An exploratory analysis of longitudinal HIV data: mutation rates, antiretroviral therapies, phylogeny, and patient outcomes

Background

HIV is a member of the retrovirus family of enveloped, single-stranded RNA viruses (Seitz, 2016). It is one of the most prevalent viral infections in the world with over 38 million individuals living with it, and greater than 650,000 HIV related deaths in 2021 (WHO, 2021). Left untreated, HIV infects and kills the majority of $CD4^+$ T cells, eventually leading to the development of acquired immunodeficiency syndrome (AIDS). $CD4^+$ T cells play many integral roles in the activation of immune responses through the secretion of specific cytokines, and the immune system is severely handicapped in their absence (Luckheeram et al., 2012). Upon the onset of AIDS, affected individuals are much more susceptible to opportunistic infections, caused by pathogens that can only infect those with compromised immune systems (Gallant, 1994). In addition to its severe impact on the immune system, HIV has also been linked to the development of osteoporosis, cancer, cardiovascular disease, renal dysfunction, liver dysfunction, and neurological diseases (Phillips et al., 2008).

There are three distinct clinical stages of HIV infection: acute stage, chronic stage, and AIDS (NIH, 2021). The acute stage is the earliest of the three stages, starting around 2 to 4 weeks after infection. The virus proliferates rapidly and causes a large drop in $CD4^+$ T cells, sometimes leading to flu-like symptoms (Levy, 1994). In the acute phase, viral titers are very high in the bloodstream and transmission risk is at its greatest (Cohen et al., 2011). Eventually, the immune system is able to control the viremia and the individual enters the chronic stage, also known as clinical latency due to the lack of observable symptoms (Zandman-Goddard & Shoenfeld, 2002). In the chronic phase, viral titers are at a low but constant level, while $CD4^+$ T

cell levels continue to steadily decline (Appay et al., 2007). This stage can last for up to a decade until the onset of AIDS, which is classified by a $CD4^+$ T cell count that is lower than 200 cells per cubic millimeter (Garcia & Guzman, 2021). An individual with AIDS is severely immunocompromised and death soon ensues due to their increased susceptibility to infections (Siegel & Lekas, 2002).

Although there is no cure for HIV/AIDS, there are numerous antiretroviral treatments (ART) that can significantly delay or even prevent the onset of AIDS (Volberding & Deeks, 2010). ARTs function by preventing replication of the virus through the inhibition of viral proteins that are integral to the HIV replication cycle, including reverse transcriptase, viral proteases, and integrase (Martinez-Cajas & Wainberg, 2008). However, HIV monotherapies are not very effective as viral rebound is observed very quickly after short term success, indicating the development of drug resistance (Rijnders & Rokx, 2019). Resistance is mainly attributed to the low genetic fidelity that is found in retroviruses; HIV has approximately 10,000 to 100,000 times the mutation rate found in eukaryotic cells (Rawson et al., 2015). These elevated mutation rates are due to the lack of proofreading mechanisms in reverse transcriptase (Lloyd et al., 2014). To decrease the likelihood of drug resistance, cocktails of several different ARTs with different mechanisms of action are used to great success (Lu et al., 2015).

Although HIV has been an intensive field of research for decades, much of the focus has been on molecular based research and there has been little exploratory analysis of longitudinal patient data. This project aims to use exploratory analysis to determine the relation between a few variables and patient outcomes: 1. Mutation rate, 2. Common HIV mutations, 3. ART used, 4. Phylogeny. Due to the complexity of HIV, there is much discussion on where future HIV

research efforts should be focused (Gallo, 2020 & Deeks et al., 2021). The results of this project can be used to direct future experimentation on significant/unusual trends in the data.

Methods

1. Mutation rates

Purpose: to explore the effects of HIV mutation rates on patient outcomes ($CD4^+$ cells and viral load)

Patients will be arranged into cohorts of high vs low mutation frequency. Mutation rate will be determined by dividing new mutations by total days. A mutational rate of over 0.01 mutations/day will be considered as high. Graphical representations of the longitudinal data will be made for each cohort and cohort outcomes for average $CD4^+$ T cell count and average viral RNA levels will be compared.

2. Common HIV mutations

Purpose: to find common mutations in patients treated with specific ARTs

Some HIV mutations are observed to arise in the majority of patients that are treated with the same drug (Shafer & Schapiro, 2009). Patient data will be arranged into groups based on drug mechanisms of action. Mutations shared across many patients in the same group ($n>5$) will be identified and compared to previous literature. Prevalent mutations across drug treatment groups will also be identified to see if any mutations are common among all patients, regardless of the ART.

3. ARTs

Purpose: to explore the efficacy of different ARTs in controlling viral titers.

Patient data will be arranged into groups based on the drugs that make up their combination therapies. Mean viral load and $CD4^+$ cell count will be compared for each group.

4. Phylogeny

Purpose: to identify relationships between HIV clades and patient outcomes

Phylogenetic trees will be generated using the Neighbor TreeMaker tool on the Los Alamos HIV Sequence Database- see Scripts/Tools section. Input data will be nucleotide sequences of viral isolates from each patient. Average viral load and $CD4^+$ T cell counts will be visualized for each clade to identify clades with increased virulence.

Scripts/Tools

1. Ushr is an R script that can be applied to longitudinal data on HIV viral load measurements. The data is visualized & fitted with a mathematical model, and the time to reach viral suppression below a threshold can be estimated (Morris, 2020). In this project, ushr will be used to visualize viral load in cohorts and individual patients.

https://github.com/SineadMorris/ushr

2. Phylogenetic trees will be constructed using the Neighbor TreeMaker tool on the Los Alamos HIV Sequence Database.

https://www.hiv.lanl.gov/components/sequence/HIV/treemaker/treemaker.html?sample_input=1

Data transparency

All longitudinal data will be acquired from Stanford University's HIV Drug Resistance Database, specifically the data under the Genotype-Clinical Outcome Correlations tab. This database has longitudinal data on HIV^+ individuals that are on ART. Recorded data includes

$CD4^+$ T cell counts, viral RNA load, nucleotide sequences (for integrase, proteases, reverse transcriptase), amino acid changes of the aforementioned proteins, and the ART drugs used. The specific data set that will be used in this study is ACTG 5257

Link:https://hivdb.stanford.edu/pages/clinicalStudyData/ACTG5257.html

References

Appay, V., Almeida, J., Sauce, D., Autran, B., & Papagno, L. (2007, May). Accelerated immune senescence and HIV-1 infection. *Experimental Gerontology*, *42*(5), 432–437. https://doi.org/10.1016/j.exger.2006.12.003

Cohen, M. S., Shaw, G. M., McMichael, A. J., & Haynes, B. F. (2011, May 19). Acute HIV-1 Infection. *New England Journal of Medicine*, *364*(20), 1943–1954. https://doi.org/10.1056/nejmra1011874

Deeks, S. G., Archin, N., Cannon, P., Collins, S., Jones, R. B., de Jong, M. A. W. P., Lambotte, O., Lamplough, R., Ndung'u, T., Sugarman, J., Tiemessen, C. T., Vandekerckhove, L., Lewin, S. R., Deeks, S., Lewin, S., de Jong, M., Ndhlovu, Z., Chomont, N., Brumme, Z., . . . Kankaka, E. N. (2021, December). Research priorities for an HIV cure: International AIDS Society Global Scientific Strategy 2021. *Nature Medicine*, *27*(12), 2085–2098. https://doi.org/10.1038/s41591-021-01590-5

Gallant, J. E. (1994, June 1). Prophylaxis for Opportunistic Infections in Patients with HIV Infection. *Annals of Internal Medicine*, *120*(11), 932. https://doi.org/10.7326/0003-4819-120-11-199406010-00006

Gallo, R. C. (2020, April). HIV/AIDS Research for the Future. *Cell Host &Amp; Microbe*, *27*(4), 499–501. https://doi.org/10.1016/j.chom.2020.03.022

Garcia, S. A., & Guzman, N. (2021, August 11). *Acquired Immune Deficiency Syndrome CD4+ Count*. National Library of Medicine. Retrieved October 16, 2022, from https://www.ncbi.nlm.nih.gov/books/NBK513289/

HIV. (2022, July 27). Retrieved October 16, 2022, from https://www.who.int/data/gho/data/themes/hiv-aids

Levy, J. A. (1994, October 17). *HIV And the Pathogenesis of AIDS*. Amer Society for Microbiology.

Lloyd, S. B., Kent, S. J., & Winnall, W. R. (2014, January). The High Cost of Fidelity. *AIDS Research and Human Retroviruses*, *30*(1), 8–16. https://doi.org/10.1089/aid.2013.0153

Lu, D. Y., Lu, T. R., Che, J. Y., Wu, H. Y., & Xu, B. (2015, May 12). New Perspectives of HIV/AIDS Therapy Study. *Recent Patents on Anti-Infective Drug Discovery*, *9*(2), 112–120. https://doi.org/10.2174/1574891x10666150109115402

Luckheeram, R. V., Zhou, R., Verma, A. D., & Xia, B. (2012). CD4+T Cells: Differentiation and Functions. *Clinical and Developmental Immunology*, *2012*, 1–12. https://doi.org/10.1155/2012/925135

Martinez-Cajas, J. L., & Wainberg, M. A. (2008). Antiretroviral Therapy. *Drugs*, *68*(1), 43–72. https://doi.org/10.2165/00003495-200868010-00004

Morris, S. E. (2020, February 11). *ushr: Understanding suppression of HIV in R - BMC Bioinformatics*. BioMed Central. Retrieved October 17, 2022, from https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-3389-x

Phillips, A. N., Neaton, J., & Lundgren, J. D. (2008, November 30). The role of HIV in serious diseases other than AIDS. *AIDS*, *22*(18), 2409–2418. https://doi.org/10.1097/qad.0b013e3283174636

Rawson, J. M. O., Landman, S. R., Reilly, C. S., & Mansky, L. M. (2015, July 10). HIV-1 and HIV-2 exhibit similar mutation frequencies and spectra in the absence of G-to-A hypermutation. *Retrovirology*, *12*(1). https://doi.org/10.1186/s12977-015-0180-6

Rijnders, B. J. A., & Rokx, C. (2019, January 2). Antiretroviral Monotherapy for HIV: Game Over or Future Perspectives? *Clinical Infectious Diseases*, *69*(9), 1506–1508. https://doi.org/10.1093/cid/ciy1136

Seitz, R. (2016). Human Immunodeficiency Virus (HIV). *Transfusion Medicine and Hemotherapy*, *43*(3), 203–222. https://doi.org/10.1159/000445852

Shafer, R., & Schapiro, J. (2009, April 1). HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. *AIDS Rev, 10*(67–84).

Siegel, K., & Lekas, H. M. (2002). AIDS as a chronic illness: psychosocial implications. *AIDS, 16*, S69–S76. https://doi.org/10.1097/00002030-200216004-00010

The Stages of HIV Infection | NIH. (2020, August 20). Retrieved October 16, 2022, from https://hivinfo.nih.gov/understanding-hiv/fact-sheets/stages-hiv-infection

Volberding, P. A., & Deeks, S. G. (2010, July). Antiretroviral therapy and management of HIV infection. *The Lancet*, *376*(9734), 49–62. https://doi.org/10.1016/s0140-6736(10)60676-9

Zandman-Goddard, G., & Shoenfeld, Y. (2002, December). HIV and autoimmunity. *Autoimmunity Reviews*, *1*(6), 329–337. https://doi.org/10.1016/s1568-9972(02)00086-1