Single Nucleotide Polymorphisms Associated with Decreased HIV Set-Point Viral Load in African Populations Modify Expression of Interferon Signaling Pathways

Riley H. Tough

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba

in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Medical Microbiology and Infectious Diseases

University of Manitoba

Winnipeg, MB

Copyright © 2025 by Riley H. Tough

Acknowledgements

First and foremost, thank you to my supervisor Dr. Paul McLaren for his constant encouragement, mentorship, and support through my graduate studies. Thank you for the incredible opportunity to be part of your lab. I wish you all the best in your new role as the Head of the Sexually Transmitted and Blood-Borne Infections Surveillance and Molecular Epidemiology Group at the JC Wilt Infectious Diseases Research Center.

Thank you as well to my co-supervisor Dr. Blake Ball for thoughtful comments, invaluable feedback, discussions on tabletop RPGs, and insight into future career opportunities. I wish you all the best in your new role as Chief of Sexually Transmitted and Blood-Borne Infections Surveillance, Research, and Innovation at the JC Wilt Infectious Diseases Research Center.

Thank you to my committee members, Dr. Xiao-Jian Yao and Dr. Geoffrey Hicks for your thoughtful feedback, valuable expertise, and insight into virology and functional genetics. Your expertise was a significant contribution to the overall success of this project.

Thank you to the department of Medical Microbiology and Infectious Diseases at the University of Manitoba for the many opportunities at personal and professional development. Specifically, thank you Ms. Angela Nelson for your invaluable role in guidance, support, and student advocacy through my graduate studies.

Thank you to members of the STBBI Surveillance and Molecular Epidemiology Group, past and present, for your support throughout my degree. In particular to Paul Jankowski, Rubendren Jamilchelvan, and Abraham Awada for our late-night problem-solving discussions, and to Rupert Capina and Bernard Abrenica of the JC Wilt Operations Center for your guidance and mentorship that helped me navigate my studentship.

Last but not least, thank you to my family throughout my graduate studies. To my incredible wife Hailey Tough, I cannot express how much I appreciate your patience, support, and kindness. To my parents, Ronald and Rhonda Tough, for cultivating my love of science, for your unwavering support, and for listening to me practice presentations countless times.

Abstract

Human immunodeficiency virus type 1 (HIV) is responsible for over 40 million deaths since its identification in the 1980s and continues to pose a significant threat to global well-being. Despite advances in prevention and treatment, the lack of a functional cure or vaccine places a heavy reliance on treatment as prevention to reduce global transmission. Our lab conducted a genome-wide association study (GWAS) in 3,879 individuals of African ancestry as part of the International Collaboration for Genomics of HIV (ICGH), identifying a novel region of chromosome 1 associated with lower of HIV set-point viral load (spVL). This region is centered on three coding genes: *PRKAB2, FMO5,* and *CHD1L*, none of which were previously associated with HIV infection. Initial bioinformatics and *in vitro* validation identified *CHD1L* as a novel HIV inhibitory factor in myeloid cells. Functional fine-mapping of the region identified a combination of three variants: rs72999655-A-G, rs7525622-G-A, and rs73004025-C-T, that collectively explained more HIV spVL variance than the lead SNP, rs59784663-A-G, suggesting a causal role. Using elastic net models trained on large, admixed cohort transcriptomes, candidate causal variants were associated with increased *CHD1L* and decreased *PRKAB2* expression in monocytes, and decreased *CHD1L* expression in whole blood. Gene expression analysis from people living with HIV (PLWH) revealed that individuals heterozygous for protective alleles exhibited reduced *CHD1L* expression and no significant change to *PRKAB2* expression in monocytes, highlighting cell-type-specific regulatory effects. No genotype-associated differences were observed in either unstimulated or CD3/CD28-stimulated CD4+ T cells. Transcriptome analysis of 11 healthy individuals showed that protective alleles were associated with differential expression of 25 genes, including significant enrichment for interferon stimulating genes in monocytes. This result was not replicated in naïve CD4+ T cells, further emphasizing myeloid-specific regulation. Together, these findings characterize a novel genomic region regulating HIV control and support a putative mechanism in which GWAS variants influence *CHD1L* and/or *PRKAB2* to alter immune activation and reduce HIV spVL. These insights advance our understanding of HIV host genetics in African populations and highlight *CHD1L* as a potential therapeutic target in chronic infection.

Table of Contents

Acknowledgements 2

Abstract 3

Table of Contents 4

List of Figures 8

List of Tables 10

List of Abbreviations 11

Author Contributions 13

Chapter 1: Background 16

The HIV epidemic 16

Human immunodeficiency virus 16

Disparities in HIV incidence 17

HIV transmission 17

Advances in HIV prevention 18

Antiretroviral therapy 19

Targeting host genes as therapeutic options 21

Host immunity to HIV 21

Innate immune response 22

Adaptive immune response 22

Dynamics of HIV infection 23

Factors influencing HIV spVL 25

Human genomics 27

Functional consequences of coding variants 27

Functional consequences of non-coding variants 28

Host genomic studies of human disease 28
Genome wide association studies 29
Interpretation of GWAS associations 30
Statistical models for identifying causal variants 30
Impact of causal variants on disease 31
Host genomics of HIV infection 34
Putative HIV host factors 37
Structure and function of PRKAB2 40
Structure and function of FMO5 40
Structure and function of CHD1L 40
Structure and function of LINC00624 41
Study rationale, hypothesis, and objectives 41
Study Rationale 41
Hypothesis 42
Chapter 2: Functionally-informed Fine-mapping Identifies Genetic Variants Linking Increased CHD1L Expression and HIV Restriction in Monocytes 44
Introduction 44
Methods 45
Results 47
Discussion 64
Chapter 3: Chromosome 1 Variants Associated with Decreased HIV Set-point Viral Load Correlate with PRKAB2 Expression Changes 66
Introduction 66
Methods 67

Results 71
Discussion 82
Chapter 4: Changes to Interferon Stimulating Genes Associates with Decreased HIV spVL in African Populations 85
Introduction 85
Methods 86
Results 94
Discussion 110
Chapter 5: Discussion 113
CHD1L is a novel HIV inhibitory factor *in vitro* 113
Variants linked to lower HIV spVL are associated with changes to *CHD1L* expression, *PRKAB2* expression, and interferon signaling 114
The interferon response is associated with reduced HIV replication *in vitro* 117
CHD1L influences the ERK1/2, PI3K-Akt, and NF-κβ signaling pathways 119
Cell-type-specificity of *CHD1L*-mediated HIV inhibition 122
Relationship between HIV spVL associated variants, HIV replication, and HIV spVL 123
The potential role of anti-*CHD1L* drugs in HIV pathogenesis 124
Study limitations 125
Study Conclusions 125
Future Directions 126
References 129
Appendices 171
Appendix 1 171
Appendix 2 174

Appendix 3 175
Appendix 4 179
Appendix 5 180
Appendix 6 181
Appendix 7 183
Appendix 8 184
Appendix 9 184

List of Figures

Figure 1. Relationship between HIV set-point viral load and disease progression.....26

Figure 2. Proposed mechanisms of how genetic variants influence HIV infection through regulatory and coding regions.....33

Figure 3. Chromosome 1 genes exhibit a high-degree of co-regulation suggesting shared regulatory mechanisms.....39

Figure 4. Variant effect predictor consequences of genetic variants in the chromosome 1 region (GRCh38; 1:147023443-148358705).....49

Figure 5. Linkage disequilibrium patterns of functional variants in the chromosome 1 region.....55

Figure 6. Functional annotations of likely causal variants.....56

Figure 7. Negative correlation between *CHD1L* expression and HIV set-point viral load.....58

Figure 8. The candidate causal variant rs72999655-A-G is associated with altered methylation of *CHD1L* in African populations.....60

Figure 9. *CHD1L* gene expression from monocytes is negatively correlated with HIV set-point viral load.....62

Figure 10. Candidate causal variants are associated with increased *CHD1L* expression in monocytes.....63

Figure 11. CHD1L and PRKAB2 expression exhibit cell-types-specific correlation patterns across immune cell subsets.....72

Figure 12. The rs59784663-G allele is associated with reduced *PRKAB2* expression in monocytes.....74

Figure 13. Alleles linked to lower HIV spVL is associated with reduced *PRKAB2* expression in whole blood.....75

Figure 14. Gene expression differences in PRKAB2$^{-/-}$ loss-of-function iPSCs.....77

Figure 15. Differentially expressed genes between PRKAB2-/- clone two and wild-type iPSCs.....81

Figure 16. HIV spVL associated variant genotypes are associated with reduced *CHD1L* expression in monocytes but no expression differences in CD4+ T cells….....96

Figure 17. Imputed CHD1L expression is associated with genotype at HIV spVL associated genetic variants.....98

Figure 18. Principal components of total gene expression variance in naïve CD4+ T cells.....100

Figure 19. Principal components of total gene expression variance in monocytes show that genotype is the major contributor to gene expression differences.....101

Figure 20. Differentially expressed genes in naïve CD4+ T cells based on genotype at HIV spVL associated variants.....102

Figure 21. Differentially expressed genes in monocytes based on genotype at HIV spVL associated variants.....102

Figure 22. Differentially expressed genes associated with genotypes linked to lower HIV spVL are associated with the immune response, cytokine signaling, and interferon signaling responses..... 105

Figure 23. Chromosome 1 genes do not directly cluster with differentially expressed genes based on HIV spVL-associated genotype.....107

Figure 24. Differentially expressed genes associated with alleles linked to lower HIV spVL exhibit significant overlap with gene expression profiles from HIV LTNPs……..109

Figure 25. Summary of the association between HIV spVL associated variants, gene expression, and HIV infection.....116

Figure 26. Association between CHD1L and the interferon signaling pathway.....122

List of Tables

Table 1. Genome-wide association studies of human immunodeficiency virus type 1....35

Table 2. Functional consequences of coding CHD1L variants.....50

Table 3. Functional consequences of coding PRKAB2 variants.....51

Table 4. Genome-wide association study variants in the chromosome 1 region that explain the majority of HIV spVL variance.....53

Table 5. Summary of studies used in this meta-analysis.....68

Table 6. DAVID enrichment of differentially expressed genes in Gene Ontology and KEGG categories.....80

Table 7. Reagents for qPCR genotyping experiments.....88

Table 8. Run settings for qPCR genotyping experiments on Viia7 system.....88

Table 9. Viability Stain Master Mix Reagents.....90

Table 10. Antibody Master Mix Reagents.....90

Table 11. Differentially expressed genes in monocytes associated with presence of rs72999655-A-G, rs73004025-C-T, and rs7525622-G-A alleles.....103

Table 12. Differentially expressed genes based on HIV spVL associated variant genotype that are also interferon stimulating genes.....119

List of Abbreviations

#

3TC: Lamivudine

A

ABC: Abacavir
ACB: African, Caribbean, Black
AIDS: Acquired Immunodeficiency Syndrome
APC: Antigen presenting cell
ART: Antiretroviral therapy

B

BF: Bayes factors
BP: base pairs

C

C1: PRKAB2-/- iPSC Clone 1
C2: PRKAB2-/- iPSC Clone 2
cDNA: complementary DNA
cGAS: Cyclic GMP-AMP synthase
CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

D

DAA: Direct-acting antiviral
DAVID: Database for Annotation, Visualization and Integrated Discovery
dbGaP: Database of Genotypes and Phenotypes
DC: Dendritic Cells
DNA: Deoxyribonucleic Acid
DTG: Dolutegravir

E

ENCODE: Encyclopedia of DNA Elements
eQTL: expression quantitative trait locus

F

FACS: Fluorescent associated cell sorting
FDR: False Discovery Rate
FTC: Emtricitabine

G

GALA II: Genes-environments and Admixture in Latino Americans
GENOA: Genetic Epidemiology Network of Arteriopathy
GEO: Gene Expression Omnibus
GO: Gene Ontology
GTEx: Genotype-Tissue Expression
GWAS: Genome Wide Association Study

H

HDA: Host directed antiviral
HDF: host dependency factor
HIV: Human Immunodeficiency Virus Type 1
HLA: Human Leukocyte Antigen

I

ICGH: International Collaboration for the Genomics of HIV
IFN: Interferon
ImmVar: Immune Variation
Indel: Insertions/deletions
INSTI: Integrase stand transfer inhibitors
iPSC: Immunopluripotent stem cell
iPSDM: Immunopluripotent stem cell-derived macrophages
ISG: interferon-stimulated gene

K

KEGG: Kyoto Encyclopedia of Genes and Genomes

L

LCL: lymphoblastoid cell line
LD: Linkage Disequilibrium
lncRNAs: Long noncoding RNAs

M

M1: Classically activated macrophages
M2: Alternatively activated macrophages
MAF: Minor allele frequency
MAGMA: Multi-marker Analysis of GenoMic Annotation
MDM: Monocyte derived macrophage
meQTL: Methylation quantitative trait loci
MESA: Multi-Ethnic Study of Atherosclerosis
MOCS: Majengo Observational Cohort Study
MSM: Men who have sex with men

N

NCBI: National Center for Biotechnology Information
NNRTI: Non-nucleoside reverse transcriptase inhibitors
NRTI: Nucleoside/nucleotide reverse transcriptase inhibitors
NTC: Non-template control

P

PAINTOR: Probabilistic Annotation INtegraTOR
PC: Principal component
PEP: Post-exposure prophylaxis
PI: Protease inhibitors
PPI: Protein-protein interaction
PrEP: Pre-exposure prophylaxis

Q

QTL: Quantitative trait loci

R

R10: RPMI-1640 supplemented with 10% FBS
RLE: Relative log expression

S

SE: standard error
SAGE: Study of African Americans, Asthma, Genes, and Environments
SNP: single-nucleotide polymorphism
spVL: Set-point viral load
STBBI: Sexually Transmitted and Blood-Borne Infection
STRING: Search Tool for the Retrieval of Interacting Genes

T

TAF: Tenofovir lafenamide
TasP: Treatment as prevention
TDF: Tenofovir disoproxil fumarate
TFBS: Transcription factor binding sites
Th1: T helper 1
Th2: T helper 2
TPM: Transcripts per million
TSS: Transcription start site

U

U=U: Undetectable equals Untransmissible

V

VEP: Variant Effect Predictor

W

WGS: Whole-genome sequencing

Author Contributions

This thesis is written using a Chapter-based format with data and results from published manuscripts. Sections detailed in this thesis have either been published or will be submitted for publication. The contributions of authors towards each section are detailed below.

Publications from this thesis

This thesis includes a post-hoc analysis of a GWAS titled “Africa-specific human genetic variation near *CHD1L* associates with HIV load” which was published in Nature in 2023[1]. The authors on this manuscript are Paul J. McLaren, Immacolata Porreca, Gennaro Iaconis, Hoi Ping Mok, Subhankar Mukhopadhyay, Emre Karakoc, Sara Cristinelli, Cristina Pomilla, István Bartha, Christian W. Thorball, **Riley H. Tough**, Paolo Angelino, Cher S. Kiar, Tommy Carstensen, Segun Fatumo, Tarryn Porter, Isobel Jarvis, William C. Skarnes, Andrew Bassett, Marianne K. DeGorter, Mohana Prasad Sathya Moorthy, Jeffrey F. Tuff, Eun-Young Kim, Miriam Walter, Lacy M. Simons, Arman Bashirova, Susan Buchbinder, Mary Carrington, Andrea Cossarizza, Andrea De Luca, James J. Goedert, David B. Goldstein, David W. Haas, Joshua T. Herbeck, Eric O. Johnson, Pontiano Kaleebu, William Kilembe, Gregory D. Kirk, Neeltje A. Kootstra, Alex H. Kral, Olivier Lambotte, Ma Luo, Simon Mallal, Javier Martinez-Picado, Laurence Meyer, José M. Miro, Pravi Moodley, Ayesha A. Motala, James I. Mullins, Kireem Nam, Niels Obel, Fraser Pirie, Francis A. Plummer, Guido Poli, Matthew A. Price, Andri Rauch, Ioannis Theodorou, Alexandra Trkola, Bruce D. Walker, Cheryl A. Winkler, Jean-François Zagury, Stephen B. Montgomery, Angela Ciuffi, Judd F. Hultquist, Steven M. Wolinsky, Gordon Dougan, Andrew M. L. Lever, Deepti Gurdasani, Harriet Groom, Manjinder S. Sandhu, and Jacques Fellay. This manuscript is not included as a thesis chapter due to intellectual property and copyright restrictions but is referenced here because it serves as the foundation of this work and I made significant contributions to the manuscript.

Specifically, I performed the gene-level analysis using the Multi-marker Analysis of GenoMic Annotation (MAGMA) framework, annotated functional consequences of chromosome 1 variants using the Encyclopedia of DNA Elements (ENCODE), identified the 95% credible set using Probabilistic Annotation INTegratOR (PAINTOR), prepared

data for publication, contributed to the statistical analysis, performed lift over of genetic variants from GRCh37 to GRCh38, annotated variants with rsIDs, generated figures for publication, and contributed to the writing and editing of the manuscript. The manuscript is available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10848312/ or https://pubmed.ncbi.nlm.nih.gov/37532928/.

Chapter Two

Chapter two was published in *Scientific Reports* in January 2025, titled "Functionally-informed fine-mapping identifies genetic variants linking increased CHD1L expression and HIV restriction in monocytes". The authors on this manuscript are **Riley H. Tough**, Paul J. McLaren, and the International Collaboration for Genomics of HIV (ICGH) consortium. In this study, Dr. Paul McLaren was involved in the conception and design, methodology, supervision, interpretation of results, funding acquisition, and revision of the article. I was responsible for aspects of the design, methodology, analysis, interpretation and visualization of results, drafting, and revision of the article. Specifically, I was performed the analysis of genetic variant consequences, functional and statistical fine-mapping, haplotype analysis, correlation of eQTLs, analysis of meQTLs, imputation of gene expression, and gene expression association analysis.

Chapter Three

Chapter three was published in *Frontiers in Genetics* in January 2025, titled "Chromosome 1 variants associated with decreased HIV set-point viral load correlate with PRKAB2 expression changes". The authors on this manuscript are **Riley H. Tough** and Paul J. McLaren. In this study, Dr. Paul McLaren participated in the conception and design, methodology, supervision, interpretation of results, funding acquisition, and revision of the article. I was responsible for the design, acquisition of data, methodology, analysis, generation and interpretation of results, visualization of results, drafting, and revision of the article.

Unpublished work at the time of thesis submission

Chapter Four

Chapter four is being prepared as a separate manuscript for publication. The authors in this work are **Riley H. Tough**, Catherine M. Card, Abraham Awada, Jeffrey F. Tuff, and

Paul J. McLaren. In this study, Dr. Paul McLaren participated in the conception and design, methodology, supervision, interpretation of results, funding acquisition, and revision of the article. Dr. Catherine Card assisted in the methodology and acquisition of data for this article. Specifically, Dr. Card developed the fluorescent associated cell sorting (FACS) panel and performed FACS on the Aria II at the JC Wilt Infectious Diseases Research Center. Mr. Tuff participated in acquisition of data for this article and assisted with FACS on the Aria II at the JC Wilt Infectious Diseases Research Center. Abraham Awada assisted in the extraction of cell-free DNA from plasma and genotyping of rs59784663-A-G. I was responsible for aspects of the design, acquisition of data, methodology, analysis, interpretation of results, visualization of results, drafting, and revision of the article. Specifically, I performed the quality control, mapping, analysis, visualization, and interpretation of microarray gene expression data. I also performed the *ex vivo* experiments including culturing primary cells, preparing cells for FACS, conducting cell stimulations, nucleic acid extractions, genotyping, cDNA generation, and qPCR analyses.

Chapter 1: Background

The HIV epidemic

Human immunodeficiency virus type 1, the causative agent of acquired immunodeficiency syndrome (AIDS), has remained a significant threat to global well-being since it was first identified in the 1980s[2,3]. When it was first identified, the average life expectancy of individuals diagnosed with AIDS was only three years[4,5]. The development of antiretroviral therapy (ART) for treatment of HIV disease has significantly reduced HIV disease progression and improved quality of life for PLWH[6]. However, as of 2024, there were 39.9 million PLWH, an estimated 1.3 million new infections, and 630,000 AIDS-related deaths[7].

To curb the pandemic, the United Nations set the 95-95-95 targets, aiming for 95% of PLWH to know their status, 95% who know their status to be accessing ART, and 95% of those on ART to have sustained viral suppression by 2025. As of 2023 global estimates, only 86% of PLWH knew their status, 77% of PLWH were accessing ART, and 93% of PLWH accessing ART were virally suppressed[8]. However, independent studies suggest that achieving these targets requires addressing disparities in treatment access and virological suppression, particularly in low and middle-income countries where suppression rates in adults range from 84-89% but only 64-74% in children and adolescents[9–11]. In addition to addressing social and clinical disparities, there is also a strong need to better understand the HIV host-pathogen interaction to guide the development of novel therapeutics and vaccine candidates.

Human immunodeficiency virus

HIV is a retrovirus belonging to the Lentivirus genus within the Retroviridae family and subdivided into M, N, O, and P groups[12,13]. The global HIV pandemic is primarily driven by group M viruses, which are subdivided into three major subtypes (A, B, and C) which account for approximately 70% of global infections[14]. There are also six additional subtypes: D, F, G, H, J, and K[15]. Recombination events between HIV subtypes have resulted in circulating and unique recombinant forms, which now make up 29% of circulating viruses[16].

The distribution of HIV subtypes is heterogeneous around the world[17]. Globally, subtype C is most prevalent, driving infections in sub-Saharan Africa, East Africa, and India. Subtype B is predominant in North America, South America, Europe, China, Japan, and Australia, while subtype A is predominant in Central Africa, Eastern Europe and Russia, and circulating recombinant forms are most prevalent in western African and Indonesia.

Disparities in HIV incidence

Globally, HIV incidence demonstrates significant disparities. Sub-Saharan Africa, home to 12% of the global population, bears over 70% of the global HIV burden[18]. Since 2002, over $94 billion has been invested into public health advances including HIV testing, treatment, and prevention across African countries[19]. This investment has yielded a 46% reduction of new infections in Western and Central Africa since 2010[7]. Despite these improvements, key populations including men who have sex with men (MSM), people who inject drugs (PWID), sex workers, transgender persons, and incarcerated persons, remain disproportionately impacted and account for more than 60% of global HIV infections amongst adults[20].

In Canada, there were 2,434 new HIV diagnoses in 2023, representing a 35.2% increase compared to 2022[21]. These new diagnoses were distributed amongst MSM (36.0%), PWID (18.2%), MSM + PWID (4.5%), heterosexual contact (39.4%), and other key populations (2.0%)[22,23]. African, Caribbean, Black (ACB), and Indigenous populations only account for 6.9% of the population collectively but represent 30% of the PLWH in Canada[24,25]. In Manitoba, HIV incidence is 19.3 per 100,000 compared to the national average of 6.1 per 100,000[23], and key populations are less likely to be in the HIV care cascade, less likely to be retained in care, less likely to be on treatment, and/or less likely to be virologically suppressed than white persons[26]. These alarming rates highlight the need for additional resources, policies, and research into the systemic inequities that affect key populations to reduce new HIV infections.

HIV transmission

HIV transmission occurs through exchange of infected bodily fluids, including blood, breast milk[27], semen, and vaginal secretions, which is spread via sexual contact, perinatal

(mother-to-child) transmission, direct inoculation, or contaminated blood products where the virus bypasses epithelial barriers. Sexual transmission remains the most common route of infection, accounting for the majority of new infections globally[7,28]. Sexual transmission can be further subcategorized based on the site of exposure, with estimated risk of HIV infection per sex act of 0.05-0.5% for vaginal tissues, 0.3-5% for lower gastrointestinal tissues, or 0.4% for upper gastrointestinal tissues[29]. Transmission risk is influenced by many factors, including viral load of the infected partner, male circumcision, co-infection of sexually transmitted infections, and presence of open wounds or mucosal disruptions[30–33].

Perinatal transmission occurs when the virus is transmitted from mother to child during breastfeeding, childbirth, or *in utero*[34]. The implementation of ART use during pregnancy, caesarean section delivery, and post-exposure prophylaxis (PEP) can significantly reduce the risk of transmission to less than 1%[35–38]. Mass adoption of these interventions has resulted in a significant decline in global perinatal transmission rates, decreasing from 42% in the 1990s to 12% in 2021, and elimination of perinatal transmission in 15 countries[39].

People who inject drugs are also at high risk for HIV infection, as sharing contaminated needles directly introduces the virus to the bloodstream and is associated with transmission rates of up to 95%[29]. Transmission risk amongst this demographic is also multi-factorial, as injection drug use is strongly correlated with inconsistent condom use, low pre-exposure prophylaxis (PrEP) use, and has been further exacerbated amongst the growing opioid crisis in North America[40,41]. The implementation of harm reduction strategies in some regions, such as needle and syringe exchange programs, has been shown to reduce HIV transmission and transmission-associated risk factors[42–45]. While there has been progress in slowing HIV transmission, further research and policies are necessary to curb the spread of HIV.

Advances in HIV prevention

Despite major scientific advances in understanding of HIV biology, there is still no vaccine that can reduce HIV infection. One of the most promising vaccine candidates to date was the RV144 HIV vaccine in Thailand, which showed 60% protection from HIV infection in

a phase III clinical trial within the first 12 months but only 31% protection after 38 months[46]. Due to the waning immunity, this vaccine was not considered successful. However, subsequent vaccine development strategies have sought to improve upon the RV144 design to improve efficacy and long-term immunity.

A modified RV144 regiment targeting HIV clade C, the HIV Vaccine Trials Network 702 trial, was investigated in South African populations and elicited a strong immune response[47]. However, the vaccine failed to show clinical benefit in the phase III clinical trial[48]. The investigators noted several factors that may have confounded vaccine efficacy including presumed genital tract inflammation, diversity of circulating HIV clades, and host genetic factors[49]. Similarly, the HIV Vaccine Trials Network 097 trial in South Africa, elicited a strong immune response in a randomized controlled double-blind study but failed to show clinical benefit in phase III clinical trials[48,49]. Although additional trials are ongoing, in the absence of a vaccine, reduction of transmission risk has relied on preventative strategies such as condom use, PrEP, and PEP[50,51].

One of the most effective prevention strategies is the adoption of treatment as prevention (TasP), a concept underscored by the Undetectable equals Untransmissible (U=U) campaign. Large cohort studies observed that there was no risk of sexual transmission in PLWH on ART with undetectable viral loads[52–55]. The U=U campaign, coupled with starting ART at the time of diagnosis, has the potential to dramatically improve adherence, reduce stigma, and health outcomes[56]. However, studies show that awareness of U=U requires additional clinical support to achieve global acceptance in all geographic regions[57,58]. Additionally, as TasP relies heavily on ART to achieve viral suppression, there are growing concerns regarding HIV resistance to current regiments.

Antiretroviral therapy

The first drug approved for HIV treatment was azidothymidine in the 1980s but antiviral resistance rapidly emerged within the first year of use[59]. To reduce the risk of antiviral resistance, treatment eventually developed into a three-drug cocktail in 1996, previously called highly active antiretroviral therapy but now commonly referred to as ART[60–62]. The drugs within ART regiments specifically target HIV proteins, known as direct-acting antivirals (DAAs). These drugs are characterized as nucleoside/nucleotide reverse

transcriptase inhibitors (NRTI), integrase inhibitors, non-nucleoside reverse transcriptase inhibitors (NNRTI), protease inhibitors (PI), integrase stand transfer inhibitors (INSTI), and/or fusion inhibitors.

Current first-line therapies typically consist of drug combinations including abacavir (ABC), lamivudine (3TC), tenofovir lafenamide (TAF), tenofovir disoproxil fumarate (TDF), emtricitabine (FTC), and one INSTI such as dolutegravir (DTG). As of 2019, the World Health Organization recommends adults and adolescents to use TDF and DTG, with 3TC or FTC as a preferred first-line regiment[63]. For children, the recommendation is the use of ABC, 3TC, and DTG[63]. However, when deciding on an appropriate treatment, specific therapeutic regiments are highly reliant on additional factors including pregnancy status, co-infection, chronic conditions, and pharmacodynamics.

The standard recommendation for PrEP consists of daily oral TDF/FTC[64] but the development of new drugs has the potential to markedly improve current regiments. Twice annual lenacapavir was compared to daily TDF/FTC treatment in over 5000 at-risk cis-gendered women and no infections occurred amongst the lenacapavir group (N=0/2134), revealing that lenacapavir exhibited better protection than TAF/FTC and TDF/FTC regiments in this study[65]. A twice annual injection allows for HIV prevention to be more discrete, can improve adherence, and has the potential to radically improve the quality of life for people taking PREP. A generic version of lenacapavir will also be made available for 120 lower income countries, greatly reducing its cost and making it more readily available to those in need[66].

However, the use of DAAs puts strong selective pressure on HIV to develop escape mutations, which can threaten their long-term effectiveness[67,68]. Both DTG and TDF, critical components of both PrEP and PEP, exhibit increasing rates of HIV resistance globally[69–71]. In fact, all major HIV drug classes exhibit some circulating resistance[72–76]. Given these trends, development of novel therapeutics strategies that are resistant to antiviral mutations is critical to ensuring the long-term efficacy of ART, PrEP, and PEP.

Targeting host genes as therapeutic options

Since HIV is an obligate, intracellular pathogen, it relies on host proteins to establish and maintain infection within a host. Therefore, limiting access to the necessary host proteins offers an alternative therapeutic strategy to DAAs. This approach uses host-directed antivirals (HDAs) and provides a greater barrier to the development of antiviral resistance than DAAs[77]. However, identifying and validating host genes essential for HIV infection but nonessential for the host is challenging.

Hundreds of genes have been implicated as HIV host dependency factors (HDFs) essential for viral replication, auxiliary factors that promote viral replication, inhibitory factors that reduce viral replication, or restriction factors that completely ablate viral replication[78–80]. Despite this, only *CCR5*, which is required for cell-mediated endocytosis of R5-tropic HIV, is the target of an approved HDA[81]. The potential of *CCR5* antagonists as HIV therapeutics first arose through observations that some individuals with multiple exposures to HIV remained seronegative[82]. Subsequent investigations identified that CD4+ T cells from these individuals were highly resistant to infection by certain strains of HIV[83]. This research led to the development of Maraviroc, an allosteric inhibitor of CCR5, which prevents R5-tropic HIV from entering HIV target cells by CCR5-mediated endocytosis[84].

There are other promising HDAs currently being investigated for therapeutic potential by repurposing drugs for chronic diseases. Ruxolitinib, an anti-cancer drug, also shows efficacy as an anti-HIV therapeutic by inhibiting the pro-inflammatory JAK-STAT pathway[85] and lovastatin which inhibits HIV replication by downregulating immunomodulatory and lipid synthesis pathways[86,87]. Other therapeutic strategies include the use of anti-inflammatory drugs to reduce HIV infection risk and slow disease progression[88,89]. These therapeutics leverage our existing understanding of the HIV host-pathogen interaction, suggesting that identification of new HIV host factors may lead to new drug development opportunities.

Host immunity to HIV

The host immune system is a complex coordination of the innate and adaptive immune responses that work together to prevent, detect, and eliminate pathogens. The innate

immune response provides a broad and generalized defense, while the adaptive immune response mounts a targeted and pathogen-specific response. These two responses are tightly regulated, as an overactive immune system can lead to autoimmune diseases where the immune response mistakes healthy cells as pathogens. In contrast, a suppressed immune system can result in an immunocompromised state, where individuals become more susceptible to severe infections.

Innate immune response

The innate immune response is comprised of physical barriers (e.g. skin, mucous membranes, bodily secretions) and immune cells that detect and eliminate foreign pathogens. Monocytes, macrophages, and dendritic cells (DCs) detect foreign pathogens using specialized pattern recognition receptors and elicit a strong antiviral response which includes interferons (IFNs) and interferon-stimulated genes (ISGs)[90].

Monocytes are short-lived, non-dividing immune cells that circulate in the blood that differentiate into macrophages or dendritic cells upon entering tissues. In response to infection, monocytes regulate the inflammatory response, capable of expressing pro-inflammatory cytokines to activate the immune response[91] and anti-inflammatory cytokines to prevent excessive damage to host tissues[92].

Macrophages are long-lived immune cells that are essential in both the innate and adaptive immune responses. Macrophages undergo polarization to become classically activated (M1) which have enhanced phagocytotic activity, antigen presenting capacity, and cytokine production, or alternatively activated (M2) which produce an anti-inflammatory response and activate tissue repair mechanisms[93,94].

Dendritic cells, like M1 macrophages, are professional antigen presenting cells (APCs) which activate the adaptive immune response. Subsets of DCs are classified as: myeloid DCs which produce a pro-inflammatory response, plasmacytoid DCs primarily produce type 1 IFNs[95] and monocyte-derived DCs have an intermediate phenotype, acting as professional APCs and producing type 1 IFNs[96].

Adaptive immune response

The adaptive immune response consists of the cellular-mediated response, regulated by T lymphocytes, and the humoral response, regulated by B lymphocytes. These immune cells use human leukocyte antigen (HLA) on the cell surface to distinguish healthy cells from pathogens and diseased cells[97]. HLA class I molecules are expressed on the surface of all nucleated cells, while HLA class II are only present on professional APCs.

Roughly 70-85% of lymphocytes are categorized as T cells, which are further subdivided into CD4+ T cells, which recognize class II HLA, and CD8+ T cells, which recognize class I HLA. Activation of CD4+ T cells cause differentiation into T helper 1 (Th1) or T helper 2 (Th2) cells. In general, Th1 cells secrete pro-inflammatory cytokines which activate macrophages and CD8+ T cells[98]. These activated CD8+ T cells identify cancerous or virally infected cells using HLA class I molecules and induce a cytotoxic response, causing destruction of the diseased cell. However, during HIV infection, persistent antigen stimulation causes CD8+ T cells to undergo constant expansion and become chronically activated, leading to T cell exhaustion and a diminished cytotoxic response[99]. In contrast to Th1 cells, Th2 secrete anti-inflammatory cytokines essential for preventing excessive host damage and activate B cells[100].

Despite only comprising 5-10% of lymphocytes, B cells are responsible for antibody secretion in response to foreign pathogens. These antibodies can be characterized as neutralizing, which bind pathogens and prevent infection of healthy cells, and non-neutralizing, which recruit immune cells or impair the viral lifecycle[101]. The development of neutralizing antibodies that broadly protect against HIV strains is a key goal in HIV vaccine development, as most vaccines provide protection through this response[102]. Over the course of HIV infection, some PLWH will naturally produce broad neutralizing antibodies but can take many years to develop[103]. Therefore, it is important to further characterize the dynamics of HIV infection and differences in the immune response amongst individuals for the development of novel therapeutic approaches.

Dynamics of HIV infection

Following an HIV transmission event, a single or small number of viruses from the transmitting host, termed founder viruses, are responsible for establishing viral infection[104]. Studies of early viral dynamics via sexual transmission suggest that HIV

virions or infected cells cross the mucosal barrier through epithelial damage caused by sexual intercourse or inflammation where DCs, monocytes, or macrophages are targeted in peripheral tissues[105,106]. Unlike DCs, which encapsulate HIV upon binding to Siglec-1 surface receptors but do not support productive infection[107], monocytes and macrophages are both susceptible to productive infection and resistant to the cytopathic effects of HIV[108,109]. During early infection, HIV exploits the motility of these cells, where infected macrophages or HIV-bound DCs travel to lymphoid tissues where it infects activated CD4+ T cells[110,111].

Once in lymphoid tissues, HIV begins rapidly replicating in activated CD4+ T cells, reaching a peak viral RNA within the first three weeks of infection[112]. As HIV replicates, the immune system responds with a strong pro-inflammatory response which limits viral replication but also causes robust activation of CD4+ T cells, thereby increasing the number of HIV susceptible cells[113]. During this time, HIV also establishes a reservoir of latently infected cells that have low or no viral transcription, allowing these cells to effectively evade the immune response[114]. This reservoir is predominately comprised of monocytes, macrophages, and activated CD4+ T cells which revert to a dormant, resting state that persist throughout the course of disease[115].

This initial immune response may be asymptomatic or have clinical manifestations such rash, and mild, or flu-like symptoms, but the high viral loads also significantly increase transmission risk during this time[116]. Following peak viremia, there is a 100-fold decrease of HIV RNA driven by the CD8+ T cell response[117] and a rebound of peripheral CD4+ T cells[118,119]. However, the CD8+ T cell response does not eliminate the viral reservoir but elicits strong selective pressure on HIV to develop escape mutations and increases viral diversity[120,121]. Eventually, an equilibrium between viral infection and the immune response results in the amount of viral RNA remaining stable for months to years, referred to as HIV spVL, as individuals enter the chronic phase of infection[122].

During chronic infection, a persistent immune response and increased viral diversity causes a shift from HIV using CCR5 as the dominant co-receptor to CXCR4[123]. This change in viral tropism results in more effective infection of Th2 cells compared to Th1 cells[124] and is associated with faster CD4+ T cell decline[125]. Without therapeutic

intervention, the progressive loss of CD4+ T cells result in an AIDS-defining illness over the course of ~10 years, with CD4+ T cells declining from a healthy range of 500 – 1500 cells/mm^3 to less than 200 cells/mm^3 [126]. However, the rate of disease progression is highly variable, with some PLWH developing AIDS within the first three years of infection and others having sustained control of HIV for over 25 years without ART[127,128].

Factors influencing HIV spVL

One of the strongest predictors of variability in HIV disease progression and transmission risk is HIV spVL, which is influenced by environmental, viral, and host genetic factors (Figure 1)[129]. Environmental factors correlated with spVL include age at HIV infection, low CD4/CD8 ratio, transmitting partner's spVL, and level of inflammation[130–133]. Viral genetic factors include infection of multiple founder viruses and viral genotype[134–136]. The two strongest host genetic factors correlated with spVL include specific HLA alleles, such as HLA-B*5701 and HLA-B*5703[137], and *CCR5Δ32* alleles[138]. However, the study of host genetic factors has been largely limited to European populations[139], creating a unique opportunity to discover novel factors associated with HIV spVL in non-European populations and significantly advancing our understanding of the HIV host-pathogen interaction.

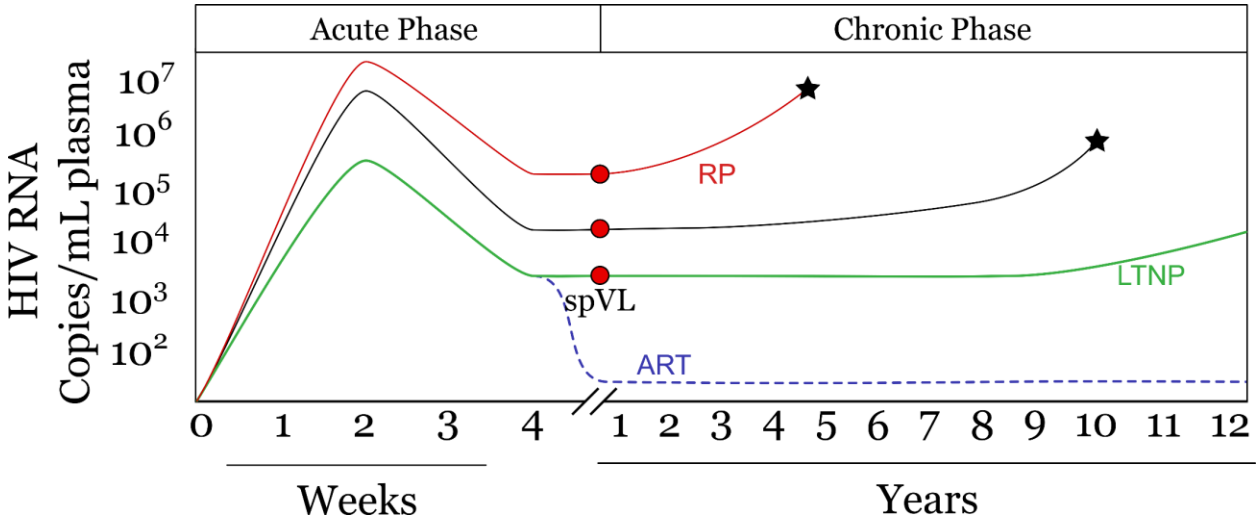


Figure 1. Relationship between HIV set-point viral load and disease progression. Plasma HIV RNA levels (copies/mL) are depicted from acute through chronic infection. After the initial peak during acute infection, individuals reach a set-point viral load (spVL; red circles) which are influenced by environmental, viral, and host genetic factors. A subset of individuals experience rapid progression (RP; red line) characterized by high spVL and earlier onset of AIDS-defining illness (black stars). Long-term non progressors (LTNP; green line) maintain low spVL for many years without therapy. The dashed blue line represents individuals achieving viral suppression through antiretroviral therapy (ART).

Human genomics

The human genome consists of approximately 3.2 billion base pairs, and variation within the DNA sequence has significant impact on the expression and function of human genes. On average, two individuals share over 99% homology between their DNA sequences but genetic variants comprised of single nucleotide polymorphisms (SNPs), insertions/deletions (indels), and structural variants influence unique individual traits[140]. By studying these variants, researchers can understand the underlying molecular biology of complex traits, including HIV susceptibility and progression.

Functional consequences of coding variants

Coding variants are located within exons and can be characterized as missense, synonymous, or nonsense variants. Missense variants change a codon within the mRNA sequence that results in the wrong amino acid being translated which can alter the structure and function of the protein and are most common type of variant associated with genetic diseases[141]. Synonymous variants also change a codon within the mRNA sequence, but due to redundancies, does not result in the change of an amino acid but may still impact the overall function of the gene[142]. Nonsense variants introduce a frameshift or premature stop codon, which results in a truncated or non-functional protein. It is estimated that genetic variants cause the complete loss-of-function of up to 20 protein-coding genes per genome[143].

When coding variants alter sequences in pattern recognition receptors, such as TLR7 (rs179008-A-T) or TLR9 (rs352140-A-G), it can result in rapid disease progression or increased susceptibility to HIV by interfering with immune signaling pathways[144,145]. Conversely, coding variants in HIV HDFs can interfere with the viral lifecycle, such as 32-bp deletion in *CCR5*, which causes a conformational change of the protein and prevents it from being expressed on the cellular membrane[146]. This variation, referred to as *CCR5Δ32*, is strongly associated with reduced viral load in heterozygous individuals and protection against R5-trophic HIV infection in homozygous alternate individuals[78,147]. Variants in the coding sequence can cause significant changes to protein structure, splicing, and function leading to altered susceptibility to diseases, differential response to medications, and overall health.

Functional consequences of non-coding variants

While coding variants can have a detrimental impact on protein function, most genetic variants (~98%) fall outside the coding region of the genome. Noncoding variants may be located between exons, called intronic variants, and in the regions between genes, called intergenic variants. A large proportion (~85%) of the noncoding genome is comprised of 5' and 3' untranslated regions, intronic regions, and/or transposable elements which can influence gene expression by altering promoter, enhancer, transcription factor binding sites (TFBS), and/or splicing of nearby genes[148,149]. Noncoding variants can also impact the folding of chromatin and DNA elements, called topologically associating domains, to influence expression of distal genes[150]. However, given that individual effect sizes for noncoding variants is small, identifying which genetic variants from large-scale genetic studies directly influence gene expression or phenotypes is challenging[151].

An important method for understanding how genetic variants alter gene expression is the identification of quantitative trait loci (QTL). This process involves identifying genetic variants associated with quantitative molecular traits, such as RNA expression quantitative trait loci (eQTLs) or methylation quantitative trait loci (meQTLs), to identify variants with regulatory potential. However, these regulatory variants can be cell-type, cell-state, and ancestry-specific, which makes assessment of large numbers of variants time-consuming and costly[152–155]. Additionally, follow-up studies of genetic associations investigating the impact of noncoding variants on gene expression have been predominantly conducted in cell lines which may not accurately represent gene expression profiles observed in primary cells[156]. Therefore, it becomes important to have scalable methods for identifying genetic variants but also consider appropriate models for functional follow-up.

Host genomic studies of human disease

When determining the association between genetic variants and a phenotype of interest, large-scale sequencing or microarray studies capture a considerable proportion of genetic variation. These studies rely on hundreds to thousands of individuals to have sufficient statistical power to determine associations with common (minor allele frequency (MAF) ≥ 5%), low frequency (0.5% ≤ MAF < 5%), or rare (MAF < 0.5%) variants.

Large-scale sequencing studies, such as whole-genome sequencing (WGS), allow researchers to directly observe the presence or absence of tens of millions of variants, insertions, deletions, and copy number variations and perform associations with traits of interest. The gold standard of these studies requires sequencing genomes to 30X coverage depth to distinguish rare variants from sequencing errors[157]. While the cost of 30X sequencing has decreased significantly over the past couple decades, these studies still have a significant cost and computational burden for analysis.

In contrast, genotyping microarrays typically identify less than one million SNPs and rely on imputation to identify the presence of additional variants. Imputation leverages linkage disequilibrium (LD) patterns, the non-random assortment of variants at different loci, to infer the presence of ungenotyped variants[158]. Variants in high LD are inherited together more frequently than random chance alone, and through imputation researchers can accurately predict the presence of tens of millions of additional variants. This process requires large-scale, high-coverage genetic sequences to ensure accurate imputation, typically sourced from WGS studies[159]. Imputation is much more cost-effective for large-scale studies but has a much lower resolution compared to whole-genome sequencing.

One of the major challenges for imputation is that host genomic studies of diseases have been conducted predominately in European populations[139]. This introduces significant bias when imputing genotypes using European reference panels with diverse populations, as human evolution and genetic bottlenecks result in population-specific variation and genetic structure[160–162]. Specifically, African populations exhibit higher levels of genetic diversity and shorter haplotype structure than Europeans, and using European reference panels for imputation can lead to the inability to capture African-specific variation in association testing[163]. Therefore, when evaluating the effect of genetic diversity on heritable traits, especially when the genetic association is population specific, requires consideration of research methodologies applicable to those specific populations.

Genome wide association studies

Genome-wide associations studies examine differences in the frequency of genomic variations across large population sizes, associated with a specific phenotype. These

studies rely on strict statistical thresholds to control for multiple comparisons, with a standard being $p < 5\times10^{-8}$ in common[164] or $p < 1\times10^{-8}$ for low frequency variants[165]. Genome-wide association studies are a valuable tool for identifying genetic associations with complex traits at the population level, but computational and functional studies are necessary to identify the underlying biological mechanisms.

Interpretation of GWAS associations

Complex human diseases result from a combination of environmental and genetic factors that influence the overall phenotype. The degree of influence of genetic factors on complex traits can vary significantly, ranging from 25% of HIV viral load[166] to over 80% in schizophrenia[167], making it difficult to draw specific biological conclusions from GWAS associations. Therefore, researchers employ fine-mapping techniques to identify causal variants to better understand underlying biological mechanisms.

When investigating GWAS associated regions, the most significant variant, referred to as the lead variant, is not likely causal for the phenotype[168]. Secondly, the nearest gene, without considering the functional context of genetic variants, may not be responsible for influencing the trait[169]. Thirdly, there can be dozens of genetic variants in LD to the lead variant, spanning multiple genes, which can mask the true association of the variant driving the phenotype changes[170]. Lastly, the majority of variants in GWAS-associated regions are noncoding, but the causal variants likely influence gene expression through regulatory regions that can be cell-type-specific, disease-specific, condition-specific, and/or ancestry-specific[171–173]. This creates a significant burden on identifying causal variants from GWAS-associated regions, as there can be thousands of variants requiring a prioritization for subsequent analysis.

Statistical models for identifying causal variants

Statistical fine-mapping approaches identify candidate causal variants from an association signal using the allele frequency, effect size, significance, and LD structure[151,174]. The four main methods for prioritizing the statistical signal include penalized regression, trans-ethnic, heuristic, and Bayesian models[175]. Penalized regression approaches investigate all genetic variants within the associated region for the likelihood of one or more causal variants. Tools such as the Least Absolute Shrinkage

and Selection Operator (LASSO) or elastic net regression models control for multiple comparison by imposing a penalty on variants, only allowing variants which have a non-zero effect on the phenotype to be included in the final model[176,177].

Trans-ethnic fine-mapping approaches take advantage of the fact that different populations have unique LD and haplotype structures, which are leveraged to identify variants that are causal across different populations. These studies rely on analyzing populations that are genetically diverse, such as European and African populations, to improve statistical confidence[178]. However, this approach also requires the assumption that the given genetic association is present within both populations.

Heuristic fine-mapping approaches prioritize the lead variant and investigate the functional impact of all variants in LD to the lead variant. This approach focuses on a relatively small subset of variants and relies on incorporating functional annotations to identify the impact of genetic variants on regulatory regions, coding sequence changes, TFBS, and gene expression to understand the biological impact of these variants.

Bayesian fine-mapping approaches calculate a posterior inclusion probability that a given variant, or combination of variants, are driving the effect. This approach requires the assumption that a single variant[179] or a fixed number of variants[180], are responsible for driving the association and a prior probability is assigned to each variant based on these assumptions. The posterior inclusion probability is then calculated using Bayes' formula for each model to determine the probability that a given variant is responsible for the genetic association. While the posterior inclusion probability can provide insight into statistically relevant genetic variants, additional follow-up studies are required to assess the functional consequences of these variants.

Impact of causal variants on disease

When assessing the impact of causal variants on a particular phenotype, it becomes important to distinguish between monogenic and complex diseases. Monogenic diseases are caused by alterations to a single gene, have large effect sizes, and typically follow Mendelian inheritance patterns (autosomal dominant, autosomal recessive, or X-linked)[181]. In contrast, GWAS have been typically applied to complex diseases which are polygenic, where genetic variants collectively contribute towards a single phenotype by

influencing expression of multiple genes[182]. Given our understanding of monogenic and polygenic diseases, there are several ways which genetic variants may influence a phenotype of interest such as: directly changing the structure or function of a gene[146], changing expression of a proximal or distal gene[183], or altering the expression of a transcription factor which alters expression of a pathway (Figure 2)[184,185]. Therefore, even with a shortlist of candidate variants, it can be difficult to distinguish the underlying biological mechanisms driving complex diseases.

Functional fine-mapping becomes necessary to determine how candidate causal variants influence protein structure, TFBS, enhancers, promoters, noncoding RNAs, mRNA stability, and mRNA splicing across a wide range of cell types and stimulation conditions. An example of this is the integration of GWAS summary statistics with large-scale multi-omics data to identify how candidate variants influence regulatory features and their subsequent target genes. Integration of epigenomic data (DNase-seq, ATAC-seq, ChIP-seq) and RNA-sequencing transcriptomics, identified that a GWAS variant, rs4810856, acts as an enhancer to three nearby genes which synergistically activated phosphorylated AKT, thereby increasing the risk of colorectal cancer[186]. This mechanism was supported by other studies which show that AKT inhibitors reduce cell proliferation and tumor progression[187]. Therefore, it becomes important to not simply focus on the gene closest to a GWAS-associated region but also consider that multiple genes may contribute towards a single phenotype.

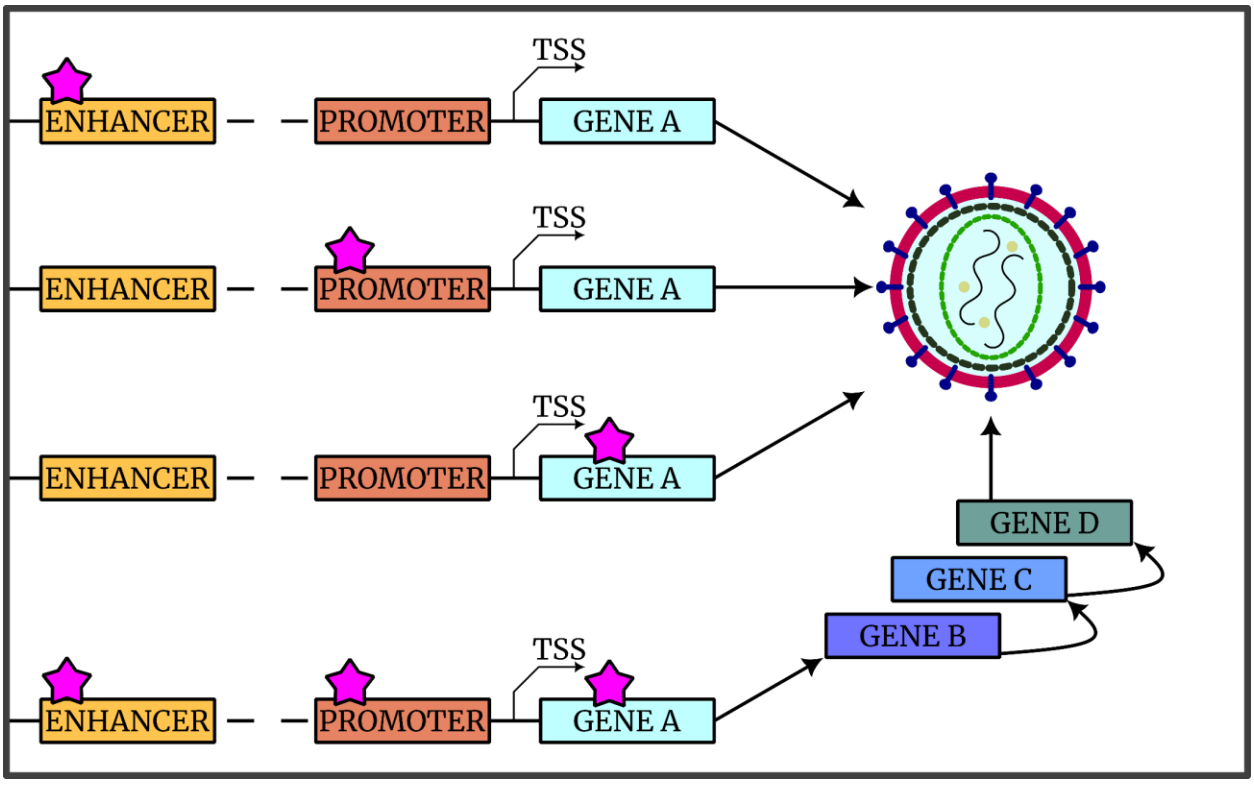


Figure 2. Proposed mechanisms of how genetic variants influence HIV infection through regulatory and coding regions.

Non-coding variants (magenta stars) within enhancer or promoter regions regulate expression of *Gene A,* directly influencing the HIV life cycle. Coding variants or intronic variants within *Gene A* alter protein structure or function, directly influencing the HIV life cycle. Variants may also indirectly affect expression of *Gene A*, which acts as a transcription factor regulating *Gene B, Gene C,* and *Gene D*, which are responsible for influencing the HIV life cycle.

Host genomics of HIV infection

The first GWAS of HIV infection was performed in 2007 after eight European and one Australian cohort established the Center for HIV/AIDS Vaccine Immunology, which included 486 patients with HIV spVL data and identified that genetic variants within the HLA region of chromosome 6 contributed to 16.1% of HIV spVL variance[188]. While the influence of *HLA-B*5701* was previously identified in candidate gene studies[189,190], this study showed that GWAS approaches can detect HIV factors associated at a population level and suggests that larger sample sizes were necessary to detect additional factors. However, subsequent HIV GWAS showed mixed results in identifying significant associations due to small cohorts, confounding between environmental factors and phenotypes, and population admixture[191–197]. Collectively, these studies recapitulated the importance of *CCR5* and the HLA locus in HIV disease but did not identify replicable, significant associations in other genomic regions (Table 1).

In 2013, 26 cohorts were consolidated into the International Consortium for the Genomics of HIV (ICGH) which allowed for a GWAS of European HIV acquisition in 6,300 PLWH and 7,200 controls to be performed[198]. In 2015, a subset of this cohort was used to investigate host genetics that influence HIV spVL in European ancestry individuals[166]. Collectively, these two studies had a significant impact on HIV host genomics as they confirmed genetic associations in the *HLA* region on chromosome 6 and identified genetic associations within the chromosome 3 region, surrounding *CCR5*, associated with control of HIV. Notably, these studies did not replicate the non-significant associations from smaller GWAS[191–197], highlighting that the *HLA* and *CCR5* regions are the most significant loci associated with HIV acquisition and progression in European populations. While these two GWAS were the largest conducted with Europeans, these analyses did not include individuals of African ancestry from the ICGH, which limited their ability to detect novel associations.

Table 1. Genome-wide association studies of human immunodeficiency virus type 1.

Study	Population	Phenotype	Lead Variant(s)[A]
Fellay et al. Science. 2007.	European N = 486	Set-point viral load	rs2395029 $p < 9.36\times10^{-12}$
Sophie Limous et al. J Infect Dis. 2009.	European Cases = 275 Controls = 1,352	Disease progression	rs2395029 $p < 6.79\times10^{-10}$
Sigrid Le Clerc et al. J Infect Dis. 2009.	European Cases = 85 Controls = 1,352	Set-point viral load	rs4118325 $p < 6.1\times10^{-7}$
Herbeck et al. J Infect Dis. 2010.	European Cases = 156 Controls = 590	Disease progression	rs6984368 $p < 8.76\times10^{-6}$
Pelak et al. Journal of Infectious Diseases. 2010.	African American N = 515	Set-point viral load	rs454422 $p < 1.49\times10^{-6}$
Van Manen et al. PLoS One. 2011.	European N = 404	AIDS-diagnosis AIDS-related deaths	rs1523635 $p < 3.50\times10^{-6}$
Lingappa et al. PLoS One. 2011.	African Cases = 496 Controls = 302	Acquisition	rs3745760 $p < 7.83\times10^{-7}$
	African N = 403	Set-point viral load	rs13111989 $p < 2\times10^{-7}$
McLaren et al. PLoS Pathog. 2013.	European Cases = 6,334 Controls = 7,247	HIV acquisition	rs4418214 $p < 3.6\times10^{-11}$
McLaren et al. Proc Natl Acad Sci USA. 2015.	European N = 6,315	Set-point viral load	rs59440261 $p < 2.0\times10^{-83}$ rs1015164

			$p < 1.5\times10^{-19}$
Wei et al. Nature. 2015.	HAN Chinese N = 288	Set-point viral load	rs947612 $p < 2.15\times10^{-6}$
	YUN Chinese N = 154		rs202072 $p < 3.\times10^{-6}$
	XIN Chinese N = 96		rs11231017 $p < 7.39\times10^{-7}$
Real et al. iScience. 2023.	European Cases = 233 Controls = 209	Disease progression	rs2261514 $p < 1.75\times10^{-6}$
McLaren et al. Nature. 2023	Africans N = 1,341 African Americans N = 2,535	Set-point viral load	rs59784663 $p < 6.4\times10^{-10}$

[A]In the event there were no significant association, only the most significant association is reported.

In 2023, we reported the findings of a GWAS meta-analysis of 11 cohorts from the ICGH, focusing on PLWH of African or African American descent[1] (see Author Contributions). The study identified a significant association on chromosome 6 in the HLA region and a novel association on chromosome 1 associated with HIV spVL. The lead variant on chromosome 1, rs59784663-A-G, and variants in LD, overlapped three protein-coding genes: *PRKAB2, FMO5,* and *CHD1L*, as well as a noncoding RNA, *LINC00624*, which had not been previously implicated in HIV disease. This region is associated with a 0.30$\log_{10}$ reduction of HIV spVL, a similar magnitude as the *CCR5Δ32* effect in heterozygous European individuals[166], but the chromosome 1 association was not present in genomic studies of European[166] or Asian[199] populations.

Preliminary statistical analysis of genetic variants in the region using a generalized gene-set analysis suggested that *CHD1L* was the causal gene. Fine-mapping of the region identified that a cluster of 17 variants encompassed the 95% credible set of candidate causal variants located downstream of *CHD1L*. However, there was no significant colocalization between *CHD1L* lymphoblastoid cell line eQTLs and HIV spVL, suggesting that the effect of causal variants may be cell-type-specific.

The bioinformatics results were validated using functional knockout and overexpression experiments performed in U2OS, immunopluripotent stem cell derived macrophages (iPSDMs), and THP-1 cells which determined *CHD1L* expression is inversely proportional to HIV replication. However, there was no effect of *CHD1L* knockout in Jurkat T cells, further emphasizing *CHD1L*-mediated HIV inhibition is cell-type-specific. Taken together, these results suggested that *CHD1L* is a novel HIV inhibitory factor but that further work is necessary to determine the association between gene variants, gene expression, and HIV spVL[1].

Putative HIV host factors

One of the challenges in determining causal genetic variants comes from shared regulatory features, where genetic variants can influence the expression of multiple genes[200]. When examining *PRKAB2, FMO5*, *CHD1L,* and *LINC00624* expression, these genes are strongly co-expressed across 34,149 public RNA-sequencing samples[201], suggesting that they are involved in similar biological processes or share similar

regulatory features (Figure 3). While our previous study investigated the role of *CHD1L* due to its proximity to the lead variant, the potential for *PRKAB2*, *FMO5*, and *LINC00624* to influence HIV replication is unclear. Therefore, it is important to determine whether these genes influence HIV-associated pathways to assess their potential to influence HIV replication and spVL.

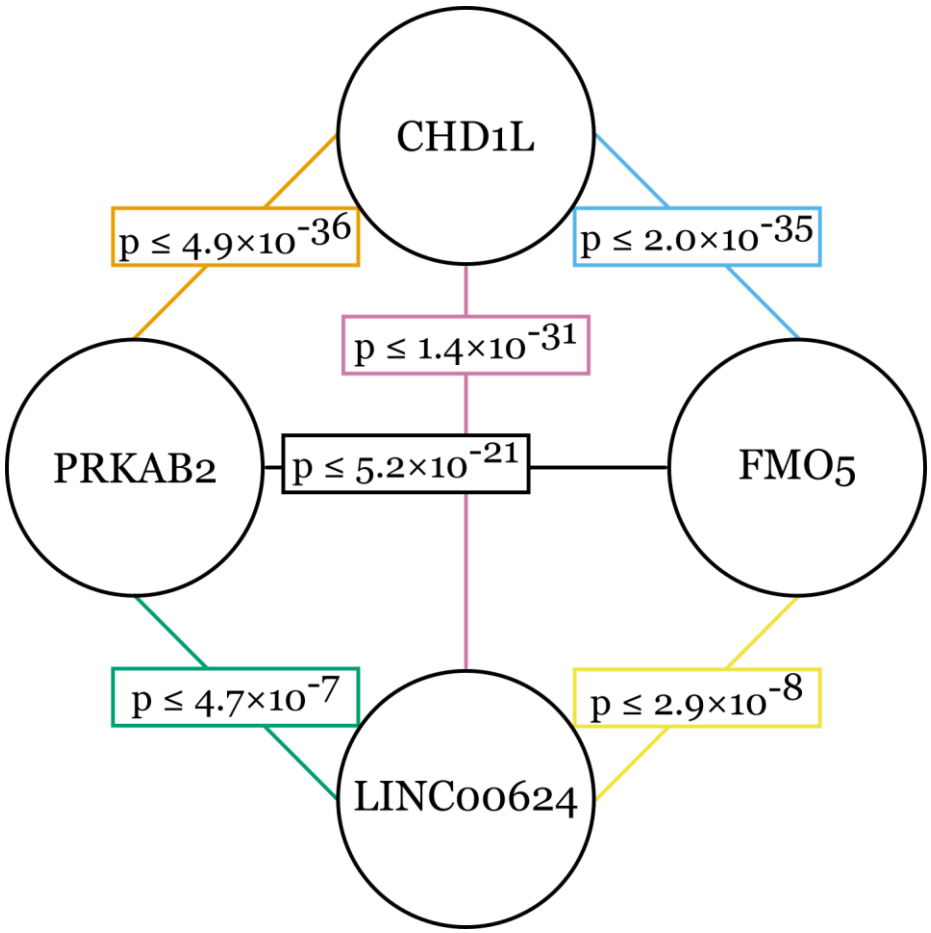


Figure 3. Chromosome 1 genes exhibit a high-degree of co-regulation suggesting shared regulatory mechanisms. Co-regulation of genes was assessed using GeneNetwork Assisted Diagnostic Optimization through www.genenetwork.nl (accessed January 10, 2025). P-values represent the significance of gene co-expression across 34,149 public RNA-sequencing values.

Structure and function of PRKAB2

The gene *PRKAB2* is located on chromosome 1 reverse strand (GRCh38; 1:147,155,106-147,172,470 Mb) and is one of the two regulatory beta subunits of the adenosine monophosphate-activated protein kinase (AMPK) complex. The AMPK complex is a heterotrimeric protein that is present in all mammalian cells and has been well characterized as a regulator of cellular metabolism. The structure of AMPK includes a catalytic subunit (α1 or α2), a scaffolding subunit (β1 or β2), and a regulatory subunit (γ1, γ2, or γ3). Together, these subunits form multiple isoforms of AMPK that are distinct in both subcellular location and function[202], and exhibit different physiological roles depending on tissue and cellular state[203].

When activated, AMPK phosphorylates enzymes which promotes catabolic processes that have been strongly implicated in regulating obesity, type 2 diabetes, cardiovascular diseases, ageing, and the immune response[204–208]. Knockout of *PRKAB2* resulted in the expression changes to over 400 genes compared to wild-type hiPSCs[209], suggesting changes to *PRKAB2* regulatory features may have significant downstream signaling consequences which include key regulators of the immune system such as the mTOR[210], PI3K-AKT[211], MAPK[212], and NF-κβ pathways[213,214].

Structure and function of FMO5

The gene *FMO5* is located on chromosome 1 reverse strand (GRCh38; 1:147,186,261-147,225,339 Mb), and upstream of *PRKAB2*. The flavin-containing monooxygenases are involved in catalyzing drugs, pesticides, toxins, and xenobiotics[215,216]. These genes have different physiological compartments, structure, and substrate specificity[217]. In *Caenorhabditis elegans*, *FMO5* is essential for the defense response and regulates inflammatory pathways[218]. This is also observed in mouse models, where *FMO5* exhibits anti-inflammatory effects through inhibition of the NF-κβ and JAK-STAT signaling pathways[219,220]. However, this effect has not been observed in humans, warranting further investigation of *FMO5* as a mediator of host defense pathways.

Structure and function of CHD1L

The gene *CHD1L* is located on the forward strand of chromosome 1 (GRCh 38; 1:147,242,684-147,295,762) and was first identified in 2008 as an oncogene in human

hepatocellular carcinoma[221]. The role of *CHD1L* has been predominately studied as a regulator of apoptotic processes, DNA repair mechanisms, chromatin remodeling, and as a transcription factor[222–224].

CHD1L acts as a transcription factor by binding promoter regions through its macro domain and activating downstream signaling pathways[225]. These downstream genes, such as *SPOCK1* and *TCTP,* also regulate key host pathways including inflammation, cell cycle regulation, apoptosis, and protein synthesis[226,227]. Knockdown of genes downstream of *CHD1L*, such as *TCTP*, have been shown to inhibit HIV replication in HeLa cells[228] suggesting that *CHD1L* may indirectly influence HIV replication through downstream pathways.

In addition to its regulatory functions, *CHD1L* complexes with PARP1 during DNA repair and chromatin remodeling, and *CHD1L* knockout inhibits downstream signaling pathways[229]. During HIV infection, PARP1 acts as an auxiliary factor through activation of HIV long terminal repeat transcription and knockdown of PARP1 significantly reduces HIV replication in monocyte-derived macrophages[230]. Therefore, it may also be important to consider *CHD1L* interactors, and not just genes downstream of *CHD1L*, which may influence HIV pathogenesis.

Structure and function of LINC00624

Long noncoding RNAs (lncRNAs) have a myriad of roles including regulating subcellular localization of proteins, generation/sequestering of siRNAs, modulating alternative splicing, and chromatin remodeling. Long Intergenic Non-Protein Coding RNA 624 (LINC00624) overlaps with the HIV spVL GWAS association on chromosome 1[1] and increases *CHD1L* expression by sequestering the HDAC6-TRIM28-ZNF354C repressor complex from binding to the *CHD1L* promoter region[231]. It is unclear whether *LINC00624* can influence the expression of *PRKAB2* or *FMO5* through this mechanism, or if this promoter specifically regulates *CHD1L* expression.

Study rationale, hypothesis, and objectives

Study Rationale

Individuals of African ancestry bear a disproportionate burden of HIV disease and are significantly underrepresented in large-scale genomic studies. A recent GWAS of HIV spVL identified a novel region of chromosome 1 associated with HIV spVL in African populations and implicated *CHD1L* as a novel HIV inhibitory factor[1]. However, the relationship between HIV spVL associated variants, *CHD1L, PRKAB2*, *FMO5*, *LINC00624,* and HIV spVL remain poorly understood. To address this gap in knowledge, this study combines detailed fine-mapping of the GWAS-associated region to characterize the relationship between HIV spVL-associated variants, HIV spVL, and gene expression profiles. Through better understanding of how host genetics influence HIV spVL, this study has a high probability of identifying novel pathways regulating HIV infection in African populations.

Hypothesis

HIV spVL associated variants within the chromosome 1 region influence expression of *PRKAB2*, *FMO5*, *LINC00624,* and/or *CHD1L* in a cell-type-specific manner. These variants are also associated with differentially expressed signaling pathways downstream of *CHD1L* and/or *PRKAB2* that are associated with HIV infection.

Objective 1. Determine which HIV spVL associated variants are driving the GWAS association

- Assess the overlap between candidate causal variants and regulatory features to prioritize variants with regulatory potential
- Develop a heuristic, functionally informed fine-mapping approach to identify candidate causal variants
- Determine the association between HIV spVL associated variants and gene expression in individuals with homozygous reference, heterozygous, and homozygous alternate alleles

Objective 2. Evaluate the relationship between chromosome 1 variants, *PRKAB2* expression, and HIV-associated pathways

- Determine whether *PRKAB2* and *CHD1L* are co-regulated by evaluating co-expression profiles in immune cells

- Determine whether GWAS candidate causal variants are associated with *PRKAB2* expression
- Determine whether PRKAB2 knockout influences expression of *CHD1L*
- Identify downstream genes and pathways regulated by *PRKAB2* and are associated with HIV infection

Objective 3. Determine the relationship between HIV spVL associated variants, *CHD1L*, *PRKAB2*, *FMO5*, *LINC00624*, and HIV spVL in *ex vivo* models

- Assess the expression of *CHD1L, PRKAB2,* and *FMO5* in PLWH based on HIV spVL associated variants
- Determine transcriptome differences in healthy individuals with different alleles at candidate causal variants
- Identify gene expression profiles in *ex vivo* immune cell subsets

Chapter 2: Functionally-informed Fine-mapping Identifies Genetic Variants Linking Increased CHD1L Expression and HIV Restriction in Monocytes

Introduction

A region on chromosome 1 surrounding *CHD1L*, *PRKAB2*, and *FMO5* was significantly associated with HIV spVL following a genome-wide association study of individuals of African ancestry[1]. This association, centered around the lead variant rs59784663-A-G (β -0.31, $p<6.4\times10^{-10}$), was not observed in previous GWAS in European[166] or Asian[199] populations. Initial investigations into the association identified *CHD1L* as the causal gene, with follow-up studies suggesting *CHD1L*-mediated HIV inhibition is cell-type-specific[1]. However, which genetic variants are driving gene expression changes and the impact on regulatory features across different cell types remains unclear.

The previous fine-mapping approach leveraged a LD reference panel consisting of 1000 Genomes African and African American samples which identified rs72694728-T-A (β=0.17, $p < 2.28\times10^{-6}$) as the most significant likely causal variant[1]. However, this variant has a high frequency in European populations (MAF ≥ 0.31)[232] and was not significant in previous European HIV spVL GWAS[166], suggesting that this variant is likely not causal for spVL variation. Given this, population admixture may have been a confounding factor in identifying candidate causal variants, as the African 1000 Genomes LD reference panel may not have accurately reflected the GWAS population[233].

Given the limitations of reference panels in admixed populations, we sought to prioritize HIV spVL associated variants based on their functional impact on nearby genes and leveraged individual-level genotypes in statistical models, thereby eliminating population admixture as a confounding factor. Here, we assess the impact of candidate causal variants using eQTLs, meQTLs, and gene expression data in peripheral blood leukocytes and monocytes to determine cell-type-specific regulatory effects and identify putative mechanisms of *CHD1L* regulation. The overall effect of HIV spVL associated variants on *CHD1L* expression was assessed using elastic net gene expression models to assess differences in whole blood and monocytes. Overall, this study progresses our

understanding of how African-specific genetic variants influence gene expression and the subsequent impact on HIV spVL.

Methods

GWAS cohort, genotyping, and quality control

Genomic data were obtained from ten independent genome-wide association studies combined under the ICGH (N=3,886). For this meta-analysis, PCs were calculated for the entire cohort using EIGENSTRAT within EIGENSOFT (v.7.2.1) to control for population stratification[234]. Significant PCs were determined using a linear regression analysis against HIV spVL at $p < 0.05$. Linkage disequilibrium of variants in the chromosome 1 region was calculated using individual-level genotype data in plink (v1.9)[235].

Functional consequences of GWAS variants

Functional consequences of HIV spVL-associated genetic variants in the chromosome 1 region (GRCh38; 1:147023443-148358705) from the ICGH cohort were assessed using the Variant Effect Predictor (VEP)[236]. Impact of variants on protein structure and functional domains was performed using mutfunc[237]. Pathogenicity of variants was determined using Sorting Intolerant from Tolerant[238], Polymorphism Phenotyping[239], and Combined Annotation Dependent Depletion[240]. Variants influencing splice sites were determined using spliceAI[241]. The percentage of genetic variants corresponding to each functional consequence was determined by dividing the number of variants within each functional category by the total number of HIV spVL-associated variants investigated (N=3,776).

Functional annotations of GWAS variants

To investigate whether the GWAS variants associated with HIV spVL overlap with functional annotations, variants in LD ($r^2 \geq 0.2$) to rs59784663-A-G were selected for follow-up. Variant annotation was performed using HaploReg v4.2 to integrate experimental annotations from over 127 epigenomes in the ENCODE and Roadmap Epigenetics projects[242–244]. Variants were annotated with H3K4me1, H3K4Me3, H3K27ac, H3K9ac, DNase, and protein binding site annotations and filtered for annotations from blood tissues to prioritize HIV-related traits (complete list Appendix 2).

Expression quantitative trait loci were included from any tissue from the Genotype-Tissue Expression Database (GTEx)[245].

Statistical prioritization of genetic variants

Variants potentially driving the genetic association with HIV spVL were filtered for at least one functional annotation from the HaploReg analysis and used as input to the BIMBAM Bayesian modeling framework[246,247]. BIMBAM computes multi-SNP Bayes factors (BF) of genetic variants against a phenotype of interest, evaluating up to N possible models. Variant dosages were calculated as the mean genotype, as recommended by developers[247]. The number of causal variants was determined by generating N models and selecting the model N, where the ($\log_{10}$BF of model N – 1) > ($\log_{10}$ BF of model N). Significance of variant combinations was defined as $\log_{10}$ BF ≥ 3.5 as defined by Servins and Stephens, 2007[246].

Identification of likely causal signals

Given that variants in the chromosome 1 region exhibited high LD patterns, instead of evaluating combinations of all genetic variants, haplotypes were generated using Haploview (v4.2) on individual-level genotypes from the ICGH cohort to identify haplotype blocks as described by Gabriel et al (MAF ≥ 0.01)[248,249]. The most significant variant from each haplotype was used to represent the strength and significance of the given association. Haplotypes (N=3) that had at least one GWAS significant variant were selected for statistical modelling. A linear model comparing rs59784663-A-G dosage against rs7525622-G-A, rs73004025-C-T, and rs72999655-A-G dosage, controlling for PC1 and sex, was used to assess the significance of likely causal variants. Significance was determined using a chi-square test against a threshold of $p < 0.05$.

Identification of CHD1L eQTLs from publicly available summary statistics

Summary statistics for eQTLs in lymphoblastoid cell lines (LCLs) were obtained from 599 individuals across six African populations within the 1000 Genomes Project (Appendix 1, https://github.com/smontgomlab/AFGR)[250]. Significant eQTLs mapped to *CHD1L* were selected and filtered for overlap with the haplotype analysis. Association between *CHD1L* eQTLs and HIV spVL β was assessed using a linear regression and two-sided spearman approximate correlation in R (v4.1.2).

Additionally, summary statistics for eQTLs in monocytes were obtained from the Multi-Ethnic Study of Atherosclerosis (MESA) cohort for 233 African Americans, 352 Hispanic, and 578 individuals of European ancestry (Appendix 1)[251]. Significant eQTLs (false discovery rate (FDR) < 0.05) from the combined population meta-analysis were filtered for overlap with the candidate causal haplotypes. Association between *CHD1L* eQTLs and HIV spVL β was tested using a linear regression model and two-sided spearman approximate correlation in R (v4.1.2).

Identification of CHD1L meQTLs from publicly available summary statistics

Summary statistics for meQTLs in peripheral blood leukocytes were obtained from the Genetic Epidemiology Network of Arteriopathy (GENOA) cohort for 961 individuals of African ancestry (Appendix 1)[252]. Variants that were significant meQTLs (FDR < 0.05) were filtered for overlap with the haplotype analysis. Hypermethylated sites were defined as $\beta \geq 0.7$ and hypomethylated as $\beta \leq 0.3$ as previously described[253]. Chromatin states of PBMCs from individuals of African (ENCFF863IVL) and European ancestry (ENCFF235OWS), as well as LCLs from individuals of African (ENCFF747OPJ) and European ancestry (ENCFF070KWN), were obtained from ENCODE[254]. Visualization of the chromatin state, *CHD1L* transcript (RefSeq hg38), and methylation sites were performed using R (v4.1.2).

Imputed CHD1L expression

Predictive gene expression models trained on monocyte eQTLs from the MESA cohort[251] (https://github.com/WheelerLab/DivPop) and applied to individual-level genotypes from the African HIV spVL GWAS (Appendix 1). Training data was limited to individuals of African Ancestry to reduce the effects of population stratification. Imputed expression of *CHD1L* was compared for individuals with allelic combinations of rs72999655-A-G, rs7525622-G-A, and rs73004025-C-T. Statistical significance was determined using Mann-Whitney U tests for genotype groups.

Results

The majority of GWAS fine-mapped variants are noncoding with small effect sizes[255], while gain-of-function and loss-of-function variants in the coding regions are associated with a more severe phenotype[256]. HIV spVL associated SNPs in the chromosome 1 region

(N=3,776) were annotated using VEP and showed that most variants were intronic (68%), overlapped noncoding transcripts (57.2%), or downstream of coding genes (24.3%), but a small subset was located within exonic regions (1.9%) (Figure 4A). There were nine variants in *CHD1L* exons distributed amongst: 3' UTR (22.2%), synonymous (22.2%), and/or missense (55.6%) variants, while *PRKAB2* had 16 variants that were 3' UTR (93.8%) or synonymous (6.2%) (Figure 4B).

Further investigation of these exonic variants identified that one *CHD1L* synonymous variant, rs72999656-C-T, was significantly associated with HIV spVL ($p < 1.18\times10^{-9}$) (Table 2). Although there are several missense variants in the *CHD1L* gene, none approached significance ($p > 0.15$) and were not evaluated further. There were no variants in *PRKAB2* exons that met statistical significance (Table 3). The lack of significant non-synonymous coding variants suggests that changes to HIV spVL is being driven by one or more noncoding variants that alter regulatory features.

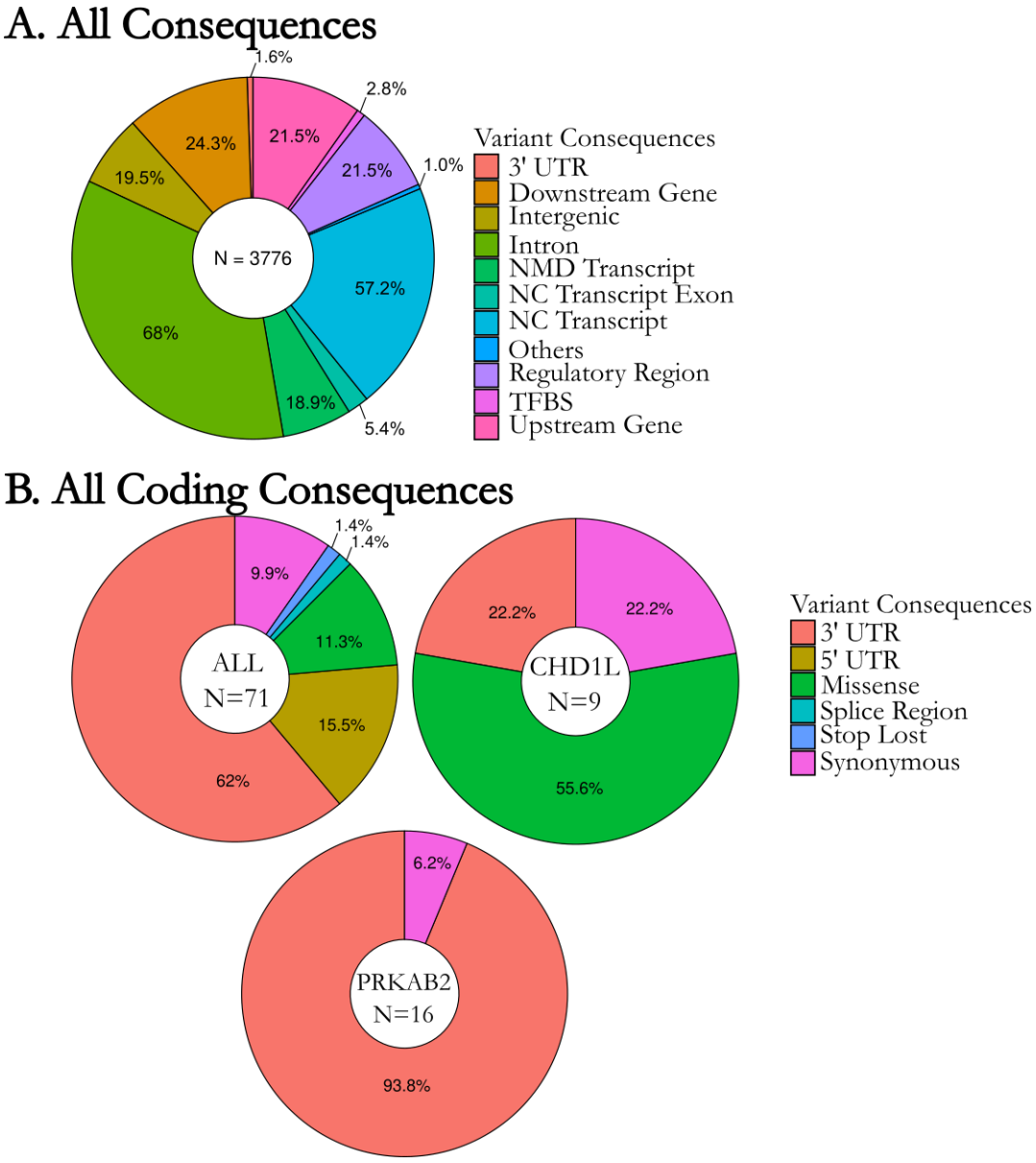


Figure 4. Variant effect predictor consequences of genetic variants in the chromosome 1 region (GRCh38; 1:147023443-148358705). A) Functional genetic consequences for all variants (N = 3776). B) Functional genetic consequences for variants that fall within exons of all genes in the region (N = 71), *CHD1L* (N = 9), and *PRKAB2* (N = 16).

Table 2. Functional consequences of coding CHD1L variants.

RSID	Effect Allele	Consequence	Beta	Effect Allele Frequency	P-value
rs72999656-C-T	C	Synonymous 3' UTR	0.30	0.96	1.2×10^{-9}
rs2275249-C-G	C	Missense 3' UTR	0.075	0.96	0.15
rs61756300-T-A	T	Missense 3' UTR	-0.047	0.94	0.24
rs11588753-G-C	G	Missense	0.032	0.92	0.40
rs4950394-G-T	G	Missense 3' UTR	0.037	0.060	0.41
rs7547279-A-C	A	Missense Synonymous 3' UTR	0.037	0.053	0.41
rs6537-G-A	A	3'UTR	0.016	0.09	0.66
rs13374920-A-C	A	Missense 3' UTR	-0.016	0.93	0.68
rs6538-G-A	G	3' UTR	-0.006	0.85	0.83

Table 3. Functional consequences of coding *PRKAB2* variants.

RSID	Effect Allele	Consequence	Beta	Effect Allele Frequency	P-value
rs72997505-C-T	C	3'UTR	0.19	0.94	6.1×10^{-6}
rs1837984-T-G	T	3'UTR	0.046	0.29	0.045
rs75047837-C-T	C	3'UTR	-0.058	0.89	0.085
rs17159890-A-C	A	3'UTR	0.045	0.89	0.16
rs116479604-C-T	C	3'UTR	-0.065	0.96	0.2
rs77014204-A-T	A	3'UTR	-0.065	0.96	0.2
rs77195331-C-G	C	3'UTR	0.068	0.97	0.25
rs10900321-C-T	C	3'UTR	0.017	0.41	0.41
rs1047140-C-G	C	3'UTR	0.024	0.14	0.43
rs79992453-G-A	G	3'UTR	-0.044	0.97	0.47
rs1442760-T-C	T	3'UTR	0.016	0.14	0.59
rs1837983-A-G	A	3'UTR	0.014	0.14	0.64
rs34838459-G-A	G	Synonymous	0.019	0.94	0.64
rs61704329-A-T	A	3'UTR	-0.02	0.96	0.7
rs6937-T-C	T	3'UTR	0.011	0.88	0.72
rs114820152-T-C	T	3'UTR	0.021	0.98	0.77

Fine-mapping of HIV spVL locus implicates multiple causal variants

A common method for fine-mapping is to prioritize the signal surrounding the lead GWAS SNP and investigate genetic variants in LD[175]. Genetic variants in LD with the lead GWAS SNP, rs59784663-A-G, were evaluated for regulatory potential using functional annotations in the ENCODE and Epigenetics Roadmap databases. Of the 121 variants in LD with rs59784663-A-G, 84 (69.4%) overlapped with at least one known or imputed histone modifications, chromatin state, DNase hypersensitivity sites, *CHD1L* eQTLs, and/or protein-bound regions.

Individual-level genotypes from 3,886 individuals for these 84 functional variants were analyzed using BIMBAM, which identified a three-SNP model ($log_{10}BF = 6.75$) that explained the majority of HIV spVL variance compared to a one-SNP model ($log_{10}BF = 6.52$) and slight improvement over a two-SNP model ($log_{10}BF = 6.74$). In the three-SNP model, the most significant combination of variants was rs72999655-A-G, rs7525622-G-A, and rs7300425-C-T (log_{10} BF=8.663), with the minor allele of each associated with reduced HIV spVL (Table 4). When comparing the effect of the three-SNP model against rs59784663-A-G dosage, the three-SNP model explained significantly more variance in HIV spVL (df=6, Sum of Sq = 13.551, $p < 0.027$). This strongly suggests that these three variants are driving changes to gene expression in the chromosome 1 region, likely through modification of the previously identified regulatory features.

Table 4. Genome-wide association study variants in the chromosome 1 region that explain the majority of HIV spVL variance.

Variant	GWAS Beta[A,B]	GWAS P[B]	r^2 to rs59784663
rs72999655-G-A	-0.30	9.09×10^{-10}	0.99
rs7525622-G-A	-0.21	3.98×10^{-9}	0.46
rs73004025-C-T	-0.29	6.36×10^{-9}	0.90

[A]Direction corresponds to the effect of the minor allele. [B]Values reflect the association between genetic variants and HIV set-point viral load in African populations from McLaren et al[1].

Lead GWAS variant tags six haplotype blocks

Although rs72999655-A-G, rs7525622-G-A, and rs7300425-C-T explain the majority of HIV spVL variance, the potential effect of linked noncoding variants driving expression changes also requires consideration. The 84 functional variants were clustered into six unique haplotypes. Among these, the candidate causal variants, rs72999655-A-G, rs7525622-G-A, and rs7300425-C-T, were the most significant GWAS variants in haplotype blocks 2, 3, and 4, respectively (Figure 5) and have regulatory potential (Figure 6). Additionally, rs72999655-A-G was in perfect LD (r^2 = 1.0) with 11 other variants, including the synonymous *CHD1L* variant rs72999656-C-T. Given the number of variants in high LD, this makes it difficult to isolate the function of a single variant through bioinformatics analysis. Therefore, we focus on identifying how candidate causal variants, and HIV spVL associated SNPs within haplotype blocks, influence *CHD1L* expression.

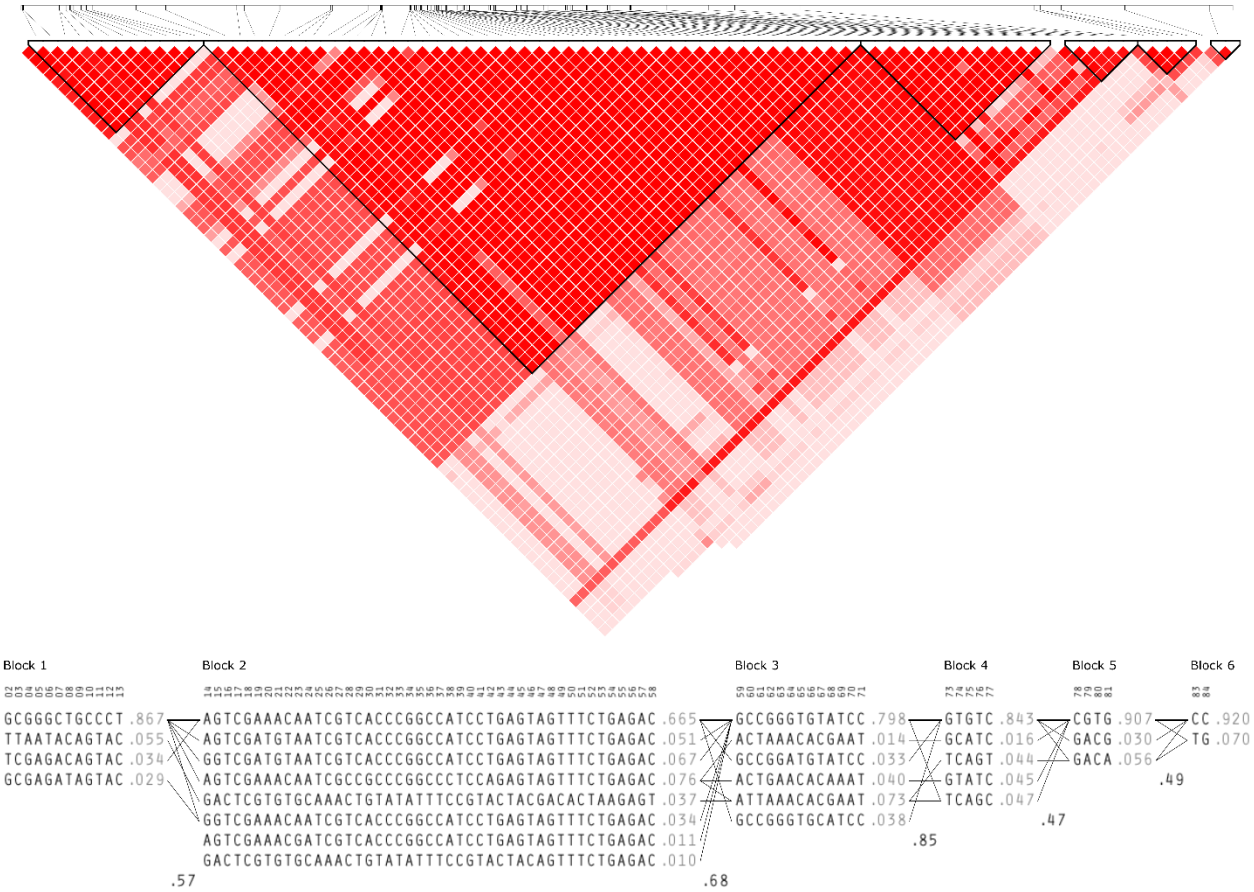


Figure 5. Linkage disequilibrium patterns of functional variants in the chromosome 1 region. The heatmap depicts LD (D') between functional chromosome 1 variants, with darker shades representing stronger LD between variants. Haplotypes are represented if they were present in at least 1% of samples.

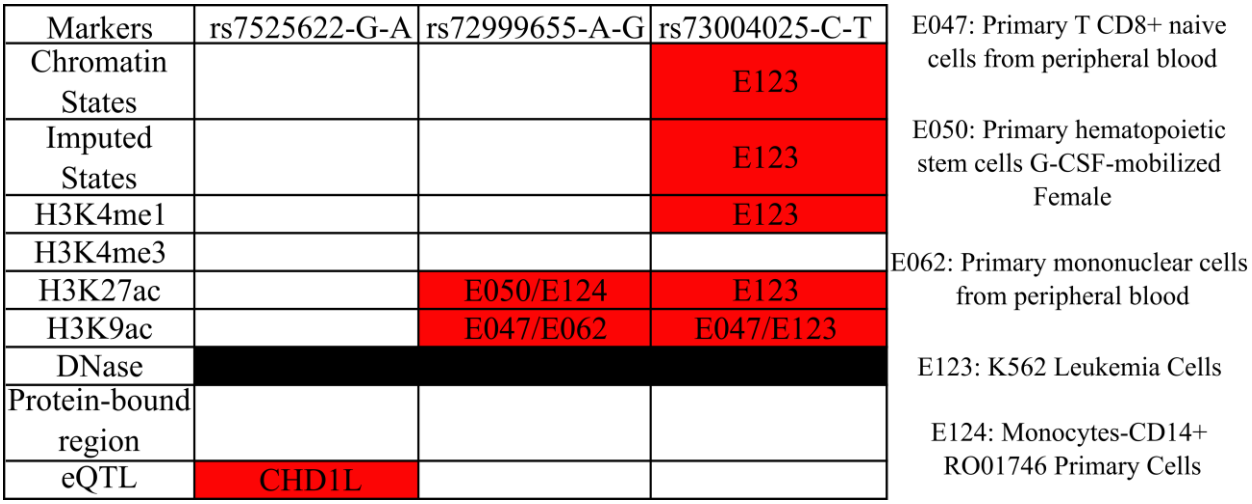


Figure 6. Functional annotations of likely causal variants. Red annotations represent overlapping functional annotations. Black bars represent untested samples. Empty bars represent no overlap. Data was retrieved from ENCODE and Roadmap Epigenetics using HaploReg (v4.2).

Increased CHD1L expression is significantly correlated with decreased HIV spVL

Summary statistics from large African eQTL studies were used to evaluate the association between variants that modify *CHD1L* expression and HIV spVL. In LCLs from the AFGR cohort, there were 14 HIV spVL associated SNPs located within the haplotype blocks that were also *CHD1L* eQTLs. These variants exhibited significant correlation between increased *CHD1L* expression and decreased HIV spVL ($\beta = -0.71$, $p < 0.012$) (Figure 7A). In monocytes from the MESA cohort, there were 18 HIV spVL associated SNPs located within the haplotype blocks that were also *CHD1L* eQTLs. These variants exhibited a significant correlation between eQTLs increasing *CHD1L* and decreased HIV spVL ($\beta = -0.05$, $p < 0.014$) (Figure 7B). These results support previous observations that increased *CHD1L* expression restricts HIV replication and suggests that while individual variant effect sizes are small, there may be a synergistic effect of candidate causal variants.

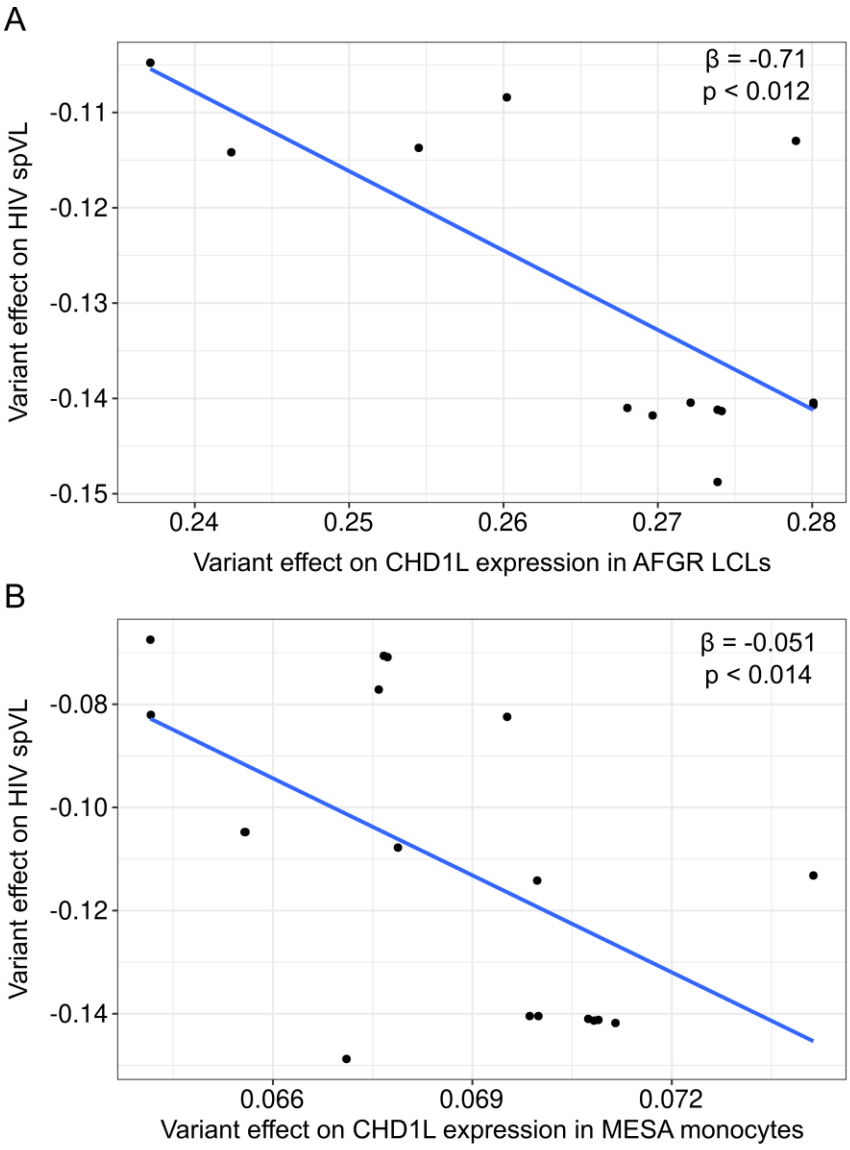


Figure 7. Negative correlation between *CHD1L* expression and HIV set-point viral load. Functional CHD1L eQTLs in haplotypes associated with likely causal GWAS variants (rs72999655-A-G, rs7525622-G-A, rs7300425-C-T), were assessed for their association with HIV spVL. (A) *CHD1L* eQTLs measured in LCLs from the African Functional Genomics Resource study[250]. (B) *CHD1L* eQTLs measured in monocytes from the Multi-Ethnic Study of Atherosclerosis cohort[251]. Blue lines indicate the fitted linear regression of HIV spVL β against *CHD1L* eQTL β. Statistical significance was determined using two-sided Spearman correlation.

Functional GWAS variants are associated with differential CHD1L TSS methylation

Next, we sought to determine whether the identified likely causal variants alter nearby regulatory features by assessing the impact of genetic variants on CpG methylation using peripheral blood leukocytes data acquired from the GENOA cohort. There were twelve variants: rs72999634-A-G, rs72999637-T-C, rs72999638-T-A, rs72999639-T-C, rs72999640-C-T, rs72999646-T-A, rs72999648-G-A, rs72999655-A-G, rs72999656-C-T, rs7526114-G-A, rs59213667-G-A, and rs7535451-A-G, where the allele associated with reduced HIV spVL was also associated tagged hypermethylation at four CpG sites (cg07572770, cg11120551, cg15022772, cg21917976) and hypomethylation at a single CpG site (cg25205988) (Figure 8, Appendix 3). While the individual impact of these variants is unclear, all associated variants were in perfect LD ($r^2 = 1.0$) with rs72999655-A-G, constituting a single biological signal.

Annotation of these five CpG sites to regulatory features in the chromosome 1 region was performed using African and European chromatin states from the ENCODE database for PBMCs and LCLs. The hypomethylated CpG site, cg25205988, overlaps with a flanking transcription start site (TSS) and active TSS, while the hypermethylated CpG sites overlap with an enhancer and flanking TSS in PBMCs. While the direction of the effect remains undetermined, hypermethylation of the *CHD1L* enhancer region may suggest that this biological signal is associated with decreased gene expression in PBMCs.

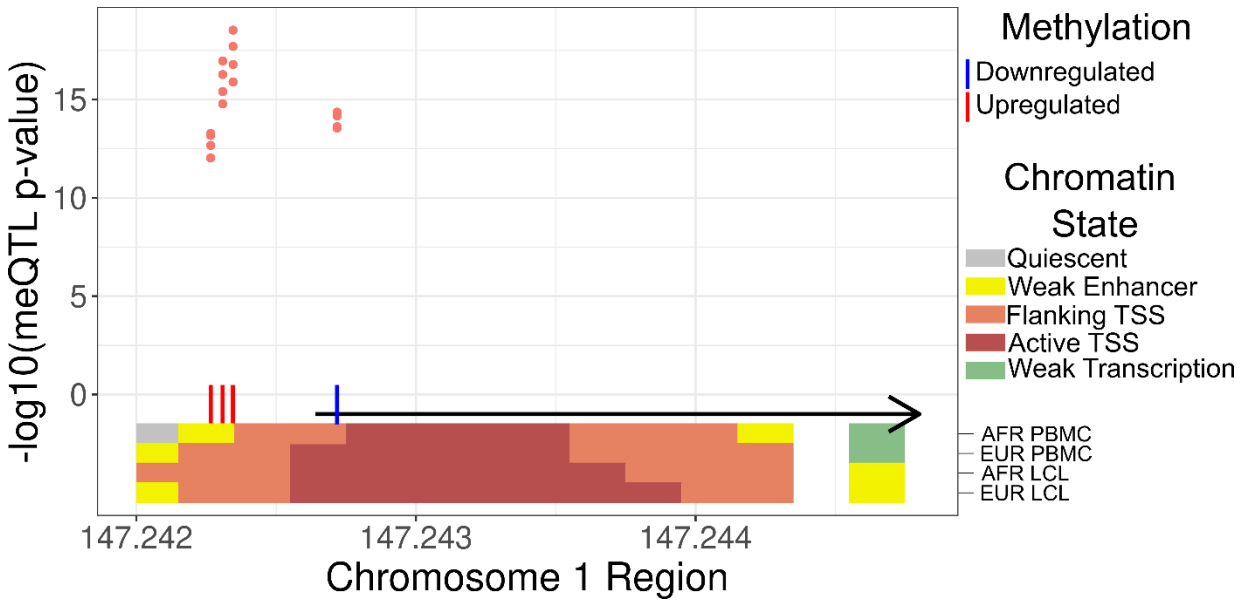


Figure 8. The candidate causal variant rs72999655-A-G is associated with altered methylation of *CHD1L* in African populations. Each data point (pink) represents a meQTL in the chromosome 1 region. The vertical bars show whether the meQTL is associated with hypermethylation (red) or hypomethylation (blue). The black arrow represents the *CHD1L* transcript from RefSeq hg38. Vertical colored bars represent chromatin states obtained from ENCODE (African PBMC: ENCFF863IVL; European PBMC: ENCFF235OWS; African LCLs: ENCFF747OPJ; European LCLs: ENCFF070KWN).

Candidate causal GWAS variants are associated with cell-type-specific CHD1L expression

Our previous analyses suggest that *CHD1L* expression differences likely occur in a cell-type-specific manner based on *CHD1L* eQTLs and meQTLs, but it is difficult to infer the cumulative effects of genetic variants on *CHD1L* expression and HIV spVL. Given that *CHD1L*-mediated HIV inhibition was characterized in THP-1 and iPSDMs[1], we chose to focus on genetically regulated *CHD1L* gene expression was imputed on individual-level genotypes from the HIV spVL GWAS using PrediXcan imputation models in monocytes from African Americans in the MESA cohort.

We observed that *CHD1L* expression was associated with decreased HIV $\log_{10}$ spVL ($\beta = -0.14$, $p < 0.031$) (Figure 9), thereby supporting previous observations that increased *CHD1L* correlates with reduced HIV infection[1]. When we analyzed PrediXcan *CHD1L* expression based on candidate causal variant alleles, individuals heterozygous for rs72999655-A-G, rs7525622-G-A, and rs73004025-C-T exhibited higher *CHD1L* expression than homozygous reference ($p < 2.7\times10^{-12}$), and homozygous alternate exhibited much higher *CHD1L* expression than homozygous reference ($p < 0.013$) from monocyte models (Figure 10).The effect of these variants in individuals with incomplete allelic combinations is less clear, supporting our fine-mapping analysis that a combination of three variants is responsible for driving the GWAS HIV spVL association (Appendix 4). Overall, these results support that HIV spVL associated SNPs influence *CHD1L* expression in myeloid cells, but further work is necessary to identify gene expression changes in other immune cells. Therefore, further investigation is needed to determine how *CHD1L* affects HIV replication in multiple primary cell types to assess the cell-type-specific variation.

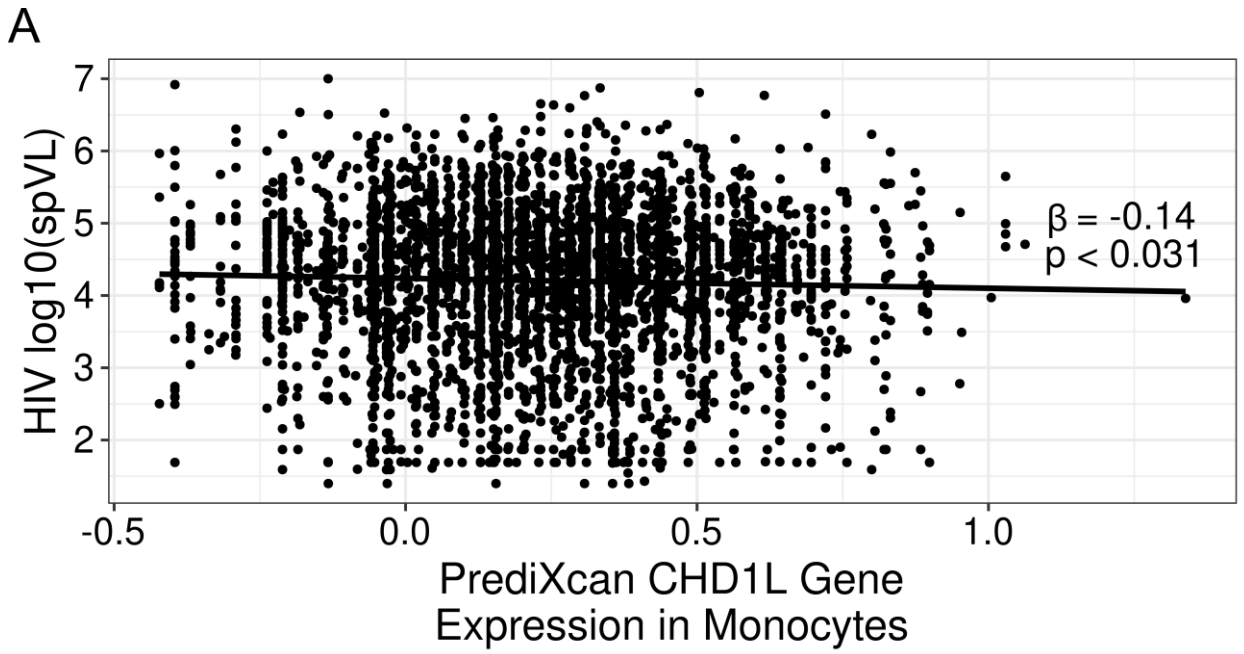


Figure 9. *CHD1L* gene expression from monocytes is negatively correlated with HIV set-point viral load. Gene expression was imputed for 3,886 individuals of African ancestry as part of the International Collaboration of Genomics for HIV using monocyte models from the Multi-Ethnic Study of Atherosclerosis cohort. The black line represents the linear regression of HIV $\log_{10}$ spVL and *CHD1L* expression.

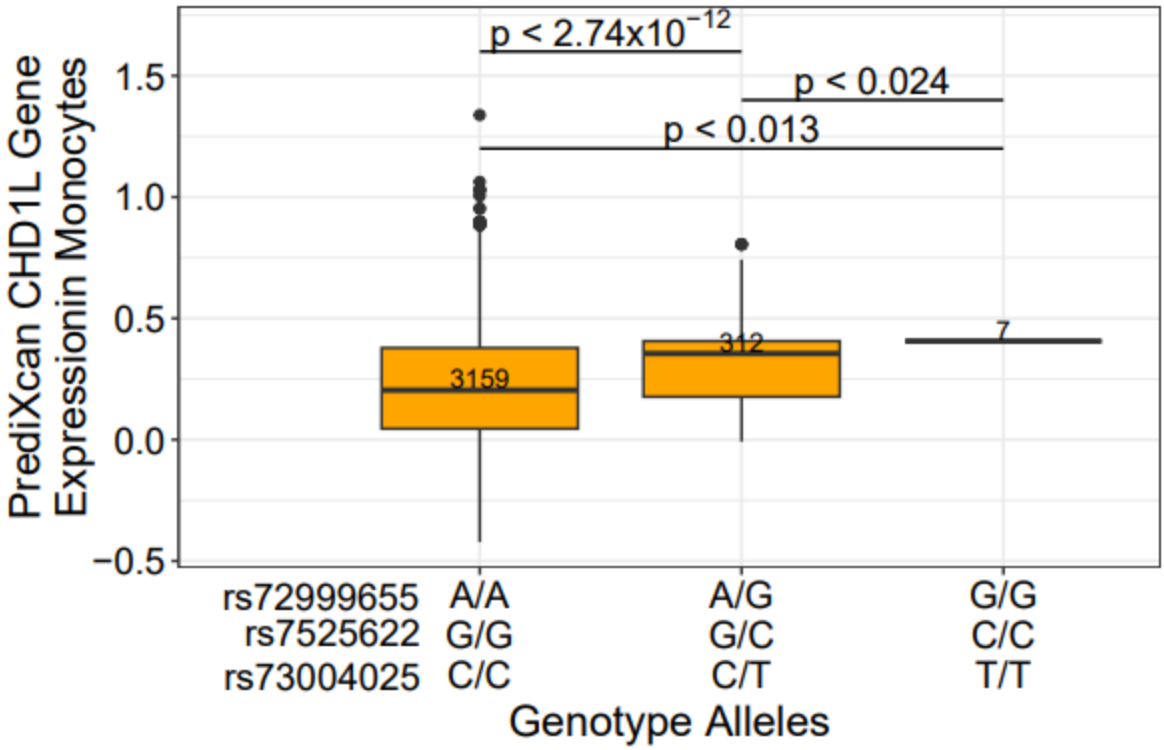


Figure 10. Candidate causal variants are associated with increased *CHD1L* expression in monocytes. *CHD1L* expression was imputed using monocyte gene expression models from the Multi-Ethnic Study of Atherosclerosis African American cohort. Individual allelic combinations are shown for homozygous reference, heterozygous, or homozygous alternate at all variants.

Discussion

Our analysis identified that the majority of HIV spVL associated variants in the chromosome 1 region are noncoding, with one synonymous variant present in the coding region that met genome-wide significance. This suggests that candidate causal variants likely impact gene expression through regulatory features as opposed to directly altering protein function or structure. Through applying a functionally informed heuristic fine mapping approach to individual-level genotypes, a combination of three variants: rs72999655-A-G, rs7525622-G-A, and rs73004025-C-T, explained significantly more variance of HIV spVL than the lead GWAS SNP, rs59784663-A-G, alone. Of note, all three variants exceed genome-wide significance, and both rs72999655-A-G and rs73004025-C-T are specific to African populations[232].

The functional impact of likely causal variants on regulatory features was investigated using chromatin states, epigenetic markers, from the ENCODE and Roadmap Epigenetics, and eQTLs from the GTEx databases. The variant rs73004025-C-T, located downstream of *CHD1L,* overlaps a promoter marker in CD8+ T cells and an enhancer, promoter, and active transcription start site in K562 cells[254]. Similarly, rs7525622-G-A, is located downstream of *CHD1L,* is associated with increased *CHD1L* expression in cultured fibroblasts and skeletal muscle (GTEx Portal accessed on 04/11/2024). The variant rs72999655-A-G overlaps with *CHD1L* transcripts (ENST00000369258.8, ESNT00000361293.10, and ENST00000369259.4), enhancer, and promoter regulatory regions in multiple immune tissues[254]. This is in line with previous observations that GWAS causal variants are most likely to influence promoters and enhancers of target genes[257,258].

Notably, we observed that rs72999655-G was also associated with hypermethylation of two CpG sites upstream of the *CHD1L* active TSS and hypomethylation of a CpG site within the *CHD1L* active TSS. Hypermethylation of promoter regions is typically associated with transcription repression, while hypomethylation is associated with increased transcriptionally activity[259,260], but recent studies identified hypermethylation can also result in increased transcriptional activity[261,262]. The strong association between these variants and HIV spVL, coupled with

their CpG modification with the *CHD1L* TSS, strongly implicate rs73004025-C-T, rs7525622-G-A, and rs72999655-A-G as candidate causal variants.

We also observed a strong negative correlation between *CHD1L* eQTLs and HIV spVL in monocytes, and a much weaker positive correlation in LCLs. This is consistent with previous findings that observed no colocalization between HIV spVL associated variants and *CHD1L* eQTLs in LCLs, further supporting that genetically regulated *CHD1L* expression is impacted in a cell-type-specific manner[1]. Additionally, imputed expression profiles from individuals with known genotypes that showed increased copies of alleles associated with lower HIV spVL are corresponded to increased *CHD1L* expression in monocytes. While the specific function of these variants have not been validated, this highlights that the chromosome 1 region is associated with changes to *CHD1L* expression, warranting further investigation in *ex vivo* studies of individuals of African ancestry.

Overall, this study provides a comprehensive analysis of genetic variants within the chromosome 1 region and the association with HIV spVL. This work provides a shortlist of candidate causal variants for additional functional studies to validate the potential of genetically regulated *CHD1L* expression and the subsequent impact on HIV infection. We highlight the need for additional studies to assess genetically regulated *CHD1L* expression in other cell types and the biological mechanism of *CHD1L*-mediated HIV inhibition.

Chapter 3: Chromosome 1 Variants Associated with Decreased HIV Set-point Viral Load Correlate with PRKAB2 Expression Changes

Introduction

The HIV host-pathogen interaction is complex, with hundreds of human genes being implicated as candidate host restriction, inhibitory, auxiliary, or dependency factors[263]. Recently, a GWAS identified a novel region of chromosome 1 surrounding *PRKAB2*, *FMO5*, and *CHD1L* to be associated with reduced HIV spVL and functional follow-up identified *CHD1L* as a novel HIV inhibitory factor in U2OS and myeloid cells[1]. However, the potential of *PRKAB2*, the β2 subunit of the AMPK complex, located upstream of *CHD1L*, was not evaluated for its effect on HIV infection.

AMPK is a heterotrimeric protein complex that is comprised of an alpha subunit (*PRKAA1/PRKAA2*), a beta subunit (*PRKAB1/PRKAB2*), and a gamma subunit (*PRKAG1/PRKAG2/PRKAG3*). This complex acts an essential regulator of cellular metabolism and during viral infection, functions as an inhibitory factor by reducing cholesterol, fatty acid, and protein synthesis of infected cells[264–266]. AMPK has also been shown to repress pro-inflammatory signaling[267,268], activate anti-inflammatory signaling in macrophages[269], and has been shown to influence HIV pathogenesis[207]. Therefore, we hypothesize that *PRKAB2* may directly or, through downstream signaling pathways, influence HIV infection.

The advent of large-scale knockdown and knockout screens have enabled researchers to screen thousands of genes as HIV factors, but have not previously implicated *CHD1L*, *PRKAB2*, or *FMO5*[228,270–272]. The lack of association suggests that these genes may not be HIV factors, that associated signaling pathways are not present in the investigated cell types (HeLa, Jurkat T, or CEM-GXRCas9 cells), or specific physiological conditions are required for their function. In contrast, analysis of closely related genes to *PRKAB2,* such as *PRKAA1*, the dominant catalytic AMPKα subunit in myeloid cells[273], revealed its function as both an HIV HDF[228] and an HIV inhibitory factor [274]. These findings highlight that changes to specific AMPK subunits can affect HIV

infection and supports investigating whether genetically regulated changes to *PRKAB2* expression are associated with HIV infection.

This study seeks to determine whether functional HIV spVL associated variants identified by the HIV spVL GWAS[1] impact *PRKAB2* expression and downstream signaling pathways relevant to HIV infection. We leverage PrediXcan whole blood gene expression models from African Americans in the Genes-environments and Admixture in Latino Americans (GALA II)/Study of African Americans, Asthma, Genes, and Environments (SAGE) cohort to assess the association between HIV spVL associated variant genotypes and *PRKAB2* expression. Using RNA-sequencing data from wild-type and CRISPR-Cas9 knockout *PRKAB2*$^{-/-}$ immunopluripotent stem cells (iPSCs), we assess whether *PRKAB2* affects *CHD1L* expression, downstream signalling pathways, and known HIV factors. The results of this study implicate *PRKAB2* as a potential HIV factor by identifying essential immune pathways altered by *PRKAB2* loss-of-function, providing additional insight into how the role of how the chromosome 1 region affects HIV infection.

Methods

Cohorts used in this study

In this study we leverage publicly available sequencing data from the Immune Variation (ImmVar) project, GTEx database, GALA II/SAGE study, Ziegler et al 2020, Kachuri et al 2023, and McLaren et al 2023 (Appendix 1) to assess the relationship between HIV spVL associated variants and *PRKAB2* expression. One of the major challenges in determining the effect of HIV spVL associated variants on gene expression is the low minor allele frequency of candidate causal variants, rs59784663-A-G (MAF = 0.048), rs72999655-A-G (MAF = 0.049), rs73004025-C-T (MAF = 0.047), rs7525622-G-A, (MAF =0.15), which require thousands of samples to identify enough individuals with homozygous alternate alleles for functional testing[232,275]. Therefore, we leveraged data from large-scale sequencing studies to obtain sufficient sample sizes of heterozygous individuals to assess differences in genetically regulated *PRKAB2* expression. A summary of cohorts and data types accessed in this meta-analysis are available in Table 5.

Table 5. Summary of studies used in this meta-analysis.

Study Cohort	Tissue	Sample Number	Data Type	Accession
Genotype-Tissue Expression Database V8	Whole blood	N= 756 white, African American, Asian, or unknown	RNA-Sequencing	www.gtexportal.org
ImmVar project	N/A	N=110 African Americans	Genotypes	phs000815.v2.p1
	Monocytes	N=110 African Americans	Microarray Expression	GSE56033
	Naïve CD4+ T cells	N=110 African Americans	Microarray Expression	GSE56034
Ziegler et al., 2020	Human iPSCs	N=12	RNA-sequencing	GSE144043
McLaren et al., 2023	N/A	N=3,886 African and African Americans	Genotypes	N/A
Kachuri et al., 2023	Whole Blood	N=757 African Americans	PrediXcan Model	N/A

Genotype data and variant imputation

Genotype data for 112 African American individuals from the ImmVar project was obtained from the database of Genotypes and Phenotypes (dbGaP; phs000815.v2.p1). Chromosome 1 variants were imputed for individuals from the ImmVar project using the 1000 Genomes reference panel with SHAPEIT[276] and IMPUTE2 [277]. Variants with an imputation score of < 0.9 and Hardy-Weinberg equilibrium threshold of $p < 1\times10^{-5}$ were removed. Genotypes were assessed at rs59784663-A-G, rs72999655-A-G, rs7525622-G-A, and rs7300425-C-T. One individual was removed from subsequent analysis due to an incomplete genotype call at rs59784663-A-G.

Gene expression databases

Gene expression profiles for individuals from the ImmVar project were evaluated to determine whether HIV spVL associated variants influence gene expression. Gene expression from naïve CD4+ T cells and monocytes, adjusted for batch effects, age, gender, and technical artifacts using principal components (PCs), were obtained from the Gene Expression Omnibus (GEO) (GSE56035). In brief, expression was profiled using Affymetrix GeneChip Human Gene ST 1.0 microarrays with background correction and normalization using the Robust Multichip Average method in Affymetrix Powertools as previously described[278]. Sample, probe, and expression quality control, as well as the adjustment for non-genetic factors, including the regression of the top 14 PCs in monocytes and top 12 PCs in naïve CD4+ T cells, are detailed in Raj et al., 2014. In monocytes and naïve CD4+ T cells, the residuals are used to compare gene expression following correction for non-genetic factors.

Gene expression, in transcripts per kilo base million, were obtained from the GTEx database v8 in whole blood tissues for all available individuals (N=756)[279]. The GTEx database v8 (N=948) includes individuals of diverse ancestry, with 84.6%, 12.9%, 1.3%, and 1.1% of individuals identifying as white, African American, Asian, or unknown, although the specific ancestries for the 756 individuals were not available for this analysis. In this dataset, whole blood was defined as gathered from the femoral vein, subclavian vein, heart, and other sites (UBERON:0013756).

Raw RNA-Seq reads and normalized RNA-Seq read counts as transcripts per million were available for wild-type and *PRKAB2*$^{-/-}$ iPSCs from the National Center for Biotechnology Information (NCBI) GEO database under the accession (GSE144043). In brief, NCBI generates RNA-seq count data using HISAT2[280] to align sequences to GCA_000001405.15. Read counts were calculated using featureCounts for runs that pass a 50% alignment threshold.

Correlation of CHD1L and PRKAB2 gene expression

Previous studies have suggested that HIV spVL associated variants in the chromosome 1 region regulate *CHD1L* expression in a cell-type-specific manner[1]. Therefore, we sought to investigate *PRKAB2* regulation across multiple immune cell types as well. Gene expression data was compared for whole blood samples from GTEx as transcripts per million (TPM) or monocytes and naïve CD4+ T cells from the ImmVar project as residuals. For each dataset, a two-sided Pearson correlation was used to determine correlation between *CHD1L* and *PRKAB2*, with $p < 0.05$ defined as statistically significant. A linear regression model, along with a 95% confidence interval, was to visualize the magnitude and direction of correlation between *PRKAB2* and *CHD1L* expression. Statistical analysis was performed using R (v4.1.2).

Imputation of CHD1L expression from whole blood models

Genomic data from 3,886 individuals of African ancestry were obtained from ten independent GWAS under the ICGH (N=3,886), as previously described (Appendix 1)[1]. Predictive gene expression models trained on African whole blood eQTLs from the GALA II/SAGE cohorts were applied to the 3,886 individual-level genotypes using PrediXcan (Appendix 1)[281]. Imputed *CHD1L* expression was compared amongst individuals with allelic combinations of functional HIV spVL associated variants: rs72999655-A-G, rs7525622-G-A, and rs73004025-C-T from the ICGH cohort. Statistical significance was determined using a Wilcoxon Rank Sum Test at $p < 0.05$.

RNA-sequencing and pathway analysis

Given that GWAS variants typically exert a small effect on gene expression changes, we sought to determine whether *PRKAB2* was associated with immune signaling pathways using expression data from *PRKAB2* knockout cells to prioritize subsequent association

testing. Differential gene expression analysis was performed using NCBI generated RNA-Seq counts on wild-type iPSCs (N=4) and *PRKAB2*$^{-/-}$ clone two (N=4) from the GEO database (GSE144043). Lowly expressed genes were filtered based on a minimum count threshold of at least 10 in a given sample. Differential expression was performed using DESeq2 using FDR < 0.05, $\alpha = 0.05$ as statistically significant. Visualization and statistical analysis were performed using R (v4.1.2).

The Database for Annotation, Visualization and Integrated Discovery (DAVID; version v2023q4) was used to perform pathway analysis and functional annotation of DEGs[282,283]. Functional enrichment was classified for Gene Ontology, HIV interaction, Reactome, and KEGG pathway terms using medium classification stringency and an enrichment score of greater than 2. Individual annotations were significant with an FDR cutoff threshold of 0.05. HIV interactions were defined through any overlap of genes with the HIV interaction category.

Results

PRKAB2 and CHD1L exhibit cell-type-specific correlation patterns

Regulatory variants can influence the expression of multiple genes which collectively contribute to complex phenotypes[186]. While previous studies have focused on genetically regulated *CHD1L* expression[1], we hypothesized that *PRKAB2* may be co-regulated due to its proximity to *CHD1L*. To address this, we analyzed the correlation between *PRKAB2* and *CHD1L* expression in immune cells from the GTEx and ImmVar databases. Gene expression from the GTEx datasets showed a strong positive correlation between *PRKAB2* and *CHD1L* expression from whole blood tissues ($r = 0.7$, $p < 3.1\times10^{-110}$) (Figure 11A). However, we observed no significant correlation in naïve CD4+ T cells ($r = 0.17$, $p < 0.081$) (Figure 11B) but a significant negative correlation in monocytes ($r = -0.29$, $p < 0.002$) (Figure 11C). These correlations suggest that *CHD1L* and *PRKAB2* likely share regulatory features in immune cells but the magnitude and direction of effect may be cell-type-specific.

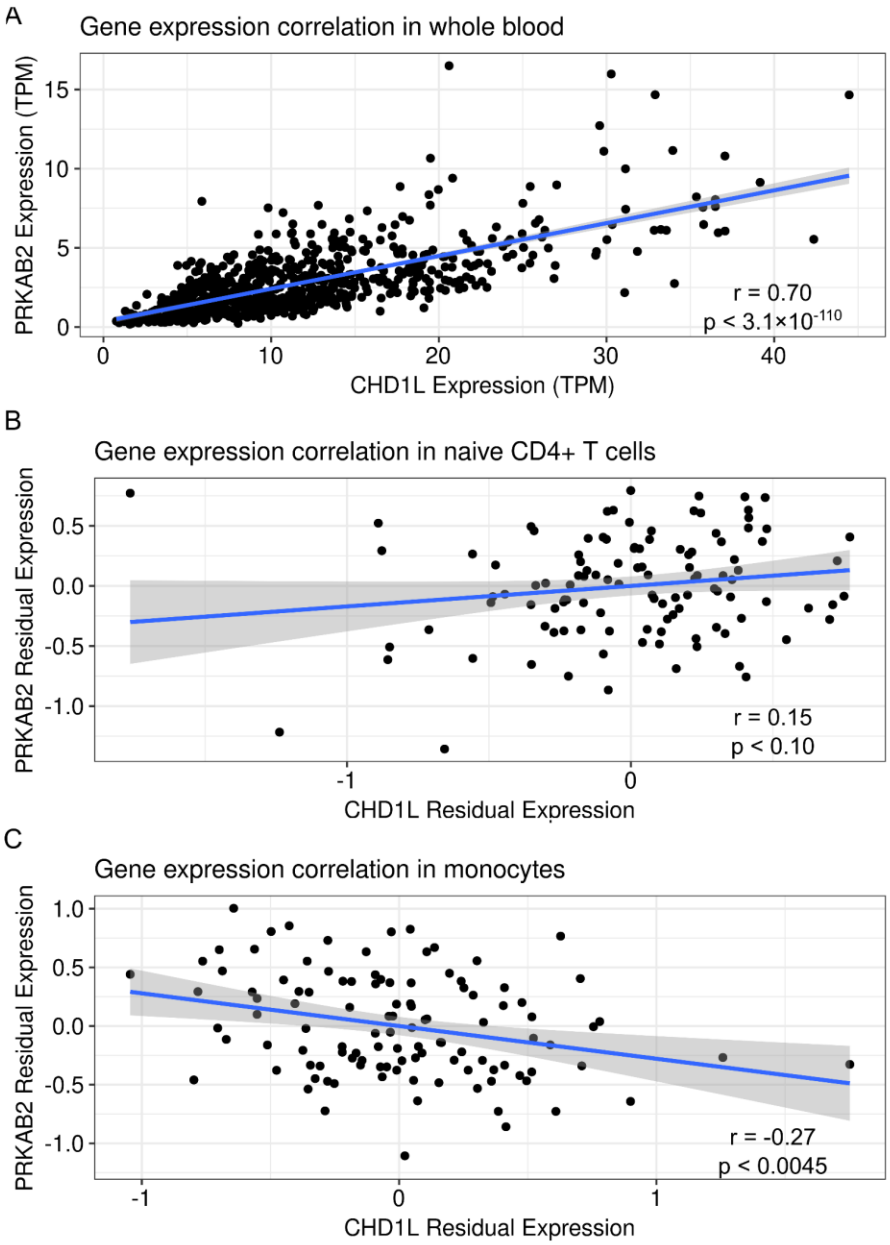


Figure 11. CHD1L and PRKAB2 expression exhibit cell-types-specific correlation patterns across immune cell subsets.

(A) Transcripts per million for *CHD1L* and *PRKAB2* expression from whole blood in the Genotype-Tissue Expression database (N = 756). (B) Residuals of *PRKAB2* and *CHD1L* expression after controlling for batch, sex, and technical variation in naïve CD4+ T cells. (C) Residuals of *PRKAB2* and *CHD1L* expression after controlling for batch, sex, and technical variation in monocytes. The blue line represents a linear regression of the values surrounded by standard error.

HIV spVL decreasing variants in the chromosome 1 region are associated with decreased PRKAB2 expression

We next assessed the impact of HIV spVL associated variants on *PRKAB2* expression to determine whether regulatory features in the chromosome 1 region influence both genes. In monocytes, individuals heterozygous for rs59784663-A-G (N=6), had a significant reduction of *PRKAB2* expression compared to homozygous reference individuals (N=103; $p < 0.036$) (Figure 12A) but no significant effect in naïve CD4+ T cells ($p < 0.56$) (Figure 12B).

We then sought to determine the effect of candidate causal variants rs72999655-A-G, rs7525622-G-A, and rs73004025-C-T on *PRKAB2* expression, and to have sufficient sample size for analysis, imputed *PRKAB2* expression from whole blood models using 3,886 individual-level genotypes from the ICGH cohort[1]. Individuals heterozygous for rs72999655-A-G, rs7525622-G-A, and rs73004025-C-T (N=312) exhibited lower *PRKAB2* expression than homozygous reference (N=3159; $p < 1.65\times10^{-26}$) and homozygous alternate individuals (N = 7) exhibited much lower *PRKAB2* expression than homozygous reference ($p < 0.00031$) (Figure 13). Overall, these results show that HIV spVL associated variants likely influence *PRKAB2* expression and highlights the need to investigate potential downstream consequences of HIV infection.

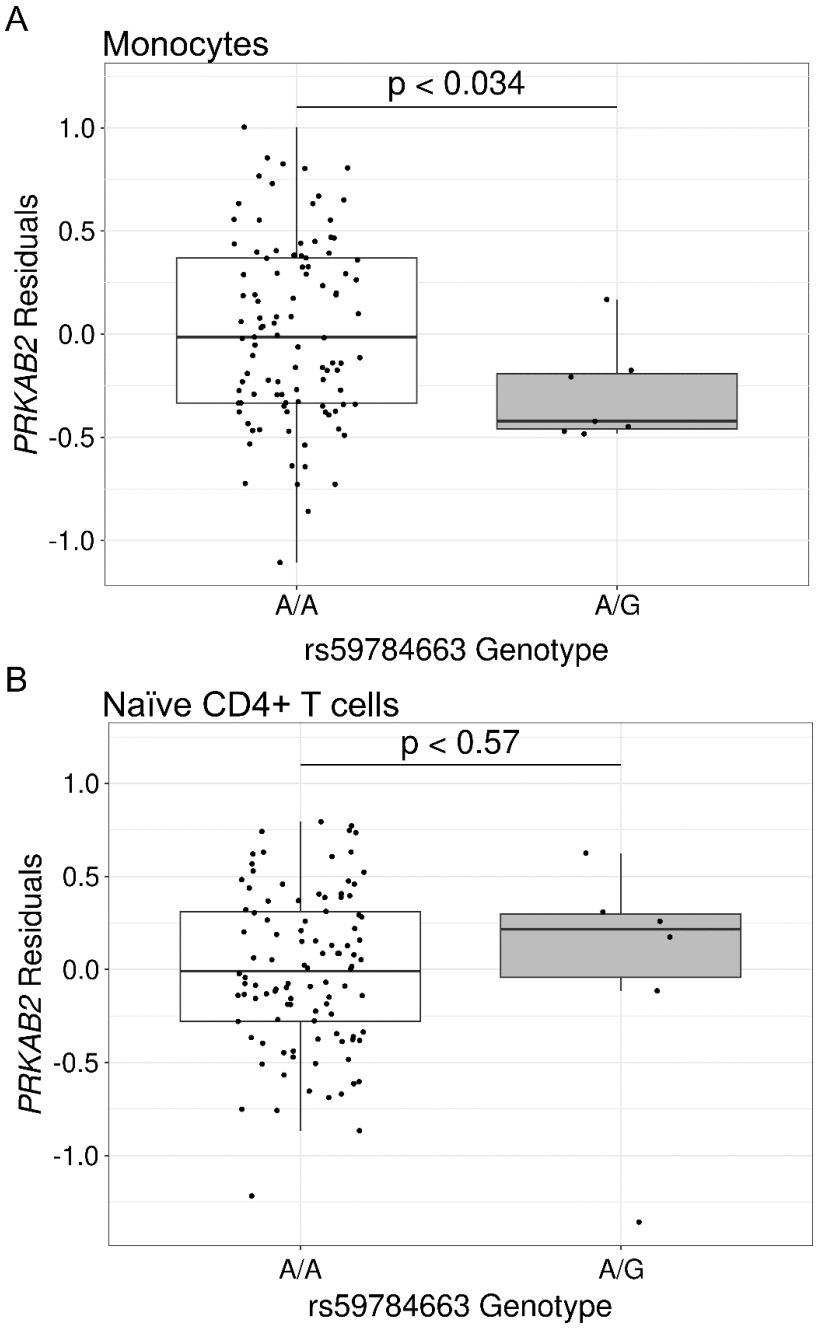


Figure 12. The rs59784663-G allele is associated with reduced *PRKAB2* expression in monocytes. Gene expression following control for non-genetic factors in African populations for A) monocytes or B) naïve CD4+ T cells. Significance was determined using a two-sided Wilcoxon Rank Sum Test.

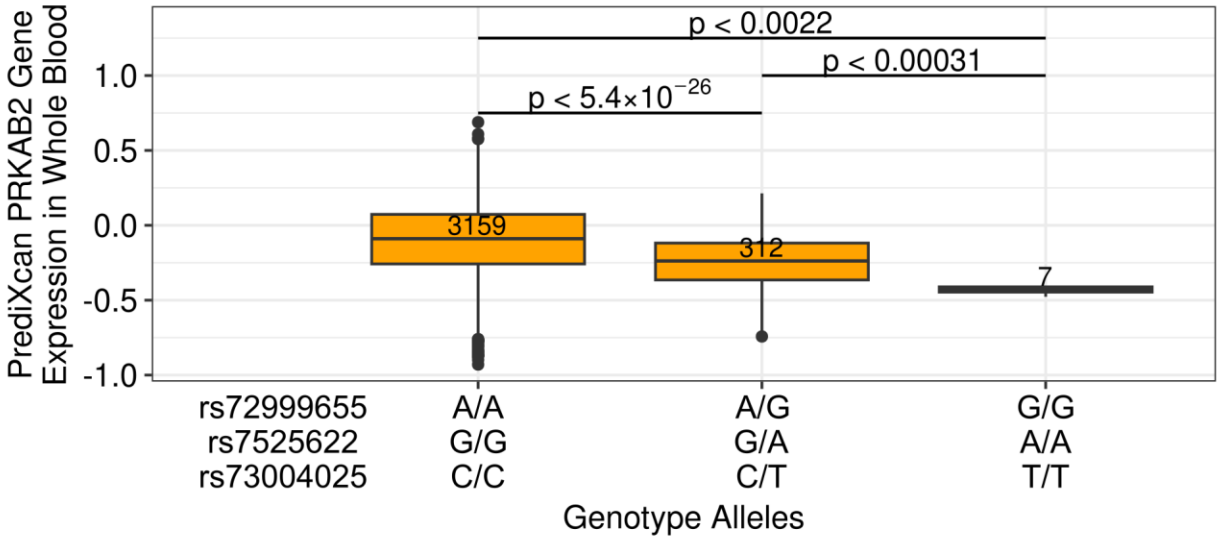


Figure 13. Alleles linked to lower HIV spVL is associated with reduced *PRKAB2* expression in whole blood. *PRKAB2* expression was imputed using whole blood models from the Genes-Environments and Admixture in Latino Asthmatics and the Study of African Americans, Asthma, Genes, and Environments cohorts. Allelic combinations are shown for individuals homozygous reference, heterozygous, or homozygous alternate at all variants.

PRKAB2$^{-/-}$ iPSC models suggest PRKAB2 does not strongly regulate CHD1L expression

Further investigation into the relationship between *PRKAB2* and *CHD1L* was performed using GEO gene expression data from *PRKAB2*$^{-/-}$ iPSC models (GSE144043; Appendix 1)[209]. In brief, PRKAB2-knockout iPSCs had been generated using two CRISPR-Cas9 guide RNAs in human iPSCs (ASE-9202, Applied StemCell) to induce frameshift mutations in *PRKAB2*$^{-/-}$ and protein loss-of-function was confirmed via Western blotting[209]. There were two clones PRKAB2$^{-/-}$ iPSC Clone 1 (C1) and PRKAB2$^{-/-}$ iPSC Clone 2 (C2) that had available gene expression data. Using this publicly available data, we observed that *PRKAB2* expression was significant reduced compared to wild-type controls ($p < 2.2\times10^{-5}$ and $p < 9.6\times10^{-7}$ for clones C1 and C2, respectively), while a larger reduction was observed between C2 compared to C1 ($p < 8.1\times10^{-5}$) (Figure 14A).

However, despite reduced *PRKAB2* expression in both clones, *CHD1L* expression was inconsistent. There was no significant difference in *CHD1L* expression between wild-type and C1 ($p < 0.068$) but a significant reduction of *CHD1L* expression observed in C2 compared to wild-type ($p < 0.0012$) (Figure 14B). The variability between clones may reflect differences in gRNA efficacy or off-target effects during generation of *PRKAB2*$^{-/-}$ knockouts. While the data is unclear in the potential of AMPKβ2-mediated *CHD1L* expression, AMPK is known to activate DNA repair pathways[284], suggesting that AMPKβ2 may alter expression of *CHD1L* only under specific physiological conditions.

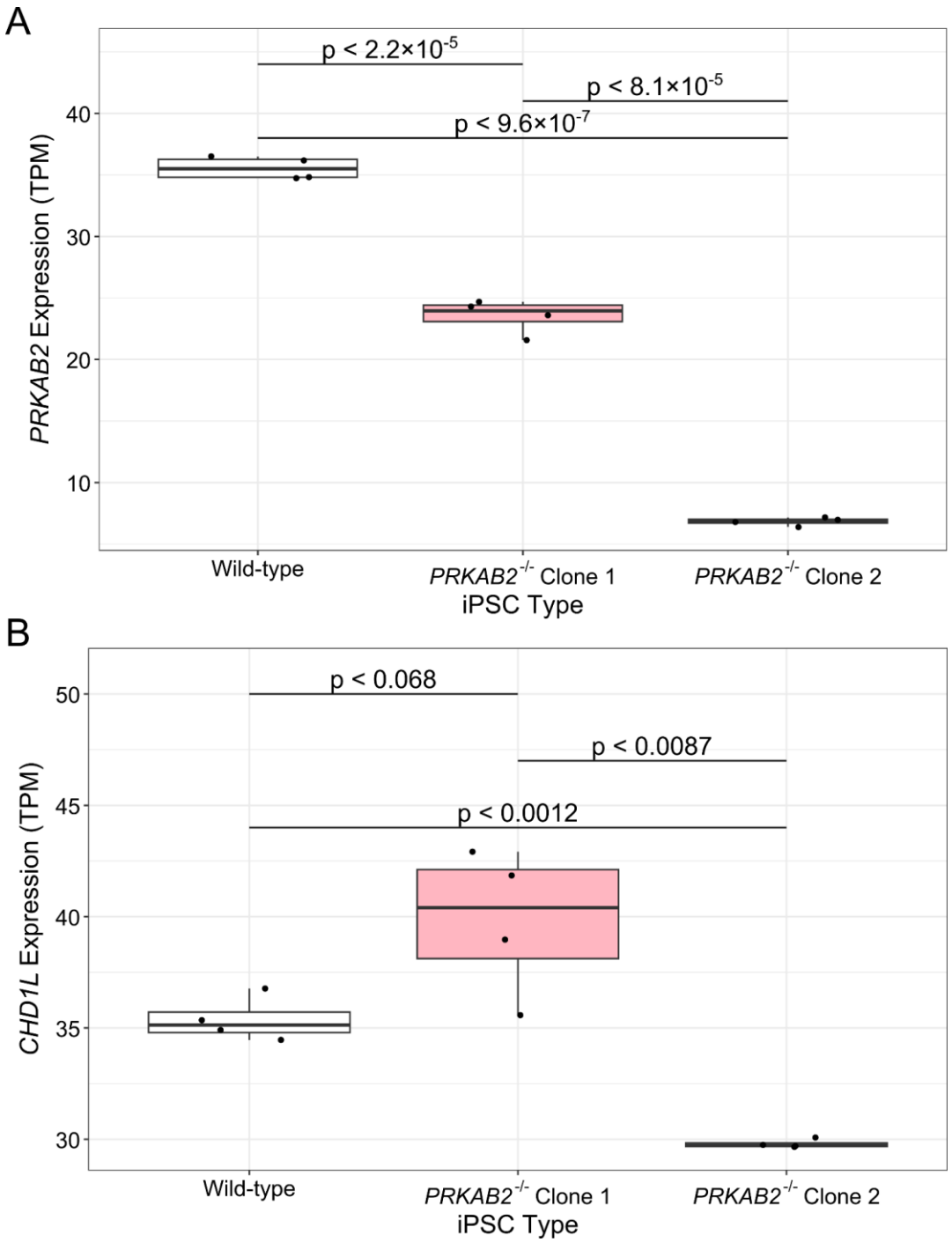


Figure 14. Gene expression differences in PRKAB2$^{-/-}$ loss-of-function iPSCs. Gene expression of A) *PRKAB2* and B) *CHD1L* in wild-type iPSCs, *PRKAB2* knockout clone 1 iPSCs, and *PRKAB2* knockout clone 2 iPSCs.

Chromosome 1 variants are associated with downstream gene expression differences

One of the major challenges in identifying changes to *PRKAB2* signaling pathways is that the functional AMPKβ2 complex regulates expression of downstream genes through phosphorylation, which is not detectable through RNA-sequencing analyses. However, AMPK regulates the activation of transcription factors[285], indicating that signaling pathways downstream of *PRKAB2* may be differentially expressed. Therefore, we leveraged gene expression profiles from the wild-type and PRKAB2$^{-/-}$ iPSCs (GSE144043) to determine potential pathways influence by *PRKAB2*.

We first sought to determine whether *PRKAB2*$^{-/-}$ C1 and C2 exhibited different gene expression profiles given the differences in *PRKAB2* and expression observed previously. We observed 58 DEGs when comparing expression between C1 and C2 suggesting that differences in gRNA efficiency or off-target effects may have influenced gene expression profiles. Additionally, when comparing DEGs in C2 and wild-type iPSCs, the most significant gene was *PRKAB2* (log_2FC = -2.4, FDR < 2.2×10^{-308}), but when comparing C1 and wild-type iPSCs, *PRKAB2* ranked 44th (log_2FC = -0.70, FDR < 7.9×10^{-59}). Therefore, we chose to analyze gene expression differences between wild-type and *PRKAB2*$^{-/-}$ C2, given that C2 exhibited a larger reduction in *PRKAB2* expression compared to C1.

There were 170 down-regulated and 145 up-regulated genes in *PRKAB2*$^{-/-}$ C2 compared to wild-type iPSCs (FDR < 0.05, absolute log2 fold change ≥ 1, Figure 15A). These genes were significantly enriched for functions related to cytokine activity (17 genes, FDR < 9.5×10^{-8}) and growth factor activity (10 genes, FDR < 0.012), as well as KEGG pathways regulating stem cell pluripotency (9 genes, FDR < 0.028; Table 6). Clustering analysis revealed two major groups: one associated with cytokine activity (enrichment score 4.24) and the other with male genitalia development (enrichment score 2.13; Figure 15B). Given the relevance of cytokine activity to HIV infection, we prioritized this group for further analysis. Among the differentially expressed genes in the cytokine activity pathway, *TNFSF9* (log_2FC = -1.5, FDR < 1.8×10^{-32}), *TGFB2* (log_2FC = -1.0, FDR < 0.000071), and *SCG2* (log_2FC = 1.003, FDR < 0.0069), stood out as known HIV

interactors using DAVID annotation. Collectively, these results show that individuals with HIV spVL associated variants exhibit reduced *PRKAB2* expression and reduced *PRKAB2* expression alters cytokine regulation, suggesting that the chromosome 1 region may regulate HIV spVL through these immune pathways.

Table 6. DAVID enrichment of differentially expressed genes in Gene Ontology and KEGG categories

Category	Term	Count	%	FDR
GOTERM MF DIRECT	cytokine activity	17	5.56	9.45×10^{-8}
GOTERM CC DIRECT	extracellular space	42	13.73	0.0093
GOTERM MF DIRECT	growth factor activity	10	3.27	0.012
KEGG PATHWAY	Signaling pathways regulating pluripotency of stem cells	9	2.941	0.029
GOTERM BP DIRECT	cell fate commitment	7	2.28	0.038
GOTERM BP DIRECT	male genitalia development	5	1.63	0.038

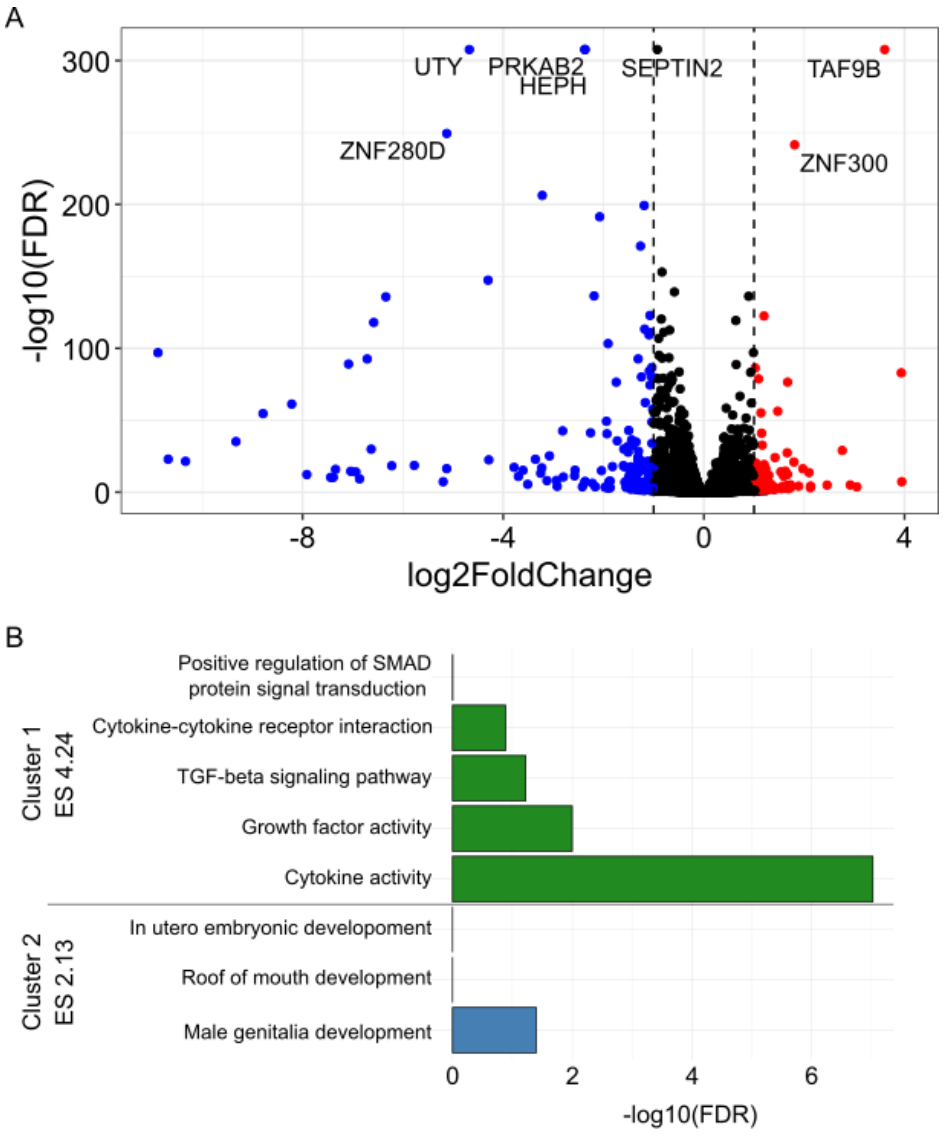


Figure 15. Differentially expressed genes between PRKAB2$^{-/-}$ clone two and wild-type iPSCs.

A) (Blue) Significant upregulated genes (FDR < 0.05, $\log_2$ fold change > 1) in *PRKAB2* knockout clone two cells. (Red) Significant downregulated genes (FDR < 0.05, $\log_2$ fold change < 1) in *PRKAB2* knockout clone two cells. B) Enrichment clustering of differentially expressed genes revealed two distinct clusters with an enrichment score of greater than 2.

Discussion

During the analysis of a recent GWAS of HIV spVL, *CHD1L* was identified to be the candidate causal gene and implicated as a novel HIV inhibitory factor[1]. However, the question remains as to why *CHD1L* was not detected in previous screens of HIV inhibitory factors[274]. While cell-type-specific effects may explain some variation, it is also likely that multiple chromosome 1 genes may collectively contribute to control HIV infection by affecting downstream signaling pathways[151]. Therefore, we chose to investigate the role of *PRKAB2*, part of the AMPKβ2 complex, as AMPK has been previously associated with HIV infection[228,264].

Our initial investigation identified that *PRKAB2* and *CHD1L* expression profiles were positively correlated in whole blood and negatively correlated in monocytes, suggesting these genes may share cell-type-specific regulatory features. This is also supported by observations that altered DNA methylation in the chromosome 1 region is associated with *CHD1L* and *PRKAB2* expression changes[286]. Individuals with protective HIV spVL associated variants exhibited reduced *PRKAB2* expression in monocytes and whole blood, but no differences were observed in naïve CD4+ T cells. While the specific variants that drive gene expression changes have not been experimentally validated, these results suggest that HIV spVL associated variants in the chromosome 1 region downregulate *PRKAB2* expression.

Previously, *CHD1L* was shown to inhibit HIV replication in IPSC-derived macrophages from KOLF2 fibroblasts[192], so we sought to leverage gene expression from iPSCs derived from fibroblasts to test whether *PRKAB2* may influence *CHD1L* expression, as AMPK regulates DNA repair pathways[284,287]. We observed that *PRKAB2* loss-of-function had no effect on *CHD1L* expression in *PRKAB2*$^{-/-}$ C1 but significantly downregulated in *PRKAB2*$^{-/-}$ C2. While both C1 and C2 were identified to have complete protein ablation[209], it is unclear whether the observed differences were driven by off-target effects, environmental, or physiological conditions. Therefore, we are unable to conclude whether HIV spVL associated variants independently influence *PRKAB2* and *CHD1L*, or if gene-gene interactions may also influence expression, but these results indicate that

further studies across different cell types and physiological conditions are necessary to validate these observations.

While our previous experiments showed that individuals with protective HIV spVL associated variants have reduced *PRKAB2* expression, the overall impact on downstream signaling pathways is unknown. This is in part because AMPKβ1 and AMPKβ2 exhibit different transcriptome profiles, but specific pathways altered by AMPKβ2 are unclear[209]. To address this, we compared the transcriptome profiles from *PRKAB2*$^{-/-}$ and wild-type iPSCs, which identified the most significant pathway altered by *PRKAB2* loss-of-function was regulation of cytokine signaling. Cytokine and chemokine levels, prior to HIV infection, can be strong predictors of disease progression[288,289] and increased AMPK activity is associated with suppression of pro-inflammatory signaling in macrophages[273]. Therefore, this provides evidence that HIV spVL associated variants in the chromosome 1 region likely influence HIV spVL through changes to cytokine signaling and immune activation pathways.

Three genes regulating cytokine activity were identified by the DAVID analysis to interact with HIV proteins: *TNFSF9, TGFB2*, and *SCG2*. The gene *SCG2*, upregulated in *PRKAB2*$^{-/-}$*,* plays an important role in regulating the immune response through the assembly and function of the major histocompatibility complex, PI3K-Akt, TGF-β, and JAK-STAT pathways[290,291]. The gene *TNFSF9*, also known as 4-1BBL and downregulated in *PRKAB2*$^{-/-}$, produces a pro-inflammatory immune response and a key component of cytotoxic CD8+ T cells, macrophages, and monocytes activation [292–294]. The gene *TGFB2,* downregulated in *PRKAB2*$^{-/-}$*,* has immunosuppressive functions and regulates activation of dendritic and T cells[295]. Collectively, these genes implicate additional pathways downstream of AMPK signaling associated with HIV pathogenesis and identifies targets for future research to elucidate mechanisms by which HIV spVL associated variants influence HIV spVL.

Overall, this study highlights that HIV spVL associated variants influence *PRKAB2* expression and raises important questions about whether this effect is synergistic, antagonistic, or independent of *CHD1L* regulation. Furthermore, as AMPK is a regulator of immune function, future studies should investigate whether chromosome 1 genes

differentially affect HIV pathogenesis during acute and chronic infection, providing insights into their potential as therapeutic targets.

Chapter 4: Changes to Interferon Stimulating Genes Associates with Decreased HIV spVL in African Populations

Introduction

The study of HIV host genetics in African populations identified a region on chromosome 1 significantly associated with HIV infection[1]. HIV spVL associated GWAS variants surrounding rs59784663-A-G have been shown to impact regulatory elements associated with *CHD1L*[275] and *PRKAB2* expression (Chapter 3). However, the cumulative effect of these variants on *CHD1L* expression has been mostly inferred from predictive models such as eQTL analyses and imputed gene expression. Previous analyses have shown that *CHD1L* acts as an inhibitory factor in myeloid cells, but not T cell lines, suggesting that genetic regulation may also be cell-type-specific. Therefore, it is crucial to characterize the relationship between these variants and *ex vivo* gene expression in HIV-relevant cell types to determine their association.

CHD1L is a DNA repair gene whose function has been predominately characterized in cancer, where it plays key roles in apoptosis, DNA repair, chromatin remodeling, and gene regulation[222,223,296]. There is substantial overlap with *CHD1L* pathways and HIV-associated pathways such as base excision repair pathways in HIV integration[297], regulation of inflammatory pathways[298], and cell-cycle regulation[228,299]. To fully understand *CHD1L*'s impact on HIV pathogenesis, it is important to investigate whether genes downstream of *CHD1L* are differentially expressed in the attempt to identify a mechanism of action.

Additionally, *PRKAB2,* the regulatory subunit of the AMPKβ2 complex, has been previously suggested to impact *CHD1L* expression and cytokine activity (Chapter 3). Both *CHD1L* and *PRKAB2* have the potential to influence downstream signaling pathways that regulate HIV replication and spVL levels. Here, we characterized the relationship between HIV spVL associated variants, *CHD1L*, and *PRKAB2* using qPCR gene expression profiles in unstimulated CD4+ T cells, CD3/CD28 stimulated CD4+ T cells, and monocytes from PLWH. Downstream signaling pathways were assessed using gene expression profiles in monocytes and CD4+ T cells from healthy individuals based on

genotype. The results of this study provide further insight into the impact of HIV spVL variants on gene expression, and suggests a vital role for *CHD1L* in interferon signaling.

Methods

The Majengo Observational Cohort Study

Plasma and cryopreserved PBMCs were available for use from African PLWH from the Majengo Observational Cohort Study (MOCS; Appendix 1). Informed consent was obtained from all participants enrolled in the MOCS cohort. The Ethics Committee of the University of Manitoba, the Ethics and Research Committee of Kenyatta National Hospital, and the Regional Ethical Review Board of Karolinska Institute approved the use ongoing genetic and immunological testing of samples through the MOCS cohort.

Plasma DNA extraction

Plasma samples were previously shipped from Nairobi, Kenya to the University of Manitoba and stored at the JC Wilt Infectious Diseases Research Center at -80ºC. Plasma samples were thawed briefly at 37ºC and centrifuged at 12,000×g to remove cellular debris and particulate matter. 200uL of supernatant was then incubated with 20uL of Proteinase K at 37ºC for 10 minutes. DNA extraction was then performed using Nucleospin® cfDNA XS kit (Macherey-Nagel). Samples were eluted in 30uL of elution buffer to increase DNA recovery and incubated at 90ºC for 8 minutes to remove residual ethanol. DNA concentration was measured using 4uL of DNA in Qubit® DNA HS kit with a Qubit 4 Flourometer and DNA purity (230/260 and 260/280) was measured a Nanodrop.

Validating Genotype of rs59784663-A-G

A subset of individuals (N=119) from the MOCS cohort had available cryopreserved plasma and PBMCs available for use. Of these, 14 individuals also had imputed genotypes available for rs59784663-A-G, rs72999655-A-G, rs7525622-G-A, and rs73004025-C-T. Genotypes for rs59784663-A-G were validated using TaqPath™ ProAMP™ Master Mix (Thermofisher Scientific) and a 20X TaqMan™ assay (TGAGAAAGTGATTTGACTGTAATGA[A/G]TAGGTAACAAATCAATTTGTAAATC) according to manufacturer's instructions (Thermofisher Scientific). Samples were analyzed using MicroAMP™ Fast Optical 96-well reaction plates (Applied Biosystems) where cell-free DNA, non-template control (NTC), or positive control DNA was added

(composition in Table 7). Plates were loaded onto a Viia7 Real-Time PCR System (Applied Biosystems) and genotype performed using *fast* run settings (Table 8).

Table 7. Reagents for qPCR genotyping experiments

Component	Test Volumes (µL)	NTC Volumes (µL)	Positive Control (µL)
TaqPath ProAmp Master Mix	5.0	5.0	5.0
TaqMan SNP Genotyping Assay	0.5	0.5	0.5
Genomic DNA, NTC, or Positive Control	10	0	3
Nuclease-Free Water	0	10	7

Table 8. Run settings for qPCR genotyping experiments on Viia7 system

Step	Temperature (°C)	Duration (seconds)	Cycles
Pre-read	60	30	1
Denature/Enzyme Activation	95	300	1
Denature	95	5	40
Anneal/Extend	60	30	
Post-Read	60	30	

Fluorescent associated cell sorting of peripheral blood mononuclear cells

Peripheral blood mononuclear cell samples were previously shipped from Nairobi, Kenya to the University of Manitoba and stored at the JC Wilt Infectious Diseases Research Center at -120ºC. Cells were briefly thawed in a 37ºC water bath, transferred to 10mL of RPMI-1640 (Thermofisher Scientific) supplemented with 10% FBS (R10; Thermofisher Scientific), and spun at 120×g for 7 minutes. The pellet was resuspended in 900uL of R10 media and cell count determined using the Muse™ Cell & Viability Kit (Millipore Sigma). Cells were incubated at 2×10^6 cells/mL in R10 media at 37ºC, 5% CO_2 overnight.

The next day, cells were washed with FACS buffer consisting of PBS, 2% fetal bovine serum (Thermofisher Scientific), 1mM EDTA, and 100ug/mL DNase (Qiagen), then centrifuged for 5 minutes at 1600rpm. Cells were incubated with a viability stain and FC block for 15 minutes at room temperature while protected from light, at a ratio of 25uL/1×10^7 cells (Table 9). Cells were then incubated with the antibody master mix for 30 minutes at 4ºC while protected from light, at a ratio of 75uL/1×10^7 cells (Table 10). Excess antibody was removed through a FACS buffer wash and centrifuged at 1600rpm for 5 minutes. Cells were resuspended in 400uL of FACS buffer and kept at 4ºC until sorting.

Antibody compensation beads were prepared using BD® Anti-Mouse CompBeads (BD Biosciences) at a ratio of 1 drop per 300uL PBS. Cells were then incubated in 1uL of corresponding antibody at 4ºC for 20 minutes while protected from light. Following incubation, 200uL PBS was added to each tube before acquisition on sorter. For live/dead compensation, one drop of ArC™ Amine Reactive Compensation Beads (Thermofisher Scientific) were added to 1uL of Live/Dead Aqua (Life Technologies) and incubated for 20 minutes at room temperature while protected from light. Negative live/dead compensation was prepared using one drop of negative control ArC™ Amine Reactive Compensation Beads (Thermofisher Scientific). Following incubation, 300uL PBS was added and samples were kept at 4ºC until sorting. Fluorescent associated cell sorting was performed using an Aria III cytometer (BD Biosciences) and analyzed using FlowJo version 10.6.1 (BD Life Sciences).

When sorting, doublets were excluded based on forward versus side scatter distributions. Monocytes were gated based on $CD3^-$, $HLA\text{-}DR^+$, $CD14^+$, $CD16^{++}$ while CD4+ T cells were gated based on $CD3^-$ and $CD4^+$.

Table 9. Viability stain master mix reagents

Reagent	Source	Volume (uL) / 1×10^7 Cells
Live/Dead Aqua	Life Tech L34957	1
Human FC Block	BD 564220	0.5
PBS	-	23.5

Table 10. Antibody master mix reagents

Marker	Flourochrome	Source	Isotype	Volume (uL) / 1×10^7 Cells
CD16	BV605	BD 563172	Mouse	4
CD4	BB515	BD 564419	Mouse	1.25
CD14	PerCPCy5.5	BD562692	Mouse	2
CD3	APCR700	BD565119	Mouse	2
HLA-DR	APCFire750	Biolegend 307658	Mouse	2
FACS Buffer	-	-	-	63.75

CD4+ T cell stimulation

Immediately following cell-sorting, CD4+ T cells washed in RPMI-1640 (Thermofisher Scientific) and incubated at 2×10^6 cells/mL in R10 media at 37ºC, 5% CO_2 overnight. The following day, CD4+ T cells were separated into two equal fractions. One fraction was stimulated using CD3/CD28 magnetic Dynabeads (Thermofisher Scientific) at a 1:1 ratio of cells to resuspended Dynabeads as per manufacturer's instructions, then incubated at 37ºC, 5% CO_2 for 8h. The other fraction had an equivalent amount of R10 media added and incubated at 37ºC, 5% CO_2 for 8h.

RNA extraction

Unstimulated CD4+T cells, CD3/CD28 stimulated CD4+T cells, and monocytes were harvested using the RNA Micro Plus kit (Qiagen). Cells were pelleted by centrifuging at 300×*g* for 5 minutes, supernatant was removed and the cell pellet was lysed by adding 350uL of Buffer RLT Plus with 1% β-mercaptoethanol. CD4+ T cells were homogenized using a QIAshredder column (Qiagen) whereas monocytes, due to their low cell number ($< 1\times10^5$), were homogenized by vortexing for 1 minute. After each centrifugation step, a new collection tube was used to limit the carry-over of contaminants. RNA was collected using 18uL of RNase-free water which resulted in collection of 16uL RNA, accounting for 2uL of dead volume.

Quantification of RNA was performed using 4uL of RNA with the Qubit RNA HS kit (Thermofisher Scientific) and integrity was assessed using the RNA HS Tapestation (Agilient) with 2uL of RNA.

cDNA generation

Following RNA extraction, up to 85pg of RNA was used to generate complementary DNA (cDNA) was synthesized using the SuperScript IV VILO™ Master Mix with ezDNase enzyme system (Thermofisher Scientific), as per manufacturer's instructions. Positive control cDNA was generated using 100ng of T Cell Leukemia (Jurkat) Total RNA (Thermofisher Scientific) was used as a positive control. Jurkat cDNA was frozen into individual aliquots and stored at -20ºC prior to use.

Relative gene expression in CD4+ T cells and monocytes from PLWH

Expression of *CHD1L, PRKAB2, FMO5,* and *LINC00624* was evaluated using TaqMan™ Gene Expression Assays (Thermofisher Scientific) with Fast Advanced Master Mix (Thermofisher Scientific), as per manufacturer's instructions. Samples were prepared in the 96-well *fast* plate format, with each gene sequenced in triplicate, on a Quantstudio 12k Flex with *fast* cycling mode (Thermofisher Scientific).

Relative gene expression was determined as described by Vandesompele et al., 2002[300] and Hellemans et al., 2007[301]. Amplification efficiency of primers was assessed using Jurkat cDNA at dilution factors of 1:1, 1:5, 1:25, 1:125, 1:625, and 1:3125, and efficiency was determined by the formula below.

$$Primer\ Efficiency\ (\%) = \left(\frac{-1}{10^{Slope} - 1}\right) x\ 100\%$$

In brief, relative gene expression was determined by the equation below where *E* represents the amplification efficiency of the primers.

$$Relative\ Gene\ Expression = \frac{(E_{GOI})^{\Delta CT\ GOI}}{Geometric\ Mean[(E_{REF})^{\Delta CT\ REF}]}$$

The expression of *CHD1L*, *PRKAB2, FMO5, and LINC00624* was normalized to *SDHA* in CD4+ T cells and monocytes[302,303]. Inter-plate variation was assessed through the expression of *TBP*, *RPL13A,* and *SDHA,* in triplicate, from cDNA aliquots generated from Jurkat Total RNA (Thermofisher Scientific) on each plate.

Genotype and gene expression sample selection

Genotype data was obtained for 112 African American individuals from the ImmVar project from dbGAP (phs000815.v2.p1; Appendix 1). Chromosome 1 variants were imputed for individuals from the ImmVar project using the 1000 Genomes reference panel with SHAPEIT2[304] and IMPUTE2[305]. Variants with a score of < 0.9 and Hardy-Weinberg equilibrium threshold of $p < 1\times10^{-5}$ were removed. Genotypes were assessed at rs59784663-A-G, rs72999655-A-G, rs7525622-G-A, and rs7300425-C-T.

Raw Affymetrix GeneChip Human Gene ST 1.0 microarrays CEL files were obtained from the GEO database (http://www.ncbi.nlm.nih.gov/geo/) for monocytes (GSE56033) and naïve CD4+ T cells (GSE56034) from African Americans (N=112) in the

ImmVar project (Appendix 1). There were 4 individuals heterozygous for rs59784663-A-G, rs72999655-A-G, rs7525622-G-A, and rs7300425-C-T alleles in the naïve CD4+ T cell and monocyte gene expression data. Up to two controls were matched to each heterozygous individual based on the following criteria: had rs59784663-A-A, rs73004025-C-C, rs7525622-G-G, and rs72999655-A-A alleles, within the same batch, same sex, and close in age to heterozygous individuals. If the total relative log expression (RLE) of a sample deviated outside of one standard deviation, the next most appropriate control was used. The final number of samples in CD4+ T cells and monocytes were 4 heterozygous individuals and 7 homozygous reference individuals.

Quality control and processing of microarray data

Raw microarray intensities underwent quantile normalization and processed using RMA from the affy (v1.80.0) package[306]. Lowly expressed and background probes were removed from expression analysis (Appendix 7). Batch effects across samples were identified and adjusted for using *ComBat* from the *sva* (v3.52.0) package[307].

In naïve CD4+ T cells, there were 33,297 probesets, of which 3,654 probes were lowly expressed, 8863 probes did not map to a gene, 2061 probes mapped to multiple genes, and 9081 probes were removed for mapping to duplicate genes, resulting in 16,801 probes for testing. In monocytes, there were 33,297 probesets, of which 4888 probes were lowly expressed, 8084 probes did not map to a gene, 1994 probes mapped to multiple genes, and 8775 probes were removed for mapping to duplicate genes, resulting in 16,269 probes for testing.

Principal components analysis was used to determine the effectiveness of matched controls and batch effect removal at limiting the effect of age, sex, and batch on gene expression. Scree plots were used to visualize the number of PCs that explained 90% of total gene expression variance for monocytes and naïve CD4+ T cells. The PCs within the 90% set for monocytes (N=11) and naïve CD4+ T cells (N=11) were regressed against batch, sex, age, and genotype to determine significant associations against $p < 0.05$. Visualization was performed using ggplot2(v3.5.0) in R (v4.1.2).

Differentially expressed genes and pathway analysis

Transcriptome changes between individuals with alleles at rs59784663-A-G, rs72999655-A-G, rs7525622-G-A, and rs7300425-C-T were assessed using linear regression models in the LIMMA (v3.50.3) package in R (v4.1.2)[308,309]. In CD4+ T cells, the gene expression model was controlled for using age, sex, and technical artifacts (PC1, PC2, PC3). In monocytes, the model was controlled for using age, sex and technical artifacts as well (PC2, PC3), but PC1 was not included as it was significantly associated with genotype. To account for multiple comparisons, p-values were adjusted for using FDR. Differentially expressed genes were defined by FDR < 0.05 and an absolute log2 fold change >1.

Functional enrichment of DEGs was performed using the DAVID 2021 database to determine candidate pathways influenced by the chromosome 1 signal[282]. Differentially expressed genes were assessed for overlap of Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome pathways, and Gene Ontology (GO) biological processes, cellular components, and molecular functions. Significant enrichment was determined against FDR < 0.05.

Protein-protein interaction (PPI) was performed using the Search Tool for the Retrieval of Interacting Genes (STRING) interaction network tool (https://string-db.org/)[310]. The interaction score was set to medium confidence and default interaction sources were used to determine the protein-protein interaction network. All DEGs with FDR < 0.05 and an absolute $\log_2$ fold change ≥ 1 were used to create the PPI network.

Overrepresentation and gene set enrichment analysis

An overrepresentation analysis and gene set enrichment analysis (GSEA; https://www.gsea-msigdb.org/gsea/index.jsp) was conducted to identify significant enrichment of genes associated with HIV spVL-associated genotypes and gene expression profiles from HIV LTNPs. Gene ranks were calculated by multiplying the $\log_2$FC and $-\log_{10}$(p-value). Overrepresentation analysis was performed using a Fisher's Exact Test based on overlapping gene lists and GSEA was performed using the fgsea package (v1.34.2) using the adaptive multilevel splitting Monte Carlo approach.

Results

Expression of chromosome 1 genes in PLWH based on genotype

HIV spVL associated variants in the chromosome 1 region have been associated with increased *CHD1L* in monocytes[275] and reduced *PRKAB2* expression in iPSCs, but it is unclear whether gene expression is differentially regulated during HIV infection. Therefore, gene expression profiles were determined from PLWH from cryopreserved PBMCs for monocytes, unstimulated CD4+ T cells, and CD3/CD28-stimulated CD4+ T cells based on HIV spVL associated variants. Individuals were genotyped for rs59784663-A-G using cell-free DNA isolated from cryopreserved plasma (Appendix 7). A subset of 9 individuals with rs59784663-A-A, rs72999655-A-A, rs7525622-G-G, and rs73004025-C-C alleles and five individuals with rs59784663-A-G, rs72999655-A-G, rs7525622-G-A, and rs73004025-C-T had sufficient cell counts, RNA quantity, and RNA quality for follow-up analyses. There was no significant difference in expression of *CHD1L* ($p < 0.39$) or *PRKAB2* ($p < 0.18$) in unstimulated CD4+ T cells based on genotype (Figure 16A), or for *CHD1L* ($p < 0.80$) or *PRKAB2* ($p < 0.93$) expression in CD3/CD28 stimulated CD4+ T cells (Figure 16B). However, in monocytes, there was a significant downregulation of *CHD1L* ($p < 0.037$) and no significant difference of *PRKAB2* ($p < 0.14$) in monocytes (Figure 16C). Expression of *FMO5* and *LINC00624* was below the level of detection in most samples and could not be assessed. These results show that HIV spVL associated variants exhibit cell-type-specific gene regulation and variants linked with lower spVL are associated with decreased *CHD1L* expression in monocytes.

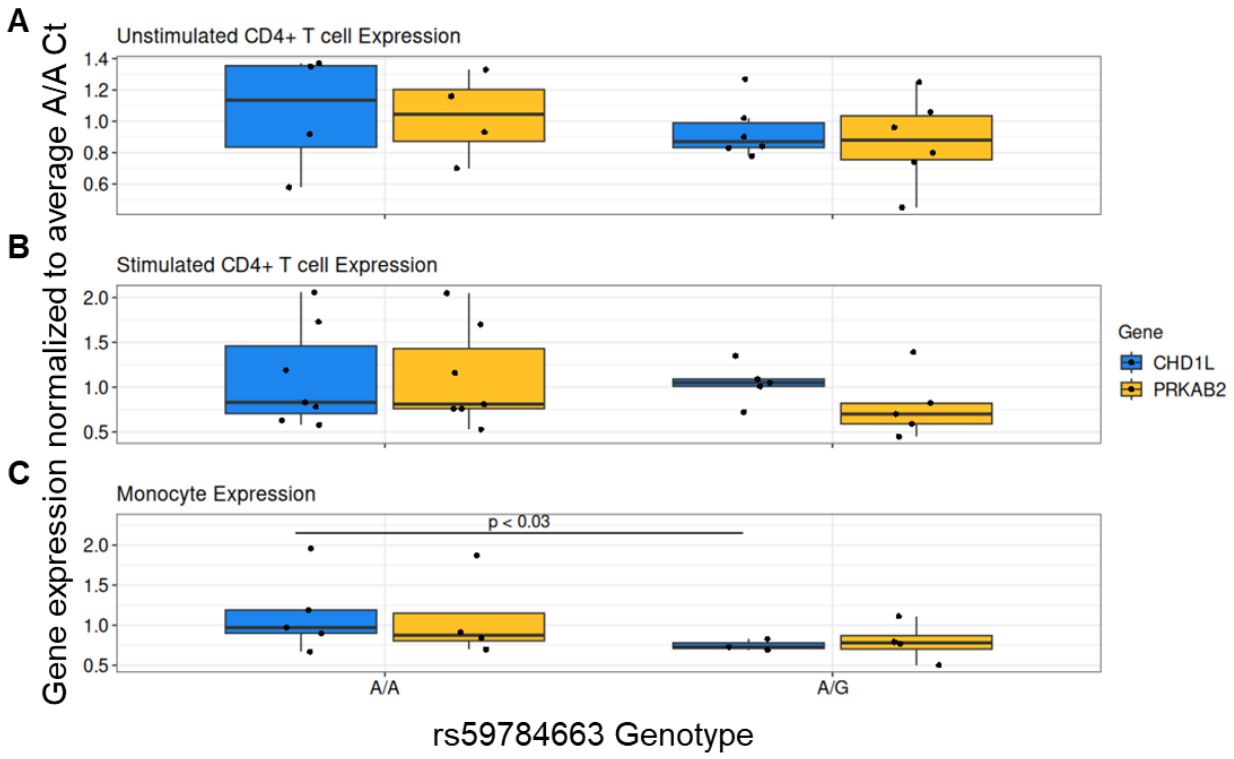


Figure 16. HIV spVL associated variant genotypes are associated with reduced *CHD1L* expression in monocytes but no expression differences in CD4+ T cells. Expression was determined for *CHD1L* (blue) and *PRKAB2* (yellow) for individuals with the rs59784663-A-A and rs59784663-A-G alleles normalized to rs59784663-A-A average expression. Gene expression was determined in unstimulated CD4+ T cells (top), CD3/CD28 stimulated CD4+ T cells (middle), and monocytes (bottom) originating from PBMCs in PLWH.

Genetically regulated *CHD1L* expression is decreased in whole blood models

Previous models of genetically regulated *CHD1L* expression predicted that individuals with HIV spVL associated variants have increased *CHD1L* expression in monocytes[275] but, in this experiment, we observed that *ex vivo* expression had the opposite effect. The previous gene imputation models were based off 233 African American individuals from the MESA cohort, and therefore, to address the discrepancies between *ex vivo* expression profiles, we leveraged expression models from 716 African American individuals in the GALA II/SAGE combined cohort to assess whether whole blood models more closely resemble the *ex vivo* data. Genetically regulated *CHD1L* expression based on HIV spVL associated variants: rs72999655-A-G, rs7525622-G-A, and rs73004025-C-T exhibited lower *CHD1L* expression in heterozygous individuals than homozygous reference ($p < 1.65\times10^{-32}$) and homozygous alternate exhibited lower *CHD1L* expression than heterozygous individuals ($p < 0.0045$) (Figure 17). This suggests that *CHD1L* expression from *ex vivo* monocytes more closely resembles whole blood expression profiles, and warrants investigation of *CHD1L*-mediated HIV inhibition in other immune cell types.

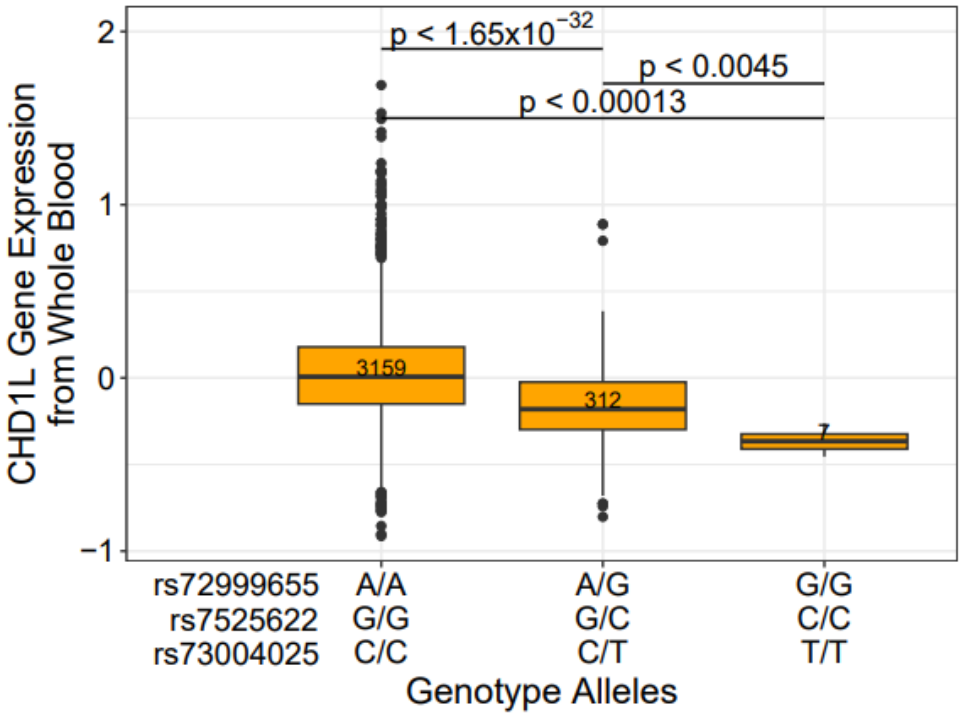


Figure 17. Imputed CHD1L expression is associated with genotype at HIV spVL associated genetic variants. CHD1L expression was imputed using whole blood models from the GALA II/SAGE cohorts. Allelic combinations are shown for individuals homozygous reference, heterozygous, or homozygous alternate at rs72999655-A-G, rs7525622-G-A, and rs73004025-C-T.

HIV spVL associated variants influence transcriptome profiles in monocytes

Next, we sought to determine whether HIV spVL associated variants are associated with transcriptome changes as *CHD1L* regulates several host pathways including protein synthesis, cell cycle regulation, and apoptosis[311]. Transcriptome changes were assessed for four individuals heterozygous for rs59784663-A-G, rs73004025-C-T, rs7525622-G-A, and rs72999656-A-G compared to seven matched controls who were homozygous reference for all alleles in monocytes (GSE56033) and naïve CD4+ T cells (GSE56034). Principal component analysis was conducted on normalized gene expression to assess the effectiveness of controlling for age, sex, and batch.

In naïve CD4+ T cells, there were no PCs significantly associated with genotype (Figure 18A, Appendix 9) and 90% of total gene expression variance was determined by the first 6 PCs (Figure 18B). In monocytes, PC1 (47.8%) was significantly associated with genotype (Figure 19A, $p < 3.3\times10^{-5}$) and the first 6 PCs accounted for 90% of total gene expression variance (Figure 19B). There was no significant association with age, sex, or batch for naïve CD4+ T cells or monocytes (Appendix 10). This supports our previous observations that candidate causal variants were associated with gene expression differences in monocytes but not naïve CD4+ T cells, suggesting these effects are also reflected at the transcriptome level.

Transcriptome analysis in naïve CD4+ T cells identified no significant gene expression differences based on HIV spVL associated variant genotypes (Figure 20). In monocytes, there were 24 genes were significantly downregulated and 1 gene was significantly upregulated in individuals heterozygous for rs72999655-A-G, rs7525622-G-A, and rs73004025-C-T compared to homozygous reference (Figure 21; Table 11). Notably, *CHD1L* ($log_2FC = -0.23$, FDR < 0.033) and *FMO5* ($log_2FC = -0.25$, FDR < 0.01) are not considered differentially expressed genes due to fold-change thresholds but are significant by FDR suggesting that genotype may have a small effect on genes in the chromosome 1 region.

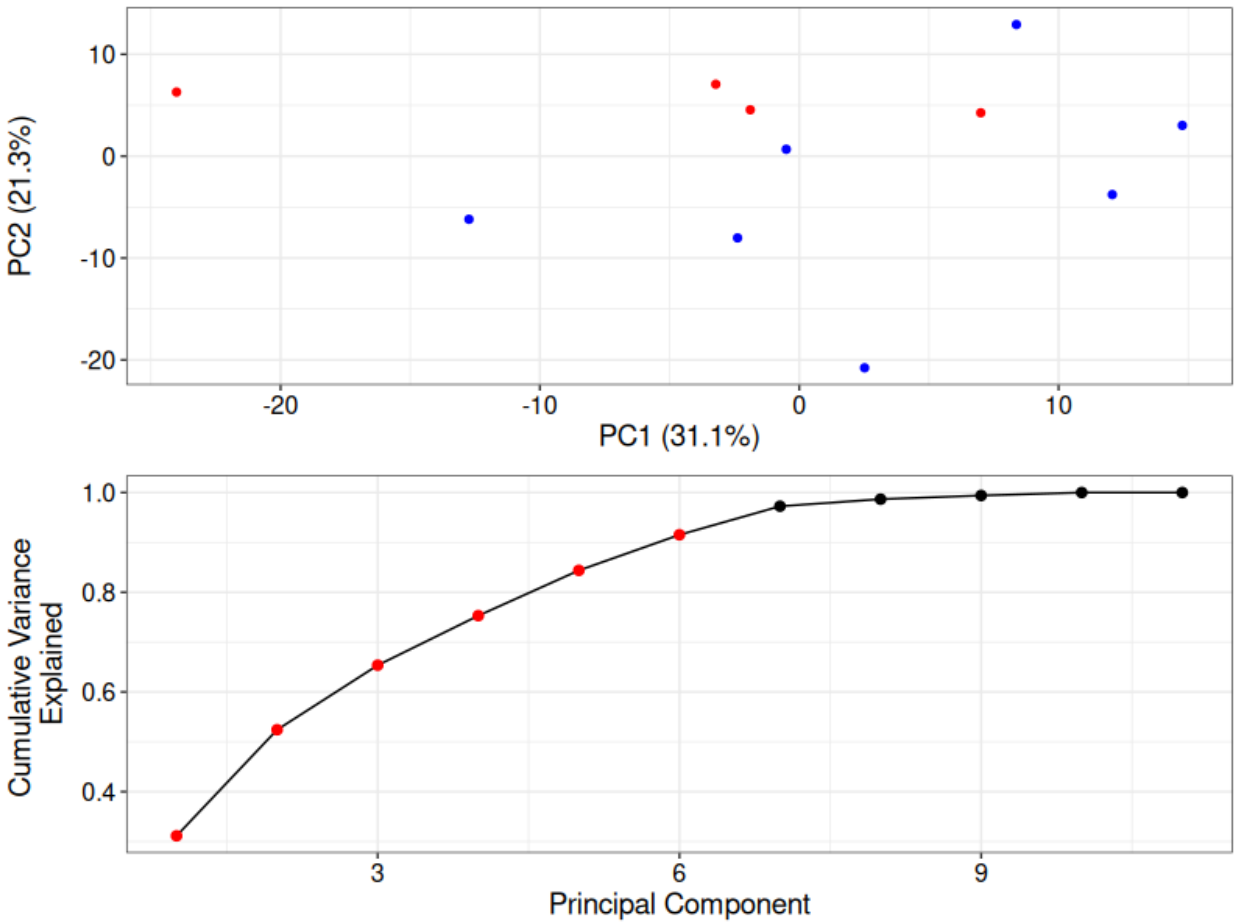


Figure 18. Principal components of total gene expression variance in naïve CD4+ T cells. (A) Principal components of log-transformed naïve CD4+ T cell gene expression. Genotypes reflect if individuals were homozygous reference (blue) or heterozygous (red) at rs59784663-A-G, rs72999655-A-G, rs73004025-C-T, and rs7525622-G-A. (B) Scree plot of the top principal components influencing gene expression in naïve CD4+ T cells. Principal components are coloured whether they are within the 90% set (red) or fall outside the 90% set (black).

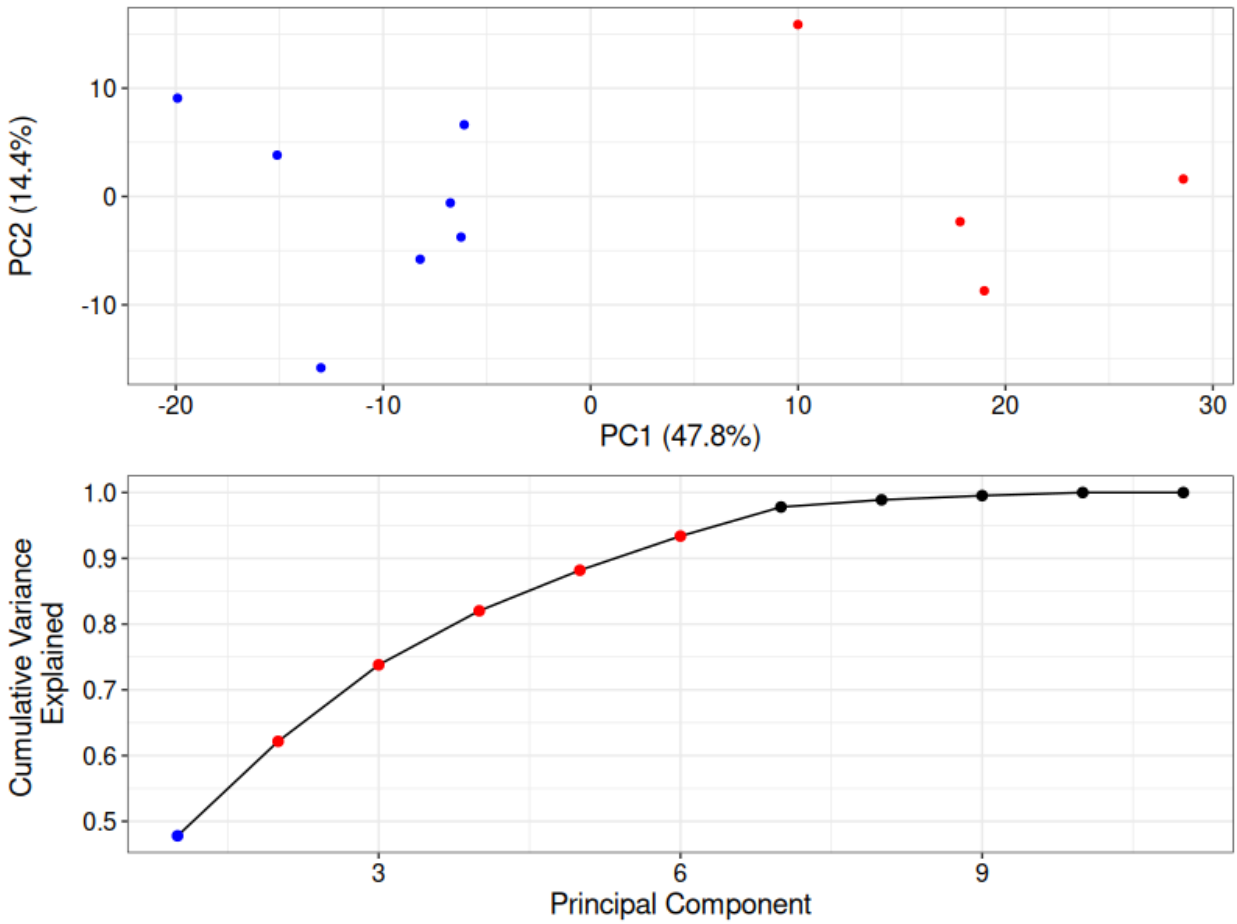


Figure 19. Principal components of total gene expression variance in monocytes show that genotype is the major contributor to gene expression differences. (A) Principal components of log-transformed monocyte gene expression. Genotypes reflect if individuals were homozygous reference (blue) or heterozygous (red) at rs59784663-A-G, rs72999655-A-G, rs73004025-C-T, and rs7525622-G-A. (B) Scree plot of the top principal components influencing gene expression in monocytes. Principal components are coloured whether they are significantly associated with genotype (blue), within the 90% set (red), or fall outside the 90% set (black).



Figure 20. Differentially expressed genes in naïve CD4+ T cells based on genotype at HIV spVL associated variants. Volcano plots were generated using FDR and $\log_2$ fold changes values from *limma*. Differentially expressed genes are significant at FDR < 0.05 and absolute $\log_2$ fold change ≥ 1.

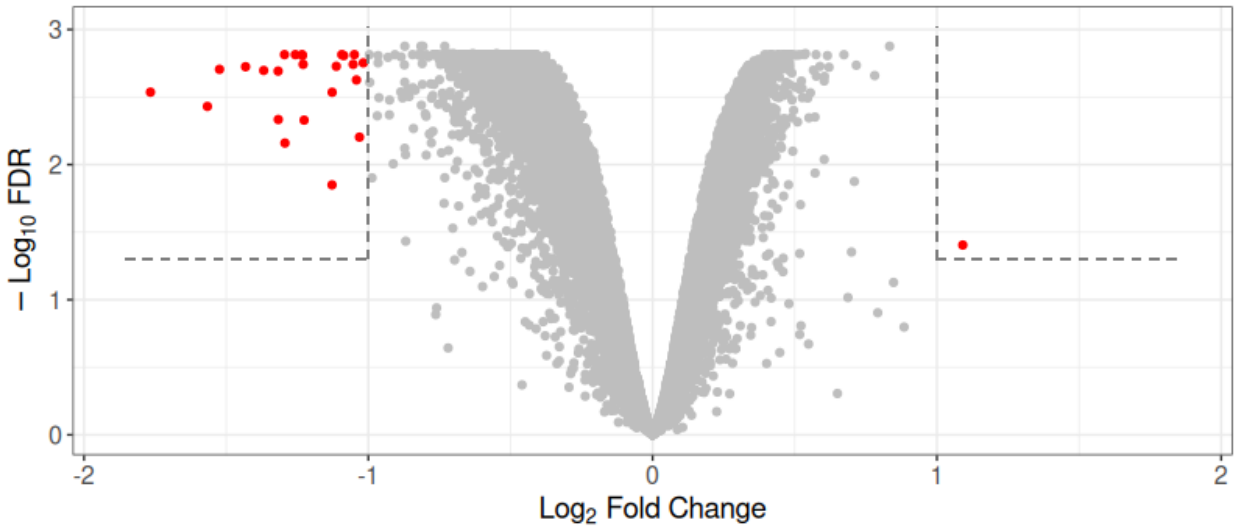


Figure 21. Differentially expressed genes in monocytes based on genotype at HIV spVL associated variants. Volcano plots were generated using FDR and $\log_2$ fold changes values from *limma*. Differentially expressed genes are depicted in red and are significant at FDR < 0.05 and absolute $\log_2$ fold change ≥ 1.

Table 11. Differentially expressed genes in monocytes associated with presence of rs72999655-A-G, rs73004025-C-T, and rs7525622-G-A alleles.

HGNC Symbol	Log_2 Fold Change	P-Value	FDR
FOSB	1.09	1.81×10^{-2}	3.94×10^{-2}
C4orf33	-1.02	3.19×10^{-5}	1.77×10^{-3}
SIGLEC1	-1.03	1.08×10^{-3}	6.26×10^{-3}
MX1	-1.04	1.10×10^{-4}	2.36×10^{-3}
MIR142	-1.05	4.19×10^{-6}	1.53×10^{-3}
IFIT5	-1.05	3.80×10^{-5}	1.81×10^{-3}
DDX60	-1.09	1.86×10^{-5}	1.56×10^{-3}
OAS3	-1.09	1.19×10^{-5}	1.53×10^{-3}
IFITM1	-1.11	4.57×10^{-5}	1.87×10^{-3}
GBP4	-1.13	1.98×10^{-4}	2.92×10^{-3}
TAS2R45	-1.13	4.15×10^{-3}	1.41×10^{-2}
GBP5	-1.23	6.04×10^{-4}	4.69×10^{-3}
IFIT3	-1.23	3.77×10^{-5}	1.81×10^{-3}
EIF2AK2	-1.23	1.82×10^{-5}	1.56×10^{-3}
GPR160	-1.23	1.63×10^{-6}	1.53×10^{-3}
RSAD2	-1.26	2.74×10^{-6}	1.53×10^{-3}
HSP90AA6P	-1.29	1.31×10^{-3}	6.93×10^{-3}
HERC5	-1.29	9.08×10^{-6}	1.53×10^{-3}
CXCL10	-1.32	5.88×10^{-4}	4.64×10^{-3}
EPSTI1	-1.32	6.89×10^{-5}	2.03×10^{-3}
IFIT2	-1.37	6.71×10^{-5}	2.01×10^{-3}
IFI44	-1.43	5.09×10^{-5}	1.89×10^{-3}
IFIT1	-1.52	5.91×10^{-5}	1.97×10^{-3}
JUP	-1.57	3.62×10^{-4}	3.71×10^{-3}
IFI44L	-1.77	1.96×10^{-4}	2.91×10^{-3}

Pathway and functional enrichment of DEGs highlight changes to interferon signaling pathways

Analysis of differentially expressed genes in monocytes was conducted using DAVID 2021 with KEGG, REACTOME, and GO terms to identify functional and pathway enrichment[282,312]. Analysis of KEGG pathways revealed enrichment of genes important for viral infections including SARS-CoV-2, Influenza A, and Hepatitis C (Figure 22A). Gene Ontology identified a downregulation of genes corresponding to defense response to viruses and antiviral innate immune response (Figure 22B), and Reactome pathways showed altered expression of cytokine and interferon signaling genes (Figure 22C). Therefore, in this dataset, individuals with genotypes associated with lower spVL also have down-regulated expression of genes involved in interferon signaling and the antiviral response. These results suggest that the chromosome 1 region may influence expression of key genes that influence the immune response and influence HIV spVL through these pathways.

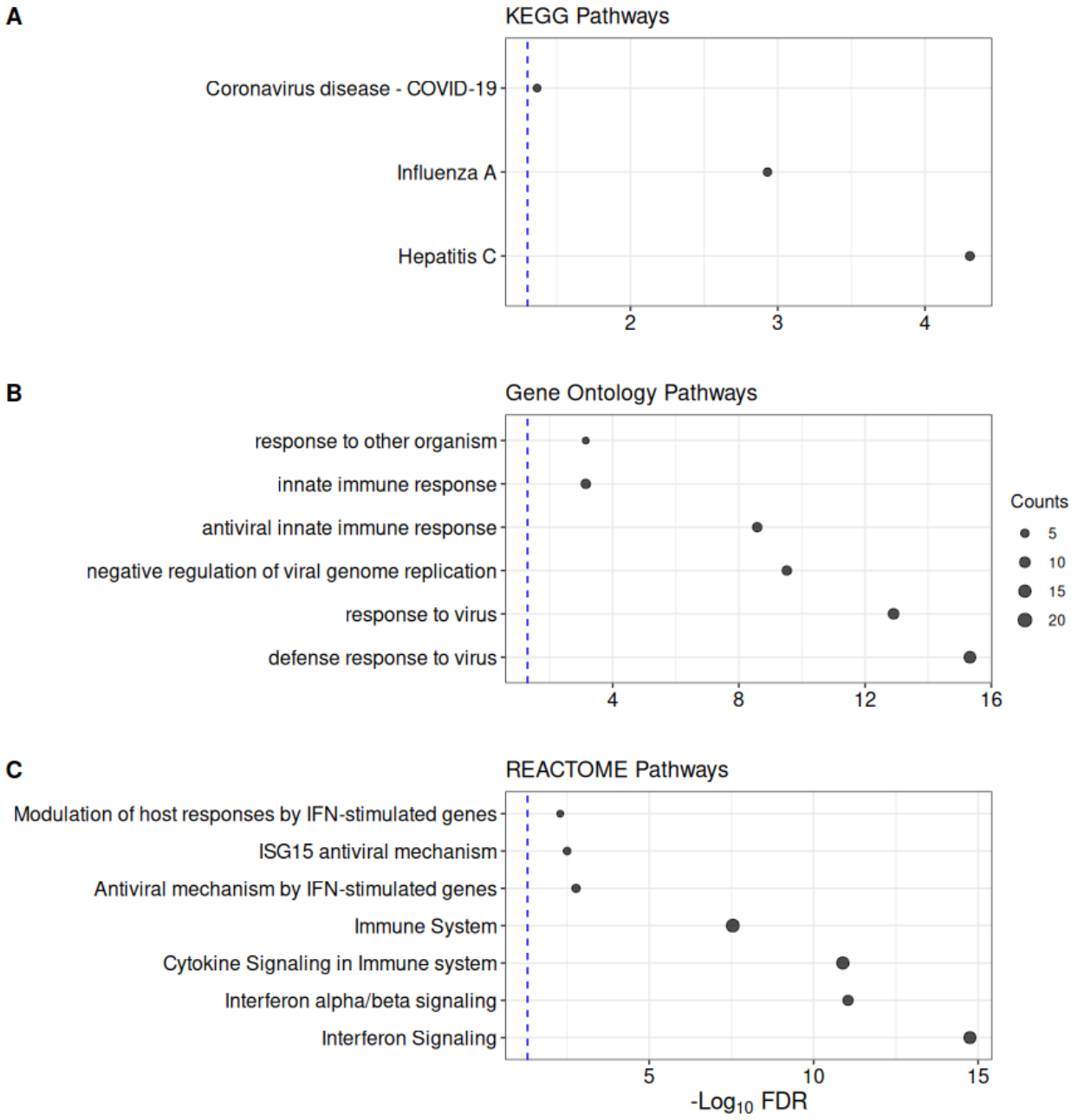


Figure 22. Differentially expressed genes associated with genotypes linked to lower HIV spVL are associated with the immune response, cytokine signaling, and interferon signaling responses. Genes were investigated for their overlap with KEGG pathways (A), Gene Ontology pathways (B), and Reactome pathways (C) using DAVID (v2025.1). Terms are present if they were below an FDR of 0.05.

CHD1L, *PRKAB2*, and *FMO5* do not cluster with DEGs

We next sought to test whether *CHD1L* is directly associated with DEGs by using the STRING database to generate a protein-protein interaction network. The DEGs network consisted of 22/26 connected nodes, 149 edges, with an enrichment p-value < 1.0×10^{-6}. When *CHD1L*, *PRKAB2*, and *FMO5* were included in this query, there were no connections between these genes and the DEGs from the analysis (Figure 23). The absence of direct associations between *CHD1L*, *PRKAB2*, or *FMO5* and the identified DEGs suggest that *CHD1L*, *PRKAB2*, or *FMO5* may exert their effects through intermediary signaling genes not captured in this dataset. An example of this is that *CHD1L* knockdown is associated with reduced phosphorylation of AKT and NF-κβ without significant changes to gene expression, but the underlying biological mechanism is unclear[313].

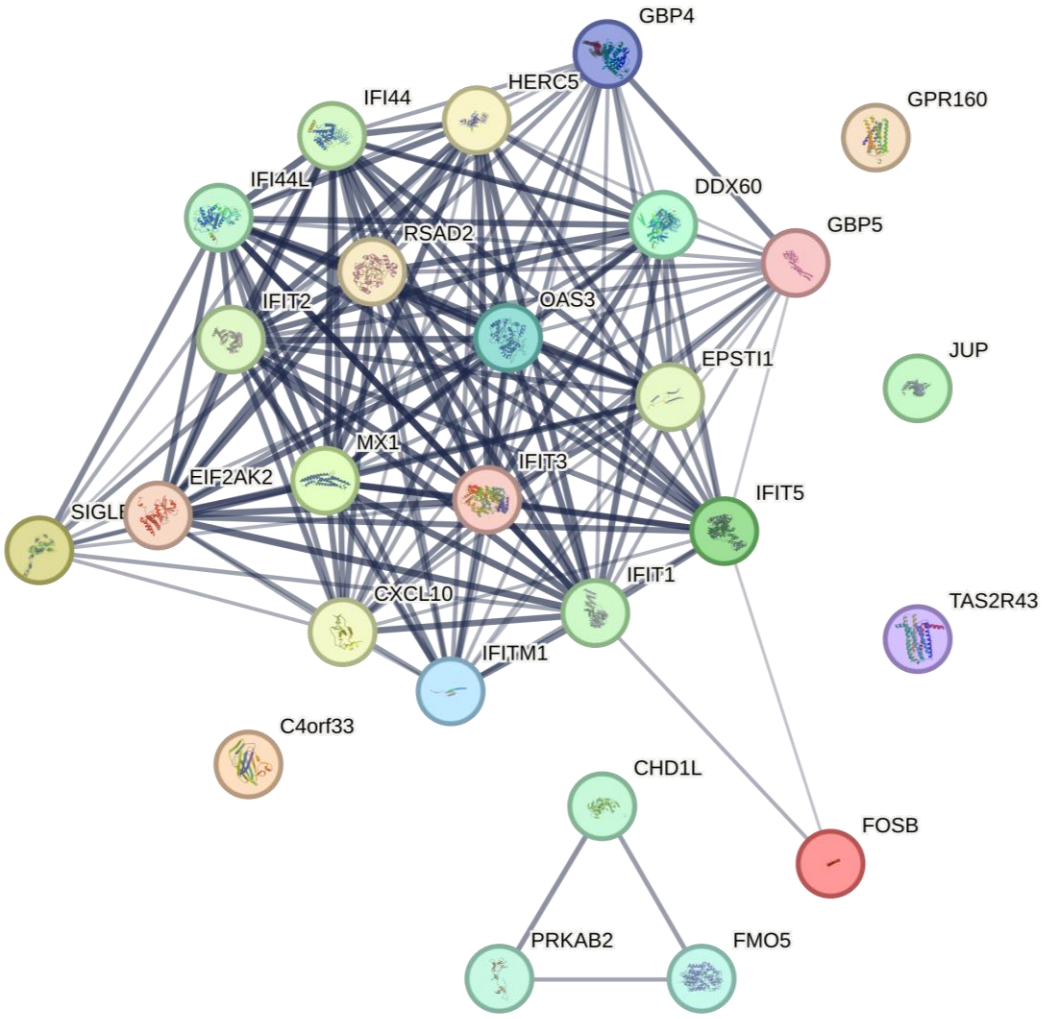


Figure 23. Chromosome 1 genes do not directly cluster with differentially expressed genes based on HIV spVL-associated genotype. The thickness of the line represents the confidence level of gene-gene interactions as determined by string-db.

Chromosome 1 gene expression profiles significantly overlap with HIV+ LTNPs transcriptomes

The biological relevance of our 25 differentially expressed genes to HIV disease was assessed using GSEA on gene lists comparing expression profiles from HIV+ LTNPs and HIV+ individuals without ART[314,315]. The two meta-analyses included contain non-overlapping samples and identified differentially expressed genes in CD4+ T cells or whole blood samples. For both whole blood ($p < 3.64\times10^{-10}$; Figure 24A) and CD4+ T cell LTNP profiles ($p < 6.75\times10^{-41}$; Figure 24B). These results suggest that the 25 DEGs associated with rs72999655-A-G, rs73004025-C-T, and rs7525622-G-A genotypes are biologically relevant to HIV disease progression and suggest the gene expression profiles resemble a protective phenotype.

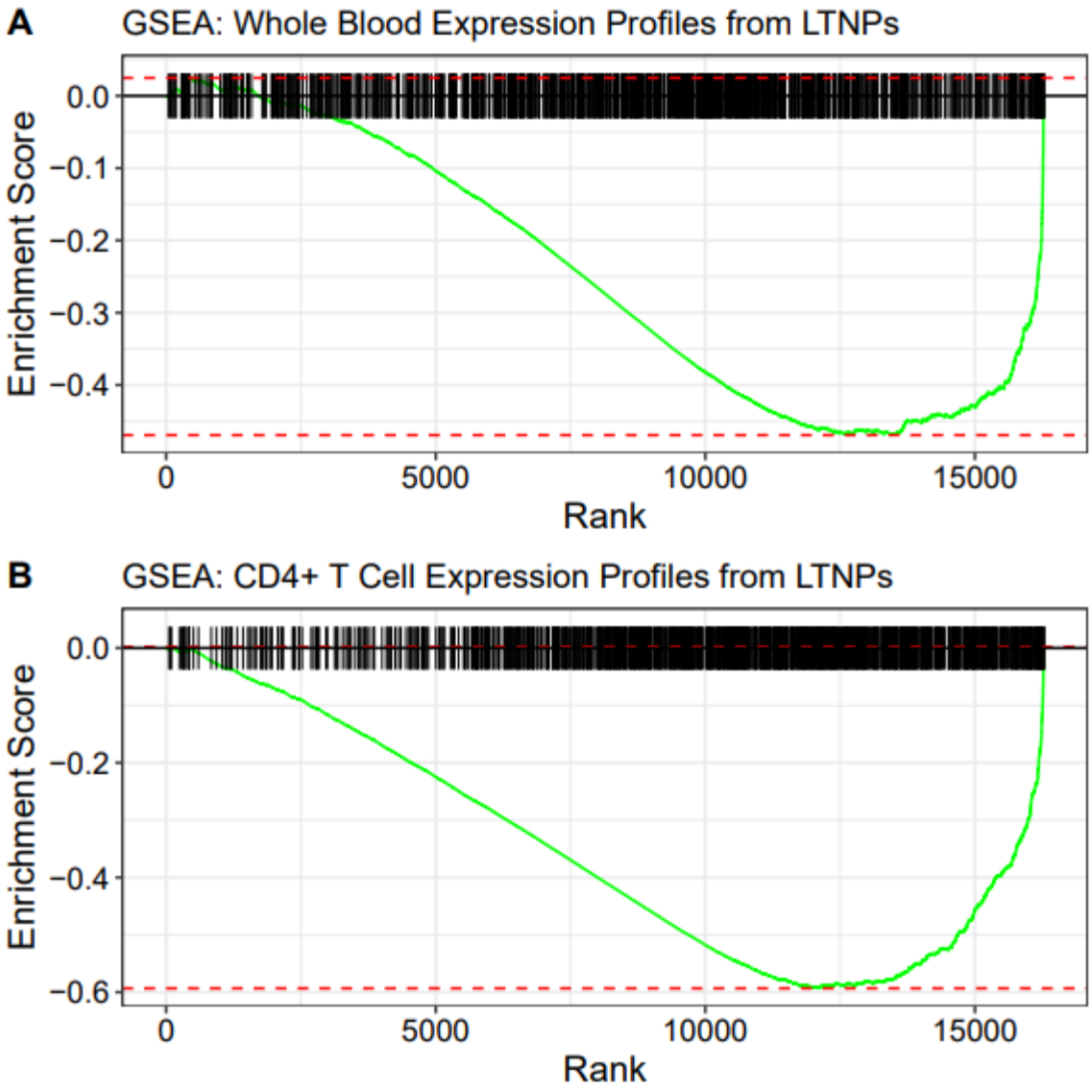


Figure 24. Differentially expressed genes associated with alleles linked to lower HIV spVL exhibit significant overlap with gene expression profiles from HIV LTNPs.

Gene sets of differentially expressed genes in A) CD4+ T cells and B) whole blood show significant overlap to genes associated with chromosome 1 genotypes. Enrichment was determined using ranked gene lists and fgsea (v1.24.1) in R. The y-axis depicts the enrichment scores of the ranked gene lists and the x-axis depicts the gene rank corresponding to fold-change and $-\log_{10}$(p-value).

Discussion

This study presents a comprehensive analysis of gene expression profiles in CD4+ T cells and monocytes, focusing on the impact of likely causal GWAS variants associated with decreased HIV spVL. The findings highlight the cell-type-specific nature of gene expression and suggest a potential mechanism for how HIV spVL associated variants likely influence HIV infection.

Initial investigations have suggested that HIV spVL associated variants are associated with increased *CHD1L* mRNA expression in monocytes[275] and that *CHD1L* expression is negatively correlated to HIV replication in myeloid cells[1]. However, here we observed that HIV spVL associated variants linked with reduced HIV spVL are associated with decreased *CHD1L* expression in monocytes and *CHD1L* expression has no significant association with genotype in CD4+ T cells. Of note, the imputed gene expression models originated from CD14+ monocyte samples in the African American MESA cohort[316], whereas here we specifically selected for $CD14^{+}$ $CD16^{++}$ monocytes from PLWH of African ancestry from Pumwani, Nairobi, Kenya (Appendix 1). Between these cohorts, there are population-level, cell-type, and disease-specific differences which may explain differences in predicted and observed gene expression.

Given these gene expression differences in monocytes, HIV spVL associated variants might affect gene expression in other cell types. We used expression models from whole blood tissues in the GALA II/SAGE cohort to investigate the impact of HIV spVL variants and variants linked to lower spVL were associated with reduced *CHD1L* expression. While there was no association with genotype and *CHD1L* expression in naïve CD4+ T cells, we did notice that *CHD1L* was reduced, but not significantly, in the physiologically active CD3/CD28-stimulated CD4+ T cells. This suggests that primary activated CD4+ T cells may also be capable of *CHD1L*-mediated HIV inhibition, but larger sample sizes will be necessary to determine the effect of genetically-regulated *CHD1L* expression in this cell type.

When we investigated transcriptome changes in individuals with the protective HIV spVL associated variants, we observed significant changes to interferon signaling pathways and the antiviral response. While *CHD1L* knockdown has been shown to

reduce inflammation signaling previously, the specific genes regulating this response is unknown[317]. This is supported by previous studies that have shown that knockdown of *CHD1L* reduces activated p-AKT, p-ERK1/2, and p-NF-κβ, key regulators of inflammation and interferon signaling[318]. In the context of HIV infection, a strong interferon response is associated with reduced HIV replication *in vitro*[319] but higher spVL during chronic infection[320,321]. This suggests that if *CHD1L* can influence the interferon response, increased *CHD1L* could inhibit HIV replication *in vitro*[1], but lower expression during chronic infection may reduce HIV-induced immune activation, improve T cell responses, and decrease the viral reservoir[322]. While these findings offer a promising model for *CHD1L*-mediated HIV spVL, further studies will be required to validate this mechanism and assess *CHD1L* regulation in multiple cell types.

This study also has limitations that need to be addressed. The first is the small sample size, particularly since the GWAS variant rs59784663-A-G has a MAF of 0.048 in African populations[232]. This low MAF requires thousands of samples to identify sufficient homozygote alternate individuals for association testing and therefore, this study is limited to testing heterozygous individuals, who display an intermediate phenotype in HIV spVL and likely have muted gene expression changes compared to homozygous alternate individuals[1]. Secondly, in assessing gene expression profiles, our previous study observed that HIV spVL associated variants are linked with changes to *PRKAB2* expression (Chapter 3). In qPCR studies, the recommended housekeeping genes in CD4+ T cells are *TBP*, *RPL13A*, and *SDHA*[323], and *SDHA* and *HPRT1* in monocytes[324]. However, changes to *PRKAB2* expression significantly alter *TBP*, *HPRT1*, and expression of ribosomal genes[209], therefore we limited our HKG selection to *SDHA*. By reducing our HKG to only *SDHA* increases variance in the qPCR data which may mask subtle gene expression differences. However, even with these limitations, the results of this study still suggest that HIV spVL associated variants are associated with changes to *CHD1L* and interferon signaling.

These findings indicate that rs72999655-A-G, rs7525622-G-A, and rs73004025-C-T are associated with downregulation of *CHD1L* and interferon stimulating genes in monocytes, but not in CD4+ T cells. This suggests a putative mechanism of *CHD1L*

expression and the immune response, requiring further studies to address the underlying biology. Overall, these results provide a solid foundation for further investigations into the host genetics of HIV control and novel mechanisms for follow-up studies.

Chapter 5: Discussion

CHD1L is a novel HIV inhibitory factor *in vitro*

Despite advances in testing, treatment, and prevention, HIV remains a global public health crisis. In the absence of a vaccine or functional cure, TasP remains an important part of HIV eradication efforts. However, drug resistance rates are increasing globally and are associated with faster disease progression, higher mortality rates, and cross-resistance to other antivirals[325,326]. Roughly 10% of individuals newly infected with HIV, prior to administration of ART, will have some level of drug resistance at the time of infection[327,328]. Therefore, there is an urgent need to improve our understanding of the HIV host-pathogen interaction and develop novel therapeutics that are less prone to antiviral resistance.

The study of HIV host genetics has predominately focused on candidate gene approaches to identify HIV restriction, inhibitory, auxiliary, and dependency factors[329]. However, investigating candidate genes in small cohorts has resulted in most associations being later classified as false positives in larger populations[330]. The only genetic associations with HIV susceptibility and disease progression which have been consistently replicated at the population level in European populations are within the *CCR5* and *HLA* loci[190]. By understanding the role of *CCR5* in HIV infection, researchers developed Maraviroc, a CCR5-antagonist, which prevents infection of R5-trophic HIV[331]. However, candidate gene approaches have not been able to identify other host genes that have led to new HIV therapeutics.

Previously, our lab conducted a GWAS with individuals of African ancestry from the ICGH cohort and identified a novel region on chromosome 1 associated with HIV spVL (Figure 24A)[1]. This region is associated with a $0.30\log_{10}$ reduction of HIV spVL, a similar magnitude as observed with individuals heterozygous for *CCR5Δ32*[166] but the association has not been observed in European HIV spVL GWAS[198]. Preliminary fine-mapping of the chromosome 1 region applied a generalized gene-set analysis to GWAS summary statistics which provided evidence that *CHD1L* was the most likely candidate gene[1]. The effect of *CHD1L*-mediated HIV inhibition was validated in U2OS, iPSDMs, and THP-1 cells which identified *CHD1L* as a novel HIV inhibitory factor (Figure 24B).

However, there was no effect of *CHD1L* knockout on p24 production or HIV infection in Jurkat T cells, indicating that *CHD1L*-mediated HIV inhibition is cell-type-specific. *CHD1L* had not been previously implicated in HIV disease, suggesting that investigation into the underlying molecular biology of *CHD1L*-mediated HIV inhibition may guide the development of novel therapeutics.

Variants linked to lower HIV spVL are associated with changes to *CHD1L* expression, *PRKAB2* expression, and interferon signaling

This project sought to expand on the initial HIV spVL GWAS[1] and characterize the relationship between HIV spVL associated variants, gene expression, and HIV pathogenesis. In Chapter 2, we applied a heuristic functionally-informed fine-mapping approach to genotypes from individuals of African ancestry in the ICGH cohort and identified a combination of three variants: rs72999655-A-G, rs7525622-G-A, and rs73004025-C-T explained the majority of the impact on HIV spVL at this locus. We imputed *CHD1L* expression for the 3,886 individuals in the ICGH cohort based on monocyte gene expression models from the MESA cohort[251], and determined that HIV spVL associated variants linked to lower spVL were associated with increased *CHD1L* expression (Figure 24C). These results support *CHD1L* as a causal GWAS gene and suggest that increased *CHD1L* is associated with lower HIV spVL.

The candidate causal variants from Chapter 2 overlapped with promoter and enhancer regions, suggesting expression of nearby genes may also be affected. In Chapter 3, we investigated the association between HIV spVL associated variants and *PRKAB2,* the regulatory β2 subunit of the AMPK complex. We imputed *PRKAB2* expression for the 3,886 individuals in the ICGH cohort based on whole blood expression models from the GALA II/SAGE cohort and identified that protective HIV spVL alleles were associated with decreased *PRKAB2* expression (Figure 24C). We then sought to determine how reduced *PRKAB2* expression might affect HIV-associated pathways using gene expression from wild-type and *PRKAB2*$^{-/-}$ iPSCs. Notably, *PRKAB2* loss-of-function was significantly associated with downregulation of genes regulating cytokine signaling and stem cell pluripotency pathways. These results suggest that the changes to HIV spVL observed in the original GWAS may be driven by altered immune signaling pathways.

In Chapter 4, we also imputed *CHD1L* expression for the 3,886 individuals of African ancestry from the ICGH cohort using GALA II/SAGE whole blood expression aamodels and observed that HIV spVL variant alleles linked to lower spVL were associated with reduced *CHD1L* and *PRKAB2* expression. To address the discrepancy between imputed *CHD1L* expression from monocyte and whole blood models, we measured gene expression differences using RNA extracted from unstimulated CD4+ T cells, CD3/CD28 stimulated CD4+ T cells, and monocytes from African PLWH in the Majengo cohort. We observed that HIV spVL associated variant alleles linked to decreased HIV spVL were associated with decreased *CHD1L* in monocytes, but no associations were observed in unstimulated or stimulated CD4+ T cells for *CHD1L* or *PRKAB2* (Figure 24D).

We then sought to determine whether HIV spVL associated variants were associated with downstream signaling pathways. We analyzed transcriptomes of healthy individuals from the ImmVar Project, based on HIV spVL associated variant genotype, in naïve CD4+ T cells and monocytes. There were no significant gene expression differences in naïve CD4+ T cells based on genotype. In monocytes, individuals with alleles linked to lower HIV spVL showed significant expression differences in 25 genes which were enriched for cytokine, interferon, and antiviral signaling (Figure 24E). Given that HIV spVL associated variants significantly both influence *CHD1L* expression and interferon signaling, this suggests a putative model that *CHD1L*-mediated HIV inhibition occurs through altering interferon and antiviral signaling pathways. However, further investigation is required to assess the discrepancy between increased *CHD1L* inhibiting HIV replication *in vitro* but decreased *CHD1L* and lower HIV spVL.

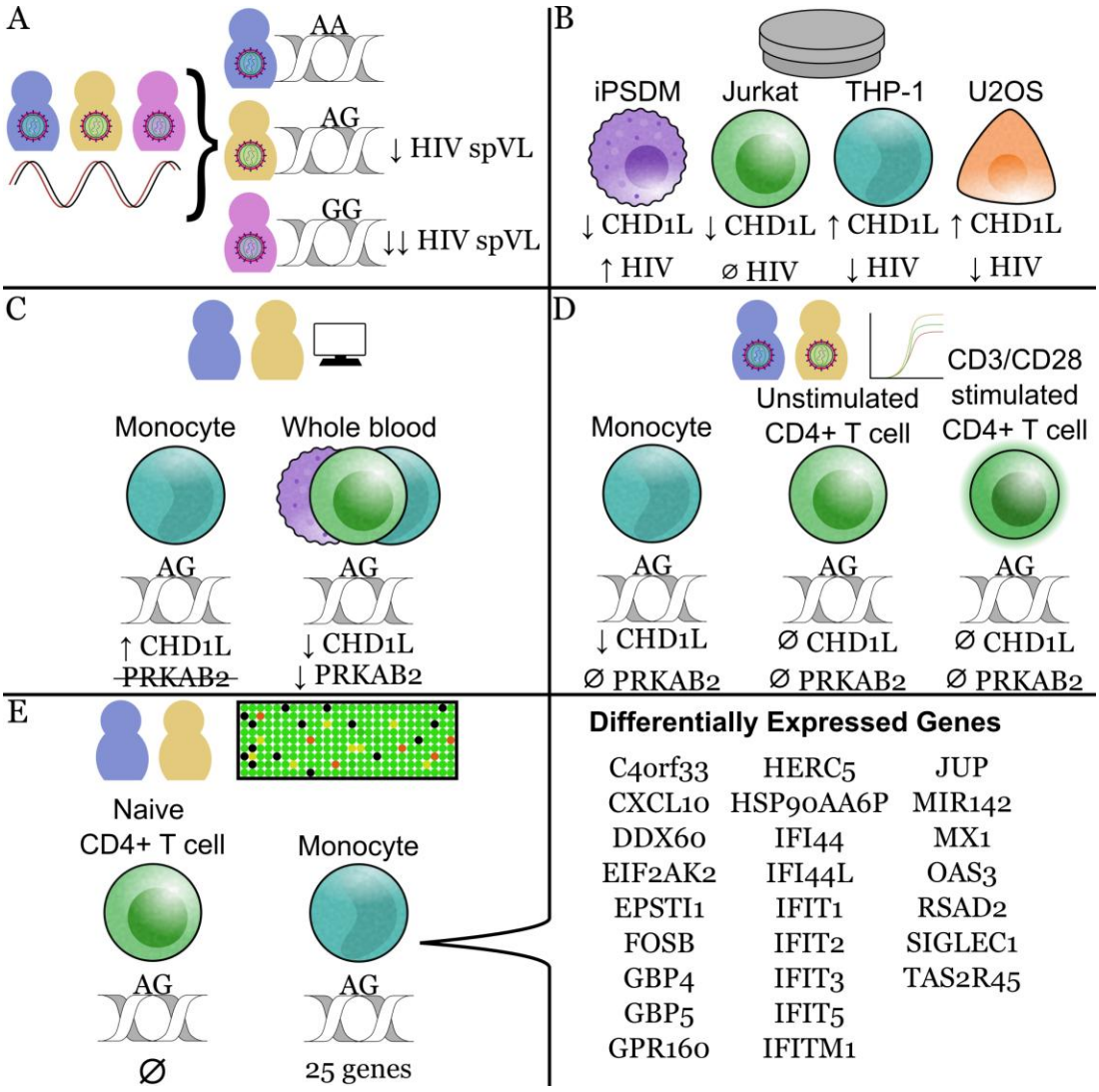


Figure 25. Summary of the association between HIV spVL associated variants, gene expression, and HIV infection. A) A GWAS of HIV spVL in African populations identified a region on chromosome 1, tagged by rs59784663-A-G, as associated with HIV spVL. B) Association between *CHD1L* expression and HIV p24 production in iPSDMs (far left), Jurkat (left), THP-1 (right), and U2OS (far right) cells. C) Imputed *CHD1L* and *PRKAB2* expression based on HIV spVL associated variant alleles from monocyte (left) or whole blood (right) expression models. D) Association between HIV spVL associated variants and gene expression from people living with HIV from the Majengo cohort in unstimulated CD4+ T cells (left), CD3/CD28 stimulated CD4+ T cells (center), and monocytes (right). E) Transcriptome analysis identified protective alleles are associated with reduced expression of ISGs.

The interferon response is associated with reduced HIV replication *in vitro*

The results of our study suggest that HIV spVL associated variants alter the cytokine and interferon response. Type I IFNs (IFN-α/IFN-β) and type II IFNs (IFN-γ) are an integral part of the immune response against viral infections and are mainly produced by monocytes, macrophages, and dendritic cells[332,333]. During viral infection, activation of cGAS, RIG-I, IFIH1/MDA5, DEAH box helicases, and TLR3/7/8/9 pattern recognition receptors signalling cascades cause the release of IFNs which induce transcription of antiviral ISGs in infected and uninfected cells.

The IFN response is heavily regulated by cofactors in the PI3K-AKT, mTOR, MAPK/ERK, and JAK-STAT pathways which regulate activation through phosphorylation. In the PI3K-AKT pathway, AKT1 is phosphorylated by PI3K which activates protein synthesis, autophagy, cell survival, and immune response pathways[334]. Activated AKT1 has been shown to influence interferon signaling by relieving EMSY-mediated ISRE repression and reduced activation of AKT1 is associated with reduced type I interferon expression[335,336].

The ERK1/2 signaling cascade is activated by MEK phosphorylation and regulates survival, differentiation, and inflammatory responses[337]. Reductions in phosphorylated ERK result in a significant reduction in IFN-α or IFN-β expression through the cGAS and JAK-STAT pathways[338,339]. Additional downstream effects of ERK1/2 signaling include activation of PARP1[340], the LKB1/AMPK axis[341], and CREB[342], which are required for interferon signaling. Specifically, activated PARP1, or caspase-cleaved PARP1 products, act as cofactors for NF-κβ in cytokine signaling[343,344] and activation of RNA polymerase III which acts as an agonist for the RIG-I and MDA5 response[345,346].

When IFNs bind to their corresponding receptors on the cell surface, they activate the JAK-STAT signaling pathway and subsequent transcription of ISGs[347]. In our transcriptome analysis, of the 25 differentially expressed genes, 22/25 are ISGs indicating a strong link between HIV spVL associated variants and the interferon response[348] (Table 12).

Table 12. Differentially expressed genes based on HIV spVL associated variant genotype that are also interferon stimulating genes.

Type I IFN ISGs	Type II IFN ISGs
C4orf33	-
CXCL10	CXCL10
DDX60	DDX60
EIF2AK2	EIF2AK2
EPSTI1	EPSTI1
FOSB	FOSB
GBP4	GBP4
GBP5	GBP5
-	GPR160
HERC5	HERC5
IFI44	IFI44
IFI44L	IFI44L
IFIT1	IFIT1
IFIT2	IFIT2
IFIT3	IFIT3
IFIT5	IFIT5
IFITM1	IFITM1
JUP	JUP
MX1	MX1
OAS3	OAS3
RSAD2	RSAD2
SIGLEC1	SIGLEC1

CHD1L influences the ERK1/2, PI3K-Akt, and NF-κβ signaling pathways

CHD1L has been well characterized as an oncogene but recent studies have shown that *CHD1L* expression is associated with inflammation and the TNF signaling pathway[317,349]. This immunomodulatory function of *CHD1L* is supported by observations that CHD1L knockdown reduces the phosphorylation of ERK1/2, AKT, and NF-κβ without affecting overall protein expression[313]. This effect is likely mediated by genes downstream of *CHD1L*, as there is currently no evidence *CHD1L* phosphorylates target genes (putative model depicted in Figure 25).

MDM2, an E3 ubiquitin ligase, is regulated by *CHD1L* and is an inhibits the tumor suppressor gene, *p53*[350]. Activated MDM2 results in increased production of activated p-AKT[351] and prevents CASP3-mediated ERK inhibition[352]. In macrophages, *MDM2* also participates in the inflammatory response and interferon signaling[353,354]. While *MDM2* has been directly implicated as an HIV inhibitory factor[355], it is also involved in activation of HIV transcription[356]. Given that *CHD1L* in regulates *MDM2* transcription, this pathway has the potential to affect HIV pathogenesis and warrants subsequent follow-up.

CHD1L also acts as a transcription factor for *SPOCK1*[225]*,* which positively regulates the PI3K-AKT-mTOR[357] and NF-κβ pathways[358]. SPOCK1-mediated AKT activation results in reduced cleavage of CASP9 and CASP3, thereby limiting the active form of these genes[225]. This is also observed in experiments that showed CHD1L knockdown was associated with significant increases in cleaved CASP3[313]. Cleaved CASP3 inactivates cGAS, MAVS, IRF3, and NF-κβ and results in the downregulation of type I IFNs[359,360]. While there has been significant investment into identifying the effects of *CHD1L*-mediated *CASP3* regulation, this has been predominately investigated in the context of cancer and the effect on viral infection remains unclear[318]. However, through this pathway, *CHD1L* expression appears to be positively correlated with type I IFN expression.

Another mechanism by which *CHD1L* may regulate the interferon response is through interactions with PARP1 during DNA repair. When a double-strand break in DNA occurs, PARP1 recruits CHD1L to induce chromatin relaxation, allowing PARP1 to interact with ATM[361]. This interaction initiates the DNA damage response through

phosphorylation of p53, SMC1, and H2AX[362], and induces expression of IFN-α and IFN-γ through downstream targets[363]. However, CHD1L knockdown also causes PARP-trapping[364], the process by which PARP1 remains bound to double-strand breaks, that induces the cGAS-STING immune signaling[365]. Therefore it is difficult to conclude how *CHD1L* may influence interferon signaling through these pathways, but it is evident that changes to *CHD1L* has significant implications during double-strand break repair. In the context of HIV infection, HIV Vpr induces DNA double-strand breaks and upregulates expression of *CXCL10, DDX60, GBP5, HERC5, IFIT1, IFIT2, IFIT3, IFITM1, IFI44L, MX1, OAS3,* and *RSAD2* in monocyte-derived macrophages[366], exhibiting significant overlap with the ISGs described in Chapter 4. Therefore, it is also important to consider *CHD1L* interactors, not just genes downstream of *CHD1L*, which may impact expression of ISGs.

Given the relationship between HIV spVL associated variants, *CHD1L* expression, and ISG expression, this suggests a putative model by which genes downstream of *CHD1L* or *CHD1L* interactors influence the interferon response. One of the limitations of this model is that changes to interferon signaling must also explain the cell-type-specificity of *CHD1L*-mediated HIV inhibition and why *CHD1L* is negatively correlated with HIV replication *in vitro* but appears to be positively correlated to HIV spVL.

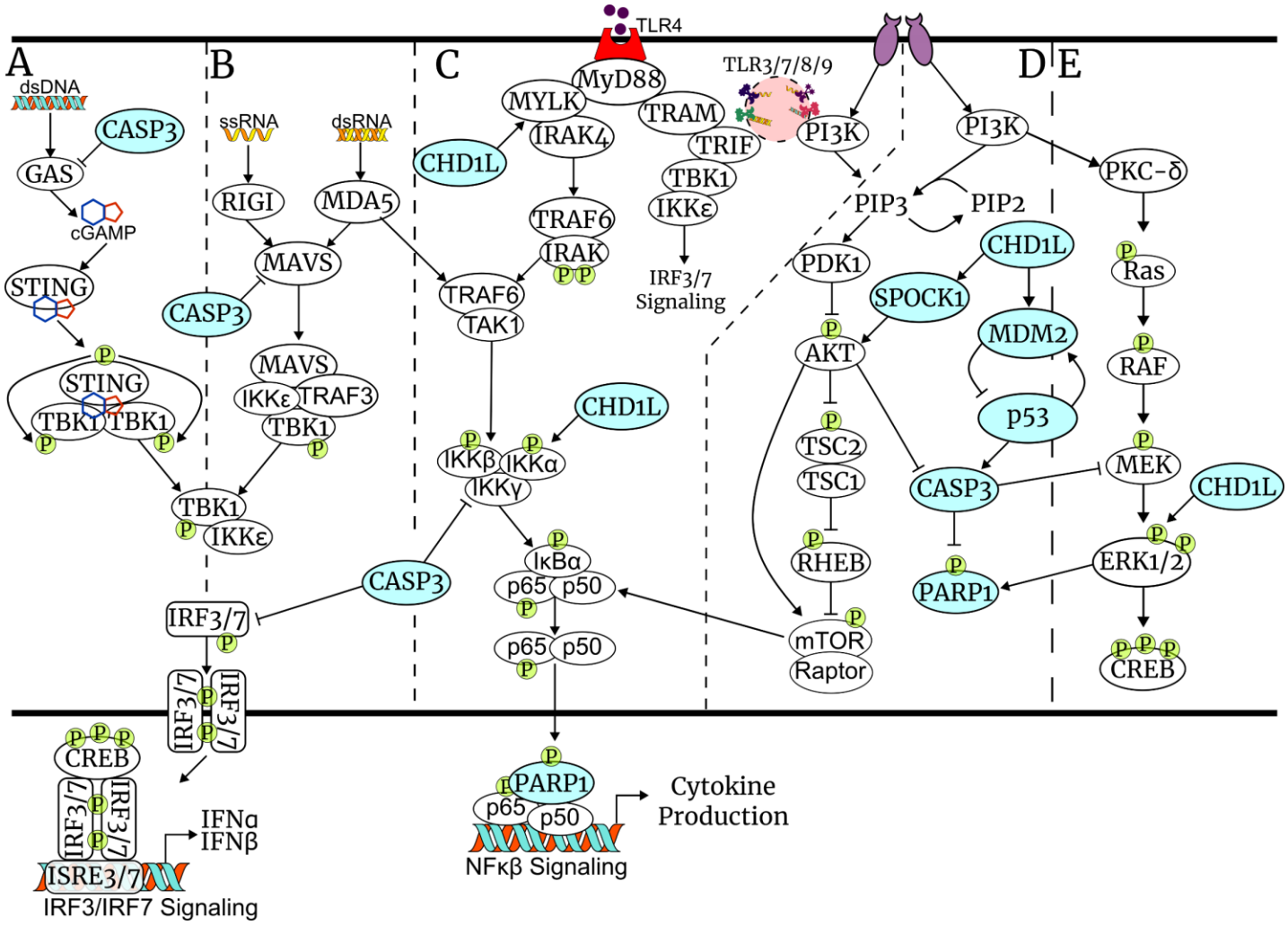


Figure 26. Association between *CHD1L* and the interferon signaling pathway. A) Genes involved in the cGAS-STING signaling pathway. B) Genes involved in ssRNA or dsRNA sensing pathways. C) Genes involved in TLR signaling pathways. D) Genes involved in the PI3K-AKT signaling pathway. E) Genes involved in the MAPK signaling pathway. Genes depicted in blue are known interactors or regulated by *CHD1L*.

Cell-type-specificity of *CHD1L*-mediated HIV inhibition

Initial investigations into *CHD1L*-mediated HIV inhibition occurs in U2OS, THP-1, and iPSDMs, but not Jurkat T cells[1]. Therefore, it follows that an effective interferon response should be present in U2OS, THP-1, and iPSDMs, but Jurkat T cells should have a deficient interferon response. Here, we examine the known literature of the interferon response in these cell types to determine whether inherent cell type differences explain the discrepancies in *CHD1L*-mediated HIV inhibition.

Firstly, the U2OS cell line is derived from an osteosarcoma derived from a 15-year-old Caucasian female individual[367]. Despite being a tumor cell line, U2OS have similar cell cycle regulation compared to normal cells, high transfection efficiency, and have been extensively used to test HIV host-pathogen interactions[368]. In response to viral infection, U2OS express interferon signaling genes with antiviral activity[369]. Specifically, *HERC*5 is an HIV inhibitory factor in U2OS cells which reduces HIV p24 production[370]. It is also important to note that the cGAS-STING signaling pathway is deficient in these cells[371], suggesting the predominant IFN response originates from the *RIGI* or *IFIH1/MDA5* pathways[372]. Overall, the literature suggests that increased expression of ISGs in U2OS cells would inhibit HIV replication.

Secondly, the THP-1 cell line is derived from monocytes from a 1-year-old Caucasian male with acute monocytic leukemia[373]. These cells have been extensively used in HIV research for studying the host-pathogen interaction and viral latency[374]. THP-1 cells have a robust interferon response that inhibits HIV replication *in vitro*[375]. These cells also have a functional cGAS-STING, *RIGI*, and *MDA5* response that restrict viral replication[376,377]. Therefore, based on previous literature, increased *CHD1L* and expression of ISGs would restrict HIV replication in this cell line.

Thirdly, we had previously investigated HIV infection in macrophages derived from the iPSC KOLF2 cell line using the CytoTune 1 Sendai method[378], where *CHD1L* knockout increased HIV p24 production[1]. While studies of HIV infection in iPSDMs are limited, as of 2024, current evidence suggests a similar response to HIV as monocyte-derived macrophages (MDM)[379]. In MDMs, HIV infection causes over a 2-fold increase of *CXCL10,DDX60, EPSTI1, HERC5, IFI44, IFI44L, IFIT1, IFIT2, IFIT3, IFITM1,JUP, MX1,*

OAS3, RSAD2, and *SIGELC1*[380], exhibiting significant overlap with our analysis in Chapter 4.

Lastly, the Jurkat cell line is an immortalized T cell line commonly used in HIV studies but have different gene expression profiles than primary CD4+ T cells[381]. Jurkat cells express very little *cGAS* expression[377] and produce a deficient interferon response that does not inhibit viral replication, suggesting that key antiviral ISGs are not expressed[375]. Other T cell lines including 293T, SupT1, and CEM cells show similar reductions of innate immunity gene profiles to Jurkat, suggesting these cell types are likely not suitable for detection of *CHD1L*-mediated HIV inhibition[382].

Taken together, these results suggest that if CHD1L-mediated HIV inhibition is dependent on interferon signaling, that no effect would be seen in Jurkat T cells but would be present in U2OS, THP-1 and iPSDMs. Therefore, follow-up studies of genetically regulated *CHD1L* expression and *CHD1L*-mediated HIV inhibition should also be investigated in non-myeloid cell lines that have a functional interferon response.

Relationship between HIV spVL associated variants, HIV replication, and HIV spVL

During infection of monocytes, macrophages, and activated CD4+ T cells, a strong type I IFN response is associated with decreased HIV replication *in vitro*[383–385]. However, *in vivo* studies suggest sustained interferon signaling is detrimental in HIV pathogenesis. In ART-negative PLWH, IFN-α treatment reduced HIV viral load after week 12, but no change in viral load was observed at week 18 compared to week 0 and a large increase in viral load from week 12 to week 18[386]. Additionally, failure of ART to suppress viral load has been associated with elevated IFN-α and *CXCL10*[387]. A chronic pro-inflammatory cytokine profile, characterized by increased levels of *CXCL10*, is associated with higher HIV viral loads during chronic infection and decreased CD4+ T cell counts[388].

When comparing gene expression profiles between PLWH based on viral load, high viral loads were associated with increased expression of *DDX60*, *EPSTI1*, *GBP4/5*, *IFI44*, *IFI44L*, *IFIT1*, *IFIT3*, *IFITM1*, *MX1*, and *OAS3*[389]. During chronic infection, ISGs such as *IFI44*, *IFI44L*, *RSAD2*, *IFITM1*, *EIF2AK2*, *MX1*, *EPSTI1*, and *IFIT3* remain

persistently upregulated, suggesting that these genes are unable to control HIV infection and are indicative of a chronic inflammatory profile[390]. Rapid HIV disease progression is also associated with elevated *EIF2AK2*, *EPSTI1*, *IFI44*, *IFI44L*, *IFITM1*, *MX1*, and *OAS3*[391], whereas HIV long term non-progressors have reduced ISG expression[392]. These findings highlight that sustained interferon signaling is associated with faster disease progression and suggests CHD1L-mediated interferon signaling could explain the differences between *in vitro* HIV replication and *in vivo* HIV spVL.

The potential role of anti-*CHD1L* drugs in HIV pathogenesis

Despite the strong association between *CHD1L* and cancer development, there are currently no approved anti-*CHD1L* therapeutics but there are candidates in development[393,394]. As we previously identified in Chapter 4, HIV spVL associated variants influence the expression of both *CHD1L* and ISGs suggesting a putative role of *CHD1L* in regulating inflammation. This hypothesis is supported by observations that *CHD1L* knockdown is associated with reduced inflammation[317] and highlights the potential of anti-inflammatory drugs for treatment of chronic HIV infection.

Anti-inflammatory drugs have been previously investigated for the potential to reduce new infections and improve chronic infection outcomes[89]. Additionally, non-steroidal anti-inflammatory drugs have been shown to improve the efficacy of NRTIs[395], indicating anti-inflammatory drugs may improve combination therapies. However, this effect is likely dependent on the specific mechanisms by which inflammation is regulated, as the anti-inflammatory prednisolone was not shown to improve HIV disease progression after 12 months of treatment[396].

In cancer, *CHD1L* expression drives increases in chromosome instability, metastasis, and decreases apoptosis of cancer cells[222]. For PLWH, the development of AIDS is also associated with the development of non-Hodgkin's lymphoma, Kaposi's sarcoma, cervical cancer, and hepatocellular carcinoma[397]. Overexpression of *CHD1L* drives resistance to cisplatin, a chemotherapy drug, contributing to multidrug resistance and poor prognosis of cancer therapy[398], suggesting anti-*CHD1L* drugs may also be effective at controlling HIV-associated comorbidities. Therefore, while there are currently

no approved anti-*CHD1L* therapeutics, our study generates new hypotheses for subsequent follow-up.

Study limitations

There are a few limitations to this study that should be acknowledged. A main limitation is the use of microarrays for conducting GWAS in African populations, as this strategy relies heavily on imputation of genetic variants from a reference panel. Here we used the 1000 Genomes Phase 3 and the African genome variation project to impute genetic variants, as opposed to the TOPMED imputation server, which may have limited our ability to detect low frequency, African-specific genetic variants[399]. Issues in imputation could partially explain the lack of colocalization between eQTLs and GWAS variants[1]. This could also be reflected in fine-mapping analyses as low frequency African-specific variants may not be reflected in the 1000 Genomes Phase 3 and the African genome variation project reference panel. However, by selecting the lead variant within haplotypes when fine-mapping, these issues are likely mitigated as the haplotypes capture a large proportion of genetic variation in the region.

Another limitation is the low minor allele frequency of the candidate causal GWAS variants which limited the ability to obtain samples from homozygous alternate individuals. The lead GWAS variant, rs59784663-A-G, has a minor allele frequency of 0.048[232] whereas likely causal variants (rs7525622-G-A, rs73004025-C-T, rs72999655-A-G) have minor allele frequencies of between 0.047 – 0.083 in African/African American populations. Identifying an individual with homozygous alternate allele at all three loci would require screening over 448 individuals, and to identify at least 10 individuals, would require sample sizes over 4,400. As heterozygous individuals exhibit an intermediate phenotype, this severely hampered our power to detect significant differences.

Study Conclusions

The study of HIV host genetics has consistently implicated the *CCR5* and *HLA* region in European populations[198,400] but GWAS have lacked sufficient sample sizes to perform similar studies in African populations. In this body of work, we identified a novel region on chromosome 1 significantly associated with HIV control. We identified that HIV spVL associated variants linked to lower spVL are associated with decreased *CHD1L*

expression in monocytes. We also identify that HIV spVL associated variants reduce expression of interferon stimulating genes, suggesting a putative model of *CHD1L*-mediated interferon signaling that mediates HIV spVL. Overall, these results identify new insights to HIV host genomics and that investigation into *CHD1L* inhibitors, as opposed to *CHD1L* activators, may have potential as host-directed antivirals for treating HIV disease.

Future Directions

In this body of work, a novel region of chromosome 1 was associated with HIV infection in African populations, signifying a major advancement in our understanding of host-pathogen interaction in diverse populations. The relationship between genetic variants, regulatory features, gene expression and HIV spVL was investigated using bioinformatics, genomics, and molecular biology approaches implicating *CHD1L* and *PRKAB2* as important contributors to HIV pathogenesis. Then, a putative mechanism was identified by which genetic variants are associated with an altered interferon signaling response. These main findings are strong candidates for future research grants and graduate student projects, specifically to validate the effect of the candidate causal variants and validate the effect of the genetic variants on the interferon response.

Validating the Effect of Candidate Causal Variants

Firstly, the mechanism by which rs72999655-A-G, rs7525622-G-A, and rs7300425-C-T influence *CHD1L* expression needs to be experimentally determined. Given that these variants do not change amino acids in coding regions but instead overlap with promoter, enhancer, and TFBS regions, subsequent studies should investigate how variants alter regulatory features and gene expression. There are several different approaches that can be done to assess this mechanism.

Firstly, the advent of CRISPR-Cas9 prime-editing now allows researchers to induce single base-pair variations into the genome[401], allowing the specific effects of rs72999655-A-G, rs7525622-G-A, and rs7300425-C-T on gene expression to be assessed. However, one of the major downfalls of this approach is that while rs72999655-A-G, rs7525622-G-A, and rs7300425-C-T are likely driving the effect, there is a non-zero possibility that it is a variant or variants within their associated haploblock that are altering

gene expression. Therefore, even after subjecting cells to many rounds of transfection to ensure the presence of cell lines heterozygous and homozygous alternate for all alleles, this approach might not accurately capture the true causal variant. Additionally, our variants of interest were identified to influence gene expression in a cell-type-specific manner, which would require inducing heterozygous and homozygous CRISPR-edited cell lines across various immune cell subsets. This process is likely to take well over a year to implement and validate, if variants within the associated haploblocks are contributing to the effect, these would be missed.

Secondly, instead of using CRISPR-Cas9 to induce to test the individual variants of interest, a dual-luciferase reporter gene assay system can directly assess the effect of individual variants, or combinations of variants, on gene expression to validate associations with enhancer and promoter sequences across multiple cell lines[402,403]. Specifically, massively parallel reporter assays (MPRA) can screen over 30,000 variants at once[404], effectively allowing researchers to prioritize an entire GWAS associated region to determine which variants alter promoters, enhancers, and TFBS. Therefore, the use of MPRA can validate our associations in Chapter two, provide additional confidence to the true causal variants, and provide strong candidates for CRISPR-Cas9 prime-editing.

Validating the Effect of Variants on Interferon Signaling

Next, a follow-up of the transcriptome analysis should be conducted in wild-type, *CHD1L* knockout, and *PRKAB2* knockout in THP-1 cells, in HIV- and HIV+ conditions, to determine the downstream effect on interferon signaling. This is important to confirm the effect of *CHD1L* and *PRKAB2* on HIV infection, as well as overlapping downstream pathways to determine the mechanism by which the chromosome 1 region influences HIV pathogenesis. While this could be done in primary cells, the use of a cell line would limit biological variability, improve statistical power, and reduce the likelihood of false positives.

Once pathways have been identified (e.g. interferon signaling), then the genes with the largest fold-changes can be prioritized for follow-up analysis. Specifically, analysis of large publicly available databases like gnomAD, which contains genotype information from over 800,000 individuals, can be assessed using the gene nonessentiality and loss-of-function identifier (gNALI) to determine whether the genes of interest are likely

nonessential for human development and well-being[405]. Additionally, by cross-referencing these candidate genes with known small molecule or clinically-approved drug targets can help accelerate the potential for identifying new HDAs against HIV infection.

However, if there are no strong candidates identified for currently-approved therapeutics, performing a natural products screen can test thousands of metabolites for their effect on both the interferon response and HIV infection. These screens have provided valuable insight into chemical compounds that restrict viral replication[406,407] and by understanding the molecular chemistry of natural products, novel therapeutics can be designed with increased efficacy and safety[408]. Through these follow-up studies, we can improve our understanding of the genetic association identified here and better inform clinical applications.

References

1. McLaren, P. J. *et al.* Africa-specific human genetic variation near CHD1L associates with HIV-1 load. *Nature* **620**, 1025–1030 (2023).

2. Barré-Sinoussi, F. *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science (New York, N.Y.)* **220**, 868–71 (1983).

3. Centers for Disease Control (CDC). Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. *MMWR. Morbidity and mortality weekly report* **30**, 305–8 (1981).

4. Bacchetti, P. *et al.* Survival patterns of the first 500 patients with AIDS in San Francisco. *The Journal of infectious diseases* **157**, 1044–7 (1988).

5. Rothenberg, R. *et al.* Survival with the acquired immunodeficiency syndrome. Experience with 5833 cases in New York City. *The New England journal of medicine* **317**, 1297–302 (1987).

6. Poorolajal, J., Hooshmand, E., Mahjub, H., Esmailnasab, N. & Jenabi, E. Survival rate of AIDS disease and mortality in HIV-infected patients: a meta-analysis. *Public health* **139**, 3–12 (2016).

7. United Nations General Assembly. 2024 global AIDS report — The Urgency of Now: AIDS at a Crossroads. https://www.unaids.org/sites/default/files/media_asset/2024-unaids-global-aids-update_en.pdf.

8. United Nations General Assembly. Global HIV & AIDS statistics — Fact sheet. https://www.unaids.org/en/resources/fact-sheet (2023).

9. McMahon, J. H., Elliott, J. H., Bertagnolio, S., Kubiak, R. & Jordan, M. R. Viral suppression after 12 months of antiretroviral therapy in low- and middle-income countries: a systematic review. *Bulletin of the World Health Organization* **91**, 377-385E (2013).

10. Mwau, M. *et al.* Scale-up of Kenya's national HIV viral load program: Findings and lessons learned. *PloS one* **13**, e0190659 (2018).

11. Han, W. M. *et al.* Global estimates of viral suppression in children and adolescents and adults on antiretroviral therapy adjusted for missing viral load measurements: a multiregional, retrospective cohort study in 31 countries. *The lancet. HIV* **8**, e766–e775 (2021).

12. Simon, F. *et al.* Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nature medicine* **4**, 1032–7 (1998).

13. Vallari, A. *et al.* Confirmation of putative HIV-1 group P in Cameroon. *Journal of virology* **85**, 1403–7 (2011).

14. Hemelaar, J. *et al.* Global and regional molecular epidemiology of HIV-1, 1990-2015: a systematic review, global survey, and trend analysis. *The Lancet. Infectious diseases* **19**, 143–155 (2019).

15. Hemelaar, J., Gouws, E., Ghys, P. D. & Osmanov, S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS (London, England)* **20**, W13-23 (2006).

16. Williams, A. *et al.* Geographic and Population Distributions of Human Immunodeficiency Virus (HIV)-1 and HIV-2 Circulating Subtypes: A Systematic Literature Review and Meta-analysis (2010-2021). *The Journal of infectious diseases* **228**, 1583–1591 (2023).

17. Santos, A. F. & Soares, M. A. HIV Genetic Diversity and Drug Resistance. *Viruses* **2**, 503–531 (2010).

18. United Nations General Assembly. UNAIDS data 2022. https://www.unaids.org/sites/default/files/media_asset/data-book-2022_en.pdf.

19. Gupta, S., Granich, R. & Williams, B. G. Update on treatment as prevention of HIV illness, death, and transmission: sub-Saharan Africa HIV financing and progress towards the 95-95-95 target. *Current opinion in HIV and AIDS* **17**, 368–373 (2022).

20. United Nations General Assembly. 2020 Global AIDS Update — Seizing the moment — Tackling entrenched inequalities to end epidemics. https://www.unaids.org/sites/default/files/media_asset/2020_global-aids-report_en.pdf (2020).

21. Public Health Agency of Canada. HIV in Canada: 2023 surveillance highlights. https://www.canada.ca/en/public-health/services/publications/diseases-conditions/hiv-2023-surveillance-highlights-infographic.html (2023).

22. Public Health Agency of Canada. Canada's progress towards ending the HIV epidemic. https://www.canada.ca/en/public-health/services/publications/diseases-conditions/canada-progress-towards-ending-hiv-epidemic.html (2024).

23. Canada, P. H. A. of. HIV in Canada: 2022 surveillance highlights. https://www.canada.ca/en/public-health/services/publications/diseases-conditions/hiv-2022-surveillance-highlights.html (2022).

24. CHABAC. African, Caribbean and Black Canadian HIV/AIDS Awareness Day – CHABAC, 2020. https://icad-cisd.com/Awareness Day Fact Sheet_2020_EN.pdf (2020).

25. Public Health Agency of Canada. Estimates of HIV incidence, prevalence and Canada's progress on meeting the 90-90-90 HIV targets. https://www.canada.ca/en/public-health/services/publications/diseases-conditions/summary-estimates-hiv-incidence-prevalence-canadas-progress-90-90-90.html (2020).

26. McClarty, L. M., Blanchard, J. F. & Becker, M. L. Leaving no one behind? An equity analysis of the HIV care cascade among a cohort of people living with HIV in Manitoba, Canada. *BMC public health* **21**, 281 (2021).

27. Broadhead, R. L. Transmission of HIV-1 in breast milk. *Tropical doctor* **26**, 100–1 (1996).

28. Beyrer, C. HIV epidemiology update and transmission factors: risks and risk contexts--16th International AIDS Conference epidemiology plenary. *Clinical*

infectious diseases : an official publication of the Infectious Diseases Society of America **44**, 981–7 (2007).

29. Shaw, G. M. & Hunter, E. HIV transmission. *Cold Spring Harbor perspectives in medicine* **2**, (2012).

30. Quinn, T. C. *et al.* Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. *The New England journal of medicine* **342**, 921–9 (2000).

31. Powers, K. A., Poole, C., Pettifor, A. E. & Cohen, M. S. Rethinking the heterosexual infectivity of HIV-1: a systematic review and meta-analysis. *The Lancet. Infectious diseases* **8**, 553–63 (2008).

32. Boily, M.-C. *et al.* Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies. *The Lancet. Infectious diseases* **9**, 118–29 (2009).

33. Secco, A. A. *et al.* Sexually transmitted infections in persons living with HIV infection and estimated HIV transmission risk: trends over time from the DC Cohort. *Sexually transmitted infections* **96**, 89–95 (2020).

34. Kourtis, A. P. & Bulterys, M. Mother-to-child transmission of HIV: pathogenesis, mechanisms and pathways. *Clinics in perinatology* **37**, 721–37, vii (2010).

35. Fiscus, S. A. *et al.* Trends in interventions to reduce perinatal human immunodeficiency virus type 1 transmission in North Carolina. *The Pediatric infectious disease journal* **21**, 664–8 (2002).

36. Warszawski, J. *et al.* Mother-to-child HIV transmission despite antiretroviral therapy in the ANRS French Perinatal Cohort. *AIDS (London, England)* **22**, 289–99 (2008).

37. Townsend, C. L. *et al.* Low rates of mother-to-child transmission of HIV following effective pregnancy interventions in the United Kingdom and Ireland, 2000-2006. *AIDS (London, England)* **22**, 973–81 (2008).

38. Koay, W. L. A. *et al.* Prevention of Perinatal HIV Transmission in an Area of High HIV Prevalence in the United States. *The Journal of pediatrics* **228**, 101–109 (2021).

39. Cardenas, M. C. *et al.* Prevention of the Vertical Transmission of HIV; A Recap of the Journey so Far. *Viruses* **15**, (2023).

40. Ayoo, K., Mikhaeil, J., Huang, A. & Wąsowicz, M. The opioid crisis in North America: facts and future lessons for Europe. *Anaesthesiology intensive therapy* **52**, 139–147 (2020).

41. Tarasuk, J. *et al.* National findings from the Tracks survey of people who inject drugs in Canada, Phase 4, 2017-2019. *Canada communicable disease report = Releve des maladies transmissibles au Canada* **46**, 138–148 (2020).

42. Abdul-Quader, A. S. *et al.* Effectiveness of structural-level needle/syringe programs to reduce HCV and HIV infection among people who inject drugs: a systematic review. *AIDS and behavior* **17**, 2878–92 (2013).

43. Aspinall, E. J. *et al.* Are needle and syringe programmes associated with a reduction in HIV transmission among people who inject drugs: a systematic review and meta-analysis. *International journal of epidemiology* **43**, 235–48 (2014).

44. Smith, M. K. *et al.* Overlap between harm reduction and HIV service utilisation among PWID in India: Implications for HIV combination prevention. *The International journal on drug policy* **57**, 111–118 (2018).

45. Yeh, P. T. *et al.* The Impact of Needle and Syringe Exchange Programs on HIV-Related Risk Behaviors in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis Examining Individual- Versus Community-Level Effects. *AIDS and behavior* **27**, 3306–3331 (2023).

46. Rerks-Ngarm, S. *et al.* Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *The New England journal of medicine* **361**, 2209–20 (2009).

47. Gray, G. E. *et al.* Immune correlates of the Thai RV144 HIV vaccine regimen in South Africa. *Science translational medicine* **11**, (2019).

48. Moodie, Z. *et al.* Analysis of the HIV Vaccine Trials Network 702 Phase 2b-3 HIV-1 Vaccine Trial in South Africa Assessing RV144 Antibody and T-Cell Correlates of HIV-1 Acquisition Risk. *The Journal of infectious diseases* **226**, 246–257 (2022).

49. Gray, G. E. *et al.* Vaccine Efficacy of ALVAC-HIV and Bivalent Subtype C gp120-MF59 in Adults. *The New England journal of medicine* **384**, 1089–1100 (2021).

50. Stover, J. & Teng, Y. The impact of condom use on the HIV epidemic. *Gates open research* **5**, 91 (2021).

51. Phanuphak, N. & Gulick, R. M. HIV treatment and prevention 2019: current standards of care. *Current opinion in HIV and AIDS* **15**, 4–12 (2020).

52. Cohen, M. S. *et al.* Prevention of HIV-1 infection with early antiretroviral therapy. *The New England journal of medicine* **365**, 493–505 (2011).

53. Rodger, A. J. *et al.* Sexual Activity Without Condoms and Risk of HIV Transmission in Serodifferent Couples When the HIV-Positive Partner Is Using Suppressive Antiretroviral Therapy. *JAMA* **316**, 171–81 (2016).

54. Bavinton, B. R. *et al.* Viral suppression and HIV transmission in serodiscordant male couples: an international, prospective, observational, cohort study. *The lancet. HIV* **5**, e438–e447 (2018).

55. Rodger, A. J. *et al.* Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): final results of a multicentre, prospective, observational study. *Lancet (London, England)* **393**, 2428–2438 (2019).

56. Granich, R. M., Gilks, C. F., Dye, C., De Cock, K. M. & Williams, B. G. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. *Lancet (London, England)* **373**, 48–57 (2009).

57. Siegel, K. & Meunier, É. Awareness and Perceived Effectiveness of HIV Treatment as Prevention Among Men Who Have Sex with Men in New York City. *AIDS and behavior* **23**, 1974–1983 (2019).

58. Okoli, C. *et al.* Undetectable equals untransmittable (U = U): awareness and associations with health outcomes among people living with HIV in 25 countries. *Sexually transmitted infections* **97**, 18–26 (2021).

59. Richman, D. D. HIV drug resistance. *AIDS research and human retroviruses* **8**, 1065–71 (1992).

60. Hirsch, M. S. & Kaplan, J. C. Treatment of human immunodeficiency virus infections. *Antimicrobial agents and chemotherapy* **31**, 839–43 (1987).

61. Collier, A. C. *et al.* Combination therapy with zidovudine, didanosine and saquinavir. *Antiviral research* **29**, 99 (1996).

62. Nadler, J. P. Early initiation of antiretroviral therapy for infection with human immunodeficiency virus: considerations in 1996. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **23**, 227–30 (1996).

63. Geneva: World Health Organization. *Consolidated Guidelines on HIV Prevention, Testing, Treatment, Service Delivery and Monitoring: Recommendations for a Public Health Approach*. (2021).

64. Tan, D. H. S. *et al.* Canadian guideline on HIV pre-exposure prophylaxis and nonoccupational postexposure prophylaxis. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* **189**, E1448–E1458 (2017).

65. Bekker, L.-G. *et al.* Twice-Yearly Lenacapavir or Daily F/TAF for HIV Prevention in Cisgender Women. *The New England journal of medicine* **391**, 1179–1192 (2024).

66. Dyer, O. Gilead to license generic lenacapavir for HIV prophylaxis in 120 lower income countries. *BMJ (Clinical research ed.)* **387**, q2200 (2024).

67. Irwin, K. K., Renzette, N., Kowalik, T. F. & Jensen, J. D. Antiviral drug resistance as an adaptive process. *Virus evolution* **2**, vew014 (2016).

68. Li, D. K. & Chung, R. T. Overview of Direct-Acting Antiviral Drugs and Drug Resistance of Hepatitis C Virus. *Methods in molecular biology (Clifton, N.J.)* **1911**, 3–32 (2019).

69. TenoRes Study Group. Global epidemiology of drug resistance after failure of WHO recommended first-line regimens for adult HIV-1 infection: a multicentre retrospective cohort study. *The Lancet. Infectious diseases* **16**, 565–575 (2016).

70. Tao, K. *et al.* Treatment Emergent Dolutegravir Resistance Mutations in Individuals Naïve to HIV-1 Integrase Inhibitors: A Rapid Scoping Review. *Viruses* **15**, (2023).

71. Chu, C. *et al.* Prevalence of Emergent Dolutegravir Resistance Mutations in People Living with HIV: A Rapid Scoping Review. *Viruses* **16**, (2024).

72. Wittkop, L. *et al.* Effect of transmitted drug resistance on virological and immunological response to initial combination antiretroviral therapy for HIV (EuroCoord-CHAIN joint project): a European multicohort study. *The Lancet. Infectious diseases* **11**, 363–71 (2011).

73. Murphy, R. A. *et al.* Dolutegravir Resistance in African Programmatic Settings Among Patients With Failure of Dolutegravir-based ART. *Open forum infectious diseases* **11**, ofae321 (2024).

74. Younger, J. *et al.* Tenofovir and emtricitabine resistance among antiretroviral-naive patients in the Canadian Observational Cohort Collaboration: implications for PrEP. *Antiviral therapy* **24**, 211–220 (2019).

75. Jennings, L. *et al.* Drug Resistance, Rather than Low Tenofovir Levels in Blood or Urine, Is Associated with Tenofovir, Emtricitabine, and Efavirenz Failure in Resource-Limited Settings. *AIDS research and human retroviruses* **38**, 455–462 (2022).

76. Temereanca, A. & Ruta, S. Strategies to overcome HIV drug resistance-current and future perspectives. *Frontiers in microbiology* **14**, 1133407 (2023).

77. Chaudhuri, S., Symons, J. A. & Deval, J. Innovation and trends in the development and approval of antiviral medicines: 1987-2017 and beyond. *Antiviral research* **155**, 76–88 (2018).

78. Huang, Y. *et al.* The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nature medicine* **2**, 1240–3 (1996).

79. Garred, P. *et al.* Dual effect of CCR5 delta 32 gene deletion in HIV-1-infected patients. Copenhagen AIDS Study Group. *Lancet (London, England)* **349**, 1884 (1997).

80. Ni, J., Wang, D. & Wang, S. The CCR5-Delta32 Genetic Polymorphism and HIV-1 Infection Susceptibility: a Meta-analysis. *Open medicine (Warsaw, Poland)* **13**, 467–474 (2018).

81. Qidwai, T. Chemokine genetic polymorphism in human health and disease. *Immunology letters* **176**, 128–38 (2016).

82. Clerici, M. *et al.* Cell-mediated immune response to human immunodeficiency virus (HIV) type 1 in seronegative homosexual men with recent sexual exposure to HIV-1. *The Journal of infectious diseases* **165**, 1012–9 (1992).

83. Paxton, W. A. *et al.* Relative resistance to HIV-1 infection of CD4 lymphocytes from persons who remain uninfected despite multiple high-risk sexual exposure. *Nature medicine* **2**, 412–7 (1996).

84. Fätkenheuer, G. *et al.* Efficacy of short-term monotherapy with maraviroc, a new CCR5 antagonist, in patients infected with HIV-1. *Nature medicine* **11**, 1170–2 (2005).

85. Haile, W. B. *et al.* The Janus kinase inhibitor ruxolitinib reduces HIV replication in human macrophages and ameliorates HIV encephalitis in a murine model. *Neurobiology of disease* **92**, 137–43 (2016).

86. del Real, G. *et al.* Statins inhibit HIV-1 infection by down-regulating Rho activity. *The Journal of experimental medicine* **200**, 541–7 (2004).

87. Montoya, C. J. *et al.* Antiretroviral effect of lovastatin on HIV-1-infected individuals without highly active antiretroviral therapy (The LIVE study): a phase-II randomized clinical trial. *Trials* **10**, 41 (2009).

88. Lajoie, J. *et al.* Using safe, affordable and accessible non-steroidal anti-inflammatory drugs to reduce the number of HIV target cells in the blood and at the female genital tract. *Journal of the International AIDS Society* **21**, e25150 (2018).

89. Cromarty, R. & Archary, D. Inflammation, HIV, and Immune Quiescence: Leveraging on Immunomodulatory Products to Reduce HIV Susceptibility. *AIDS research and treatment* **2020**, 8672850 (2020).

90. Schneider, W. M., Chevillotte, M. D. & Rice, C. M. Interferon-stimulated genes: a complex web of host defenses. *Annual review of immunology* **32**, 513–45 (2014).

91. Boyette, L. B. *et al.* Phenotype, function, and differentiation potential of human monocyte subsets. *PloS one* **12**, e0176460 (2017).

92. Dinarello, C. A. Proinflammatory cytokines. *Chest* **118**, 503–8 (2000).

93. Ivashkiv, L. B. Epigenetic regulation of macrophage polarization and function. *Trends in immunology* **34**, 216–23 (2013).

94. Martinez, F. O. & Gordon, S. The M1 and M2 paradigm of macrophage activation: time for reassessment. *F1000prime reports* **6**, 13 (2014).

95. Chistiakov, D. A., Sobenin, I. A., Orekhov, A. N. & Bobryshev, Y. V. Myeloid dendritic cells: Development, functions, and role in atherosclerotic inflammation. *Immunobiology* **220**, 833–44 (2015).

96. Segura, E. *et al.* Human inflammatory dendritic cells induce Th17 cell differentiation. *Immunity* **38**, 336–48 (2013).

97. Hudson, L. E. & Allen, R. L. Leukocyte Ig-Like Receptors - A Model for MHC Class I Disease Associations. *Frontiers in immunology* **7**, 281 (2016).

98. Berger, A. Th1 and Th2 responses: what are they? *BMJ (Clinical research ed.)* **321**, 424 (2000).

99. Demers, K. R. *et al.* Temporal Dynamics of CD8+ T Cell Effector Responses during Primary HIV Infection. *PLoS pathogens* **12**, e1005805 (2016).

100. Butcher, M. J. & Zhu, J. Recent advances in understanding the Th1/Th2 effector choice. *Faculty reviews* **10**, 30 (2021).

101. Chandler, T. L., Yang, A., Otero, C. E., Permar, S. R. & Caddy, S. L. Protective mechanisms of nonneutralizing antiviral antibodies. *PLoS pathogens* **19**, e1011670 (2023).

102. Haynes, B. F. *et al.* Strategies for HIV-1 vaccines that induce broadly neutralizing antibodies. *Nature reviews. Immunology* **23**, 142–158 (2023).

103. Hraber, P. *et al.* Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. *AIDS (London, England)* **28**, 163–9 (2014).

104. Sagar, M. HIV-1 transmission biology: selection and characteristics of infecting viruses. *The Journal of infectious diseases* **202 Suppl**, S289-96 (2010).

105. Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 7552–7 (2008).

106. Hladik, F. & McElrath, M. J. Setting the stage: host invasion by HIV. *Nature reviews. Immunology* **8**, 447–57 (2008).

107. Martín-Moreno, A. & Muñoz-Fernández, M. A. Dendritic Cells, the Double Agent in the War Against HIV-1. *Frontiers in immunology* **10**, 2485 (2019).

108. Lee, B., Sharron, M., Montaner, L. J., Weissman, D. & Doms, R. W. Quantification of CD4, CCR5, and CXCR4 levels on lymphocyte subsets, dendritic cells, and differentially conditioned monocyte-derived macrophages. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 5215–20 (1999).

109. Le Douce, V., Herbein, G., Rohr, O. & Schwartz, C. Molecular mechanisms of HIV-1 persistence in the monocyte-macrophage lineage. *Retrovirology* **7**, 32 (2010).

110. Wu, L. & KewalRamani, V. N. Dendritic-cell interactions with HIV: infection and viral dissemination. *Nature reviews. Immunology* **6**, 859–68 (2006).

111. Gutiérrez-Martínez, E. *et al.* Actin-regulated Siglec-1 nanoclustering influences HIV-1 capture and virus-containing compartment formation in dendritic cells. *eLife* **12**, (2023).

112. Robb, M. L. *et al.* Prospective Study of Acute HIV-1 Infection in Adults in East Africa and Thailand. *The New England journal of medicine* **374**, 2120–30 (2016).

113. Stacey, A. R. *et al.* Induction of a striking systemic cytokine cascade prior to peak viremia in acute human immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections. *Journal of virology* **83**, 3719–33 (2009).

114. Whitney, J. B. *et al.* Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature* **512**, 74–7 (2014).

115. Chavez, L., Calvanese, V. & Verdin, E. HIV Latency Is Established Directly and Early in Both Resting and Activated Primary CD4 T Cells. *PLoS pathogens* **11**, e1004955 (2015).

116. Fideli US *et al.* Virologic and immunologic determinants of heterosexual transmission of human immunodeficiency virus type 1 in Africa. *AIDS research and human retroviruses* **17**, 901–10 (2001).

117. Borrow, P., Lewicki, H., Hahn, B. H., Shaw, G. M. & Oldstone, M. B. Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *Journal of virology* **68**, 6103–10 (1994).

118. Mellors, J. W. *et al.* Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science (New York, N.Y.)* **272**, 1167–70 (1996).

119. Lyles, R. H. *et al.* Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. Multicenter AIDS Cohort Study. *The Journal of infectious diseases* **181**, 872–80 (2000).

120. Wilson, J. D. *et al.* Direct visualization of HIV-1-specific cytotoxic T lymphocytes during primary infection. *AIDS (London, England)* **14**, 225–33 (2000).

121. Bernardin, F., Kong, D., Peddada, L., Baxter-Lowe, L. A. & Delwart, E. Human immunodeficiency virus mutations during the first month of infection are

preferentially found in known cytotoxic T-lymphocyte epitopes. *Journal of virology* **79**, 11523–8 (2005).

122. Rodríguez, B. *et al.* Predictive value of plasma HIV RNA level on rate of CD4 T-cell decline in untreated HIV infection. *JAMA* **296**, 1498–506 (2006).

123. Connell, B. J. *et al.* Immune activation correlates with and predicts CXCR4 co-receptor tropism switch in HIV-1 infection. *Scientific reports* **10**, 15866 (2020).

124. Moonis, M., Lee, B., Bailer, R. T., Luo, Q. & Montaner, L. J. CCR5 and CXCR4 expression correlated with X4 and R5 HIV-1 infection yet not sustained replication in Th1 and Th2 cells. *AIDS (London, England)* **15**, 1941–9 (2001).

125. Connor, R. I., Sheridan, K. E., Ceradini, D., Choe, S. & Landau, N. R. Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. *The Journal of experimental medicine* **185**, 621–8 (1997).

126. Schneider, E. *et al.* Revised surveillance case definitions for HIV infection among adults, adolescents, and children aged <18 months and for HIV infection and AIDS among children aged 18 months to <13 years--United States, 2008. *MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports* **57**, 1–12 (2008).

127. Deeks, S. G. & Walker, B. D. Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity* **27**, 406–16 (2007).

128. Olson, A. D. *et al.* Evaluation of rapid progressors in HIV infection as an extreme phenotype. *Journal of acquired immune deficiency syndromes (1999)* **67**, 15–21 (2014).

129. Tough, R. H. & McLaren, P. J. Interaction of the Host and Viral Genome and Their Influence on HIV Disease. *Frontiers in genetics* **9**, 720 (2018).

130. Roberts, L. *et al.* Genital tract inflammation during early HIV-1 infection predicts higher plasma viral load set point in women. *The Journal of infectious diseases* **205**, 194–203 (2012).

131. Hollingsworth, T. D. *et al.* HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda. *PLoS pathogens* **6**, e1000876 (2010).

132. Goodreau, S. M. *et al.* Why does age at HIV infection correlate with set point viral load? An evolutionary hypothesis. *Epidemics* **41**, 100629 (2022).

133. Ron, R. *et al.* CD4/CD8 Ratio During Human Immunodeficiency Virus Treatment: Time for Routine Monitoring? *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **76**, 1688–1696 (2023).

134. Janes, H. *et al.* HIV-1 infections with multiple founders are associated with higher viral loads than infections with single founders. *Nature medicine* **21**, 1139–41 (2015).

135. Alizon, S. *et al.* Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. *PLoS pathogens* **6**, e1001123 (2010).

136. Blanquart, F. *et al.* Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. *PLoS biology* **15**, e2001855 (2017).

137. Carrington, M. & O'Brien, S. J. The Influence of *HLA* Genotype on AIDS. *Annual Review of Medicine* **54**, 535–551 (2003).

138. Dean, M. *et al.* 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. *Science (New York, N.Y.)* **273**, 1856–62 (1996).

139. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).

140. Collins, F. S. & Mansoura, M. K. The Human Genome Project. Revealing the shared inheritance of all humankind. *Cancer* **91**, 221–5 (2001).

141. Stenson, P. D. *et al.* The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic

diagnosis and next-generation sequencing studies. *Human genetics* **136**, 665–677 (2017).

142. Vihinen, M. When a Synonymous Variant Is Nonsynonymous. *Genes* **13**, (2022).

143. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science (New York, N.Y.)* **335**, 823–8 (2012).

144. Bochud, P.-Y. *et al.* Polymorphisms in Toll-like receptor 9 influence the clinical course of HIV-1 infection. *AIDS (London, England)* **21**, 441–6 (2007).

145. Shaikh, N., Nirmalkar, A. & Thakar, M. Polymorphisms in Toll-Like Receptors (TLRs)-7 and 9 Genes in Indian Population with Progressive and Nonprogressive HIV-1 Infection. *AIDS research and human retroviruses* **35**, 577–582 (2019).

146. Liu, R. *et al.* Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367–77 (1996).

147. Quillent, C. *et al.* HIV-1-resistance phenotype conferred by combination of two separate inherited mutations of CCR5 gene. *Lancet (London, England)* **351**, 14–8 (1998).

148. Sakharkar, M. K., Chow, V. T. K. & Kangueane, P. Distributions of exons and introns in the human genome. *In silico biology* **4**, 387–93 (2004).

149. de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A. & Pollock, D. D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS genetics* **7**, e1002384 (2011).

150. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–80 (2012).

151. Broekema, R. V, Bakker, O. B. & Jonkers, I. H. A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open biology* **10**, 190221 (2020).

152. Lu, Q., Powles, R. L., Wang, Q., He, B. J. & Zhao, H. Integrative Tissue-Specific Functional Annotations in the Human Genome Provide Novel Insights on Many

Complex Traits and Improve Signal Prioritization in Genome Wide Association Studies. *PLoS genetics* **12**, e1005947 (2016).

153. He, L., Loika, Y. & Kulminski, A. M. Allele-specific analysis reveals exon- and cell-type-specific regulatory effects of Alzheimer's disease-associated genetic variants. *Translational psychiatry* **12**, 163 (2022).

154. Deng, Q. *et al.* Single-cell chromatin accessibility profiling of cell-state-specific gene regulatory programs during mouse organogenesis. *Frontiers in neuroscience* **17**, 1170355 (2023).

155. Pala, M. *et al.* Population- and individual-specific regulatory variation in Sardinia. *Nature genetics* **49**, 700–707 (2017).

156. Fillmore, G. C. *et al.* Gene expression profiling of cell lines derived from T-cell malignancies. *FEBS letters* **522**, 183–8 (2002).

157. Yang, L. A Practical Guide for Structural Variation Detection in the Human Genome. *Current protocols in human genetics* **107**, e103 (2020).

158. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nature reviews. Genetics* **11**, 499–511 (2010).

159. Mauleekoonphairoj, J. *et al.* A diverse ancestrally-matched reference panel increases genotype imputation accuracy in a underrepresented population. *Scientific reports* **13**, 12360 (2023).

160. Huang, L. *et al.* Genotype-imputation accuracy across worldwide human populations. *American journal of human genetics* **84**, 235–50 (2009).

161. Lin, Y. *et al.* Genotype imputation for Han Chinese population using Haplotype Reference Consortium as reference. *Human genetics* **137**, 431–436 (2018).

162. Bergström, A., Stringer, C., Hajdinjak, M., Scerri, E. M. L. & Skoglund, P. Origins of modern human ancestry. *Nature* **590**, 229–237 (2021).

163. Carlson, C. S. *et al.* Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS biology* **11**, e1001661 (2013).

164. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M. J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genetic epidemiology* **32**, 381–5 (2008).

165. Fadista, J., Manning, A. K., Florez, J. C. & Groop, L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *European journal of human genetics : EJHG* **24**, 1202–5 (2016).

166. McLaren, P. J. *et al.* Polymorphisms of large effect explain the majority of the host genetic contribution to variation of HIV-1 virus load. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 14658–63 (2015).

167. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Archives of general psychiatry* **60**, 1187–92 (2003).

168. van de Bunt, M. *et al.* Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS Loci. *PLoS genetics* **11**, e1005535 (2015).

169. Chua, E. H. Z., Yasar, S. & Harmston, N. The importance of considering regulatory domains in genome-wide analyses - the nearest gene is often wrong! *Biology open* **11**, (2022).

170. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *American journal of human genetics* **101**, 5–22 (2017).

171. Kumar, V., Wijmenga, C. & Withoff, S. From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases. *Seminars in immunopathology* **34**, 567–80 (2012).

172. Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–43 (2015).

173. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).

174. Li, B. & Ritchie, M. D. From GWAS to Gene: Transcriptome-Wide Association Studies and Other Methods to Functionally Understand GWAS Discoveries. *Frontiers in genetics* **12**, 713230 (2021).

175. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature reviews. Genetics* **19**, 491–504 (2018).

176. Waldmann, P., Mészáros, G., Gredler, B., Fuerst, C. & Sölkner, J. Evaluation of the lasso and the elastic net in genome-wide association studies. *Frontiers in genetics* **4**, 270 (2013).

177. Yang, S. *et al.* Prioritizing genetic variants in GWAS with lasso using permutation-assisted tuning. *Bioinformatics (Oxford, England)* **36**, 3811–3817 (2020).

178. Asimit, J. L., Hatzikotoulas, K., McCarthy, M., Morris, A. P. & Zeggini, E. Trans-ethnic study design approaches for fine-mapping. *European journal of human genetics : EJHG* **24**, 1330–6 (2016).

179. Hormozdiari, F., Kostem, E., Kang, E. Y., Pasaniuc, B. & Eskin, E. Identifying causal variants at loci with multiple signals of association. *Genetics* **198**, 497–508 (2014).

180. Chen, W. *et al.* Fine Mapping Causal Variants with an Approximate Bayesian Method Using Marginal Test Statistics. *Genetics* **200**, 719–36 (2015).

181. Risch, N. J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847–56 (2000).

182. Hindorff, L. A., Gillanders, E. M. & Manolio, T. A. Genetic architecture of cancer and other complex diseases: lessons learned and future directions. *Carcinogenesis* **32**, 945–54 (2011).

183. Weedon, M. N. *et al.* Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nature genetics* **46**, 61–64 (2014).

184. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *The New England journal of medicine* **373**, 895–907 (2015).

185. Li, X., Lappalainen, T. & Bussemaker, H. J. Identifying genetic regulatory variants that affect transcription factor activity. *Cell genomics* **3**, 100382 (2023).

186. Ying, P. *et al.* Genome-wide enhancer-gene regulatory maps link causal variants to target genes underlying human cancer risk. *Nature communications* **14**, 5958 (2023).

187. Malkomes, P. *et al.* Selective AKT Inhibition by MK-2206 Represses Colorectal Cancer-Initiating Stem Cells. *Annals of surgical oncology* **23**, 2849–57 (2016).

188. Fellay, J. *et al.* A whole-genome association study of major determinants for host control of HIV-1. *Science (New York, N.Y.)* **317**, 944–7 (2007).

189. Harrer, T. *et al.* Cytotoxic T lymphocytes in asymptomatic long-term nonprogressing HIV-1 infection. Breadth and specificity of the response and relation to in vivo viral quasispecies in a person with prolonged infection and low viral load. *Journal of immunology (Baltimore, Md. : 1950)* **156**, 2616–23 (1996).

190. Kaslow, R. A. *et al.* Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nature medicine* **2**, 405–11 (1996).

191. Limou, S. *et al.* Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *The Journal of infectious diseases* **199**, 419–26 (2009).

192. Le Clerc, S. *et al.* Genomewide association study of a rapid progression cohort identifies new susceptibility alleles for AIDS (ANRS Genomewide Association Study 03). *The Journal of infectious diseases* **200**, 1194–201 (2009).

193. Herbeck, J. T. *et al.* Multistage genomewide association study identifies a locus at 1q41 associated with rate of HIV-1 disease progression to clinical AIDS. *The Journal of infectious diseases* **201**, 618–26 (2010).

194. Pelak, K. *et al.* Host determinants of HIV-1 control in African Americans. *The Journal of infectious diseases* **201**, 1141–9 (2010).

195. van Manen, D. *et al.* Genome-wide association scan in HIV-1-infected individuals identifying variants influencing disease course. *PloS one* **6**, e22208 (2011).

196. Lingappa, J. R. *et al.* 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 (2011).

197. Vilhjálmsson, B. J. & Nordborg, M. The nature of confounding in genome-wide association studies. *Nature reviews. Genetics* **14**, 1–2 (2013).

198. McLaren, P. J. *et al.* Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls. *PLoS pathogens* **9**, e1003515 (2013).

199. Wei, Z. *et al.* Genome-wide association studies of HIV-1 host control in ethnically diverse chinese populations. *Scientific Reports* **5**, 1–9 (2015).

200. Lutz, S., Brion, C., Kliebhan, M. & Albert, F. W. DNA variants affecting the expression of numerous genes in trans have diverse mechanisms of action and evolutionary histories. *PLoS genetics* **15**, e1008375 (2019).

201. Deelen, P. *et al.* Improving the diagnostic yield of exome- sequencing by predicting gene-phenotype associations using large-scale gene expression analysis. *Nature communications* **10**, 2837 (2019).

202. Wu, J. *et al.* Chemoproteomic analysis of intertissue and interspecies isoform diversity of AMP-activated protein kinase (AMPK). *The Journal of biological chemistry* **288**, 35904–12 (2013).

203. Hardie, D. G. Molecular Pathways: Is AMPK a Friend or a Foe in Cancer? *Clinical cancer research : an official journal of the American Association for Cancer Research* **21**, 3836–40 (2015).

204. Carling, D. AMPK signalling in health and disease. *Current opinion in cell biology* **45**, 31–37 (2017).

205. Saravia, J., Raynor, J. L., Chapman, N. M., Lim, S. A. & Chi, H. Signaling networks in immunometabolism. *Cell research* **30**, 328–342 (2020).

206. Wu, S. & Zou, M.-H. AMPK, Mitochondrial Function, and Cardiovascular Disease. *International journal of molecular sciences* **21**, (2020).

207. Bhutta, M. S., Gallo, E. S. & Borenstein, R. Multifaceted Role of AMPK in Viral Infections. *Cells* **10**, (2021).

208. Li, J. *et al.* LncRNAs are involved in regulating ageing and age-related disease through the adenosine monophosphate-activated protein kinase signalling pathway. *Genes & diseases* **11**, 101042 (2024).

209. Ziegler, N. *et al.* AMPKβ1 and AMPKβ2 define an isoform-specific gene signature in human pluripotent stem cells, differentially mediating cardiac lineage specification. *The Journal of biological chemistry* **295**, 17659–17671 (2020).

210. Sun, D. *et al.* Aspirin disrupts the mTOR-Raptor complex and potentiates the anti-cancer activities of sorafenib via mTORC1 inhibition. *Cancer letters* **406**, 105–115 (2017).

211. Han, F. *et al.* The critical role of AMPK in driving Akt activation under stress, tumorigenesis and drug resistance. *Nature communications* **9**, 4728 (2018).

212. Yuan, J., Dong, X., Yap, J. & Hu, J. The MAPK and AMPK signalings: interplay and implication in targeted cancer therapy. *Journal of hematology & oncology* **13**, 113 (2020).

213. Salminen, A., Hyttinen, J. M. T. & Kaarniranta, K. AMP-activated protein kinase inhibits NF-κB signaling and inflammation: impact on healthspan and lifespan. *Journal of molecular medicine (Berlin, Germany)* **89**, 667–76 (2011).

214. Prasad, S. & Tyagi, A. K. Curcumin and its analogues: a potential natural compound against HIV infection and AIDS. *Food & function* **6**, 3412–9 (2015).

215. Koukouritaki, S. B., Simpson, P., Yeung, C. K., Rettie, A. E. & Hines, R. N. Human hepatic flavin-containing monooxygenases 1 (FMO1) and 3 (FMO3) developmental expression. *Pediatric research* **51**, 236–43 (2002).

216. Cashman, J. R. & Zhang, J. Human flavin-containing monooxygenases. *Annual review of pharmacology and toxicology* **46**, 65–100 (2006).

217. Krueger, S. K. & Williams, D. E. Mammalian flavin-containing monooxygenases: structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacology & therapeutics* **106**, 357–87 (2005).

218. Wani, K. A. *et al.* NHR-49/PPAR-α and HLH-30/TFEB cooperate for C. elegans host defense via a flavin-containing monooxygenase. *eLife* **10**, (2021).

219. Kong, L. *et al.* Alcoholic fatty liver disease inhibited the co-expression of Fmo5 and PPARα to activate the NF-κB signaling pathway, thereby reducing liver injury via inducing gut microbiota disturbance. *Journal of experimental & clinical cancer research : CR* **40**, 18 (2021).

220. Phillips, I. R., Veeravalli, S., Khadayate, S. & Shephard, E. A. Metabolomic and transcriptomic analyses of Fmo5-/- mice reveal roles for flavin-containing monooxygenase 5 (FMO5) in NRF2-mediated oxidative stress response, unfolded protein response, lipid homeostasis, and carbohydrate and one-carbon metabolism. *PloS one* **18**, e0286692 (2023).

221. Ma, N.-F. *et al.* Isolation and characterization of a novel oncogene, amplified in liver cancer 1, within a commonly amplified region at 1q21 in hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* **47**, 503–10 (2008).

222. Chen, L. *et al.* Chromodomain helicase/adenosine triphosphatase DNA binding protein 1-like (CHD1l) gene suppresses the nucleus-to-mitochondria translocation of nur77 to sustain hepatocellular carcinoma cell survival. *Hepatology (Baltimore, Md.)* **50**, 122–9 (2009).

223. Ahel, D. *et al.* Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1. *Science (New York, N.Y.)* **325**, 1240–3 (2009).

224. Liu, M. *et al.* CHD1L promotes lineage reversion of hepatocellular carcinoma through opening chromatin for key developmental transcription factors. *Hepatology (Baltimore, Md.)* **63**, 1544–59 (2016).

225. Li, Y. *et al.* SPOCK1 is regulated by CHD1L and blocks apoptosis and promotes HCC cell invasiveness and metastasis in mice. *Gastroenterology* **144**, 179-191.e4 (2013).

226. Chan, T. H. M. *et al.* Translationally controlled tumor protein induces mitotic defects and chromosome missegregation in hepatocellular carcinoma development. *Hepatology (Baltimore, Md.)* **55**, 491–505 (2012).

227. Bommer, U.-A. & Thiele, B.-J. The translationally controlled tumour protein (TCTP). *The international journal of biochemistry & cell biology* **36**, 379–85 (2004).

228. Zhou, H. *et al.* Genome-scale RNAi screen for host factors required for HIV replication. *Cell host & microbe* **4**, 495–504 (2008).

229. Jiang, B. H. *et al.* CHD1L Regulated PARP1-Driven Pluripotency and Chromatin Remodeling during the Early-Stage Cell Reprogramming. *Stem Cells* **33**, 2961–2972 (2015).

230. Rom, S., Reichenbach, N. L., Dykstra, H. & Persidsky, Y. The dual action of poly(ADP-ribose) polymerase -1 (PARP-1) inhibition in HIV-1 infection: HIV-1 LTR inhibition and diminution in Rho GTPase activity. *Frontiers in microbiology* **6**, 878 (2015).

231. Li, Z. *et al.* Gain of LINC00624 Enhances Liver Cancer Progression by Disrupting the Histone Deacetylase 6/Tripartite Motif Containing 28/Zinc Finger Protein 354C Corepressor Complex. *Hepatology (Baltimore, Md.)* **73**, 1764–1782 (2021).

232. Chen, S. *et al.* A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**, 92–100 (2024).

233. Rosenberg, N. A. *et al.* Genome-wide association studies in diverse populations. *Nature reviews. Genetics* **11**, 356–66 (2010).

234. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904–9 (2006).

235. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).

236. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 1–14 (2016).

237. Wagih, O. *et al.* A resource of variant effect predictions of single nucleotide variants in model organisms. *Molecular systems biology* **14**, e8430 (2018).

238. Ng, P. C. & Henikoff, S. Predicting deleterious amino acid substitutions. *Genome research* **11**, 863–74 (2001).

239. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature methods* **7**, 248–9 (2010).

240. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research* **47**, D886–D894 (2019).

241. Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535-548.e24 (2019).

242. Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic acids research* **44**, D877-81 (2016).

243. Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Research* **46**, D794–D801 (2018).

244. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–30 (2015).

245. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (New York, N.Y.)* **348**, 648–60 (2015).

246. Servin, B. & Stephens, M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS genetics* **3**, e114 (2007).

247. Guan, Y. & Stephens, M. Practical issues in imputation-based association mapping. *PLoS genetics* **4**, e1000279 (2008).

248. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics (Oxford, England)* **21**, 263–5 (2005).

249. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science (New York, N.Y.)* **296**, 2225–9 (2002).

250. DeGorter, M. K. *et al.* Transcriptomics and chromatin accessibility in multiple African population samples. *bioRxiv : the preprint server for biology* (2023) doi:10.1101/2023.11.04.564839.

251. Mogil, L. S. *et al.* Genetic architecture of gene expression traits across diverse populations. *PLoS genetics* **14**, e1007586 (2018).

252. Shang, L. *et al.* meQTL mapping in the GENOA study reveals genetic determinants of DNA methylation in African Americans. *Nature communications* **14**, 2711 (2023).

253. Bundo, M. *et al.* A systematic evaluation of whole genome amplification of bisulfite-modified DNA. *Clinical epigenetics* **4**, 22 (2012).

254. Luo, Y. *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic acids research* **48**, D882–D889 (2020).

255. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nature genetics* **51**, 1339–1348 (2019).

256. Claussnitzer, M. *et al.* A brief history of human disease genetics. *Nature* **577**, 179–189 (2020).

257. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science (New York, N.Y.)* **337**, 1190–5 (2012).

258. McManus, J. N. J., Lovelett, R. J., Lowengrub, D. & Christensen, S. A unifying statistical framework to discover disease genes from GWASs. *Cell genomics* **3**, 100264 (2023).

259. Boyes, J. & Bird, A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* **64**, 1123–34 (1991).

260. Newell-Price, J., Clark, A. J. & King, P. DNA methylation and silencing of gene expression. *Trends in endocrinology and metabolism: TEM* **11**, 142–8 (2000).

261. Chatterjee, R. & Vinson, C. CpG methylation recruits sequence specific transcription factors essential for tissue specific gene expression. *Biochimica et biophysica acta* **1819**, 763–70 (2012).

262. Wu, Y. *et al.* Promoter Hypermethylation Promotes the Binding of Transcription Factor NFATc1, Triggering Oncogenic Gene Activation in Pancreatic Cancer. *Cancers* **13**, (2021).

263. Hiatt, J. *et al.* A functional map of HIV-host interactions in primary human T cells. *Nature communications* **13**, 1752 (2022).

264. Zhang, H. S. & Wu, M. R. SIRT1 regulates Tat-induced HIV-1 transactivation through activating AMP-activated protein kinase. *Virus Research* **146**, 51–57 (2009).

265. Zhang, H.-S., Wu, T.-C., Sang, W.-W. & Ruan, Z. MiR-217 is involved in Tat-induced HIV-1 long terminal repeat (LTR) transactivation by down-regulation of SIRT1. *Biochimica et biophysica acta* **1823**, 1017–23 (2012).

266. Moreira, D. *et al.* AMP-activated Protein Kinase As a Target For Pathogens: Friends Or Foes? *Current drug targets* **17**, 942–53 (2016).

267. Jeong, H. W. *et al.* Berberine suppresses proinflammatory responses through AMPK activation in macrophages. *American journal of physiology. Endocrinology and metabolism* **296**, E955-64 (2009).

268. Rutherford, C. *et al.* Phosphorylation of Janus kinase 1 (JAK1) by AMP-activated protein kinase (AMPK) links energy sensing to anti-inflammatory signaling. *Science signaling* **9**, ra109 (2016).

269. Sag, D., Carling, D., Stout, R. D. & Suttles, J. Adenosine 5'-monophosphate-activated protein kinase promotes macrophage polarization to an anti-inflammatory functional phenotype. *Journal of immunology (Baltimore, Md. : 1950)* **181**, 8633–41 (2008).

270. Brass, A. L. *et al.* Identification of host proteins required for HIV infection through a functional genomic screen. *Science (New York, N.Y.)* **319**, 921–6 (2008).

271. Yeung, M. L., Houzet, L., Yedavalli, V. S. R. K. & Jeang, K. T. A genome-wide short hairpin RNA screening of Jurkat T-cells for human proteins contributing to productive HIV-1 replication. *Journal of Biological Chemistry* **284**, 19463–19473 (2009).

272. Park, R. J. *et al.* A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. *Nature genetics* **49**, 193–203 (2017).

273. Sag, D., Carling, D., Stout, R. D. & Suttles, J. Adenosine 5'-monophosphate-activated protein kinase promotes macrophage polarization to an anti-inflammatory functional phenotype. *Journal of immunology (Baltimore, Md. : 1950)* **181**, 8633–41 (2008).

274. Gélinas, J.-F., Gill, D. R. & Hyde, S. C. Multiple Inhibitory Factors Act in the Late Phase of HIV-1 Replication: a Systematic Review of the Literature. *Microbiology and molecular biology reviews : MMBR* **82**, (2018).

275. Tough, R. H., McLaren, P. J. & International Collaboration for the Genomics of HIV. Functionally-informed fine-mapping identifies genetic variants linking increased CHD1L expression and HIV restriction in monocytes. *Scientific reports* **15**, 2325 (2025).

276. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nature methods* **9**, 179–81 (2011).

277. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics* **5**, e1000529 (2009).

278. Raj, T. *et al.* Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science (New York, N.Y.)* **344**, 519–23 (2014).

279. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (New York, N.Y.)* **369**, 1318–1330 (2020).

280. Koga, Y. *et al.* Roles of Cyclic AMP Response Element Binding Activation in the ERK1/2 and p38 MAPK Signalling Pathway in Central Nervous System, Cardiovascular System, Osteoclast Differentiation and Mucin and Cytokine Production. *International journal of molecular sciences* **20**, (2019).

281. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics* **47**, 1091–8 (2015).

282. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44–57 (2009).

283. Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic acids research* **50**, W216–W221 (2022).

284. Szewczuk, M., Boguszewska, K., Kaźmierczak-Barańska, J. & Karwowski, B. T. The role of AMPK in metabolism and its influence on DNA damage repair. *Molecular biology reports* **47**, 9075–9086 (2020).

285. Cantó, C. & Auwerx, J. AMP-activated protein kinase and its downstream transcriptional pathways. *Cellular and molecular life sciences : CMLS* **67**, 3407–23 (2010).

286. Pahlevan Kakhki, M. *et al.* A genetic-epigenetic interplay at 1q21.1 locus underlies CHD1L-mediated vulnerability to primary progressive multiple sclerosis. *Nature communications* **15**, 6419 (2024).

287. Wu, C. L. *et al.* Role of AMPK in UVB-induced DNA damage repair and growth control. *Oncogene* **32**, 2682–9 (2013).

288. Huang, X. *et al.* Cytokine cascade and networks among MSM HIV seroconverters: implications for early immunotherapy. *Scientific reports* **6**, 36234 (2016).

289. Ngcobo, S. *et al.* Pre-infection plasma cytokines and chemokines as predictors of HIV disease progression. *Scientific reports* **12**, 2437 (2022).

290. Wang, H. *et al.* SCG2 is a Prognostic Biomarker Associated With Immune Infiltration and Macrophage Polarization in Colorectal Cancer. *Frontiers in cell and developmental biology* **9**, 795133 (2021).

291. Steinfass, T. *et al.* Secretogranin II influences the assembly and function of MHC class I in melanoma. *Experimental hematology & oncology* **12**, 29 (2023).

292. Ju, S.-W. *et al.* A functional anti-human 4-1BB ligand monoclonal antibody that enhances proliferation of monocytes by reverse signaling of 4-1BBL. *Hybridoma and hybridomics* **22**, 333–8 (2003).

293. Wang, C. *et al.* 4-1BBL induces TNF receptor-associated factor 1-dependent Bim modulation in human T cells and is a critical component in the costimulation-dependent rescue of functionally impaired HIV-specific CD8 T cells. *Journal of immunology (Baltimore, Md. : 1950)* **179**, 8252–63 (2007).

294. Olofsson, P. S. *et al.* CD137 is expressed in human atherosclerosis and promotes development of plaque inflammation in hypercholesterolemic mice. *Circulation* **117**, 1292–301 (2008).

295. Tu, L. *et al.* TGF-β2 interfering oligonucleotides used as adjuvants for microbial vaccines. *Journal of leukocyte biology* **108**, 1673–1692 (2020).

296. Liu, M. *et al.* CHD1L promotes lineage reversion of hepatocellular carcinoma through opening chromatin for key developmental transcription factors. *Hepatology (Baltimore, Md.)* **63**, 1544–59 (2016).

297. Espeseth, A. S. *et al.* siRNA screening of a targeted library of DNA repair factors in HIV infection reveals a role for base excision repair in HIV integration. *PloS one* **6**, e17612 (2011).

298. Pitha, P. M. Innate antiviral response: role in HIV-1 infection. *Viruses* **3**, 1179–203 (2011).

299. Ha, H. C. *et al.* Poly(ADP-ribose) polymerase-1 is required for efficient HIV-1 integration. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 3364–8 (2001).

300. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology* **3**, RESEARCH0034 (2002).

301. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome biology* **8**, R19 (2007).

302. Mane, V. P., Heuer, M. A., Hillyer, P., Navarro, M. B. & Rabin, R. L. Systematic method for determining an ideal housekeeping gene for real-time PCR analysis. *Journal of biomolecular techniques : JBT* **19**, 342–7 (2008).

303. Ledderose, C., Heyn, J., Limbeck, E. & Kreth, S. Selection of reliable reference genes for quantitative real-time PCR in human T cells and neutrophils. *BMC research notes* **4**, 427 (2011).

304. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nature methods* **9**, 179–81 (2011).

305. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics* **5**, e1000529 (2009).

306. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics (Oxford, England)* **20**, 307–15 (2004).

307. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics (Oxford, England)* **8**, 118–27 (2007).

308. Smyth, G. K., Michaud, J. & Scott, H. S. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics (Oxford, England)* **21**, 2067–75 (2005).

309. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology* **3**, Article3 (2004).

310. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research* **45**, D362–D368 (2017).

311. Xiong, X., Lai, X., Li, A., Liu, Z. & Ma, N. Diversity roles of CHD1L in normal cell function and tumorigenesis. *Biomarker Research* **9**, 16 (2021).

312. Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic acids research* **50**, W216–W221 (2022).

313. Fan, G.-T., Ling, Z.-H., He, Z.-W., Wu, S.-J. & Zhou, G.-X. Suppressing CHD1L reduces the proliferation and chemoresistance in osteosarcoma. *Biochemical and biophysical research communications* **554**, 214–221 (2021).

314. Zhang, L.-L. *et al.* Transcriptomic meta-analysis identifies gene expression characteristics in various samples of HIV-infected patients with nonprogressive disease. *J Transl Med* **15**, 191 (2017).

315. Lee, S. Y., Park, Y. K., Yoon, C.-H., Kim, K. & Kim, K.-C. Meta-analysis of gene expression profiles in long-term non-progressors infected with HIV-1. *BMC Med Genomics* **12**, 3 (2019).

316. Liu, Y. *et al.* Methylomics of gene expression in human monocytes. *Human molecular genetics* **22**, 5065–74 (2013).

317. Zhang, X. & Sun, Y. Chromodomain Helicase DNA Binding Protein 1-like, a negative regulator of Forkhead box O3a, promotes the proliferation and migration of Angiotensin II-induced vascular smooth muscle cells. *Bioengineered* **13**, 2597–2609 (2022).

318. Cheng, W., Su, Y. & Xu, F. CHD1L: a novel oncogene. *Molecular cancer* **12**, 170 (2013).

319. Poli, G., Orenstein, J. M., Kinter, A., Folks, T. M. & Fauci, A. S. Interferon-alpha but not AZT suppresses HIV expression in chronically infected cell lines. *Science (New York, N.Y.)* **244**, 575–7 (1989).

320. Herbeuval, J.-P. & Shearer, G. M. HIV-1 immunopathogenesis: how good interferon turns bad. *Clinical immunology (Orlando, Fla.)* **123**, 121–8 (2007).

321. Chang, J. J. *et al.* Higher expression of several interferon-stimulated genes in HIV-1-infected females after adjusting for the level of viral replication. *The Journal of infectious diseases* **208**, 830–8 (2013).

322. Cheng, L. *et al.* Blocking type I interferon signaling enhances T cell recovery and reduces HIV-1 reservoirs. *The Journal of clinical investigation* **127**, 269–279 (2017).

323. Ledderose, C., Heyn, J., Limbeck, E. & Kreth, S. Selection of reliable reference genes for quantitative real-time PCR in human T cells and neutrophils. *BMC research notes* **4**, 427 (2011).

324. Mane, V. P., Heuer, M. A., Hillyer, P., Navarro, M. B. & Rabin, R. L. Systematic method for determining an ideal housekeeping gene for real-time PCR analysis. *Journal of biomolecular techniques : JBT* **19**, 342–7 (2008).

325. Scarpaci, J. L. Help-seeking behavior, use, and satisfaction among frequent primary care users in Santiago de Chile. *Journal of health and social behavior* **29**, 199–213 (1988).

326. Puertas, M. C. *et al.* Pan-resistant HIV-1 emergence in the era of integrase strand-transfer inhibitors: a case report. *The Lancet. Microbe* **1**, e130–e135 (2020).

327. Afonso, J. M., Bello, G., Guimarães, M. L., Sojka, M. & Morgado, M. G. HIV-1 genetic diversity and transmitted drug resistance mutations among patients from the North, Central and South regions of Angola. *PloS one* **7**, e42996 (2012).

328. Rowley, C. F. *et al.* Sharp increase in rates of HIV transmitted drug resistance at antenatal clinics in Botswana demonstrates the need for routine surveillance. *The Journal of antimicrobial chemotherapy* **71**, 1361–6 (2016).

329. An, P. & Winkler, C. A. Host genes associated with HIV/AIDS: advances in gene discovery. *Trends in genetics : TIG* **26**, 119–31 (2010).

330. McLaren, P. J. & Fellay, J. HIV-1 and human genetic variation. *Nature reviews. Genetics* **22**, 645–657 (2021).

331. Wood, A. & Armour, D. The Discovery of the CCR5 Receptor Antagonist, UK-427,857, A New Agent for the Treatment of HIV Infection and AIDS. in *Antiviral Chemistry and Chemotherapy* vol. 20 239–271 (2005).

332. Ivashkiv, L. B. & Donlin, L. T. Regulation of type I interferon responses. *Nature Reviews Immunology* **14**, 36–49 (2014).

333. Ali, S. *et al.* Sources of Type I Interferons in Infectious Immunity: Plasmacytoid Dendritic Cells Not Always in the Driver's Seat. *Frontiers in immunology* **10**, 778 (2019).

334. Ersahin, T., Tuncbag, N. & Cetin-Atalay, R. The PI3K/AKT/mTOR interactive pathway. *Molecular bioSystems* **11**, 1946–54 (2015).

335. Kaur, S. *et al.* Role of the Akt pathway in mRNA translation of interferon-stimulated genes. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 4808–13 (2008).

336. Ezell, S. A. *et al.* The protein kinase Akt1 regulates the interferon response through phosphorylation of the transcriptional repressor EMSY. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E613-21 (2012).

337. Wortzel, I. & Seger, R. The ERK Cascade: Distinct Functions within Various Subcellular Organelles. *Genes & cancer* **2**, 195–209 (2011).

338. David, M. *et al.* Requirement for MAP kinase (ERK2) activity in interferon alpha- and interferon beta-stimulated gene expression through STAT proteins. *Science (New York, N.Y.)* **269**, 1721–3 (1995).

339. Li, Y. *et al.* Role of p38alpha Map kinase in Type I interferon signaling. *The Journal of biological chemistry* **279**, 970–9 (2004).

340. Cohen-Armon, M. *et al.* DNA-independent PARP-1 activation by phosphorylated ERK2 increases Elk1 activity: a link to histone acetylation. *Molecular cell* **25**, 297–308 (2007).

341. Kawashima, I. *et al.* Negative regulation of the LKB1/AMPK pathway by ERK in human acute myeloid leukemia cells. *Experimental hematology* **43**, 524–33.e1 (2015).

342. Koga, Y. *et al.* Roles of Cyclic AMP Response Element Binding Activation in the ERK1/2 and p38 MAPK Signalling Pathway in Central Nervous System, Cardiovascular System, Osteoclast Differentiation and Mucin and Cytokine Production. *International journal of molecular sciences* **20**, (2019).

343. Pétrilli, V. *et al.* Noncleavable poly(ADP-ribose) polymerase-1 regulates the inflammation response in mice. *The Journal of clinical investigation* **114**, 1072–81 (2004).

344. Erener, S. *et al.* Inflammasome-activated caspase 7 cleaves PARP1 to enhance the expression of a subset of NF-κB target genes. *Molecular cell* **46**, 200–11 (2012).

345. Chiu, Y.-H., Macmillan, J. B. & Chen, Z. J. RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell* **138**, 576–91 (2009).

346. Chen, Q. *et al.* Truncated PARP1 mediates ADP-ribosylation of RNA polymerase III for apoptosis. *Cell discovery* **8**, 3 (2022).

347. Poli, G., Orenstein, J. M., Kinter, A., Folks, T. M. & Fauci, A. S. Interferon-alpha but not AZT suppresses HIV expression in chronically infected cell lines. *Science (New York, N.Y.)* **244**, 575–7 (1989).

348. Rusinova, I. *et al.* Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res* **41**, D1040-6 (2013).

349. Kang, Y. Y. *et al.* Genome-wide scanning for CHD1L gene in papillary thyroid carcinoma complicated with type 2 diabetes mellitus. *Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico* **23**, 2536–2547 (2021).

350. Wang, W. *et al.* CHD1L promotes cell cycle progression and cell motility by up-regulating MDM2 in breast cancer. *American journal of translational research* **11**, 1581–1592 (2019).

351. Singh, S. *et al.* Human oncoprotein MDM2 activates the Akt signaling pathway through an interaction with the repressor element-1 silencing transcription factor conferring a survival advantage to cancer cells. *Cell death and differentiation* **20**, 558–66 (2013).

352. Marchetti, A. *et al.* p53 can inhibit cell proliferation through caspase-mediated cleavage of ERK2/MAPK. *Cell death and differentiation* **11**, 596–607 (2004).

353. Sun, D., Qian, H., Li, J. & Xing, P. Targeting MDM2 in malignancies is a promising strategy for overcoming resistance to anticancer immunotherapy. *Journal of biomedical science* **31**, 17 (2024).

354. Wu, K. K.-L. *et al.* MDM2 induces pro-inflammatory and glycolytic responses in M1 macrophages by integrating iNOS-nitric oxide and HIF-1α pathways in mice. *Nature communications* **15**, 8624 (2024).

355. Izumi, T. *et al.* MDM2 is a novel E3 ligase for HIV-1 Vif. *Retrovirology* **6**, 1 (2009).

356. Breton, Y. *et al.* Expression of MDM2 in Macrophages Promotes the Early Postentry Steps of HIV-1 Infection through Inhibition of p53. *Journal of virology* **93**, (2019).

357. Xu, M. *et al.* SPOCK1/SIX1axis promotes breast cancer progression by activating AKT/mTOR signaling. *Aging* **13**, 1032–1050 (2020).

358. Cui, X. *et al.* SPOCK1 promotes metastasis in pancreatic cancer via NF-κB-dependent epithelial-mesenchymal transition by interacting with IκB-α. *Cellular oncology (Dordrecht, Netherlands)* **45**, 69–84 (2022).

359. Ning, X. *et al.* Apoptotic Caspases Suppress Type I Interferon Production via the Cleavage of cGAS, MAVS, and IRF3. *Molecular cell* **74**, 19-31.e7 (2019).

360. Wu, D. *et al.* Apoptotic caspase inhibits innate immune signaling by cleaving NF-κBs in both Mammals and Flies. *Cell death & disease* **13**, 731 (2022).

361. Páhi, Z. G., Borsos, B. N., Pantazi, V., Ujfaludi, Z. & Pankotai, T. PARylation During Transcription: Insights into the Fine-Tuning Mechanism and Regulation. *Cancers* **12**, (2020).

362. Haince, J.-F. *et al.* Ataxia telangiectasia mutated (ATM) signaling network is modulated by a novel poly(ADP-ribose)-dependent pathway in the early response to DNA-damaging agents. *The Journal of biological chemistry* **282**, 16441–53 (2007).

363. Brzostek-Racine, S., Gordon, C., Van Scoy, S. & Reich, N. C. The DNA damage response induces IFN. *Journal of immunology (Baltimore, Md. : 1950)* **187**, 5336–45 (2011).

364. Verma, P. *et al.* ALC1 links chromatin accessibility to PARP inhibitor response in homologous recombination-deficient cells. *Nature cell biology* **23**, 160–171 (2021).

365. Kim, C., Wang, X.-D. & Yu, Y. PARP1 inhibitors trigger innate immunity via PARP1 trapping-induced DNA damage response. *eLife* **9**, (2020).

366. Zahoor, M. A. *et al.* HIV-1 Vpr induces interferon-stimulated genes in human monocyte-derived macrophages. *PloS one* **9**, e106418 (2014).

367. Pontén, J. & Saksela, E. Two established in vitro cell lines from human mesenchymal tumours. *International journal of cancer* **2**, 434–47 (1967).

368. Sandoval, C., Nisson, K. & Fregoso, O. I. HIV-1 Vpr-induced DNA damage activates NF-κB through ATM-NEMO independent of cell cycle arrest. *mBio* **15**, e0024024 (2024).

369. Lu, Y. *et al.* Histone deacetylase 4 promotes type I interferon signaling, restricts DNA viruses, and is degraded via vaccinia virus protein C6. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 11997–12006 (2019).

370. Woods, M. W. *et al.* Human HERC5 restricts an early stage of HIV-1 assembly by a mechanism correlating with the ISGylation of Gag. *Retrovirology* **8**, 95 (2011).

371. Deschamps, T. & Kalamvoki, M. Impaired STING Pathway in Human Osteosarcoma U2OS Cells Contributes to the Growth of ICP0-Null Mutant Herpes Simplex Virus. *Journal of virology* **91**, (2017).

372. Jiang, X. *et al.* Ubiquitin-induced oligomerization of the RNA sensors RIG-I and MDA5 activates antiviral innate immune response. *Immunity* **36**, 959–73 (2012).

373. Tsuchiya, S. *et al.* Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). *International journal of cancer* **26**, 171–6 (1980).

374. Kisaka, J. K., Rauch, D., Griffith, M. & Kyei, G. B. A macrophage-cell model of HIV latency reveals the unusual importance of the bromodomain axis. *Virology journal* **21**, 80 (2024).

375. Goujon, C. & Malim, M. H. Characterization of the alpha interferon-induced postentry block to HIV-1 infection in primary human macrophages and T cells. *Journal of virology* **84**, 9254–66 (2010).

376. Cingöz, O. & Goff, S. P. HIV-1 Is a Poor Inducer of Innate Immune Responses. *mBio* **10**, (2019).

377. Elsner, C. *et al.* Absence of cGAS-mediated type I IFN responses in HIV-1-infected T cells. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 19475–19486 (2020).

378. Lieu, P. T., Fontes, A., Vemuri, M. C. & Macarthur, C. C. Generation of induced pluripotent stem cells with CytoTune, a non-integrating Sendai virus. *Methods in molecular biology (Clifton, N.J.)* **997**, 45–56 (2013).

379. Yang, Q. *et al.* Macrophages derived from human induced pluripotent stem cells (iPSCs) serve as a high-fidelity cellular model for investigating HIV-1, dengue, and influenza viruses. *Journal of virology* **98**, e0156323 (2024).

380. Taylor, J. P. *et al.* CRISPR/Cas9 knockout of USP18 enhances type I IFN responsiveness and restricts HIV-1 infection in macrophages. *Journal of leukocyte biology* **103**, 1225–40 (2018).

381. Gao, L. *et al.* Identifying noncoding risk variants using disease-relevant gene regulatory networks. *Nature communications* **9**, 702 (2018).

382. Rausell, A. *et al.* Innate immune defects in HIV permissive cell lines. *Retrovirology* **13**, 43 (2016).

383. Cobos Jiménez, V. *et al.* Differential expression of HIV-1 interfering factors in monocyte-derived macrophages stimulated with polarizing cytokines or interferons. *Scientific reports* **2**, 763 (2012).

384. Dragin, L. *et al.* Interferon block to HIV-1 transduction in macrophages despite SAMHD1 degradation and high deoxynucleoside triphosphates supply. *Retrovirology* **10**, 30 (2013).

385. Itell, H. L., Humes, D. & Overbaugh, J. Several cell-intrinsic effectors drive type I interferon-mediated restriction of HIV-1 in primary CD4 + T cells. *bioRxiv : the preprint server for biology* (2023) doi:10.1101/2023.02.07.527545.

386. Asmuth, D. M. *et al.* Safety, tolerability, and mechanisms of antiretroviral activity of pegylated interferon Alfa-2a in HIV-1-monoinfected participants: a phase II clinical trial. *The Journal of infectious diseases* **201**, 1686–96 (2010).

387. Stylianou, E., Aukrust, P., Bendtzen, K., Müller, F. & Frøland, S. S. Interferons and interferon (IFN)-inducible protein 10 during highly active anti-retroviral therapy (HAART)-possible immunosuppressive role of IFN-alpha in HIV infection. *Clinical and experimental immunology* **119**, 479–85 (2000).

388. Valverde-Villegas, J. M. *et al.* High CXCL10/IP-10 levels are a hallmark in the clinical evolution of the HIV infection. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **57**, 51–58 (2018).

389. Rotger, M. *et al.* Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals. *PLoS pathogens* **6**, e1000781 (2010).

390. Mackelprang, R. D. *et al.* Upregulation of IFN-stimulated genes persists beyond the transitory broad immunologic changes of acute HIV-1 infection. *iScience* **26**, 106454 (2023).

391. Bosinger, S. E. *et al.* Global genomic analysis reveals rapid control of a robust innate response in SIV-infected sooty mangabeys. *The Journal of clinical investigation* **119**, 3556–72 (2009).

392. Lee, S. Y., Park, Y. K., Yoon, C.-H., Kim, K. & Kim, K.-C. Meta-analysis of gene expression profiles in long-term non-progressors infected with HIV-1. *BMC medical genomics* **12**, 3 (2019).

393. Abbott, J. M. *et al.* First-in-Class Inhibitors of Oncogenic CHD1L with Preclinical Activity against Colorectal Cancer. *Molecular cancer therapeutics* **19**, 1598–1612 (2020).

394. Clune, S. *et al.* The validation of new CHD1L inhibitors as a therapeutic strategy for cancer. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* **170**, 116037 (2024).

395. Clemente, M. I. *et al.* Non-steroidal anti-inflammatory drugs increase the antiretroviral activity of nucleoside reverse transcriptase inhibitors in HIV type-1-infected T-lymphocytes: role of multidrug resistance protein 4. *Antiviral therapy* **14**, 1101–11 (2009).

396. Kasang, C. *et al.* Effects of Prednisolone on Disease Progression in Antiretroviral-Untreated HIV Infection: A 2-Year Randomized, Double-Blind Placebo-Controlled Clinical Trial. *PloS one* **11**, e0146678 (2016).

397. Yarchoan, R. & Uldrick, T. S. HIV-Associated Cancers and Related Diseases. *The New England journal of medicine* **378**, 1029–1041 (2018).

398. Li, Y. *et al.* CHD1L contributes to cisplatin resistance by upregulating the ABCB1-NF-κB axis in human non-small-cell lung cancer. *Cell death & disease* **10**, 99 (2019).

399. Sengupta, D. *et al.* Performance and accuracy evaluation of reference panels for genotype imputation in sub-Saharan African populations. *Cell genomics* **3**, 100332 (2023).

400. Lane, J. *et al.* A genome-wide association study of resistance to HIV infection in highly exposed uninfected individuals with hemophilia A. *Human Molecular Genetics* **22**, 1903–1910 (2013).

401. Kantor, A., McClements, M. E. & MacLaren, R. E. CRISPR-Cas9 DNA Base-Editing and Prime-Editing. *International journal of molecular sciences* **21**, (2020).

402. Singh, M. K. *et al.* A Non-Coding Variant in SLC15A4 Modulates Enhancer Activity and Lysosomal Deacidification Linked to Lupus Susceptibility. *Frontiers in lupus* **1**, (2023).

403. Wang, Z. *et al.* An autoimmune pleiotropic SNP modulates IRF5 alternative promoter usage through ZBTB3-mediated chromatin looping. *Nature communications* **14**, 1208 (2023).

404. Tewhey, R. *et al.* Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. *Cell* **165**, 1519–1529 (2016).

405. Tough RH, Liu X, Enns E, Van Domselaar G, M. P. Gene Nonessentiality and LoF Identifier (gNALI): Identifying Nonessential Host Genes from Viral Host Dependency Factor Studies.

406. Cary, D. C. & Peterlin, B. M. Natural Products and HIV/AIDS. *AIDS research and human retroviruses* **34**, 31–38 (2018).

407. Perovic, V. *et al.* Exploring the Antiviral Potential of Natural Compounds against Influenza: A Combined Computational and Experimental Approach. *International journal of molecular sciences* **25**, (2024).

408. Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., International Natural Product Sciences Taskforce & Supuran, C. T. Natural products in drug discovery: advances and opportunities. *Nature reviews. Drug discovery* **20**, 200–216 (2021).

409. De Jager, P. L. *et al.* ImmVar project: Insights and design considerations for future studies of 'healthy' immune variation. *Seminars in immunology* **27**, 51–7 (2015).

410. Daniels, P. R. *et al.* Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. *The American journal of medicine* **116**, 676–81 (2004).

411. Piot, P. *et al.* Heterosexual transmission of HIV. *AIDS (London, England)* **1**, 199–206 (1987).

412. Plummer, F. A. *et al.* Epidemiology of chancroid and Haemophilus ducreyi in Nairobi, Kenya. *Lancet (London, England)* **2**, 1293–5 (1983).

413. Fowke, K. R. *et al.* Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. *Lancet (London, England)* **348**, 1347–51 (1996).

414. Luo, M. *et al.* A genetic polymorphism of FREM1 is associated with resistance against HIV infection in the Pumwani sex worker cohort. *Journal of virology* **86**, 11899–905 (2012).

Appendices

Appendix 1

Cohorts used in this thesis work

The International Collaboration for the Genomics of HIV

The ICGH is a consortium of studies and cohorts that have amalgamated population-level human genetic variation data to identify mechanisms of HIV host genetics and HIV pathogenesis[1]. At the time of this thesis, the ICGH (N=3,879) consists of 2535 African American individuals living in the USA and 1,344 individuals of African ancestry from Kenya, Rwanda, South Africa, Uganda, Zambia, and African ancestry living in Switzerland. These individuals have HIV set-point viral load data which was defined as the mean $\log_{10}$ RNA copies per mL of plasma during chronic infection without the use of antiretroviral therapy.

The African Functional Genomics Resource

The AFGR is a repository of data that includes functional genomics data from LCLs derived from 599 donors across Gambian in Western Divisions (N = 112), Esan in Nigeria (N = 97), Maasai in Kinyawa, Kenya (N=164), Mende in Sierra Leone (N=83), Yoruba in Ibadan, Nigeria (N = 41), and Luhya in Webuye, Kenya (N = 96). The LCLs originated from samples stored at the Coriell Institute for Medical Research and arose from PBMCs transformed using Epstein-Barr virus. This repository contains gene expression data from RNA-sequencing, eQTLs, splice QTLs, ATAC-seq, and chromatin accessibility QTLs to aid researchers in determining the functional impact of genetic variants on gene expression and phenotypes of interest.

The Immune Variation project

The ImmVar project is a collaborative effort to characterize how genetic and environmental factors influence the immune response in healthy individuals[409]. This cohort consists of healthy individuals from Boston, Massachusetts, aged 18 to 50, who self-reported with no autoimmune, neurologic, metabolic, or chronic infectious. We prioritized a subset of African American individuals (N=110) who had available genotype and gene expression data. Primary blood mononuclear cells from these individuals were

sorted into monocytes based on CD14+ CD16- and naïve CD4+ T cells were CD4+ CD62L+.

The Genetic Epidemiology Network of Arteriopathy study

The GENOA study was designed to investigate the genetics of hypertension and is comprised of African Americans and European Americans from Rochester, Minnesota[410]. In our study, we incorporated data from 1,032 African Americans who had available genotype, eQTL, and meQTL data. In brief, gene expression was determined in lymphoblastoid cell lines using the Affymetrix Human Transcriptome Array 2.0. For identification of meQTLs, genomic DNA from peripheral blood leukocytes was analyzed using Illumina Infinium HumanMethylationEPIC BeadChips.

The Majengo Observational Cohort Study

MOCS was established in 1985 as a longitudinal study of epidemiology and immunology of female sex workers in Pumwani, Nairobi, Kenya[411,412]. Participants received STI testing upon enrollment and subsequent testing and treatment bi-annually. Individuals within this cohort had repeated exposure to HIV, however some individuals displayed reduced susceptibility to HIV infection[413]. Through the U.S. President's Emergency Plan for AIDS Relief program, ART was made available to HIV-infected study participants whose CD4+ T cell counts were below 200 cells/mm^3 after 2003[414]. Follow-up data was used to confirm selected samples were ART-naïve at the time of collection.

The Multi-Ethnic Study of Atherosclerosis study

The MESA study includes 6,814 individuals and was established to investigate the risk factors and disease progression of cardiovascular disease. A subset of these individuals that self-identified as African American with available genotype and gene expression data were utilized in our study. We leveraged a PrediXcan model based on monocytes isolated from peripheral blood mononuclear cells using positive magnetic selection with anti-CD14 beads. RNA was extracted using the AllPrep DNA/RNA Mini Kit and gene expression was determined using the Illumina HumanHT-12 v4 Expression BeadChip.

The Genes-Environments and Admixture in Latino Asthmatics II and Study of African Americans, Asthma, Genes, and Environments study

The GALA II study consists of 2,481 Latinos and Blacks with asthma between the ages of 8-17 from the United States and Puerto Rico. The SAGE study consist of 1556 (920 with asthma, 636 healthy controls) African American individuals between the ages of 8-21 and exclusion criteria included chronic illness, pregnancy, and smoking within 1 year of the start date. The total number of self-identified African Americans in this study were 757 that were recruited from the San Francisco Bay Area in California. In this group, 433 (57.2%) have been diagnosed with asthma and the median interquartile age is 16. Total RNA was isolated from PAXgene tubes and depleted for globin before sequencing using a Illumina TruSeq Stranded mRNA Sample Preparation Kit. Gene expression was determined using the TOPMed RNA-seq pipeline and GRCh38 alignment.

PRKAB2 knockout human immunopluripotent stem cells study

In Chapter 3, we leveraged gene expression data from human iPSCs (ASE-9202, Applied StemCell)[209] to investigate the relationship between HIV spVL associated variants and *PRKAB2* expression. The PRKAB2 knockout was generated using two CRISPR-Cas9 guide RNAs which induced two frameshift mutations into the *PRKAB2* gene. Frameshift mutations were validated using sequencing and protein loss-of-function was confirmed using Western blotting. For *PRKAB2*, there were two separate clones (C1 and C2) that were used in the study's downstream analyses. Gene expression profiles were determined using RNA-sequencing on total RNA against GRCH37.

Appendix 2

Blood tissues with DNase, histone ChIP-Seq, and chromatin states from the ENCODE and Roadmap Epigenomics Consortiums

Epigenome ID	Tissue Description
E029	Primary monocytes from peripheral blood
E030	Primary neutrophils from peripheral blood
E031	Primary B cells from cord blood
E032	Primary B cells from peripheral blood
E033	Primary T cells from cord blood
E034	Primary T cells from peripheral blood
E035	Primary hematopoietic stem cells
E036	Primary hematopoietic stem cells short term culture
E037	Primary T helper memory cells from peripheral blood 2
E038	Primary T helper naive cells from peripheral blood
E039	Primary T helper naive cells from peripheral blood
E040	Primary T helper memory cells from peripheral blood 1
E041	Primary T helper cells PMA-I stimulated
E042	Primary T helper 17 cells PMA-I stimulated
E043	Primary T helper cells from peripheral blood
E044	Primary T regulatory cells from peripheral blood
E045	Primary T cells effector/memory enriched from peripheral blood
E046	Primary Natural Killer cells from peripheral blood
E047	Primary T CD8+ naive cells from peripheral blood
E048	Primary T CD8+ memory cells from peripheral blood
E049	Mesenchymal Stem Cell Derived Chondrocyte Cultured Cells
E050	Primary hematopoietic stem cells G-CSF-mobilized Female
E051	Primary hematopoietic stem cells G-CSF-mobilized Male
E062	Primary mononuclear cells from peripheral blood
E115	Dnd41 T Cell Leukemia Cell Line
E116	GM12878 Lymphoblastoid Cells
E123	K562 Leukemia Cells
E124	Monocytes-CD14+ RO01746 Primary Cells

Appendix 3

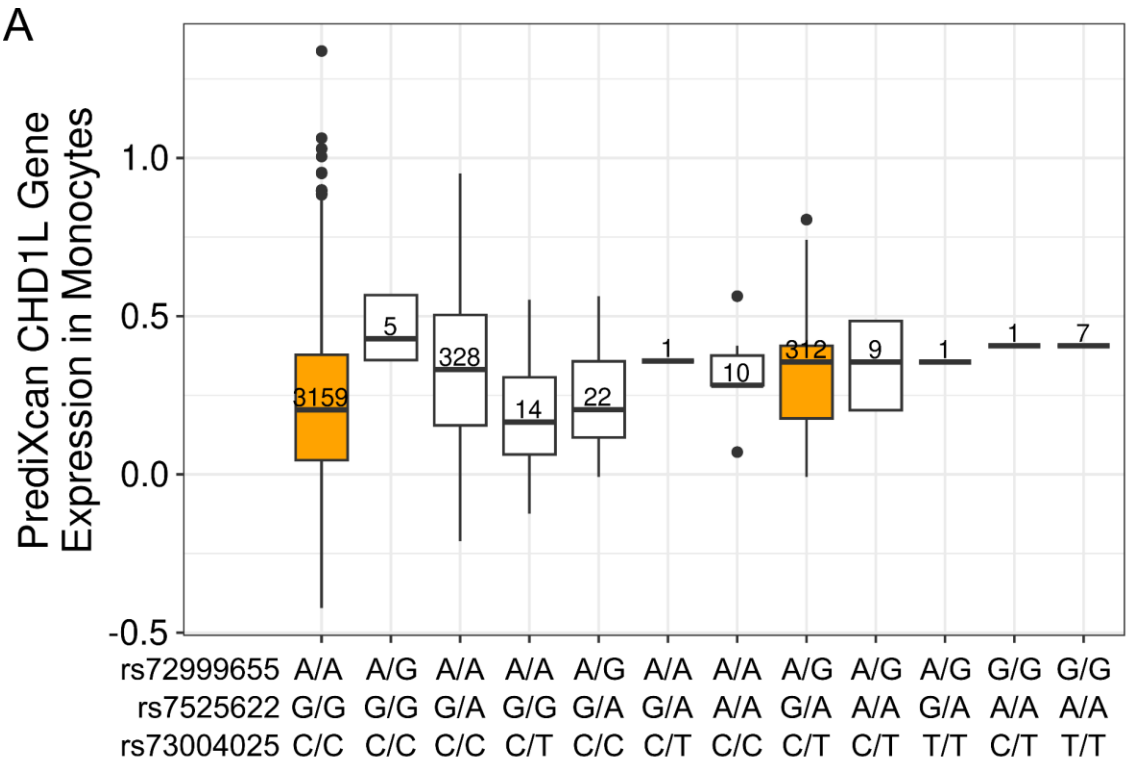
CHR	hg38 Pos	RSID	Reference Allele	Alternate Allele	ChromHMM Markers	Imputed ChromHMM Markers	Methylation Markers	DNase Markers	Protein-bound Region Markers	Known eQTL
1	147242563	rs72997590	C	T	33	33	33	12	3	FALSE
1	147242441	rs72997588	G	T	33	33	33	12	2	FALSE
1	147080339	rs12127461	C	G	29	32	30	1	0	FALSE
1	147073025	rs79468247	G	T	26	33	29	1	0	FALSE
1	147241985	rs72997586	G	A	22	30	30	0	0	FALSE
1	147085526	rs12752148	T	C	16	33	29	1	0	TRUE
1	147245310	rs59402704	C	T	18	26	30	1	1	FALSE
1	147628766	rs115475998	C	G	16	25	23	0	0	FALSE
1	147246776	rs12568866	A	C	12	15	28	0	0	TRUE
1	147546944	rs80296981	G	A	8	13	17	0	0	FALSE
1	147617540	rs114757780	C	T	6	13	16	0	0	FALSE
1	147250328	rs12402309	T	A	4	1	20	0	0	TRUE
1	147250809	rs72999623	G	C	2	0	20	0	0	FALSE
1	147248432	rs111602294	T	C	1	0	20	0	0	FALSE
1	147248435	rs113662441	C	G	1	0	20	0	0	FALSE
1	147223639	rs58276129	T	A	3	1	17	0	0	FALSE
1	147223640	rs60565795	C	A	3	1	17	0	0	FALSE
1	147174868	rs72997514	G	C	0	1	19	0	0	FALSE

1	147292702	rs7519713	C	T	2	10	6	2	0	FALSE
1	147249458	rs112922479	C	T	1	0	17	0	0	FALSE
1	147240938	rs58820837	C	T	4	5	9	0	0	FALSE
1	147082598	rs12129691	C	T	1	0	11	0	0	FALSE
1	147274523	rs72999646	T	A	1	2	9	0	0	FALSE
1	147052335	rs12124149	G	T	3	1	3	2	2	FALSE
1	147082773	rs71510558	C	A	0	0	11	0	0	FALSE
1	147577392	rs73007970	G	A	2	4	5	0	0	FALSE
1	147223310	rs7527846	A	C	0	0	11	0	0	FALSE
1	147239832	rs60063927	C	A	1	3	5	0	0	FALSE
1	147265213	rs72999640	C	T	2	2	5	0	0	FALSE
1	147106556	rs143409350	A	G	2	1	3	1	1	FALSE
1	147217234	rs6696468	C	T	2	2	4	0	0	FALSE
1	147264152	rs72999639	T	C	2	2	4	0	0	FALSE
1	147052817	rs78891176	C	T	3	2	3	0	0	FALSE
1	147252083	rs59689147	A	T	0	0	7	0	0	FALSE
1	147275229	rs72999648	G	A	0	0	7	0	0	FALSE
1	147222864	rs7525225	A	G	0	0	7	0	0	FALSE
1	147379251	rs34602039	G	A	1	1	3	0	0	TRUE
1	147331900	rs6677679	C	T	1	0	4	0	0	TRUE
1	147074795	rs12737860	C	A	0	0	5	0	0	FALSE
1	147237263	rs1932979	A	G	0	1	3	0	0	TRUE
1	147236836	rs68080346	T	C	0	0	2	0	2	TRUE
1	147075155	rs12129419	T	C	0	0	4	0	0	FALSE

1	147236823	rs60107727	G	A	0	0	2	0	2	FALSE
1	147259242	rs72999634	A	G	0	0	4	0	0	FALSE
1	147262674	rs72999637	T	C	0	0	4	0	0	FALSE
1	147262876	rs72999638	T	A	0	0	4	0	0	FALSE
1	147342019	rs73001613	G	T	1	0	3	0	0	FALSE
1	147422945	rs73004025	C	T	1	1	2	0	0	FALSE
1	147076727	rs12743228	G	A	0	0	3	0	0	FALSE
1	147289796	rs59606861	G	A	0	0	2	0	0	TRUE
1	147330922	rs66486822	T	A	0	0	2	0	0	TRUE
1	147330943	rs72692962	C	A	0	0	2	0	0	TRUE
1	147236954	rs72997571	C	T	0	0	2	0	1	FALSE
1	147284111	rs72999655	A	G	0	0	3	0	0	FALSE
1	147287690	rs72999656	C	T	0	0	3	0	0	FALSE
1	147276904	rs7535451	A	G	0	0	3	0	0	FALSE
1	147200161	rs12027579	A	T	0	0	2	0	0	FALSE
1	147200163	rs12027580	A	G	0	0	2	0	0	FALSE
1	147534278	rs17160351	C	G	0	0	2	0	0	FALSE
1	147277728	rs59213667	G	A	0	0	2	0	0	FALSE
1	147157884	rs72997505	C	T	1	0	1	0	0	FALSE
1	147199068	rs72997527	A	G	0	0	2	0	0	FALSE
1	147238980	rs72997577	C	T	0	0	2	0	0	FALSE
1	147537150	rs74797765	T	C	0	0	2	0	0	FALSE
1	147120640	rs75226023	G	A	0	0	2	0	0	FALSE
1	147302425	rs7525622	G	A	0	0	1	0	0	TRUE

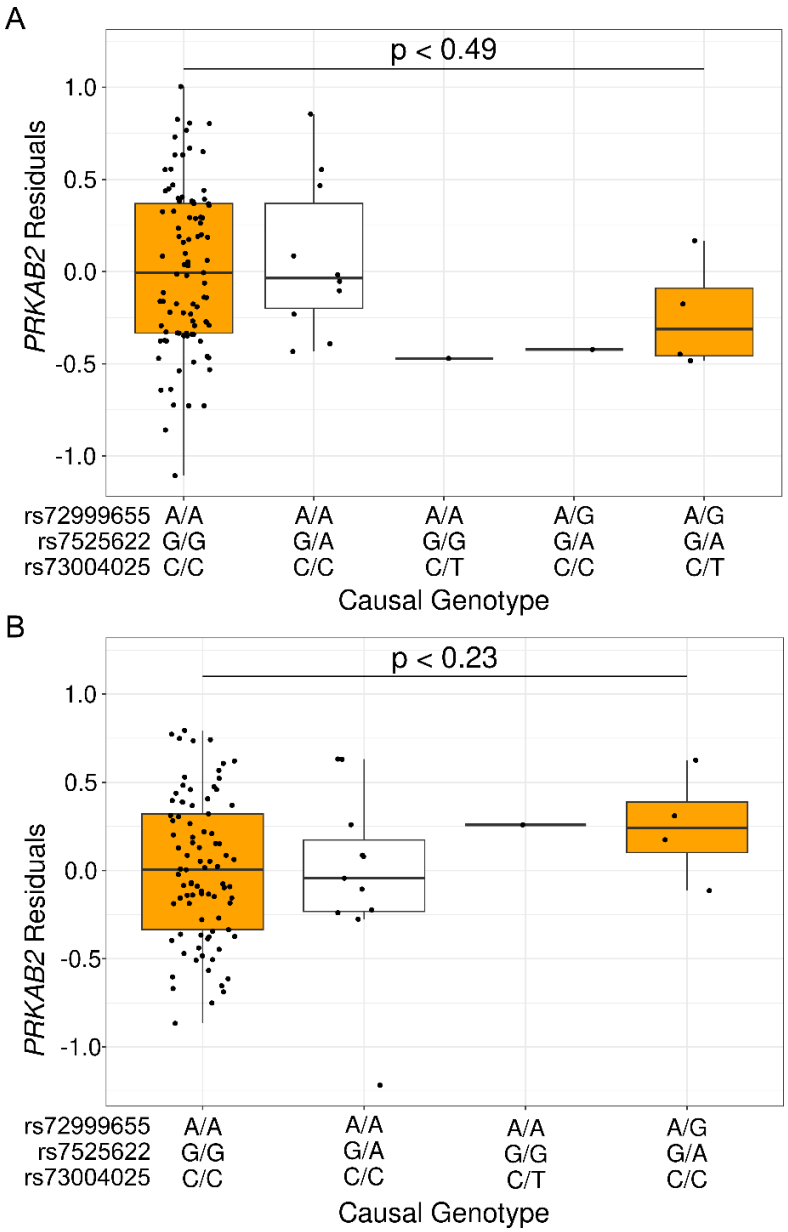
1	147262113	rs7526114	G	A	0	0	2	0	0	FALSE
1	147537149	rs75379534	G	A	0	0	2	0	0	FALSE
1	147310982	rs11239980	G	A	0	0	0	0	0	TRUE
1	147069895	rs115805773	G	A	0	0	1	0	0	FALSE
1	147069963	rs12118734	G	A	0	0	1	0	0	FALSE
1	147051741	rs12723253	C	T	0	0	1	0	0	FALSE
1	147341608	rs2353984	C	T	0	0	1	0	0	FALSE
1	147358974	rs2353999	T	C	0	0	1	0	0	FALSE
1	147311853	rs4950398	G	A	0	0	0	0	0	TRUE
1	147313199	rs58941651	T	C	0	0	0	0	0	TRUE
1	147300364	rs59180647	C	T	0	0	0	0	0	TRUE
1	147154442	rs72995597	T	C	0	0	1	0	0	FALSE
1	147254047	rs72999628	G	A	0	0	1	0	0	FALSE
1	147254521	rs72999632	T	C	0	0	1	0	0	FALSE
1	147321430	rs7520661	A	G	0	0	0	0	0	TRUE
1	147320958	rs7520841	T	C	0	0	0	0	0	TRUE
1	147391631	rs7522804	T	G	0	0	1	0	0	FALSE
1	147314360	rs885305	G	A	0	0	0	0	0	TRUE

Appendix 4



Imputed *CHD1L* expression for all GWAS participants. Orange boxplots represent individuals which are homozygous reference, heterozygous, or homozygous alternate at all alleles. *CHD1L* expression was imputed using monocyte models from the African American subset of the MESA cohort.

Appendix 5



Gene expression of *PRKAB2* based on candidate causal GWAS alleles in African American populations. Gene expression is the residuals following control for non-genetic factors in African populations for A) monocytes or B) naïve CD4+ T cells. Gene expression is expressed as residuals following control for batch effects, age, gender, and technical artifacts using the top 14 PCs in monocytes or 12 PCs in naïve CD4+ T cells. Significance was determined using a two-sided Wilcoxon Rank Sum Test.

Appendix 6

Differentially expressed genes in *PRKAB2*$^{-/-}$ immunopluripotent stem cells

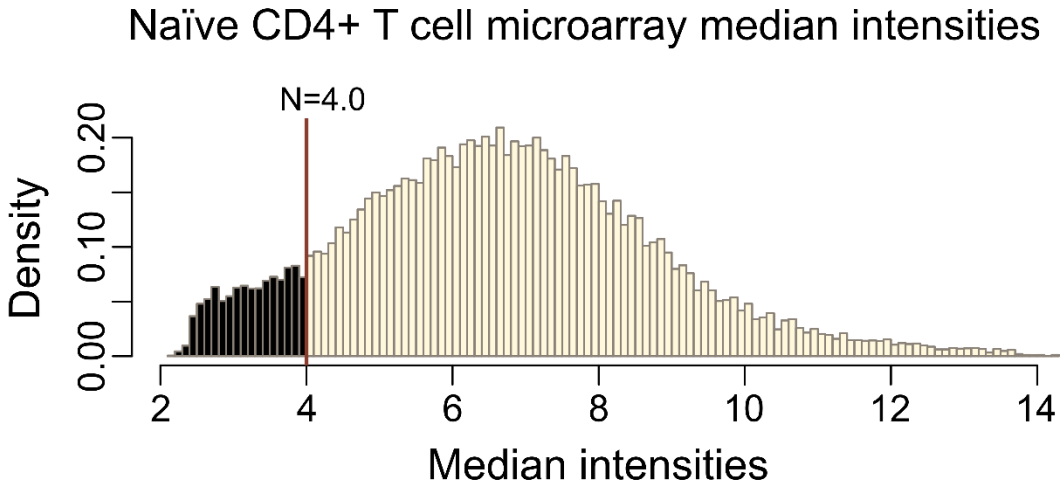
Entrez GeneID	HGNC Symbol	log2 Fold Change[A]	FDR
79784	MYH14	-9.33	7.76×10^{-36}
10656	KHDRBS3	-8.79	2.63×10^{-55}
5565	PRKAB2	-2.37	2.2×10^{-308}
5142	PDE4B	-1.91	5.63×10^{-104}
3266	ERAS	-1.91	1.53×10^{-3}
4057	LTF	-1.89	4.79×10^{-8}
639	PRDM1	-1.87	2.01×10^{-8}
8744	TNFSF9	-1.51	1.83×10^{-32}
6482	ST3GAL1	-1.44	4.70×10^{-37}
3937	LCP2	-1.31	5.89×10^{-3}
64786	TBC1D15	-1.26	7.88×10^{-172}
2119	ETV5	-1.24	9.29×10^{-81}
3579	CXCR2	-1.24	6.84×10^{-3}
3783	KCNN4	-1.23	7.99×10^{-3}
9182	RASSF9	-1.2	8.76×10^{-3}
963	CD53	-1.15	7.76×10^{-3}
10397	NDRG1	-1.11	3.06×10^{-22}
847	CAT	-1.09	5.79×10^{-85}
55329	MNS1	-1.08	1.92×10^{-13}
2318	FLNC	-1.07	3.20×10^{-75}
727897	MUC5B	-1.07	2.05×10^{-2}
1026	CDKN1A	-1.04	8.63×10^{-81}
718	C3	-1.04	1.31×10^{-34}
54843	SYTL2	-1.01	6.54×10^{-19}
7042	TGFB2	-1.01	7.12×10^{-5}
137902	PXDNL	1	4.67×10^{-4}
26998	FETUB	1.02	4.62×10^{-2}
7857	SCG2	1.03	6.93×10^{-3}
2571	GAD1	1.08	4.80×10^{-3}

3684	ITGAM	1.11	1.18×10^{-7}
8558	CDK10	1.2	2.87×10^{-123}
4489	MT1A	1.28	1.03×10^{-2}
3007	H1-3	1.37	6.30×10^{-3}
50506	DUOX2	1.55	1.80×10^{-3}
29119	CTNNA3	1.67	3.53×10^{-77}
345611	IRGM	2.92	1.12×10^{-5}

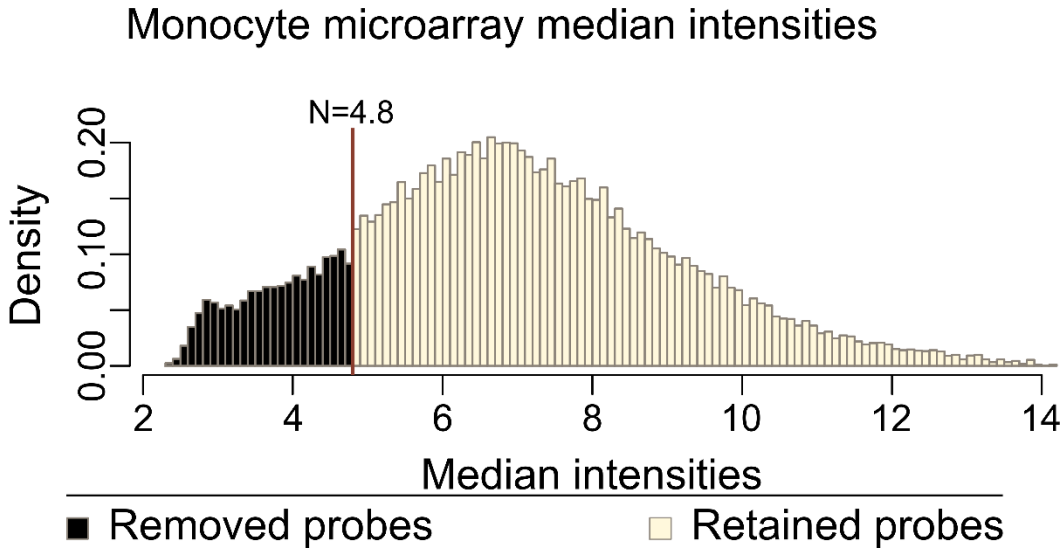
[A]Negative $\log_2$ fold change corresponds to a reduction in *PRKAB2*$^{-/-}$.

Appendix 7

A



B



Histogram of microarray median probe intensities for A) naïve CD4+ T cells and B) monocytes. Probes with a median intensity less than the given threshold (black) were removed while the remaining probes (yellow) were retained for transcriptome analysis. Threshold was determined where values flattened to the left of the intensities peak.

Appendix 8

Association between covariates and naïve CD4+ T cell microarray expression following quality control and normalization.

Principal Component	Genotype P	Sex P	Age P	Batch P
PC1	0.23	0.88	0.94	0.83
PC2	0.14	0.96	0.63	0.99
PC3	0.05	0.84	0.97	1.00
PC4	0.15	0.96	0.79	1.00
PC5	0.62	0.89	0.75	1.00
PC6	0.83	0.99	0.96	1.00

Appendix 9

Association between covariates and monocyte microarray expression following quality control and normalization.

Principal Component	Genotype P	Sex P	Age P	Batch P
PC1	3.3×10^{-5}	0.91	0.95	0.93
PC2	0.67	0.87	0.84	1.00
PC3	0.96	0.91	0.95	0.99
PC4	0.58	0.96	0.90	1.00
PC5	0.54	0.88	0.74	1.00
PC6	0.68	0.98	0.55	0.98