 1, human protein interaction database at NCBI. Nucleic Acids Res 37, D417-422.

18. Chan, E.Y., Qian, W.J., Diamond, D.L., Liu, T., Gritsenko, M.A., Monroe, M.E., Camp, D.G., 2nd, Smith, R.D., and Katze, M.G. (2007). Quantitative analysis of human immunodeficiency virus type 1-infected CD4+ cell proteome: dysregulated cell cycle

progression and nuclear transport coincide with robust virus production. J Virol 81, 7571-7583.

19. Haas, D.W., Ribaudo, H.J., Kim, R.B., Tierney, C., Wilkinson, G.R., Gulick, R.M., Clifford, D.B., Hulgan, T., Marzolini, C., and Acosta, E.P. (2004). Pharmacogenetics of efavirenz and central nervous system side effects: an Adult AIDS Clinical Trials Group study. AIDS 18, 2391-2400.

20. Mallal, S., Nolan, D., Witt, C., Masel, G., Martin, A.M., Moore, C., Sayer, D., Castley, A., Mamotte, C., Maxwell, D., et al. (2002). Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. Lancet 359, 727-732.

21. Lubomirov, R., di Iulio, J., Fayet, A., Colombo, S., Martinez, R., Marzolini, C., Furrer, H., Vernazza, P., Calmy, A., Cavassini, M., et al. (2010). ADME pharmacogenetics: investigation of the pharmacokinetics of the antiretroviral agent lopinavir coformulated with ritonavir. Pharmacogenet Genomics 20, 217-230.

22. Dietrich, L.G., Thorball, C.W., Ryom, L., Burkhalter, F., Hasse, B., Thurnheer, M.C., Weisser, M., Schmid, P., Bernasconi, E., Darling, K.E.A., et al. (2021). Rapid Progression of Kidney Dysfunction in People Living With HIV: Use of Polygenic and Data Collection on Adverse Events of Anti-HIV Drugs (D:A:D) Risk Scores. J Infect Dis 223, 2145-2153.

23. Cheng, H., Sewda, A., Marquez-Luna, C., White, S.R., Whitney, B.M., Williams-Nguyen, J., Nance, R.M., Lee, W.J., Kitahata, M.M., Saag, M.S., et al. (2020). Genetic architecture of cardiometabolic risks in people living with HIV. BMC Med 18, 288.

24. UNAIDS. (2022). Global HIV & AIDS Statistics — Fact Sheet. URL: https://www.unaids.org/en/resources/fact-sheet. Accessed: 14 December, 2022

25. Celum, C., Wald, A., Lingappa, J.R., Magaret, A.S., Wang, R.S., Mugo, N., Mujugira, A., Baeten, J.M., Mullins, J.I., Hughes, J.P., et al. (2010). Acyclovir and transmission of HIV-1 from persons infected with HIV-1 and HSV-2. N Engl J Med 362, 427-439.

26. Baeten, J.M., Donnell, D., Ndase, P., Mugo, N.R., Campbell, J.D., Wangisi, J., Tappero, J.W., Bukusi, E.A., Cohen, C.R., Katabira, E., et al. (2012). Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. N Engl J Med 367, 399-410.

Chapter 2. GENE EXPRESSION ANALYSIS OF HOST SUSCEPTIBILITY IN HIV-1 SERODIFFERENT COUPLES

2.1 ABSTRACT

HIV-1 exposed seronegative (HESN) individuals may have unique characteristics that alter susceptibility to HIV-1 infection. In this study, we sought to identify genes and pathways present in HESN that are associated with their resistance to HIV. We utilized stored data and biospecimens from HIV-1 serodifferent heterosexual couple cohorts. In the data, couples' HIV-1 exposure was quantified based on unprotected sex frequency and viral load of the partner living with HIV-1. We compared peripheral blood gene expression between 15 highly-exposed HESN (25 samples) and 18 HIV-seroconverters prior to infection (21 samples). We found *PTPRC* (encoding CD45 antigen), and interferon-response pathways had significantly higher expression among individuals who went on to acquire HIV-infection and thus may be a signature for increased acquisition risk.

2.2 INTRODUCTION

A well-documented feature of HIV-1 infection is that some individuals have high levels of exposure to HIV-1 yet remain uninfected[1-5]. Over the last decade, efforts have been made to explain their apparent lower susceptibility to HIV-1 in order to better understand correlates of protection[6]. One challenge for such studies is the difficulty of reliably quantifying HIV-1 exposure, potentially leading to misclassifying a lack of HIV-1 exposure as reduced susceptibility to HIV-1 infection. Here, we use epidemiologic data, including the frequency of unprotected sex and HIV-1 viral load of the infected partner, from HIV-1 serodifferent heterosexual couples (e.g., one partner living with HIV-1 and the other not at enrollment) in order to identify persons with high levels of exposure to HIV-1[7]. Based on this epidemiologic characterization, we compared peripheral blood gene expression profiles between HIV-1 seroconverters (sampled prior to HIV-infection) and highly HIV-1-exposed participants who remained seronegative (HESN) to identify genes and pathways associated with risk of HIV-1 infection.

2.3 METHODS

2.3.1 *Study subjects*

This study included HIV-1 seroconverters and HESN participants from Soweto, South Africa and Kampala, Uganda who were enrolled in the Partners in Prevention HSV/HIV Transmission Study (Partners cohort)[8] or the Couples Observational Study (COS cohort)[9]. All participants were HIV-1 seronegative at enrollment, had an enrolled heterosexual, regular sex partner living with HIV-1, and had HIV-1 serology testing every three months for the duration of follow-up (24-months for Partners cohort, and 12-months for the COS cohort).

Collection of whole blood for gene expression analyses every 6 months was implemented late in the Partners cohort follow-up, and throughout follow-up for the COS cohort. HIV-1 seroconverters and their partners had plasma HIV-1 RNA sequenced to confirm that the virus was linked by sequence to the virus from the partner living with HIV-1 who was enrolled in the study. Estimated dates of HIV-1 infection were determined retrospectively by reverse transcriptase polymerase chain reaction (RT-PCR) testing of plasma collected at visits prior to seroconversion as previously described[10]. HIV-1 seroconverters were included in the study if they had at least one whole blood RNA sample collected at a visit prior to detection of HIV-1 RNA.

For each HIV-1 seroconverter case, we identified all HESNs from the same cohort, study site, and with the same sex as the seroconverter and selected the HESN control as the participant with the longest follow-up and highest HIV-1 exposure. HIV exposure was quantified based on male circumcision status of male HESNs, self-reported frequency of unprotected sex acts with their enrolled partner, and the plasma HIV-1 RNA level of that partner, as previously described[7].

2.3.2 *Study samples*

Peripheral blood RNA samples (PAXgene, Preanalytic Inc.) were collected every 6 months, with HIV-1 status assessed by rapid HIV enzyme immunoassay (EIA) (Alere Determine, Organics Ltd) every three months, with positive rapid tests confirmed by HIV ELISA, and HIV-1 Western Blot. For HESN subjects, RNA was extracted from each PAXgene sample collected during follow-up. For HIV-1 seroconverters, RNA was extracted from PAXgene

samples collected at any follow-up visit through the last visit where the participant remained HIV-1 RT-PCR negative.

2.3.3 *mRNA expression microarrays*

Microarray analysis was performed as previously described[11]. Briefly, RNA was isolated using RNeasy Micro Kits (Qiagen, Hombrechtikon, Switzerland). The quantity and quality of the RNA were evaluated using a NanoDrop 2000c (NanoDrop Technologies, Willington, DE) and an Experion Electrophoresis System (Bio-Rad, Hercules, CA). Total RNA (50 ng) was amplified using Illumina TotalPrep RNA amplification kits (Life Technologies, Grand Island, NY). Biotinylated cRNA (750 ng) was hybridized onto IlluminaHumanV3 BeadChips and quantified using an Illumina BeadStation 500GX scanner and Illumina BeadScan software.

The microarray data were pre-processed including filters to remove probes that were less reliable and less informative, and data from arrays with unusually low median intensity, low variability, and low Pearson correlation coefficient relative to the other arrays. The data were also quantile normalized to make the observations across arrays comparable, and log2-transformed to symmetrize gene expression distributions.

2.3.4 *Statistical analyses*

Differential gene expression analysis was conducted to identify genes associated with risk of HIV-1 acquisition. To avoid misclassifying individuals with little or no HIV exposure as biologically resistant to HIV, we used an extreme phenotypes design comparing gene expression signatures of individuals who were highly HIV-exposed but remained seronegative to HIV seroconverters. We applied a linear model and an empirical Bayes

method to improve power by shrinking estimated gene variance towards a pooled estimate through the Limma[12] package (https://git.bioconductor.org/packages/limma), and with inter-subject correlation to account for repeated samples from the same subject. Models were adjusted for age, sex, and study site. Multiple testing was accounted for using the False Discovery Rate (FDR)[13]. Gene set enrichment analysis (GSEA) was used to identify gene sets associated with HIV-1 acquisition based on the Hallmark gene sets from the Molecular Signatures Database[14] (MSigDB)(http://www.gsea-msigdb.org/). A normalized enrichment score was generated from GSEA to quantify the distribution of gene set across a list of genes ranked by fold-change in differential gene expression analyses.

2.3.5 *Human subjects*

Informed consent including for research into the effect of genes on HIV infection was obtained from all participants. Study protocols describing this were approved by human subject research committees at the University of Washington, local study sites and affiliated institutions.

2.4 RESULTS

We identified 25 whole blood RNA samples from 15 HESN, and 21 RNA samples from 18 HIV-1 seroconverters. HESN and seroconverters were similar in age, sex, country of residence, and study cohort (**Table 2.1**). HESN and seroconverters were predominantly female (87% and 78%, respectively) and most were enrolled in Uganda (93% and 83%, respectively). HESN had higher HIV-1 exposure than seroconverters. Among HESN, half reported unprotected sex with their partner throughout the study, while only a third of those who went on to seroconvert reported unprotected sexual activity. The median plasma HIV-1 RNA

of the partners of HESNs was 5.3 log10 copies/mL (IQR 5.2 to 5.5), which was higher than the median log HIV-1 RNA of partners of HIV-seroconverters (4.7, IQR 4.2 to 5.0 log10 copies/mL).

A total of 336 probes emerged with evidence based on FDR < 0.05 that they are differentially expressed between HESN and seroconverters, including 125 that were upregulated in HESN and 211 that were downregulated in HESN. *PTPRC* (Protein Tyrosine Phosphatase Receptor Type C) (2 probes, adjusted *p*-value [adj.*p*] = 6.6×10^{-10}, log2 fold-change [logFC] = 1.7; adj.*p* = 5.0×10^{-10}, logFC = 1.5), *SPATA13* (Spermatogenesis Associated 13) (adj.*p* = 5.1×10^{-10}, logFC = 1.1), *RORA* (RAR Related Orphan Receptor A) (adj.*p* = 1.1×10^{-9}, logFC = 0.7), and *RCSD1* (RCSD Domain Containing 1) (adj.*p* = 2.6×10^{-9}, logFC = 0.7) had the largest fold changes magnitude and smallest adjusted *p*-values (**Figure 2.1**).

PTPRC encodes the CD45 antigen, a transmembrane protein on differentiated hematopoietic cells involved in immune activation particularly in lymphocytes[15]. The other highly ranked genes are also involved in immune response: *RORA* encodes a nuclear receptor and impacts T-helper cell differentiation to Th17 cells[16]. *SPATA13* is a guanine nucleotide exchange factor that has been associated with diverse neurologic, pulmonary, and other outcomes. *RCSD1* encodes a protein which regulates actin filament assembly in several lymphoid organs[17]. Additional differentially expressed genes that have been associated with HIV-1 outcomes in other studies included *NFAT5*[18] and *SH3KBP1*[19] (**Figure 2.3**). Sensitivity analyses adjusted for the HIV-1 infected-partners' HIV-1 RNA levels did not change the significance or magnitude of the associations for genes described here.

Analysis of 50 gene sets previously defined for specific biological pathways identified 21 that were associated with HIV-1 acquisition with FDR < 0.05. In particular, several immune or inflammation-related pathways had small adjusted *p*-values (**Figure 2.4**) and large normalized enrichment scores, including interferon-α response (adj.$p = 1.8x10^{-3}$, normalized enrichment score [NES] = 2.6), interferon-γ response (adj.$p = 1.8x10^{-3}$, NES = 2.5), IL6-JAK-STAT3 signaling (adj.$p = 1.8x10^{-3}$, NES = 1.9), allograft rejection (adj.$p = 1.8x10^{-3}$, NES = 1.9), and complement (adj.$p = 1.8x10^{-3}$, NES = 1.7).

Given their large fold change, small *p*-values, and co-regulation, we were particularly interested in whether the two *PTPRC* probes most strongly associated with HIV-1 acquisition could link HIV-1 infection risk to levels of specific CD45 isoforms. CD45, the protein product of *PTPRC*, has eight isoforms: alternative splicing of exons 4, 5, and 6 differentiates these isoforms (exons 2 and 3 are included in all isoforms)[15]. The full microarray dataset included four *PTPRC* probes annealing to distinct genic regions with two probes (IDs: 2600408 and 6180288) associated with significant fold-change in our analysis, the third probe (ID: 6650193) with a marginally significant fold-change, and the fourth probe (ID: 870095) with a non-significant fold-change. We aligned the four probe sequences to the *PTPRC* sequence through the Basic Local Alignment Search Tool (BLAST), and found that the high fold-change probes, 2600408 and 6180288, mapped to exon 2 and exon 3, respectively (**Figure 2.5**); while the non-significant probe, 870095, mapped to intron 3; and the marginally significant probe, 6650193, mapped to exon 6. These probes do not discriminate distinct CD45 isoforms.

2.5 DISCUSSION

Our study found a signature of differentially expressed genes (DEG) in whole blood that differentiates HESN with exposure to a partner living with HIV and seroconverters, who were sampled prior to infection. Most notably, individuals who went on to acquire HIV exhibited higher expression of *PTPRC* (**Figure 2.2**) prior to HIV-1 infection. We also found strong associations with gene sets involved in interferon-α and -γ responses.

PTPRC has previously been shown to be important for T cell activation and HIV-1 infection. Two variants, C77G and A54G, were identified in *PTPRC* exon 4. The C77G mutation was observed in UK residents and is known to cause abnormal splicing with altered expression of CD45RA in activated or memory T cell[20]. The A54G missense mutation was observed in Ugandans[21]. Both variants have been associated with increased acquisition of HIV-1. Data from humanized mice show that the frequency of human CD45 positive cells is significantly associated with increased HIV-1 susceptibility[22]. Our data add to this context providing a direct link between elevated CD45 expression and increased risk of HIV-1 acquisition.

We found that genes stimulated by Interferon-α and Interferon-γ are upregulated in samples from seroconverters which were obtained prior to seroconversion, suggesting elevated interferon responses may increase susceptibility to HIV-1 infection. Interferon-α and -γ are cytokines elicited in response to infections and environmental stressors, and directly stimulate antiviral responses and immunoregulatory functions such as immune cell recruitment. Although type 1 interferon can contribute to controlling HIV-1 infection, it may also facilitate infection by increasing target cell activation. This may explain why we

observed elevated interferon responses in seroconverters prior to infection and is consistent with the hypothesis that low baseline immune activation, or immune quiescence, is associated with protection from HIV acquisition[23].

Previous data suggested that interferon production may be partly regulated by CD45, the protein product of *PTPRC*. CD45 was reported to be necessary for type 1 interferon production in dendritic cells[24]. Together, elevated CD45 and interferon responses may indicate a state of increased immune activation with more target cells available for HIV infection.

Notably, variation in the *PTPRC* gene was not associated with risk of HIV-1 infection in our whole genome sequencing study[25], suggesting that this differentially expressed gene signature may represent an epigenetic pathway. However, the statistical power of our analysis was not high enough to detect associations between *PTPRC* expression and host genetic variants.

We were not able to perform qPCR or quantify *PTPRC* protein levels due to lack of sample remainders or additional aliquots of the appropriate samples from the relevant participant visits thus, limiting our ability to use these assays for further confirmation of our results. Another limitation in our study is that difference in partner HIV RNA levels in two groups could have confounded the association between *PTPRC* and HIV status if higher HIV RNA levels were associated with lower *PTPRC* expression, which is unlikely. We've performed sensitivity analyses in which models are also adjusted for partner HIV RNA levels, which

should remove a part of the residual confounding. For *PTPRC*, adjustment for the infected partners' HIV RNA levels did not significantly change the significance or magnitude the associations. Other limitations include possible social desirability bias resulting in inaccurate response to the sexual behavior questions and inaccurate HIV exposure quantification. The result may not be generalizable to other mode of virus exposure. Last but not least, during the study follow-up, our participants were provided with condoms, safe-sex information, and intensive counseling on infection risk reduction. This protective effort could reduce risky behavior and ultimately lead to a decrease in infection pressure. However, this could make it difficult to identify HESN individuals. Individuals who intrinsically resist infection may be missed in subject selection when they already have low virus exposure. Therefore, we may lose a chance to discover some potential protective factors. Further evaluation of the relationship between protein-level changes, such as cell surface CD45 expression, and genetic variation in *PTPRC* expression is warranted for better understanding correlates of HIV-1 susceptibility.

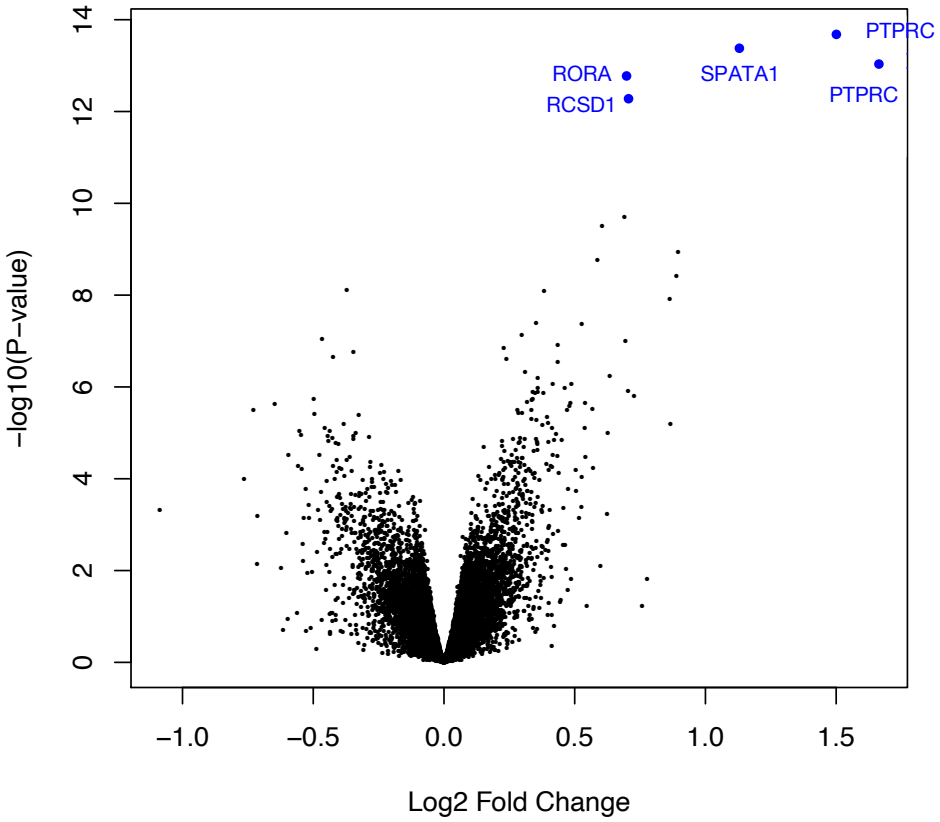
2.6 TABLES & FIGURES

Table 2.1 Descriptive Statistics* of Study Subjects.

Characteristics	HESN (n=15)	Seroconverter (n=18)
Female	13 (86.6%)	14 (77.8%)
Age (years)	24 [21, 25]	27 [21, 30]
Country of residence:		
Uganda	14 (93%)	15 (83%)
South Africa	1 (7%)	3 (17%)
Study cohort:		
COS	10 (66.7%)	12 (66.7%)
Partners	5 (33.3%)	6 (33.3%)
Partner's HIV-1 log viral load at the enrollment visit	5.3 [5.2, 5.5]	4.7 [4.2, 5.0]
Ever reported unprotected sex with enrolled partner from the last visit	8 (53.3%)	6 (33.3%)

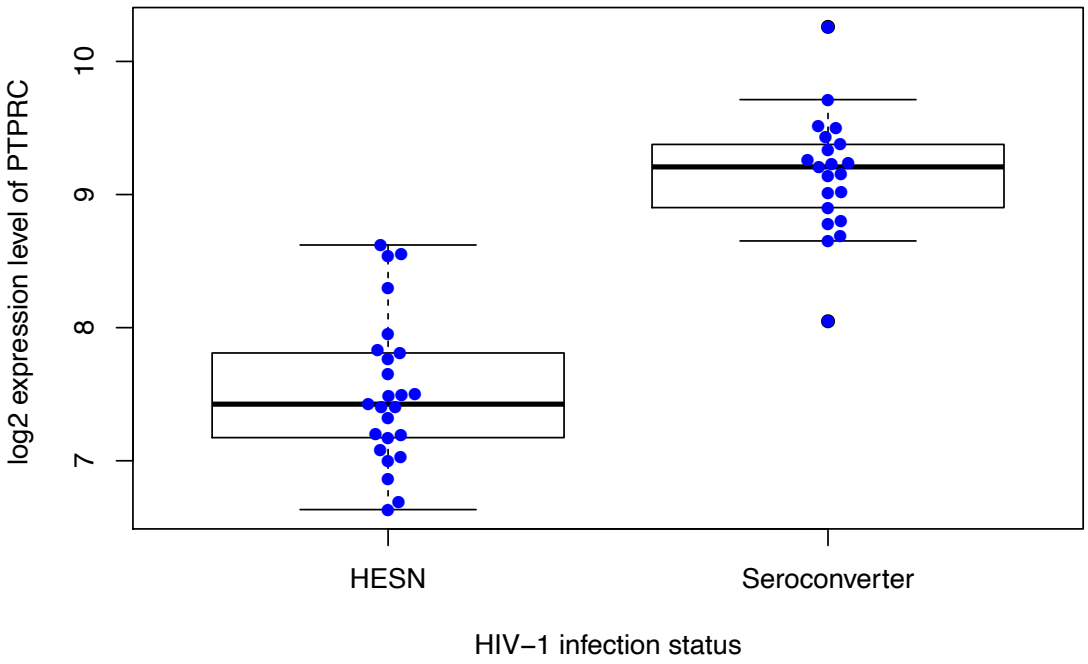
* Numbers (percentages) for categorical variables. Median [interquartile range (IQR)] for continuous variables.
Abbreviation: COS- Couples Observational Study; Partners- Partners in Prevention HSV/HIV Transmission Study; HESN- subjects who are high HIV-1 exposed but remained seronegative through the full two-year follow-up of the study; Seroconverter- subjects who are HIV-1 exposed and experienced HIV-1 seroconversion during follow-up.

Figure 2.1 Volcano Plot of Differentially Expressed Signals.



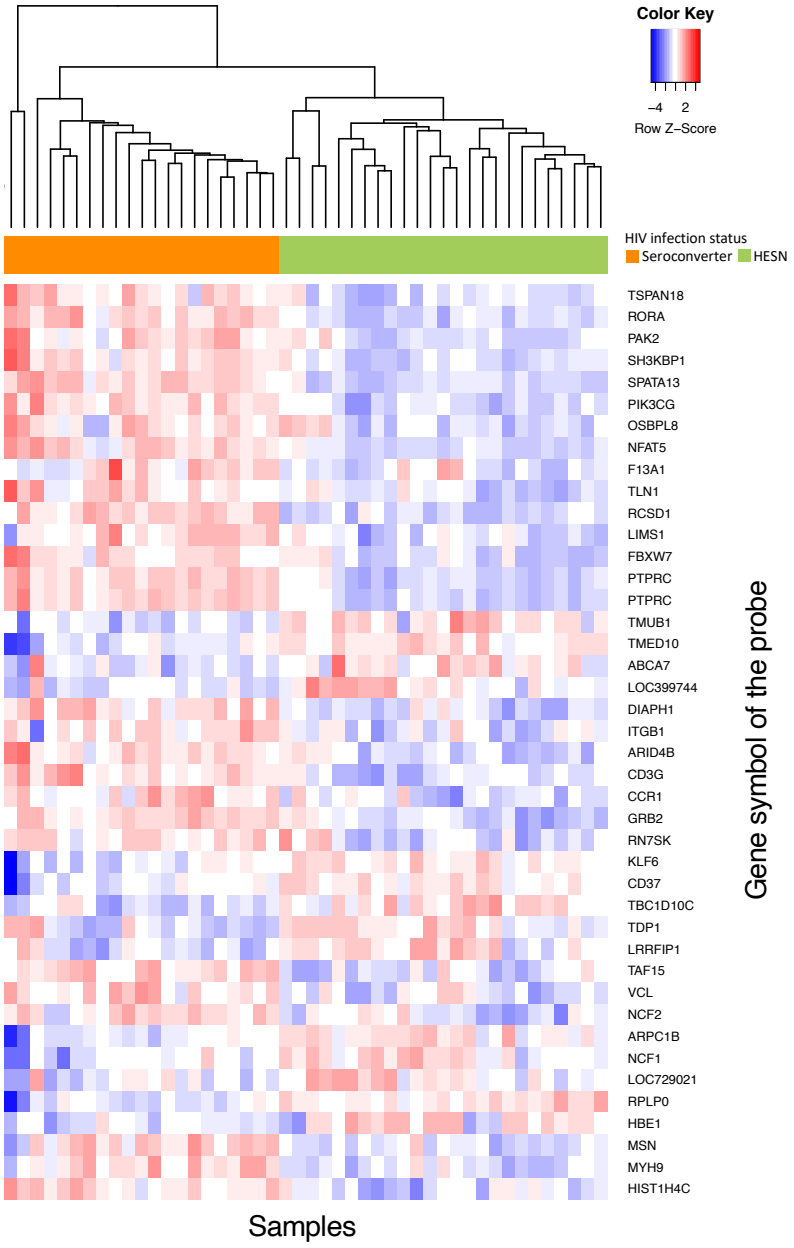
Volcano plot indicating the log 2-fold-changes and adjusted *p*-values for all tested probes. The top five signals with greater fold changes and small *p*-values are annotated with the human gene names in blue. Probe ID: *PTPRC* — 6180288 (adj.*p* = 6.6×10^{-10}, logFC = 1.7), 2600408(adj.*p* = 5.0×10^{-10}, logFC = 1.5); *SPATA13*—1570300(adj.*p* = 5.10×10^{-10}, logFC = 1.1); *RORA*—1110180(adj.*p* = 1.1×10^{-9}, logFC = 0.7); *RCSD1*—1170703(adj.*p* = 2.6×10^{-9}, logFC = 0.7).

Figure 2.2 Seroconverter Exhibited Higher Expression of *PTPRC*.



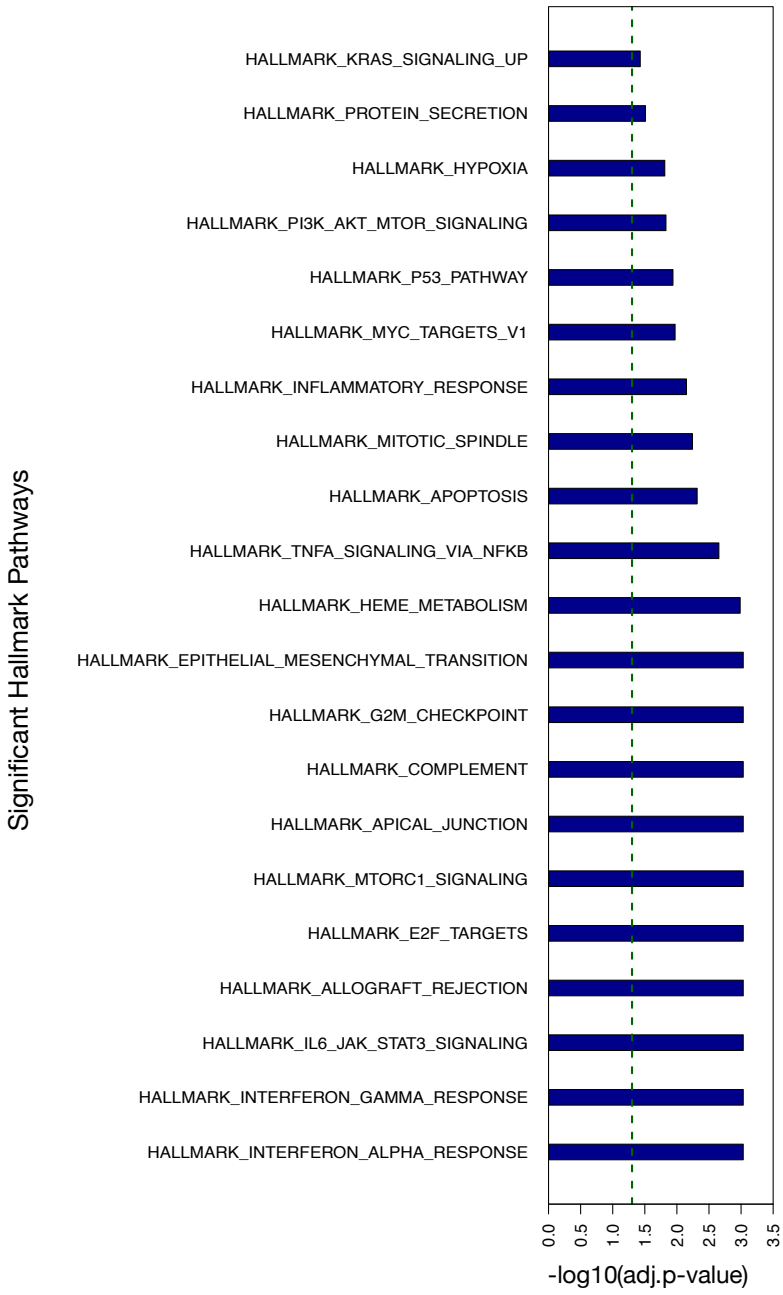
Boxplot with beeswarm plot superimposed for *PTPRC* gene expression between HIV-1 seroconverter and HIV-1 exposed seronegative (HESN). Higher expression of *PTPRC* gene is observed in seroconverter compared to HESN.

Figure 2.3 Heatmap Showing Unique Biomarkers of Seroconverter Status.



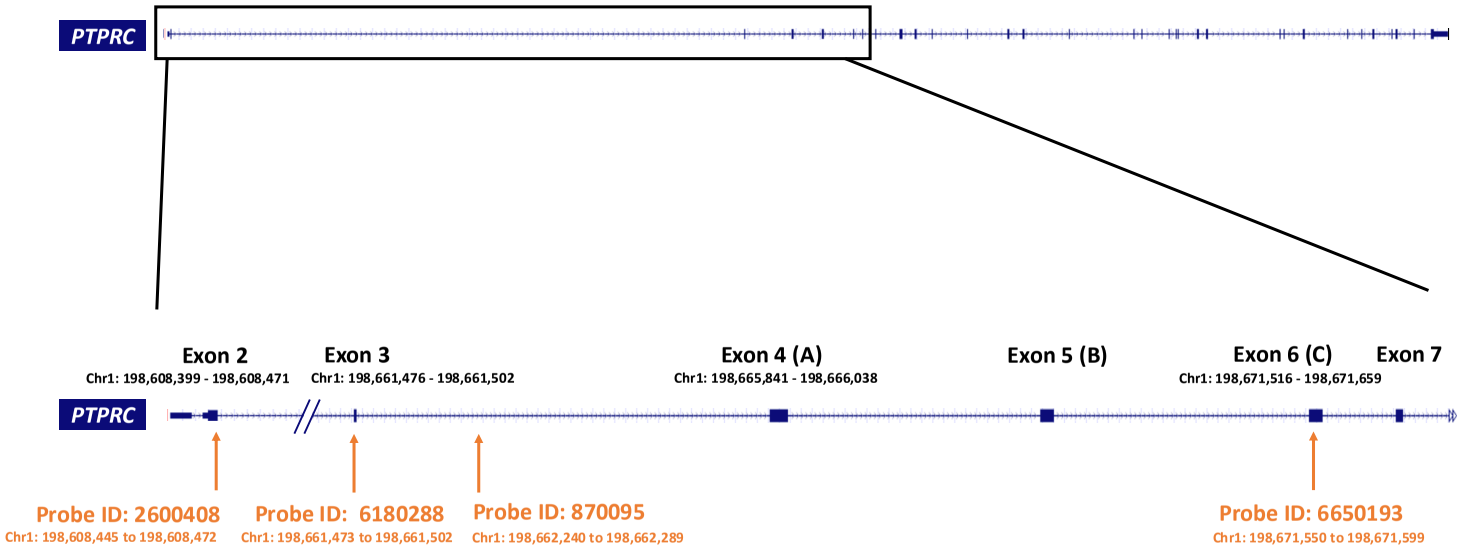
Heatmap containing the top 40 differentially expressed genes (DEGs) with an absolute log2-fold change > 0.5 and adjusted *p*-value less than 0.05 under FDR correction. Cell color represents the gene expression level—higher expression in red and lower expression in blue. Samples are annotated with HIV infectious status, with green indicating HESN and orange indicating seroconverter.

Figure 2.4 Bar Plot of Enriched Pathways in Seroconverters.



Bar plot indicating the normalized enrichment score of significant Hallmark pathways with adjusted *p*-values less than 0.05. A positive value indicates an enrichment in seroconverters, and a negative value indicates enrichment in HESN. Several immune or inflammation-related pathways are identified with high scores, including interferon responses and JAK-STAT signaling.

Figure 2.5 Probes' Locations on *PTPRC* Gene.



We used the Basic Local Alignment Search Tool (BLAST) to align four probe sequences with the GRCh37 reference genome. All of the four probe sequences mapped to the *PTPRC* gene region on chromosome 1.

17

2.7 REFERENCES

1. Choi, R.Y., Levinson, P., Guthrie, B.L., Lohman-Payne, B., Bosire, R., Liu, A.Y., Hirbod, T., Kiarie, J., Overbaugh, J., John-Stewart, G., et al. (2012). Cervicovaginal HIV-1-neutralizing immunoglobulin A detected among HIV-1-exposed seronegative female partners in HIV-1-discordant couples. AIDS 26, 2155-2163.

2. Lajoie, J., Juno, J., Burgener, A., Rahman, S., Mogk, K., Wachihi, C., Mwanjewe, J., Plummer, F.A., Kimani, J., Ball, T.B., et al. (2012). A distinct cytokine and chemokine profile at the genital mucosa is associated with HIV-1 protection among HIV-exposed seronegative commercial sex workers. Mucosal Immunol 5, 277-287.

3. Burgener, A., Mogk, K., Westmacott, G., Plummer, F., Ball, B., Broliden, K., and Hasselrot, K. (2012). Salivary basic proline-rich proteins are elevated in HIV-exposed seronegative men who have sex with men. AIDS 26, 1857-1867.

4. Holditch, S.J., Eriksson, E.M., Tarosso, L.F., Kuebler, P.J., Kallas, E.G., Nielsen, E.K., Wiznia, A.A., Rosenberg, M.G., and Nixon, D.F. (2011). Decay kinetics of HIV-1 specific T cell responses in vertically HIV-1 exposed seronegative infants. Front Immunol 2, 94.

5. Vince, N., Bashirova, A.A., Lied, A., Gao, X., Dorrell, L., McLaren, P.J., Fellay, J., and Carrington, M. (2014). HLA class I and KIR genes do not protect against HIV type 1 infection in highly exposed uninfected individuals with hemophilia A. J Infect Dis 210, 1047-1051.

6. Horton, R.E., McLaren, P.J., Fowke, K., Kimani, J., and Ball, T.B. (2010). Cohorts for the study of HIV-1-exposed but uninfected individuals: benefits and limitations. J Infect Dis 202 Suppl 3, S377-381.

7. Mackelprang, R.D., Baeten, J.M., Donnell, D., Celum, C., Farquhar, C., de Bruyn, G., Essex, M., McElrath, M.J., Nakku-Joloba, E., Lingappa, J.R., et al. (2012). Quantifying ongoing HIV-1 exposure in HIV-1-serodiscordant couples to identify individuals with potential host resistance to HIV-1. J Infect Dis 206, 1299-1308.

8. Celum, C., Wald, A., Lingappa, J.R., Magaret, A.S., Wang, R.S., Mugo, N., Mujugira, A., Baeten, J.M., Mullins, J.I., Hughes, J.P., et al. (2010). Acyclovir and transmission of HIV-1 from persons infected with HIV-1 and HSV-2. N Engl J Med 362, 427-439.

9. Lingappa, J.R., Petrovski, S., Kahle, E., Fellay, J., Shianna, K., McElrath, M.J., Thomas, K.K., Baeten, J.M., Celum, C., Wald, A., et al. (2011). Genomewide association study for determinants of HIV-1 acquisition and viral set point in HIV-1 serodiscordant couples with quantified virus exposure. PLoS One 6, e28632.

10. Lingappa, J.R., Thomas, K.K., Hughes, J.P., Baeten, J.M., Wald, A., Farquhar, C., de Bruyn, G., Fife, K.H., Campbell, M.S., Kapiga, S., et al. (2013). Partner characteristics predicting HIV-1 set point in sexually acquired HIV-1 among African seroconverters. AIDS Res Hum Retroviruses 29, 164-171.

11. Caskey, M., Lefebvre, F., Filali-Mouhim, A., Cameron, M.J., Goulet, J.P., Haddad, E.K., Breton, G., Trumpfheller, C., Pollak, S., Shimeliovich, I., et al. (2011). Synthetic double-stranded RNA induces innate immune responses similar to a live viral vaccine in humans. J Exp Med 208, 2357-2366.

12. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47.

13. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 57, 289-300.

14. Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 1, 417-425.

15. Tchilian, E.Z., and Beverley, P.C. (2006). Altered CD45 expression and disease. Trends Immunol 27, 146-153.

16. Wang, R., Campbell, S., Amir, M., Mosure, S.A., Bassette, M.A., Eliason, A., Sundrud, M.S., Kamenecka, T.M., and Solt, L.A. (2021). Genetic and pharmacological inhibition of the nuclear receptor RORalpha regulates TH17 driven inflammatory disorders. Nat Commun 12, 76.

17. Eyers, C.E., McNeill, H., Knebel, A., Morrice, N., Arthur, S.J., Cuenda, A., and Cohen, P. (2005). The phosphorylation of CapZ-interacting protein (CapZIP) by stress-activated protein kinases triggers its dissociation from CapZ. Biochem J 389, 127-135.

18. Ranjbar, S., Tsytsykova, A.V., Lee, S.K., Rajsbaum, R., Falvo, J.V., Lieberman, J., Shankar, P., and Goldfeld, A.E. (2006). NFAT5 regulates HIV-1 in primary monocytes via a highly conserved long terminal repeat site. PLoS Pathog 2, e130.

19. Landi, A., Vermeire, J., Iannucci, V., Vanderstraeten, H., Naessens, E., Bentahir, M., and Verhasselt, B. (2014). Genome-wide shRNA screening identifies host factors involved in early endocytic events for HIV-1-induced CD4 down-regulation. Retrovirology 11, 118.

20. Tchilian, E.Z., Wallace, D.L., Dawes, R., Imami, N., Burton, C., Gotch, F., and Beverley, P.C. (2001). A point mutation in CD45 may be associated with an increased risk of HIV-1 infection. AIDS 15, 1892-1894.

21. Stanton, T., Boxall, S., Bennett, A., Kaleebu, P., Watera, C., Whitworth, J., French, N., Dawes, R., Hill, A.V., Bodmer, W., et al. (2004). CD45 variant alleles: possibly increased frequency of a novel exon 4 CD45 polymorphism in HIV seropositive Ugandans. Immunogenetics 56, 107-110.

22. Nguyen, P.V., Wessels, J.M., Mueller, K., Vahedi, F., Anipindi, V., Verschoor, C.P., Chew, M., Deshiere, A., Karniychuk, U., Mazzulli, T., et al. (2017). Frequency of Human CD45+ Target Cells is a Key Determinant of Intravaginal HIV-1 Infection in Humanized Mice. Sci Rep 7, 15263.

23. Card, C.M., Ball, T.B., and Fowke, K.R. (2013). Immune quiescence: a model of protection against HIV infection. Retrovirology 10, 141.

24. Montoya, M., Dawes, R., Reid, D., Lee, L.N., Piercy, J., Borrow, P., Tchilian, E.Z., and Beverley, P.C. (2006). CD45 is required for type I IFN production by dendritic cells. Eur J Immunol 36, 2150-2158.

25. Mackelprang, R.D., Bamshad, M.J., Chong, J.X., Hou, X., Buckingham, K.J., Shively, K., deBruyn, G., Mugo, N.R., Mullins, J.I., McElrath, M.J., et al. (2017). Whole genome sequencing of extreme phenotypes identifies variants in CD101 and UBE2V1 associated with increased risk of sexually acquired HIV-1. PLoS Pathog 13, e1006703.

Chapter 3. HOST TRANSCRIPTOME SIGNATURE DIFFERENTIATES ART-TREATED VIRALLY SUPPRESSED HIV-INFECTION FROM UNINFECTED INDIVIDUALS

3.1 ABSTRACT

Objective: Despite the capacity of antiretroviral therapy (ART) to suppress HIV-1, ART-treated persons living with HIV-1 with undetectable plasma HIV-1 have higher comorbidity rates than people without HIV-1. This may, in part, be due to expression of specific interferon-stimulated gene (ISGs) driving chronic immune activation, which then increases the risk of adverse outcomes. This study is part of the large effort to further understand this biological process. Here, we hypothesize that a subset of ISGs is upregulated in people living with HIV-1 who are on suppressive ART compared to people without HIV-1.

Methods: Using RNA-seq, we measured genome-wide gene expression of peripheral blood mononuclear cells (PBMCs) collected from 18 ART-treated individuals documented with undetectable plasma HIV, and 12 uninfected controls. We performed differential gene expression analysis, and gene-set enrichment analysis (GSEA) to evaluate whether a gene or *a priori* sets of genes in common pathways have similar differences between treated HIV-1 infection and uninfected controls.

Results: The most significant differentially expressed genes were *CD8A* and *CD8B*, which had ~2.5-fold higher expression in ART-suppressed participants compared to uninfected controls. Differential expression of these genes likely reflects increased activity of cytotoxic T cells. Gene sets involved in immune activation, such as interferon responses, TNF-α signaling, inflammatory responses, and allograft rejection were also enriched during treated

infection. ISGs such as *IFI27*, *CXCL10*, *OASL*, *IFI35* and *MX1* were among the genes that contributed most to the interferon response enrichment signals.

Conclusions: We found that expression of specific ISGs may differentiate individuals living with virally suppressed HIV-1 infection from those without HIV-1. This suggests that ISGs may contribute to chronic immune activation underlying higher co-morbidity of PLWH even after viral suppression with ART. The ISGs identified here are a starting point to identify direct targets to reduce the morbidity of HIV-associated chronic inflammation.

3.1 INTRODUCTION

Antiretroviral therapy (ART) suppresses HIV-1 and extends life. However, people living with HIV (PLWH) who are treated with ART with documented undetectable plasma viral loads have higher comorbidity rates than people without HIV-1[1]. In the US, the number of PLWHs who are 50 years old and above increased by about 100, 000 from 2010 to 2014, and continues to grow[2]. As the burden of age-related mortality and morbidity increases, it is imperative that we understand the biological process underlying these adverse health outcomes in order to improve outcomes of this aging population with HIV-1.

One possible biological mechanism to explain this higher comorbidity is that expression of specific ISGs drives chronic immune activation in PLWH. (**Figure 3.1**) Multiple studies have documented an association between chronic immune activation and adverse health outcome, including dysregulation or exhaustion of the immune system, CD4+ T cell depletion, AIDS progression, non-AIDS-defining illness, and death[3-7]. The association of ISG expression as a mediator of chronic immune activation is supported by previous animal and human studies. Studies of SIV infection first noted that, although expression of specific ISGs increases dramatically in acute infection, expression of these genes declines to nearly baseline levels in the low-morbidity context of non-pathogenic chronic infection (such as occurs in sooty mangabeys)[8; 9]. In contrast, for pathogenic SIV infection, such as that observed in rhesus macaques, upregulation of ISGs persists into chronic infection[8; 9] with a parallel increase in morbidity and mortality. Studies of human subjects also report that ISGs expression may persist from acute infection to chronic infection despite the decline in HIV

viral load from peak levels during acute infection to the lower set-point during chronic infection.

Our long-term goal is to identify ISGs that directly drive chronic immune activation in PLWH using three stages of analysis: 1) identify ISGs differentially expressed between HIV-infected individuals who are virally suppressed on ART versus HIV-uninfected controls. The identified ISGs could be associated with chronic immune activation, ART usage, or HIV infection. 2) identify which of those ISGs continue to be expressed in PLWH who are virally suppressed on ART compared to PLWH not on ART; 3) test the association between ISGs and functional immunological markers. My focus here is to address the first of these stages.

We compared gene expression, including ISGs and others, in peripheral blood mononuclear cells (PBMCs) from individuals with virally suppressed HIV-1 infection after ART initiation (denoted as the ART-suppressed group hereafter) versus HIV-1 uninfected controls. We aim to identify genes and pathways that were differentially expressed in the two groups, which may be associated with chronic infection, ART usage, and viral infection. These data can then be utilized to identify ISGs that mediate chronic immune activation solely.

3.2 METHODS

3.2.1 *Study subjects and samples*

Study subjects were selected from two HIV-1 serodifferent heterosexual couples cohorts: Partners PrEP clinical trial[10] (Partners PrEP cohort) and the Couples Observational Study (COS cohort)[11]. The Partners PrEP cohort was a randomized placebo-controlled clinical trial,

which tested whether tenofovir (TDF) or tenofovir-emtricitabine (TDF-FTC) provided to an HIV-infected participant could prevent HIV-1 acquisition. COS was an observational study followed a similar recruitment strategy as Partners PrEP.

All subjects from the Partners PrEP trial that contributed to this study of ISG expression were sampled from the placebo arm of the trial. We identified PLWH who initiated ART after study enrollment, had PBMCs samples collected after ART suppression, and having undetectable plasma HIV-1 RNA levels after ART initiation. We identified HIV-uninfected controls by matching the HIV-infected subjects on age and biological sex and did not report any unprotected sex act or have STIs or HSV2 infections that could elevate ISG expression and inflammation responses.

Visit-level data used in this analysis include age, biological sex, region of residence, study cohort, CD4+ T cell count at the time of ART initiation, CD4+ T cell count at successive visits after ART initiation when PBMCs were collected, HIV-1 RNA levels at ART initiation and at post-ART visits when PBMCs were collected.

3.2.2 *RNA-sequencing and preprocessing*

PBMCs from PLWH and HIV-uninfected controls were evaluated through RNA sequencing. PBMCs RNA was isolated by RNA Mini-prep kits for standard input or Micro-prep kits (Qiagen) for low inputs. RNA-Seq libraries were constructed using TruSeq Stranded Total RNA kits (Illumina) and Nextera XT adaptors (Clontech). Libraries were then sequenced using an Illumina HiSeq2500 platform.

Raw FASTQ files were analyzed by removing the adaptor sequences, low-quality entries, and low-read length entries using SKEWER[12]. Low-quality entries were identified with an average Phred quality score[13; 14] less than 30, meaning that the probability that the base is incorrectly called is less than 1 in 1000. Low-read length entries were identified with read length less than 36 nucleotides, which may not be uniquely matched to a location within the genome. This trimmed and filtered FASTQ file was aligned to the reference genome using HISAT[15]. We used the *Homo sapiens* NCBI reference genome assembly version GRCh38 as the reference sequence. Reads were sorted based on chromosome number and genome position was coordinated using SAMtools[16]. Last, the count matrix , which contains expression level with genes in rows and samples in columns, was generated using featureCounts[17] to downstream statistical analysis.

3.2.3 *Statistical Analysis*

Differential gene expression analysis was used to compare ART-suppressed PLWH versus HIV-uninfected controls, or vice versa. We applied linear models with an empirical Bayes method, which protects against false positive signals by shrinking estimated gene variances towards a pooled estimate, using Limma[18]. Models were adjusted for age, sex, and study site. Multiple testing was accounted for using the False Discovery Rate (FDR)[19]. Gene set enrichment analysis (GSEA) was performed to identify pathway or gene sets associated with HIV-1 acquisition through fGSEA[20]. Pre-defined gene sets were obtained Hallmark gene sets from the Molecular Signatures Database[21] (MSigDB).

3.2.4 *Human subjects*

Informed consent, including for genetic research on HIV infection, was obtained from all participants. Study protocols were approved by human subject research committees at the University of Washington, local study sites and affiliated institutions.

3.3 RESULTS

We identified 18 ART-treated PLWH with viral suppression and 12 HIV-1-uninfected controls. ART-suppressed participants and HIV-uninfected controls were similar in age and sex (**Table 3.1**) but were imbalanced in the country of residence. ART-suppressed group and uninfected group were mostly female (67% respectively) and aged from 20-60. Participants were enrolled in Kenya, Uganda, and South Africa. The mean plasma HIV-1 RNA of ART-suppressed subjects was 4.87 log10 copies/mL before the initiation of ART then decreasing level below the limit of detection, which indicating viral suppression. The Partners PrEP trial required CD4+ T cell counts >200 cells at the time of enrollment. The mean CD4+ T cell counts were 300 cells/mm^3 before the initiating ART and 434 cells/mm^3 post-ART.

A total of 594 genes were differentially expressed between the ART-suppressed and HIV-1 uninfected groups based on FDR<0.05. Among the differentially expressed genes, 254 of them had an absolute fold-change greater than 1.5. *CD8A* (CD8a molecule) (adj. $p = 1.8\text{x}10^{-5}$, fold change = 2.5), *CD8B* (CD8b molecule) (adj. $p = 7.2\text{x}10^{-5}$, fold change = 2.5), *CRTAM* (cytotoxic and regulatory T cell molecule) (adj. $p = 1.6\text{x}10^{-4}$, fold change = 2.2), *ACVR1C* (activin A receptor type 1C) (adj. $p = 2.0\text{x}10^{-4}$, fold change = 0.3), *MSC* (musculin) (adj. p =

$2.0x10^{-4}$, fold change =2.6) had the largest fold changes and smallest adjusted *p*-values (**Figure 3.2**).

CD8A and *CD8B* had higher expression among post-ART participants (**Figure 3.3**). These genes encode the alpha and beta chain of the CD8 glycoprotein which is located on cytotoxic T lymphocytes and mediates recognition of antigens displayed by class I MHC molecules on antigen presenting cells (APCs).

Analysis of 50 well-curated Hallmark genes sets identified 17 gene sets that were enriched in with the post-ART state with FDR<0.05 (**Figure 3.4**). In particular, several immune or inflammation-related pathways had small adjusted *p*-values and large normalized enrichment scores (NES, magnitude of enrichment normalized based on the number of genes in the gene sets). These include interferon-α and β responses (adj. $p = 7.9x10^{-8}$ and $1.5x10^{-13}$, normalized enrichment scores [NES] = 2.3 and 2.4, respectively), TNF-α signaling, inflammatory responses, and allograft rejection (adj. $p = 1.0x10^{-8}$, $1.0x10^{-4}$ and $1.6x10^{-9}$, NES = 2.1, 1.9, and 2.3, respectively). Altogether, these results suggest presence of an activated immune state in HIV-infected individuals compared to HIV-uninfected controls. Genes that contribute most to the enrichment signal of interferon responses include ISGs such as *IFI27*, *CXCL10*, *OASL*, *IFI35*, and *MX1* (**Figure 3.5**).

3.4 Discussion

Our study found a transcriptome signature in PBMCs. The transcriptome signature discriminates HIV-uninfected individuals from PLWH who are virally suppressed due to

ART. Post-ART individuals exhibited more than two-fold higher expression of *CD8A* and *CD8B.* Immune and inflammation-related pathways were enriched in virally suppressed HIV-1 infection, including interferon-α and -γ responses. Notably, several ISGs, such as *IFI27*, *CXCL10*, *OASL*, *IFI35*, and *MX1* contributed most to the enrichment signals of interferon responses.

As would be expected, *CD8A* and *CD8B,* which are genes involved in CD8+ T cell immunity, and therefore actively expressed during HIV-infection, showed a higher expression in PLWH compared to HIV-uninfected participants. CD8+ T cell numbers have been documented throughout the course of HIV infection (reviewed in reference[22]) CD8+ T cell counts remain elevated in treated HIV patients who are pharmacologically HIV suppressed. The upregulation of CD8+ T cell counts can be associated with adverse health events, such as non-AIDS morbidities. It echoes our result that the PLWH display residual immune dysregulation despite effective ART. It may also hint that the observed increase in *CD8A* and *CD8B* expression may be driven by higher cell counts instead of elevated expression in individual cells.

We also observed that interferon responses are elevated among ART-treated PLWH. The relationship between interferon expression and HIV-1 infection has received increased scrutiny recently. There is some controversy whether interferon and ISGs hinder or facilitate HIV-1 progression. During acute infection, ISGs encode anti-HIV-1 restriction factors that function to inhibit viral replication. However, in chronic infection, ISGs also show persistent upregulation, which may be associated with adverse health outcomes[23].

We re-visit our initial study hypothesis that there are ISGs differentially expressed between ART-suppressed group and HIV-1 uninfected group. We did not observe significantly differentially expressed ISGs in our gene-level analysis. However, pathway-level analysis reveals a subset of ISGs contribute to the upregulated interferon response pathways. Some of the identified ISGs were reported to be associated with HIV/SIV outcomes in other studies. *IFI27* and *MX1* are identified to be rapidly upregulated in pathogenic SIV infection and to persist into chronic infection[8; 9]. *IFI35* and *MX1* expression have been reported in association with plasma HIV-1 set point during chronic infection[24]. Plasma levels of *CXCL10* have been reported directedly related to HIV-associated neurocognitive disorder in women with chronic HIV infection[25]. This suggested that this subset of ISGs appears to be associated with chronic infection. A recent study compared the CD4+ T cell gene expression between post-ART controllers and early viral rebounders in participants who have stopped therapy[26]. It identified ISGs including *ISG15*, *XAF1*, *TRIM25*, and *USP18* associated with the rebound. We did not identify these specific ISGs as part of a differentially expressed signature for immune activation.

The study has some limitations. First, the small sample size may reduce the robustness of the results. Previous power calculations estimated that 25 participants per group could provide 80% power to identify a 1.5-fold-change in expression with a coverage depth of 10 and coefficients of variation of counts per group of 0.4. We were unfortunately not able to achieve the sample size. The study should be considered as a pilot study, and the results from our small sample should be confirmed in larger studies. Second, the bulk RNA-seq may not be

able to account for cellular heterogeneity. The observed higher CD8A and CD8B expression among people living with HIV-1 on ART may come from a higher proportion of CD8+ cells in PBMCs samples instead of cellular upregulation.

Further studies are needed to identify ISGs that contribute to chronic immune activation in PLWH. The current observed differential gene expression between post-ART and uninfected controls may not be solely due to the chronic state of inflammation and immune activation. It is possible that some of this signature comes from effects of ART usage or HIV infection itself. Follow-up studies are needed to dissect these observations one-step further. We can conduct a longitudinal comparison between pre-ART and post-ART time points for the same individuals to identify ISGs driven by ART use and HIV replication. We can also conduct association test between ISG expression and clinical HIV or immune activation parameters to identify ISGs driven by HIV infection and inflammation. In addition, we can perform immunophenotyping to determine specific cell populations.

In summary, we found that genes involved in CD8+ T cell responses and multiple immune-related pathways were upregulated in virally suppressed HIV-1 infection compared to HIV-uninfected controls. They demonstrate evidence of sustained inflammation among virally suppressed PLWH compared to HIV-uninfected people. Notably, we also identified several ISGs contributing to interferon response enrichment signals. Our findings provide support and incremental evidence on previously non-human primate and human studies, which explore the relationship between ISGs and HIV clinical outcomes/chronic activation. Further studies are still needed to distinguish ISGs related viral infection from ISGs that contribute

to or are influenced by chronic immune activation. Those ISGs that drive chronic immune activation could then serve as potential candidate for a targeted intervention that could be used alongside ART to reduce inflammation during HIV-1 infection and restore healthy immune homeostasis.

3.6 TABLES & FIGURES

Figure 3.1 Study Background and Motivation.

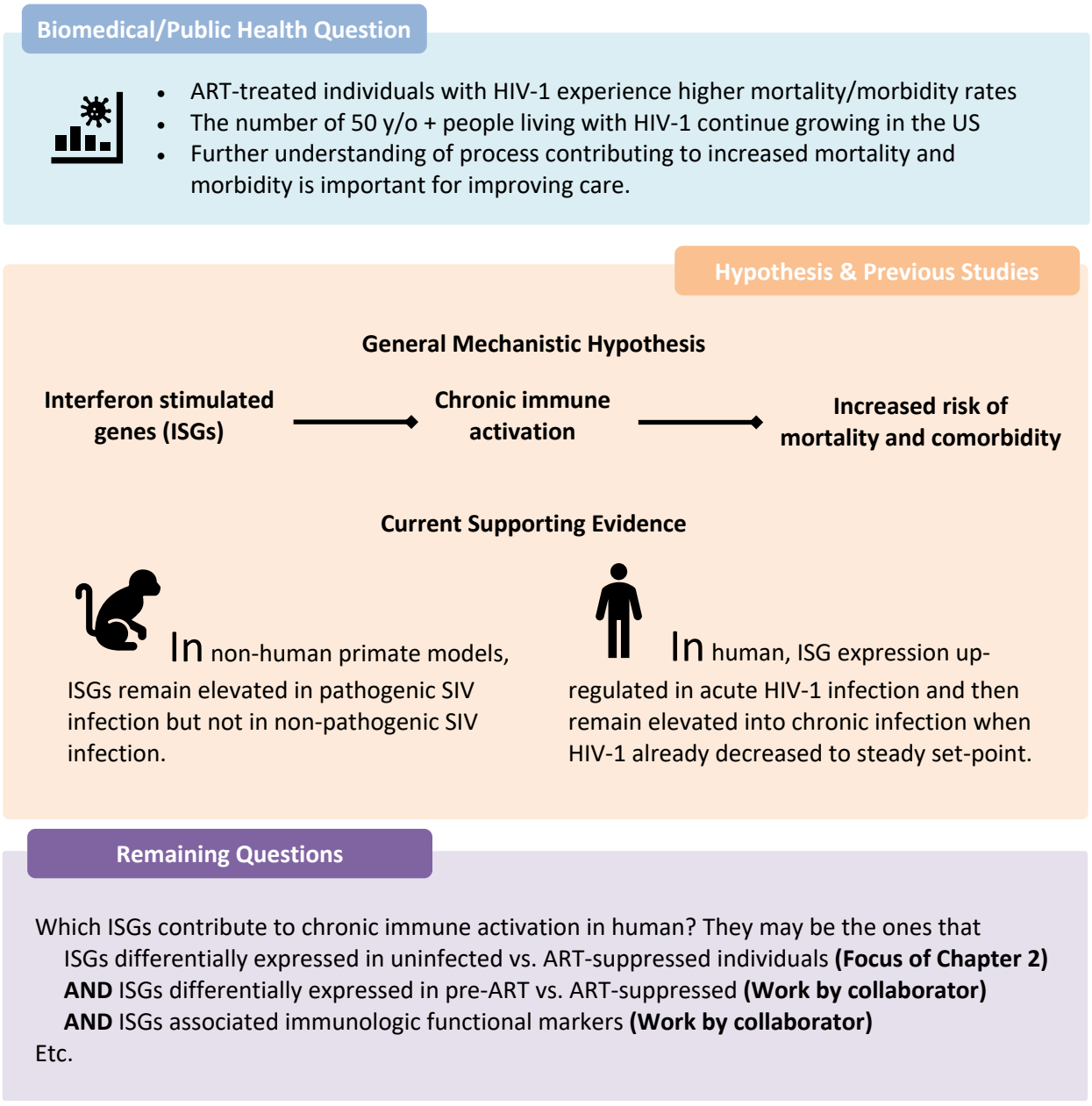
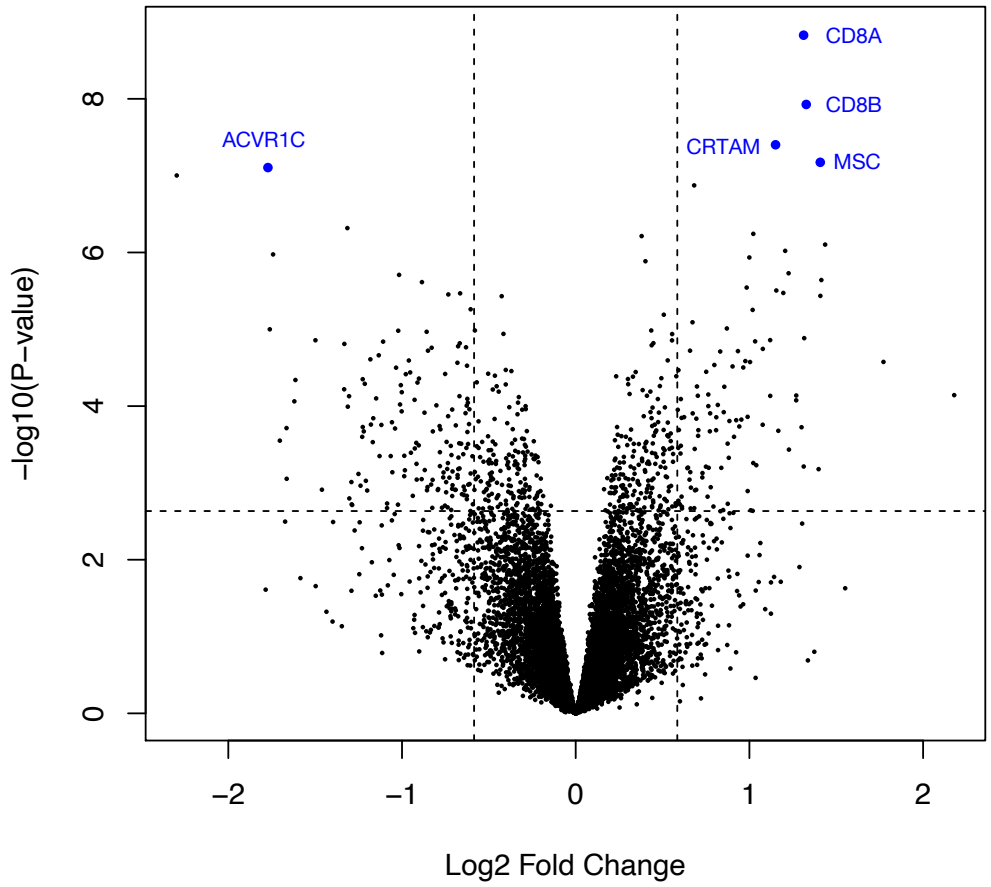


Table 3.1 Descriptive Statistics* of Study Subjects.

Characteristics	ART-suppressed (n=18)	Uninfected (n=12)
Female	12 (66.7%)	8 (66.7%)
Age (years)	32.0 (20.1, 53.4)	36.4 (20.0, 55.7)
Country of residence:		
Kenya	7	1
South Africa	2	3
Uganda	9	8
Cohort:		
COS	7	8
PrEP	11	4
HIV-1 viral load (log10 copies/mL):		
Pre-ART	4.87 (4.65, 5.01)	
Post-ART	0.40 (0, 0.96)	
CD 4 T cell counts (cells/mm^3):		
Pre-ART	300 (247, 356)	
Post-ART	434 (359, 519)	

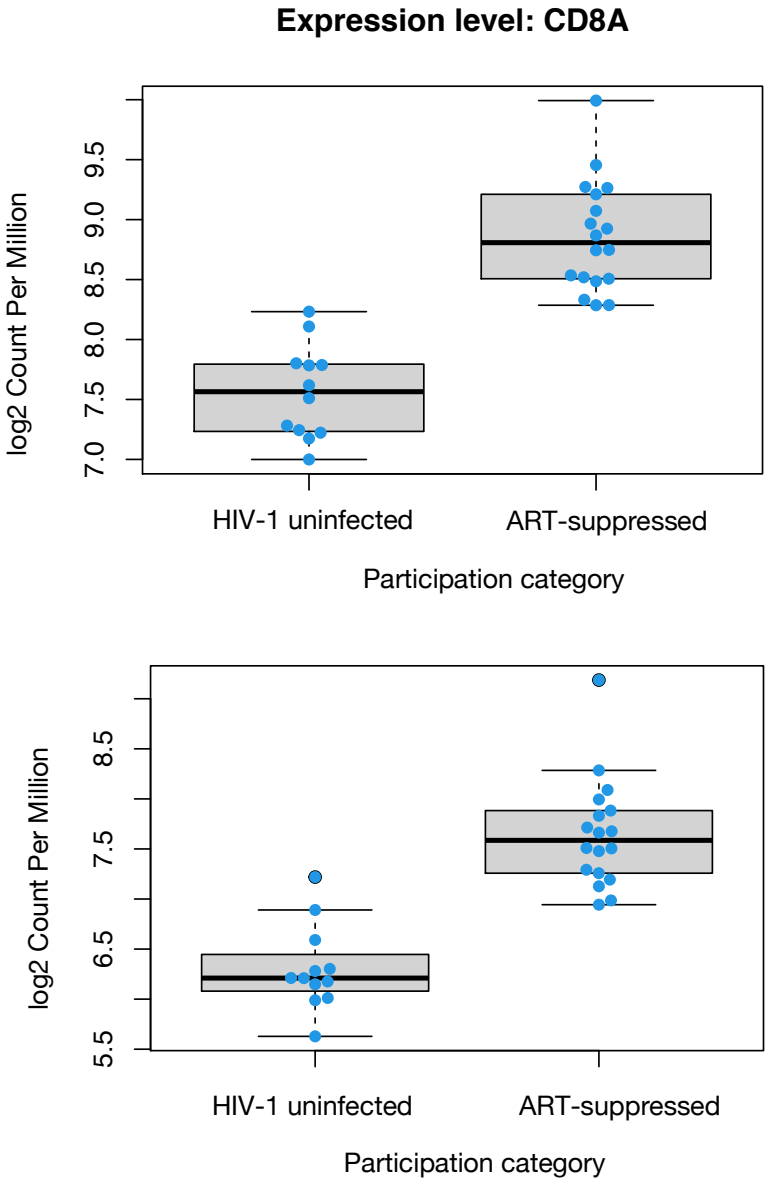
* Numbers (percentages) for categorical variables. Mean (Minimum, maximum) for continuous variables.
Abbreviation: COS- Couples Observational Study; PrEP- Partners PrEP cohort

Figure 3.2 Volcano Plot of Differentially Expressed Genes.



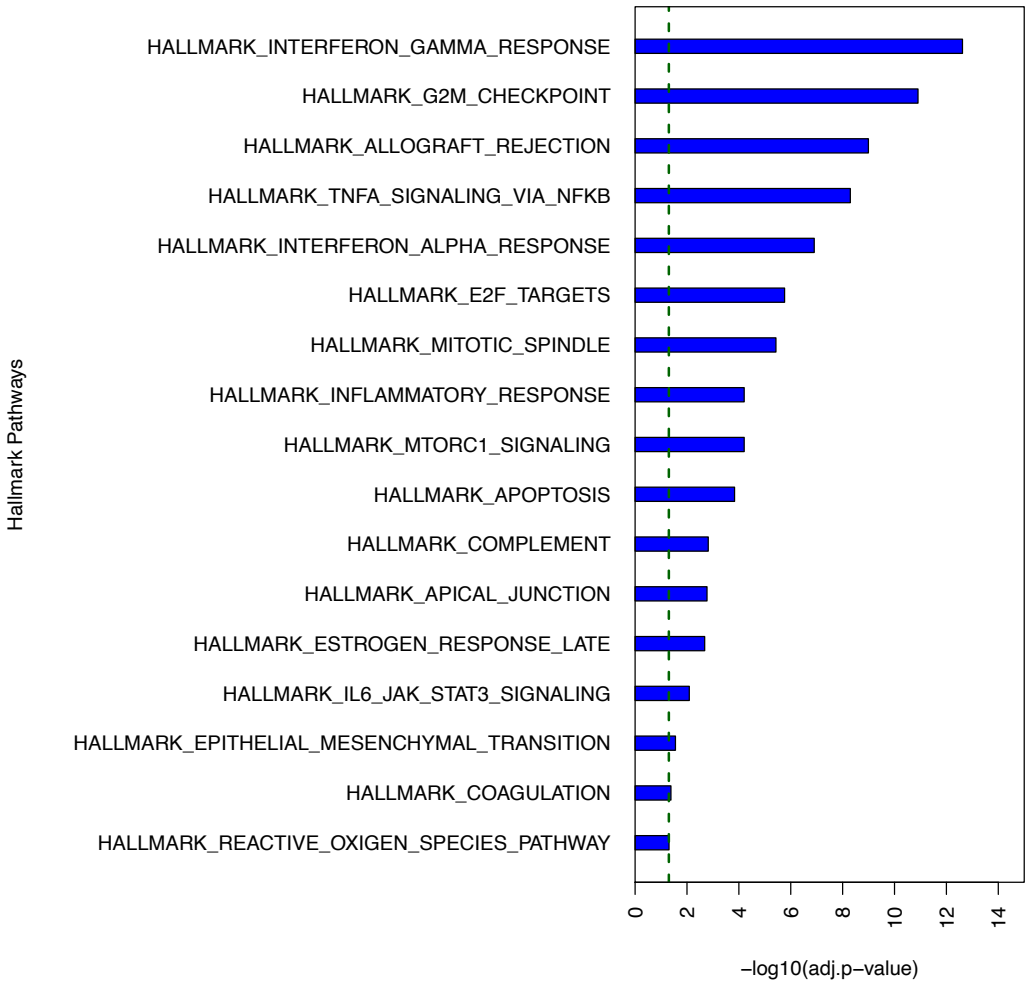
Volcano plot displaying the log 2-fold-changes and adjusted *p*-values for all tested genes. The top five signals with both large observed fold changes and small *p*-values are annotated with the human gene names in blue.

Figure 3.3 Post-ART Individuals Exhibited Higher Expression of *CD8A* and *CD8B*.



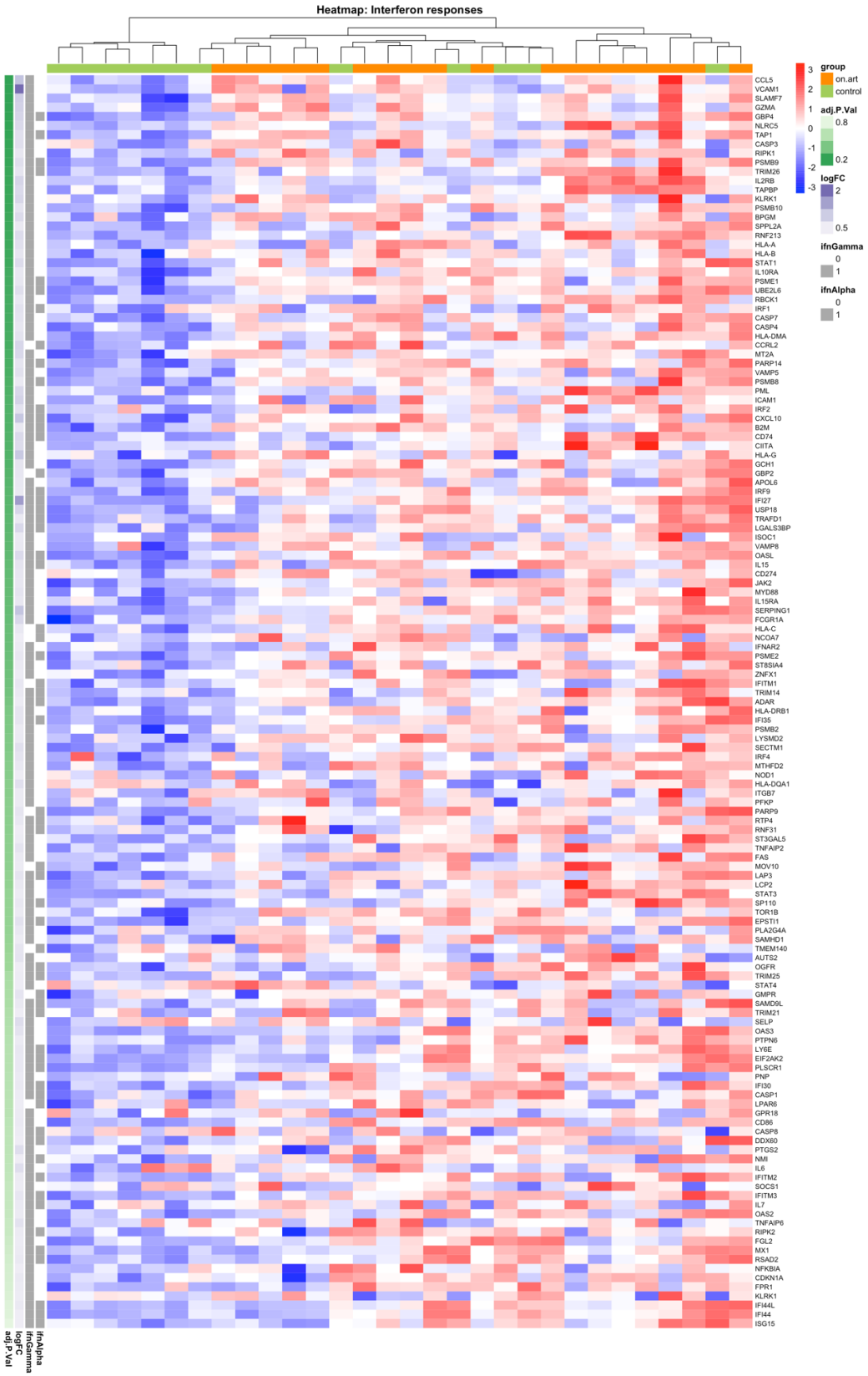
Boxplot with beeswarm plot superimposed for *CD8A* (Upper panel) and *CD8B* (lower panel) expression between post-ART subjects and uninfected controls. Higher expression of *CD8A* and *CD8B* were observed in post-ART compared to HIV-uninfected controls.

Figure 3.4 Bar Plot of Enriched Pathways in Virally Suppressed HIV-1 Infection.



Bar plot indicating the adjusted *p*-values of significant Hallmark pathways from GSEA-based pathway analysis of identified DEG. Several immune or inflammation-related pathways are identified, including interferon responses.

Figure 3.5 Heatmap of Leading Edges in Interferon Responses Pathways.



(Cont.) Heatmap containing the leading-edge genes (genes that contribute most to enrichment signals) of interferon-alpha and gamma response pathways. Cell color represents the gene expression level—higher expression in red and lower expression in blue. Samples are annotated with participants' group, with green indicating uninfected controls and orange indicating post-ART. Genes are annotated regarding whether they are involved in interferon-alpha and gamma response pathways, respectively, FDR-adjusted *p*-value, and log2 Fold change from differential gene expression analysis.

3.7 REFERENCES

1. Marcus, J.L., Leyden, W.A., Alexeeff, S.E., Anderson, A.N., Hechter, R.C., Hu, H., Lam, J.O., Towner, W.J., Yuan, Q., Horberg, M.A., et al. (2020). Comparison of Overall and Comorbidity-Free Life Expectancy Between Insured Adults With and Without HIV Infection, 2000-2016. JAMA Netw Open 3, e207954.

2. Centers for Disease Control and Prevention. HIV Surveillance Report. URL: https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html. Accessed: 14 December, 2022

3. Sandler, N.G., Wand, H., Roque, A., Law, M., Nason, M.C., Nixon, D.E., Pedersen, C., Ruxrungtham, K., Lewin, S.R., Emery, S., et al. (2011). Plasma levels of soluble CD14 independently predict mortality in HIV infection. J Infect Dis 203, 780-790.

4. Hunt, P.W., Cao, H.L., Muzoora, C., Ssewanyana, I., Bennett, J., Emenyonu, N., Kembabazi, A., Neilands, T.B., Bangsberg, D.R., Deeks, S.G., et al. (2011). Impact of CD8+ T-cell activation on CD4+ T-cell recovery and mortality in HIV-infected Ugandans initiating antiretroviral therapy. AIDS 25, 2123-2131.

5. Lyons, J.L., Uno, H., Ancuta, P., Kamat, A., Moore, D.J., Singer, E.J., Morgello, S., and Gabuzda, D. (2011). Plasma sCD14 is a biomarker associated with impaired neurocognitive test performance in attention and learning domains in HIV infection. J Acquir Immune Defic Syndr 57, 371-379.

6. Tincati, C., Bellistri, G.M., Casana, M., Merlini, E., Comi, L., Bai, F., Sinigaglia, E., Cristina, M., Carpani, G., Bini, T., et al. (2009). CD8+ hyperactivation and senescence correlate with early carotid intima-media thickness in HIV+ patients with no cardiovascular disease. J Acquir Immune Defic Syndr 51, 642-644.

7. Piconi, S., Trabattoni, D., Gori, A., Parisotto, S., Magni, C., Meraviglia, P., Bandera, A., Capetti, A., Rizzardini, G., and Clerici, M. (2010). Immune activation, apoptosis, and Treg activity are associated with persistently reduced CD4+ T-cell counts during antiretroviral therapy. AIDS 24, 1991-2000.

8. Jacquelin, B., Mayau, V., Targat, B., Liovat, A.S., Kunkel, D., Petitjean, G., Dillies, M.A., Roques, P., Butor, C., Silvestri, G., et al. (2009). Nonpathogenic SIV infection of African green monkeys induces a strong but rapidly controlled type I IFN response. J Clin Invest 119, 3544-3555.

9. Bosinger, S.E., Li, Q., Gordon, S.N., Klatt, N.R., Duan, L., Xu, L., Francella, N., Sidahmed, A., Smith, A.J., Cramer, E.M., et al. (2009). Global genomic analysis reveals rapid control of a robust innate response in SIV-infected sooty mangabeys. J Clin Invest 119, 3556-3572.

10. Baeten, J.M., Donnell, D., Ndase, P., Mugo, N.R., Campbell, J.D., Wangisi, J., Tappero, J.W., Bukusi, E.A., Cohen, C.R., Katabira, E., et al. (2012). Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. N Engl J Med 367, 399-410.

11. Lingappa, J.R., Petrovski, S., Kahle, E., Fellay, J., Shianna, K., McElrath, M.J., Thomas, K.K., Baeten, J.M., Celum, C., Wald, A., et al. (2011). Genomewide association study for determinants of HIV-1 acquisition and viral set point in HIV-1 serodiscordant couples with quantified virus exposure. PLoS One 6, e28632.

12. Jiang, H., Lei, R., Ding, S.W., and Zhu, S. (2014). Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15, 182.

13. Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8, 175-185.

14. Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8, 186-194.

15. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12, 357-360.

16. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

17. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923-930.

18. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47.

19. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 57, 289-300.

20. Korotkevich, G., Sukhov, V., and Sergushichev, A. (2019). Fast gene set enrichment analysis. bioRxiv.

21. Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 1, 417-425.

22. Mudd, J.C., and Lederman, M.M. (2014). CD8 T cell persistence in treated HIV infection. Curr Opin HIV AIDS 9, 500-505.

23. Doyle, T., Goujon, C., and Malim, M.H. (2015). HIV-1 and interferons: who's interfering with whom? Nat Rev Microbiol 13, 403-413.

24. Rotger, M., Dang, K.K., Fellay, J., Heinzen, E.L., Feng, S., Descombes, P., Shianna, K.V., Ge, D., Gunthard, H.F., Goldstein, D.B., et al. (2010). Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals. PLoS Pathog 6, e1000781.

25. Burlacu, R., Umlauf, A., Marcotte, T.D., Soontornniyomkij, B., Diaconu, C.C., Bulacu-Talnariu, A., Temereanca, A., Ruta, S.M., Letendre, S., Ene, L., et al. (2020). Plasma CXCL10 correlates with HAND in HIV-infected women. J Neurovirol 26, 23-31.

26. Zacharopoulou, P., Marchi, E., Ogbe, A., Robinson, N., Brown, H., Jones, M., Parolini, L., Pace, M., Grayson, N., Kaleebu, P., et al. (2022). Expression of type I interferon-associated genes at antiretroviral therapy interruption predicts HIV virological rebound. Sci Rep 12, 462.

Chapter 4. INFORMED CONSENT IN GENETIC/GENOMIC STUDIES IN AFRICA: A SYSTEMATIC REVIEW OF POTENTIAL BIOETHICAL ISSUES

4.1 ABSTRACT

Objective: International collaboration on genetic/genomic studies has expanded dramatically, yet, the application of traditional ethical principles, most often through informed consent, may be challenging when conducted in culturally diverse settings. Recommendations for conducting ethical international collaborations offer best practices ranging from assessing comprehension to strategies for community engagement yet, adherence to best practices and whether these represent the scope of possible ethical issues raised by international studies remains unknown.

Methods: We performed a systematic literature review of empirical studies that explored stakeholders' perspectives on informed consent in the context of international genetic/genomic research, primarily in countries of Africa. Twenty-three relevant articles were identified and reviewed for five topics: (1) comprehension, (2) voluntary participation, (3) consent elements, (4) consent model, and (5) community engagement.

Results: Different levels of understanding and recall over different consent elements, and multiple factors associated with comprehension were observed. Voluntary participation could be influenced by misconception, monetary and healthcare compensation, and previously established trust with the research team. Many participants show less interest in certain consent elements, such as biobanking, data sharing, and future use, but more in the section on benefits. Returning results had potential advantages to build trust between

participants and the research team; however, the most appropriate way to return results is a topic under discussion. There was also conversation over consent models, broad, tired, and dynamic, on their balance between autonomy and practicality. Finally, many articles ascertained the value of community engagement and encouraged researchers to consider local cultural beliefs, social stigma, and decision-making habits when designing the consent process.

Conclusion: We observed an increased interest in the ethical conduct of the informed consent process under international genetics/genomics studies in recent decades. Awareness and attention to these issues are needed to develop more appropriate consent strategies that could further safeguard the right and welfare of local participants and support the smooth conduct of research.

4.2 INTRODUCTION

Genetics and genomic studies have experienced growth and development in recent decades due to scientific advances and the falling cost of sequencing technologies. In global collaborations between countries, researchers recruit subjects and collect biosamples in one place and then rely on genetic analyses conducted outside of the specimens' countries-of-origin[1; 2]. In this process, several critical ethical concerns have been raised about the ability to ethically engage, recruit and consent study participants[3-5].

Informed consent ensures research participants are aware of the risk and potential benefits and make voluntary decisions about participation. It is based on the ethical principle of autonomy and is required by the U.S. Federal Policy for the Protection of Human Subjects[6] 45 CFR Part 46, referred to as the Common Rule. The concept of informed consent in human subject research originated from the Nuremberg Trials following World War II. Nazi defendants were charged with crimes against humanity and war crimes for unspeakable medical experiments on concentration camp inmates. In 1947, the verdict of the International Military Tribunal in the case of United States of America vs. Karl Brandt et al. set out ten principles of human experimentation, known as The Nuremberg Code, which later were codified into laws regulating the ethical conduct of human subject research. Voluntary consent from human subjects is the first and foremost principle in the code[7].

When invited to participate in a research study, prospective research participants or their legal representatives are provided key information about the study in question so they may voluntarily decide whether or not to participate in the study. Essential elements of the

consent include (1) a declaration that the study involves research, (2) a statement of research purpose, (3) the expected duration of participation, (4) a description of research procedures, (5) a description of potential risks or discomforts, (6) a description of potential benefit, (7) a disclosure of alternative procedure or treatment that might benefit the subjects (8) a description of any compensation, (9) a confidentiality statement, (10) a statement of voluntary participation and withdrawal, and (11) contact information for pertinent questions[8]. Additional elements of special consideration for genetic and genomic research include a long-term, uncertain privacy risk given the stability of DNA overtime and broad sharing of samples or data, and group harm given the implication of genetic information for families and communities, among others[9]. To ensure that participation is voluntary, researchers are obligated to confirm that each prospective participant understands the information provided and that the invitation to participate (i.e., recruitment procedures, incentives, etc.) is not coerced.

Research conducted in international settings must contend with variation in levels of educational attainment and common modes of communication, which raise concerns about comprehension of study information and thus voluntariness of research participation. For example, researchers from the International Clinical Research Center (ICRC) at the University of Washington, who have field work experiences in Sub-Sahara Africa, have noted that some participants did not finish high-school and their local language, Kiswahili, has an oral tradition[10]. Such social and cultural contexts of research raise the possibility that participants might practicably find voiced-over and multimedia-based study information

easier to comprehend as compared to traditional text-based modes, especially given the complexity of genomic research, data sharing, and other relevant research details.

In this study, we hypothesize that culturally diverse contexts may raise additional ELSI concerns in the informed consent process for genetic/genomic studies. We aim to identify and describe those ELSI challenges by performing a systematic literature review of empirical studies that explored stakeholders' perspectives on informed consent process in the context of genetic and genomic research in Africa.

4.3 METHODS

This systematic literature review sought to identify bioethical issues raised in the informed consent process in international genetic/genomic studies, primarily Africa. The review is conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guidelines[11]. As such, this review specifies information sources, search strategy, eligibility criteria, selected articles, and details data collection, coding, and synthesis. The protocol for the systematic review is not registered.

4.3.1 *Data sources and search strategy*

Based on differences in content coverage, three search platforms were used to maximize the number of possibly relevant articles. PubMed MEDLINE (National Library of Medicine), Web of Sciences (Clarivate), and Google Scholar were searched. PubMed accesses databases on life science and biomedical topics. Web of Sciences accesses databases on multidisciplinary disciplines including science, technology, social sciences, arts, and humanity. Google Scholar also accesses multidisciplinary fields.

For PubMed, the advanced search interface was used with the search syntax ("informed consent") AND (genetics OR genomics) AND (Africa) AND (ethics). Web of Science advanced search was queried using the search terms ALL=(("informed consent") AND (genetics OR genomics) AND (Africa) AND (ethics)). No filtration on publication year, date, language, and article types were applied. The search strategy yielded 267 and 28 records for PubMed and Web of Sciences, respectively. The first 50 “hits” in Google scholar were included because they were most relevant among the million hits returned by the search engine. Searches were conducted on April 26th, 2022. Records were exported and saved to a Microsoft Excel workbook.

4.3.2 *Eligibility Criteria*

Eligible studies were included if they met the following criteria:

(i) published in peer-reviewed journals in English language.

(ii) having full text available.

(iii) empirical study including case-study, qualitative interview/focus-group discussion, and survey/questionnaire.

(iv) address ELSI issues in the informed consent process related to host genetic/genomic research.

(v) were conducted or topically about international research in Africa.

The articles were excluded if they were

(i) abstracts, conference proceedings, review opinion papers, or document analyses.

4.3.3 *Literature screening*

We employed a double screening approach. Two members (YL, HK) of the research team independently screened the search results to ensure the consistency and reproducibility of the application of eligibility criteria.

Two-levels of screening was conducted, and an Excel spreadsheet was used to track articles. The literature search initially yielded 345 articles. After removing 76 duplicate articles, 269 remained for further screening. These articles were first screened based on their titles and abstracts for keywords to identify whether they address ethical issues (as opposed to presenting non-ethics-related scientific/biomedical discoveries), informed consent, and genetics or genomics research. This first-level screening process resulted in an exclusion of 184 articles, leaving 85 that were read in full to verify eligibility. Next, the 85 were screened based on full text read to determine if they were empirical studies, conducted in Africa settings, and addressed consent in host genetics/genomic research. The second-level screening process resulted in the exclusion of 62 additional articles, leaving 23 articles for data extraction, coding, and analysis (**Figure 4.1**).

4.3.4 *Data extraction, coding, and analysis*

Upon the collection of articles, we first used a tracking spreadsheet to summarize the characteristics of the included articles. We organized the articles by year of publication, study design, study subjects, study sites, main funding, and any recommendations for improvement if proposed.

We generated a codebook to support the analysis of consent within the articles (**Table 4.2**). We created initial codes of issues drawing from the *H3Africa Guideline for Informed Consent*[12] and the opinion article *Informed Consent in Genomics and Genetic Research*[9]. During the data abstraction and coding process, the coding framework was iteratively revised with existing issue topics edited for clarity and new issue topics added.

Each included article was retrieved for full text and coded. One member (YL) of the research team extracted issues using the defined coding framework, and one member (HK) reviewed the coding to ensure the appropriateness and consistency of the code application. Coding discrepancies were resolved through discussion and consensus between the two members. Data were summarized in a spreadsheet in wide format, with each article as an entry and each issue topic as a column.

4.4 RESULTS

Results are presented in three parts: first, an overview of the characteristics of the included articles; second, the frequency of the issue topics across articles; and third, key descriptive findings from each issue topic.

4.4.1 *Characteristics of the reviewed articles*

Twenty-three articles were included in the review (**Figure 4.1**). Articles spanned between 2006 and 2022. Thirteen (57%) were published in 2018 or later, and only four were published before 2010. The studies were conducted in ten countries in Africa, including South Africa (8), Ghana (4), Nigeria (3), Gambia (2), Tanzania (2), Kenya (2), Cameroon,

Ethiopia, Mali, and Uganda. The majority of the studies had authorship collaborations with researchers from the United States and United Kingdom.

The studies employed multiple non-mutually exclusive categories of designs: seventeen (17) were semi-structured/in-depth interviews, nine (9) were focus-group discussions, four (4) were surveys, and two (2) were case observations. They involved various key stakeholders including recruited local participants in genetic/genomic research, non-participants who were potentially eligible for specific studies, research staff, fieldworkers, ethics board members, and community members - community leader, community advisory board, community liaison group.

4.4.2 *Frequency of issue topics across articles*

Participants' *comprehension* of the information provided in the consent appeared most often across reviewed articles. *Comprehension* appeared in thirteen (13) of the reviewed articles, followed by *voluntary participation* in twelve (12) articles, *consent element* in eleven (11) articles, *community engagement* in nine (9) articles, and *consent models* in two (2) articles. (**Table 4.2**)

4.4.3 *Description of topics and issues raised about informed consent in international genetic/genomic studies*

Comprehension

Of the 13 studies that assessed participants' comprehension, study participants demonstrated different levels of understanding and recollection of information provided in the consent process. Participants reported a high level of comprehension of study goals and

study procedures while reported difficulty in understanding of biobanking and future use (see "consent element" section for more discussion)[13-16].

Concept of gene and heritability. One feature of genetic/genomic research is that the consent process may utilize several technical terms or describe a technical process. One study reported that the indigenous population, Yoruba, in Nigeria has its own words, proverbs, and aphorisms for genetic or heritability-related concepts[17]. With the local dialect, concepts such as the transmission of traits by "blood," dominant and recessive traits, and penetrance could be understood appropriately. The study further suggested that incorporating indigenous linguistic and cultural concepts that describe genes or heritability may help improve comprehension and accuracy in the consent process.

Factors influencing understanding. Seven studies described factors that could influence participants' comprehension and multiple factors have been reported including: time research staff spending with the study participants in the process[18]; adequacy of information provided and quality of the interaction[18]; the age of participant[18]; literacy level[16; 18]; whether this is an iterative learning process[19]; whether there are comparable concepts in the local language[20; 21]; and whether the local language has a standard written form[20]. Accordingly, these studies proposed possible ways to improve understanding during the consent process including holding group sessions before the individual consenting process; giving Q&A opportunities to study participants[18]; providing intensive and continued training to research staff about research and consent element, attitude and communication[19; 20; 22]; explaining the element in simple, accessible language[18; 20]; including practical examples[20; 23]; utilizing visual

aids including pictures, videos, diagram[23; 24]; planning the consent process in considering local and population needs[18]; and planning re-consenting phase if the study has long time span[18].

Voluntary participation

Two qualitative studies reported that the majority of their participants had not experienced pressure to participate and that they understood that they could withdraw from the study[15; 16]. However, it is worth noting that in two studies, participants made their decision based on trust and previous positive experience with research rather than genuine voluntary participation [23; 25].

Reasons for participation/non-participation. Two studies reported that monetary compensation and free healthcare services were common reasons for participation[23; 26]. The specific reason varied between adults and youth: adults decided to participate based on altruism and the common good, while youth tended to make their decisions based on personal benefit, preferences, and family welfare[27]. Time conflicts, negative feeling of taking information away but not addressing local issues, and aggravation of stigmatization associated with the familial nature of diseases were among the reasons for non-participation[26; 28].

Misconception. Participants showed a misconception of research benefits. Some of them believed the research could identify the cause of illness or help identify a treatment; however, this was not true for genetic epidemiology studies[29]. Participants commonly interpreted

research as medical care (known as therapeutic misconception)[30]. Additionally, in a case-control genetic study, participants showed diagnostic misconceptions (i.e., participants interpreted research to be a diagnostic test, and join the study in order to learn about their health status) among healthy controls and therapeutic misconceptions among cases[31]. A suggestion to overcome this issue is to discuss the difference between research and medical care in simple and clear term[23].

Consent elements

Studies reported that participants had different levels of interest in each consent element. One study showed that participants considered the study benefit as the most important information to know[23], while another study reported less interest to the learn about the study itself[26]. This lack of interest in the study is consistent with the observation that many participants did not ask questions about study aspects such as archiving samples, future use or data sharing, all of which researchers considered to be complex concepts that are difficult to understand initially[14; 30]. The authors, therefore, suggested that the consent process focus on the potential benefit of proposed research to participant and community.

Biobanking. Besides less interest in biobanking, multiple considerations emerged around this topic. One study found that participants had a number of questions such as what is meant by biobanking or cell immortalization; what are the individual and community benefit[24]? Another study questioned, from the researchers' standpoint: what is the competence of the research ethics committee; how to measure and explain the risk vs. benefit of biobanking; which consent model is most appropriate; and how to manage and sustain storage?[32].

Data sharing. One study of genomic dataset sharing found that participants might not understand data sharing from consent language stating, "if this identified a particular cluster of disease… then data ought only to be shared for such research…"[33]. Instead, the study found that participants gave their consent based mainly on trust.

Future use. Like biobanking and data sharing, interview studies of participants and research staff reported that, similar to biobanking and data sharing, participants and staff were less aware of possible future secondary uses and the breadth of the generated data[13; 25].

Returning results. One study reported that research staff did not discuss the issue of returning results with study participants, despite that this was required by the local ethics guideline[34]. There was an on-going discussion regarding this issue. On the one-hand, returning results is highly preferred. A study reported that participants might show an expectation and gratitude for receiving results[14]. They believed that the results could fulfill their curiosity of knowledge and benefit them when the result is clinically significant or actionable. In addition, absence of feedback in prior studies could discourage community members from future research participation. On the other hand, returning results should be treated with caution. The same study also pointed out that most of the study results would not have direct benefit or could not be understood without further specialized analysis. It further summarized barriers to returning results such as the absence of a well-trained genetic counselor to appropriately deliver the results.

Consent model

Articles discussed different consent models including broad, tiered, and dynamic models of consent. Some researchers argued that one-time broad consent is preferred because re-consenting is resource-intensive and thus likely impractical, while others suggested that broad consent could become more compatible with recent technological advances in remote communication[32]. Another perspective was skeptical of broad consent, arguing that keeping the community updated could be an effective way to sustain a trust relationship between participants and researchers[14].

Community Engagement

Among the nine articles that discuss community engagement, they all valued community engagement and encouraged researchers to consider its application when designing the consent process.

Engagement Structure. Community engagement has been discussed at various levels and various stages of the consent process. One study summarized three levels of community engagement: high level including international/national/institution decision makers; peer level including researchers in biomedical field; and community level including patients, member of supporting groups, community advisory board, field workers, general public, etc[22]. Several studies highlighted the key role of field workers[20; 24; 30], whose understanding, attitude and communication were key influences on the ethical practice of consent locally.

Cultural beliefs. Developing a consent process may also need to consider common beliefs in culture, community, and society. For example, there were cultural beliefs about the withdrawal of biosamples[22]. One study reported that some traditional communities believed that if their blood and hair remained in a fridge, they would linger in this life. Or they worried that their blood could be used in witchcraft. It was also observed that participants became agitated during blood draw because of their belief in the power of amulets[14]. Other examples were associated with beliefs about heritability[22], where participants believed that it would be disrespectful to ancestors if they explained inheritance.

Social stigma. Social stigma may also impact the process of obtaining informed consent. One study focused on the stigma around podoconiosis, also known as mossy food[28]. Despite the fact that podoconiosis is caused by long-term red clay soil exposure, it is widely believed in endemic areas of Ethiopia that podoconiosis run in families and is an untreatable condition. Based on these beliefs, community members worried that participation in a genetic study might reinforce the view that the disease is familial. Furthermore, the staff of a local organization working to prevent environmental exposure to mossy foot, also reported fear that discussing its familial nature might disappoint local patients who had already been stigmatized and threaten established trust. This research team later demonstrated that addressing the community's feedback on issues like this prior to the conduct of the actual study could benefit the development and application of the consent process.

Shared decision making. Permission by elders, community leaders, husbands, healthcare providers and or religious leaders might be necessary prior to the enrollment and consent

of individuals in some communities[22; 35]; however, views varied as to whether such permission is needed in practice. For example, in Wolaita, Ethiopia, it was more common for people to receive information through local community gatherings, group discussions, and consultations instead of one-to-one discussion[23]. Therefore, community consent might be preferred as a practical issue. In contrast, participants from another study in Igbo-Ora, Nigeria, reported that no one should seek permission from community elders, probably because of the local community's view that elders are not considered essential in personal decision making about research participation[16].

4.5 DISCUSSION

This systematic review outlines the state of current empirical bioethical research on informed consent in international collaborative genetic/genomic studies in Sub-Saharan Africa. There is an increasing trend in the number of published articles discussing ethics, suggesting growing interest in the ethical conduct of consent within the past few decades. Frequently raised topics and issues include comprehension, voluntary participation, consent elements, consent model, and community engagement.

A good number of included studies (n=9) were supported by or collaborated with the Human Heredity and Health in Africa (H3Africa) initiative. As a partnership between the African Society for Human Genetics, the U.S. National Institutes of Health (NIH), and U.K. Welcome Trust, the initiative was launched in 2010 with part of its mission to understand ELSI questions that accompany genetic/genomic research in Africa[36]. Our review found that H3Africa studies tended to include a wide range of emerging issue topics, such as community

engagement, shared decision-making, misconception in voluntary participation, comprehension, and recall, etc. This is not particularly surprising given that this initiative has made a consistent effort to access a wealth of expertise in bioethics and policy of genetics and genomic research in recent years. Valuable lessons can be drawn from their experiences and model. For example, it is particularly noteworthy that they foster local professional leadership, and actively integrate community values in genetic and genomic research through engagement at all-levels.

Among the reviewed studies, there are several topics and issues that have received less attention. Despite intensive discussion in the United States, few articles discuss data sharing and the consent process: for example, whether sharing genomic or phenotypic data or study results through open or access-restricted databases; how to sign the consent when participants are illiterate; whether the consent is electronic-based or paper-based; and whether to adopt a governance approach rather than individual consent as the way to protect participants.

Returning to the initial motive for this review – utilizing multi-media-based informed consent with voice-over to facilitate participants' comprehension –, we identified studies that echo this point by suggesting researchers include visual aids, such as pictures, videos, or diagrams, when designing the consent[20; 23]. Other studies also demonstrate that it would be helpful to understand if the consent process utilizes a simple, accessible local language, especially when there are comparable concepts in genetics and heritability[17; 18; 21]. One additional consideration is that a community may not be familiar with multi-media formats,

thus inadvertently creating an additional barrier. In the U. S., electronic informed consent has been implemented in the All of Us Research Program with post-consent assessments reporting optimal levels of participant understanding of differences between research and medical care, voluntary participation, and withdrawal, but less understanding of privacy risks[37]. In addition, the study showed that some words used in the consent might not be familiar to low-health-literacy participants, therefore, suggesting a participant-centered consent development approach. Taken together, the design of a multi-media-based consent process for international genetic/genomic study may also want to consider co-creating the process with the local community members to anticipate contextual factors.

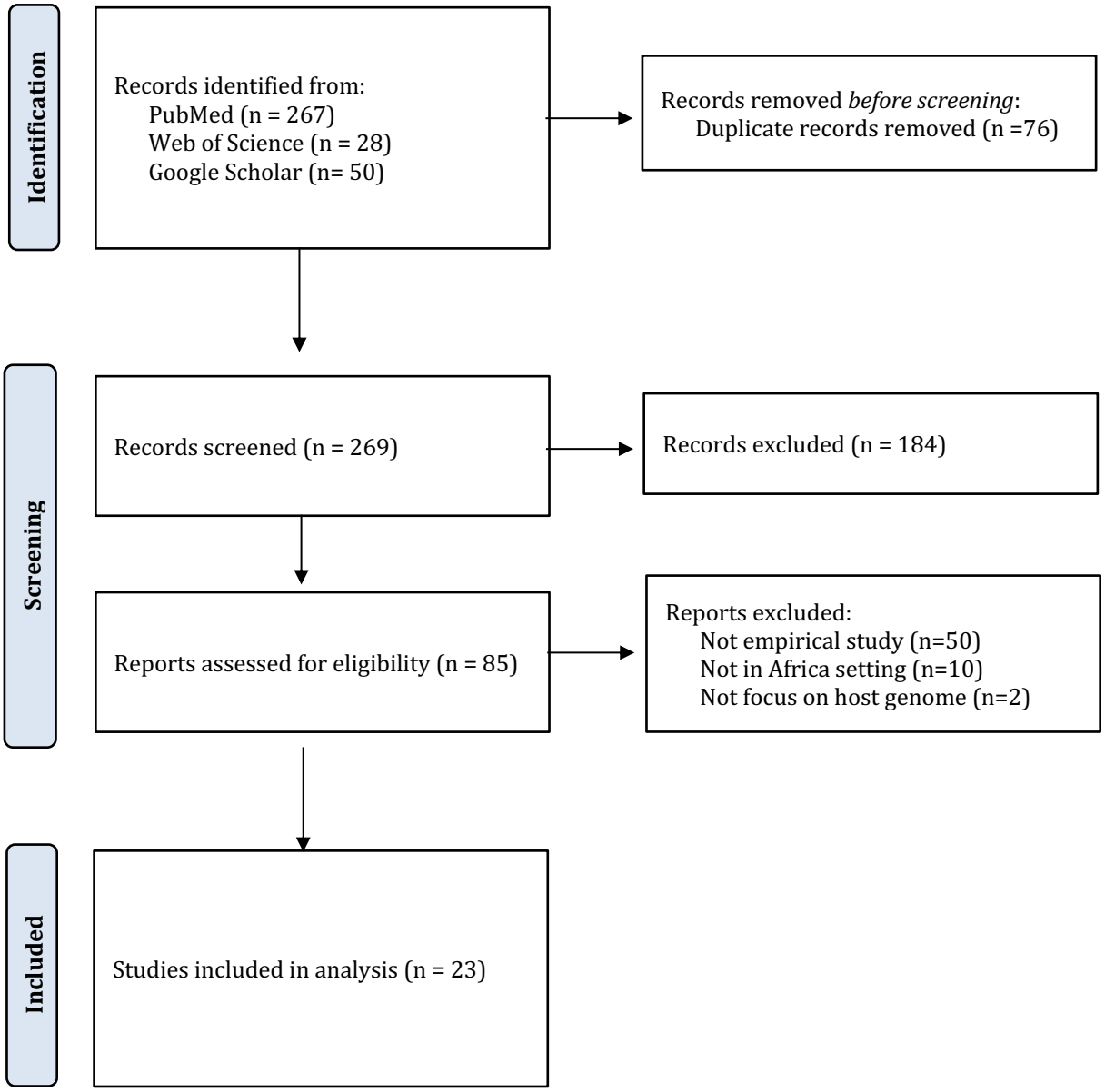
This study contains the following limitations. First, all of the reviewed studies were conducted in sub-Saharan Africa, and findings might not generalize to other geographical setting of the world, such as South America, Asia, the Pacific, etc. Second, although we attempted to include a comprehensive list of literature within the review, we may still have missed potentially relevant study articles. During the screening, we removed an article if its title and abstract focused exclusively on scientific discovery, but ELSI issues might have been discussed in some part of the main text. Third, some reviewed studies may suffer from sampling bias. Some reviewed studies were studies of general residents who had not participated previously in genetic/genomic studies. Their perspectives more likely express attitudes or preferences that may differ from those of research subjects who had previously undergone an informed consent process. In addition, few of the reviewed studies included individuals who declined participation in genetic/genomic studies.

Future work may give closer attention to disease-specific issues. For example, studies of host- and viral-genomics in infectious disease are increasing. Phylogenetic analysis of viral genomic data, such as for HIV-1, could suggest a pathogen transmission path between individuals. A study of HIV-1 serodifferent couples found that 30% of the seroconversion events that occurred during study follow-up were not attributable to the enrolled seropositive partner suggesting the seroconverting partner was not monogamous[38]. Disclosure of such information to participants may yield negative consequences such as loss of trust between partners, relationship dissolution, and potentially domestic violence between couples. This raises ethical concerns about the potential risks from disclosure of participant viral genomic data.

Another direction for future work is the possible role of community consent. As opposed to individual-based consent, community consent is a group-level governance strategy to safeguard the autonomy of the whole community. For instance, in the context of recruiting indigenous groups or tribes in the U.S. for genomic studies, individual consent obtained from tribal members living outside of their community may ignore group privacy risks when the community is small and an individual's genome is sufficiently representative for scientific inferences about the larger group[39]. Researchers should consider, beyond the individual-level, how their research may impact socially or politically impacts the very communities they seek to help.

4.6 TABLES & FIGURES

Figure 4.1 PRISMA Flowchart.



The PRISMA flow diagram illustrates study selection steps in this systematic review.

Table 4.1 List of Included Studies (n = 23).

No.	Article	Year	Authors	Design	Subjects	Sites
1	Participant recall and understandings of information on biobanking and future genomic research: experiences from a multi-disease community-based health screening and biobank platform in rural South Africa[13]	2022	Luthuli et al.	Semi-structured interview	Recruited Vukuzazi* study participants, non- participants but are potentially eligible, and research staff (n=39)	KwaZulu-Natal, South Africa
2	Perspectives and ethical considerations for return of genetics and genomics research results: a qualitative study of genomics researchers in Uganda[34]	2021	Ochieng et al.	In-depth interview	Researchers involved in genomics studies (n=30)	Uganda
3	Qualitative study of comprehension of heritability in genomics studies among the Yoruba in Nigeria[17]	2020	Taiwo et al.	Focus group discussion and key informant interview	Yoruba residents in Ibadan who had not participated in any genomics research (n=140)	Ibadan, Nigeria
4	Participant understanding of informed consent in a multidisease community-based health screening and biobank platform in rural South Africa[26]	2020	Ngwenya et al.	Semi-structured interview	Recruited Vukuzazi* study participants, non- participants but are potentially eligible, and research staff (n=39)	KwaZulu-Natal, South Africa
5	A qualitative study on aspects of consent for genomic research in communities with low literacy[18]	2020	Bukini et al.	Focus-group discussion (n=30), in-depth interview (n=13), and participants observation	Recruited participants in Sickle Cell Disease study	Tanzania
6	Informed consent in genomic research and biobanking: taking feedback of findings seriously[14]	2020	Tindana et al.	Focus-group discussion, individual interview (n=21), and deliberative workshops	Implementers, reviewers of research, and research participants and community members (n=126)	Ghana
7	Exploring the Role of Shared Decision Making in the Consent Process for Pediatric Genomics	2019	Bukini et al.	Individual interview (n=14), and focus-group discussion (n=49)	Recruited participants in Sickle Cell Disease study (n=64)	Ghana, Cameroon, and Tanzania

	Research in Cameroon, Tanzania, and Ghana[35]					
8	Rules of engagement: perspectives on stakeholder engagement for genomic biobanking research in South Africa[22]	2018	Staunton et al.	In-depth interview	Researchers involved in genomic and biobanking studies and/or community engagement (n=30)	South African
9	Relative solidarity: Conceptualising communal participation in genomic research among potential research participants in a developing Sub-Saharan African setting[27]	2018	Ogunrin et al.	Focus Group Discussion	Community members registered at and attend research facility (potential but not current/past participants in genomic research)	South west Nigeria
10	Predictors of consent to cell line creation and immortalisation in a South African schizophrenia genomics study[24]	2018	Campbell et al.	Survey	Recruited SAX (The Genomics of Schizophrenia in South African Xhosa People) participants (n=760 cases, n=760 controls)	South Africa
11	Using iterative learning to improve understanding during the informed consent process in a South African psychiatric genomics study[19]	2017	Campbell et al.	University of California, San Diego Brief Assessment of Capacity to Consent Questionnaire (UBACC)	Recruited SAX (The Genomics of Schizophrenia in South African Xhosa People) participants (n=1056)	Eastern and Western Cape, South Africa
12	"I passed the test!" Evidence of diagnostic misconception in the recruitment of population controls for an H3Africa genomic study in Cape Town, South Africa[31]	2017	Masiye et al.	In-depth interview(n=34), participant observations(n=57)	Recruited participants in RHDGen study (identify genetics risk factor for Rheumatic Heart Disease), and research staff	Cape Town, South Africa
13	"It's all about trust": reflections of researchers on the complexity and controversy surrounding biobanking in South Africa[32]	2016	Moodley et al.	In-depth interview	Medical and scientific researchers, biobanking experts and governance experts working across different disciplines: virology, hematology, immunology, pathology, HIV and Tuberculosis (n=21)	Western Cape, Gauteng, Kwa-Zulu Natal, South Africa
14	Understandings of genomic research in developing countries: a	2015	Traore et al.	Semi-structured interview (n=64), focus-group	Participant's parents and researchers of MalariaGEN study	Mali

	qualitative study of the views of MalariaGEN participants in Mali[21]			discussion (n=5), observation of consent processes	(which enrolled young children with parental consent into a study to identify genetic factors affecting immune responses to malaria in 11 African and two Asian countries)	
15	Exploring researchers' experiences of working with a researcher-driven, population-specific community advisory board in a South African schizophrenia genomics study[20]	2015	Campbell et al.	Meeting discussion	SAX (The Genomics of Schizophrenia in South African Xhosa People) Community advisory board (n=10)	South Africa
16	Knowing who to trust: exploring the role of 'ethical metadata' in mediating risk of harm in collaborative genomics research in Africa[33]	2014	de Vries et al.	Semi-structured interview	Fieldworkers, researchers, and ethics committee member in MalariaGEN study (n=49), a study of genetic factors influencing immune responses to malaria	Kenya, Gambia, and the UK
17	Voluntary participation and comprehension of informed consent in a genetic epidemiological study of breast cancer in Nigeria[15]	2014	Marshall et al.	Survey (n=215) and in-depth interview (n=17)	Recruited participants in breast cancer genetic epidemiology study	Ibadan, Nigeria
18	Seeking consent to genetic and genomic research in a rural Ghanaian setting: a qualitative study of the MalariaGEN experience[25]	2012	Tindana et al.	In-depth interview, focus-group discussion, observation of consent process	Participants and researchers in MalariaGEN study (n=84), a study of genetic factors influencing immune responses to malaria	Ghana
19	Participants' perceptions of research benefits in an African genetic epidemiology study[29]	2011	Appiah-poku et al.	In-depth interview	Research participants in tuberculosis genetic epidemiology study (n=25 cases and n=25 controls)	Kumasi, Ghana
20	Experiences with community engagement and informed consent in a genetic cohort study of severe childhood diseases in Kenya[30]	2010	March et al.	Case observation (Documents of community engagement, staff meeting, training, etc.)	Researchers and community liaison group at KWTP (collaboration between the Kenya Medical Research	Kenya

					Institute and the Wellcome Trust, UK)	
21	Tailoring consent to context: designing an appropriate consent process for a biomedical study in a low income setting[23]	2009	Tekola et al.	Focus group discussion and in-depth interview	Podoconiosis patients and non-patients in the community, fieldworkers, researchers, staff of the local non-governmental organization working on prevention and treatment of podoconiosis, and community leaders (n=46)	South Ethiopia
22	Impact of social stigma on the process of obtaining informed consent for genetic research on podoconiosis: a qualitative study[28]	2009	Tekola et al.	Focus-group discussion (n=4), and in-depth interview (n=25)	Community members, fieldworkers, researchers and staff of the Mossy Foot Treatment and Prevention Association (MFTPA) working on prevention and treatment of podoconiosis.	Gambia
23	Voluntary participation and informed consent to international genetic research[16]	2006	Marshall et al.	Survey questionnaire	Participants in hypertension genetic epidemiology research	Igbo-Ora, Nigeria (n=307), United States (n=348)

Table 4.2 Frequency of Identified Issue Topics.

Issue topics	No. of Articles containing
Comprehension	13
Recall	2
Language	1
Concept of gene and heritability	2
Factors influencing understanding	6
Voluntary participation	12
Reason for (non-)participation	4
Perception of risk and benefit	2
Misconception	4
Consent element	11
Study purpose	1
Study objective	1
Repository and biobanking	6
Future use	2
Returning results	3
Data sharing	2
Consent model	2
Community engagement	9
Engagement structure	1
Cultural Beliefs	3
Social stigma	1
Shared decision making	5

4.7 REFERENCES

1. Ginsburg, G.S. (2019). A Global Collaborative to Advance Genomic Medicine. Am J Hum Genet 104, 407-409.

2. The H3Africa Consortium, Rotimi, C., Abayomi, A., Abimiku, A., Adabayeri, V.M., Adebamowo, C., Adebiyi, E., Ademola, A.D., Adeyemo, A., Adu, D., et al. (2014). Enabling the Genomic Revolution in Africa: H3Africa is Developing Capacity for Health-related Genomics Research in Africa. Science 344, 1346-1348.

3. Adebamowo, S.N., Francis, V., Tambo, E., Diallo, S.H., Landoure, G., Nembaware, V., Dareng, E., Muhamed, B., Odutola, M., Akeredolu, T., et al. (2018). Implementation of genomics research in Africa: challenges and recommendations. Glob Health Action 11, 1419033.

4. Wright, G.E., Koornhof, P.G., Adeyemo, A.A., and Tiffin, N. (2013). Ethical and legal implications of whole genome and whole exome sequencing in African populations. BMC Med Ethics 14, 21.

5. Tiffin, N. (2014). Unique considerations for advancing genomic medicine in African populations. Per Med 11, 187-196.

6. 45 C.F.R. § 46. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html. Accessed: 14 December, 2022

7. Bazzano, L.A., Durant, J., and Brantley, P.R. (2021). A Modern History of Informed Consent and the Role of Key Information. Ochsner J 21, 81-85.

8. U.S. Department of Health & Human Services Office for Human Research Protection. (2016). Informed Consent Checklist (1998).

9. McGuire, A.L., and Beskow, L.M. (2010). Informed consent in genomics and genetic research. Annu Rev Genomics Hum Genet 11, 361-381.

10. Scheub, H. (1985). A review of African oral traditions and literature. African Studies Review 28, 1-72.

11. Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Rev Esp Cardiol (Engl Ed) 74, 790-799.

12. H3Africa Working Group on Ethics and Regulatory Issues. (2017). H3Africa Guideline for Informed Consent. URL: https://h3africa.org/wp-content/uploads/2018/05/H3A%202017%20Revised%20IC%20guideline%20for%20SC%2020_10_2017.pdf. Accessed: 14 December, 2022

13. Luthuli, M., Ngwenya, N., Gumede, D., Gunda, R., Gareta, D., Koole, O., Siedner, M.J., Wong, E.B., and Seeley, J. (2022). Participant recall and understandings of information on biobanking and future genomic research: experiences from a multi-disease community-based health screening and biobank platform in rural South Africa. BMC Med Ethics 23, 43.

14. Tindana, P., Depuur, C., de Vries, J., Seeley, J., and Parker, M. (2020). Informed consent in genomic research and biobanking: taking feedback of findings seriously. Glob Bioeth 31, 200-215.

15. Marshall, P.A., Adebamowo, C.A., Adeyemo, A.A., Ogundiran, T.O., Strenski, T., Zhou, J., and Rotimi, C.N. (2014). Voluntary participation and comprehension of informed consent in a genetic epidemiological study of breast cancer in Nigeria. BMC Med Ethics 15, 38.

16. Marshall, P.A., Adebamowo, C.A., Adeyemo, A.A., Ogundiran, T.O., Vekich, M., Strenski, T., Zhou, J., Prewitt, T.E., Cooper, R.S., and Rotimi, C.N. (2006). Voluntary participation and informed consent to international genetic research. Am J Public Health 96, 1989-1995.

17. Taiwo, R.O., Ipadeola, J., Yusuf, T., Fagbohunlu, F., Jenfa, G., Adebamowo, S.N., Adebamowo, C.A., and Indigene Study as part of the, H.A.C. (2020). Qualitative study of comprehension of heritability in genomics studies among the Yoruba in Nigeria. BMC Med Ethics 21, 124.

18. Bukini, D., Mbekenga, C., Nkya, S., Purvis, L., McCurdy, S., Parker, M., and Makani, J. (2020). A qualitative study on aspects of consent for genomic research in communities with low literacy. BMC Med Ethics 21, 48.

19. Campbell, M.M., Susser, E., Mall, S., Mqulwana, S.G., Mndini, M.M., Ntola, O.A., Nagdee, M., Zingela, Z., Van Wyk, S., and Stein, D.J. (2017). Using iterative learning to improve

understanding during the informed consent process in a South African psychiatric genomics study. PLoS One 12, e0188466.

20. Campbell, M.M., Susser, E., de Vries, J., Baldinger, A., Sibeko, G., Mndini, M.M., Mqulwana, S.G., Ntola, O.A., Ramesar, R.S., and Stein, D.J. (2015). Exploring researchers' experiences of working with a researcher-driven, population-specific community advisory board in a South African schizophrenia genomics study. BMC Med Ethics 16, 45.

21. Traore, K., Bull, S., Niare, A., Konate, S., Thera, M.A., Kwiatkowski, D., Parker, M., and Doumbo, O.K. (2015). Understandings of genomic research in developing countries: a qualitative study of the views of MalariaGEN participants in Mali. BMC Med Ethics 16, 42.

22. Staunton, C., Tindana, P., Hendricks, M., and Moodley, K. (2018). Rules of engagement: perspectives on stakeholder engagement for genomic biobanking research in South Africa. BMC Med Ethics 19, 13.

23. Tekola, F., Bull, S.J., Farsides, B., Newport, M.J., Adeyemo, A., Rotimi, C.N., and Davey, G. (2009). Tailoring consent to context: designing an appropriate consent process for a biomedical study in a low income setting. PLoS Negl Trop Dis 3, e482.

24. Campbell, M.M., de Vries, J., Mqulwana, S.G., Mndini, M.M., Ntola, O.A., Jonker, D., Malan, M., Pretorius, A., Zingela, Z., Van Wyk, S., et al. (2018). Predictors of consent to cell line creation and immortalisation in a South African schizophrenia genomics study. BMC Med Ethics 19, 72.

25. Tindana, P., Bull, S., Amenga-Etego, L., de Vries, J., Aborigo, R., Koram, K., Kwiatkowski, D., and Parker, M. (2012). Seeking consent to genetic and genomic research in a rural Ghanaian setting: a qualitative study of the MalariaGEN experience. BMC Med Ethics 13, 15.

26. Ngwenya, N., Luthuli, M., Gunda, R., Gumede, N.A., Adeagbo, O., Nkosi, B., Gareta, D., Koole, O., Siedner, M., Wong, E.B., et al. (2020). Participant understanding of informed consent in a multidisease community-based health screening and biobank platform in rural South Africa. Int Health 12, 560-566.

27. Ogunrin, O., Woolfall, K., Gabbay, M., and Frith, L. (2018). Relative solidarity: Conceptualising communal participation in genomic research among potential research participants in a developing Sub-Saharan African setting. PLoS One 13, e0195171.

28. Tekola, F., Bull, S., Farsides, B., Newport, M.J., Adeyemo, A., Rotimi, C.N., and Davey, G. (2009). Impact of social stigma on the process of obtaining informed consent for genetic research on podoconiosis: a qualitative study. BMC Med Ethics 10, 13.

29. Appiah-Poku, J., Newton, S., and Kass, N. (2011). Participants' perceptions of research benefits in an African genetic epidemiology study. Dev World Bioeth 11, 128-135.

30. Marsh, V.M., Kamuya, D.M., Mlamba, A.M., Williams, T.N., and Molyneux, S.S. (2010). Experiences with community engagement and informed consent in a genetic cohort study of severe childhood diseases in Kenya. BMC Med Ethics 11, 13.

31. Masiye, F., Mayosi, B., and de Vries, J. (2017). "I passed the test!" Evidence of diagnostic misconception in the recruitment of population controls for an H3Africa genomic study in Cape Town, South Africa. BMC Med Ethics 18, 12.

32. Moodley, K., and Singh, S. (2016). "It's all about trust": reflections of researchers on the complexity and controversy surrounding biobanking in South Africa. BMC Med Ethics 17, 57.

33. de Vries, J., Williams, T.N., Bojang, K., Kwiatkowski, D.P., Fitzpatrick, R., and Parker, M. (2014). Knowing who to trust: exploring the role of 'ethical metadata' in mediating risk of harm in collaborative genomics research in Africa. BMC Med Ethics 15, 62.

34. Ochieng, J., Kwagala, B., Barugahare, J., Mwaka, E., Ekusai-Sebatta, D., Ali, J., and Sewankambo, N.K. (2021). Perspectives and ethical considerations for return of genetics and genomics research results: a qualitative study of genomics researchers in Uganda. BMC Med Ethics 22, 154.

35. Bukini, D., deVries, J., Treadwell, M., Anie, K., Dennis-Antwi, J., Kamga, K.K., McCurdy, S., Ohene-Frempong, K., Makani, J., and Wonkam, A. (2019). Exploring the Role of Shared Decision Making in the Consent Process for Pediatric Genomics Research in Cameroon, Tanzania, and Ghana. AJOB Empir Bioeth 10, 182-189.

36. H3Africa Working Group. (2011). Harnessing Genomic Technologies Toward Improving Health in Africa: Opportunities and Challenges. Recommendations for the H3Africa Initiative to the Wellcome Trust and the National Institutes of Health.

37. Doerr, M., Moore, S., Barone, V., Sutherland, S., Bot, B.M., Suver, C., and Wilbanks, J. (2021). Assessment of the All of Us research program's informed consent process. AJOB Empir Bioeth 12, 72-83.

38. Cohen, M.S., Chen, Y.Q., McCauley, M., Gamble, T., Hosseinipour, M.C., Kumarasamy, N., Hakim, J.G., Kumwenda, J., Grinsztejn, B., Pilotto, J.H., et al. (2016). Antiretroviral Therapy for the Prevention of HIV-1 Transmission. N Engl J Med 375, 830-839.

39. Tsosie, K.S., Yracheta, J.M., and Dickenson, D. (2019). Overvaluing individual consent ignores risks to tribal participants. Nat Rev Genet 20, 497-498.

VITA

Yunqi earned her Bachelor of Sciences degree in Nutritional Sciences from Cornell University in 2016. She then received her Master of Science degree in Human Genetics at the University of Pittsburgh Graduate School of Public Health under the supervision of Profs. Daniel E. Weeks and John R. Shaffer in 2018. In the same year, she moved to Seattle and started to pursue her Doctor of Philosophy degree at the Institute for Public Health Genetics, and the Certificate in Statistical Genetics at the University of Washington, expected to graduate in 2022. During the program, Yunqi served as teaching assistant in graduate-level course, Statistical Genetics: Mendelian Traits, and undergraduate-level course, Biostatistics for the Health Sciences. In the last two summer quarters, she also participated in industrial internships at Merck Early Discovery Genetics group and Genentech Departments of Cancer Immunology and Human Genetics. Yunqi's research interests intersect with genetics epidemiology, statistical genetics, and the implications of genetics to society.