academicJournals

Vol. 8(44), pp. 3710-3715, 29 October, 2014
DOI: 10.5897/AJMR12.1569
Article Number: 1EEC21348500
ISSN 1996-0808

http://www.academicjournals.org/AJMR

African Journal of Microbiology Research

Full Length Research Paper

An investigation into the comparison of three human immunodeficiency virus (HIV) drug resistance interpretation algorithms

Yashik Singh[1,2*] and Maurice Mars[1]

[1]Department of Telehealth, Nelson R Mandela School of Medicine, Durban, South Africa.
[2]School of Information Systems and Technology, University of Kwa-Zulu Natal, South Africa.

Received 16 July, 2012; Accepted 6 October, 2014

Human immunodeficiency virus (HIV) drug resistance is caused by mutations in the patient's human immunodeficiency virus genome that renders antiretroviral (ARV) drugs less effective. Drug resistance not only results in the patient being more vulnerable to opportunistic infections, but also may increase the spread of resistant strains of HIV. Interpretation computer algorithms may be used to determine which ARV drug(s) the patient are resistant to, by analyzing the mutations that occurred in the patient's HIV genome, instead of using expensive time consuming phenotypic laboratory tests. There are many different interpretation algorithms, but they often provide different resistance measures, even if applied to the same resistance profile. The aim of this study was to compare the latest versions of three HIV drug resistance interpretation algorithms in order to determine the extent of discrepancies between them. 2926 protease and 1981 reverse transcriptase subtype B sequences where obtained from the Stanford HIV-db genotype-phenotype correlation database. These sequences were pre-processed and the latest rules of the ANRS, HIV-db and REGA algorithms applied to them. The results were then compared with each other. These results indicate that although the accuracy of REGA, ANRS and HIV-db are similar, a deeper analysis of the results indicates that the interpretation algorithms are different. There need to be a mechanism of providing a single interpretation for a resistance profile of a genome. This may be created by collating the strengths of each of the interpretation algorithms.

Key words: Bioinformatics, human immunodeficiency virus (HIV), REGA, Agence Nationale de Recherches sur le SIDA (ANRS), HIV-db, genotype, interpretation algorithms.

INTRODUCTION

The human immunodeficiency virus (HIV) is a roughly spherically shaped lentivirus with a diameter of about 120 nm (McGovern et al., 2002). HIV infects helper T-cells, macrophage and dendritic cells (Cunningham et al., 2010). The two known strains of HIV are HIV-1 and HIV-2. Most HIV infection is attributed to HIV-1 (Gilbert et al., 2003). In 2009 the World Health Organisation reported that there were 2.6 million new cases of HIV infection.

*Corresponding author. E-mail: singhy@ukzn.ac.za. Tel: +27829408032. Fax: + 27312604737.

Thirty million of the 33.3 million infected with HIV live in low- and middle-income countries (WHO, 2010).

HIV infection may be managed with antiretroviral (ARV) drugs usually in the form of highly active antiretroviral therapy (HAART), which comprises of a regimen of three drugs from at least two of the following five drug classes (Bartlett et al., 2004; Mitton, 2000; Pierret, 2007; Bartlett et al., 2004; Pierret, 2007): Reverse transcriptase inhibitors (RTI), non-reverse transcriptase inhibitors (NRTI), protease inhibitors (PI), integrase inhibitors (II) and fusion inhibitors (FI). If untreated HIV-1 eventually develop into AIDS (Migueles and Connors, 2010). Several factors contribute to the success or failure of HAART including poor treatment, stage of the disease, drug potency, patient adherence, achievable drug levels, drug resistance and toxic effects of the drug. Of these factors, drug resistance is arguably the most critical (Tang and Shafer, 2012; Yashik and Maurice, 2012). The ability of HIV to mutate and reproduce itself in the presence of antiretroviral drugs is called HIV drug resistance. The three common pathways that lead to the development of HIV anti-retroviral drug resistance are high replication rates, selective pressure and initial infection by resistant strains of HIV.

These three pathways cause mutations in the HIV genome that render the ARV drugs less effective. These mutations in the HIV genome cause structural changes in the HIV genome resulting in the inability of the ARVS to stereotypically or chemically block binding sites required for the reproduction of HIV. Drug resistance not only results in the patient been more vulnerable to opportunistic infections, but may also increase the spread of resistant strains of HIV.

Testing for HIV resistance may consist of wet or dry chemistry laboratory tests, or by employing electronic computerized algorithms (Jaideep et al., 2003). Computer based interpretation algorithms using genomes can also be used to predict HIV drug resistance. These interpretation algorithms can be generally divided into one of two groups: those based on known domain knowledge, that is, they are based on the fact that certain combinations of known genome mutations cause unequivocal resistance, and those not based on predefined domain knowledge. These algorithms include machine learning and statistical methods.

Domain knowledge interpretation algorithms are based on scientific and published interactions between certain mutations and/or combination of mutations with resistance. This means that all computational decisions concerning resistance are based on known mutation-resistance rules found in published scientific literature. REGA, Agence Nationale de Recherches sur le SIDA (ANRS) and Stanford's HIV-db algorithm (de Oliveria et al., 2005b) are three examples of publically available domain knowledge interpretation algorithms. These algorithms are used widely and are regarded as goal standards.

REGA and ANRS classify ARV resistance according to three levels viz. susceptible, intermediate and resistant. Susceptible is indicative of the fact that a particular ARV drug will be effective against HIV. This means that the patient will respond to treatment with that particular ARV. Intermediate indicates that the ARV drug is partially effective.

In this case, the patient will respond to treatment with that particular drug, but it will not suppress the growth of HIV effectively. It is classified as resistant, if the ARV is not effective at all and treatment with this ARV will lead to virological failure. HIV-db classifies HIV resistance according to five levels: susceptible, potential low-level resistance, low-level resistance, intermediate resistance and high-level resistance. These algorithms employ Boolean based rules, some with penalties, and predict resistance by determining which mutations are present and/or absent.

Many different pattern recognition and machine learning algorithms have been applied to find a predicable correlation between genotypic and phenotypic data (called virtual phenotyping (Hales et al., 2006). Machine learning may be used to develop a model that predicts virological response. Machine learning is an artificial intelligence computer science technique that tries to find a mathematical model that maps between inputs and outputs of a domain problem. The efficiency of these algorithms are usually measured against REGA, ANRS and HIV-db.

Virtual phenotyping is growing in popularity and Kuritzkes et al. (2002) supports virtual phenotype as a tool for interpreting viral genotypes. The following are some of the algorithms that have been applied: least absolute shrinkage and selection operator (LASSO), ridge regression, neural networks like multilayer perceptron (MLP), principle component analysis, support vector machines (SVMs), linear regression models, hidden markov models, decision trees and multiple correspondence analysis (Tang et al., 2012).

These interpretation algorithms were however developed using different datasets, subtypes, analysis on drug-naive and -experienced patients etc. All these differences have led to the creation of many different interpretation algorithms. Initially, studies suggested that the interpretation algorithms produce different resistance measures even if applied to the same resistance profile. However, after subsequent changes in interpretation rules, literature suggested low discordance between interpretation algorithms.

Jaideep and others (2003) studied four interpretation algorithms (ANRS-3-02, TRUGENE VGI-6, Rega 5.5 and HIVdb-8-02) and concluded that there was a discrepancy in interpretations in 33% of all resistance profiles tested. The most discordant were NRTI's. De Luca et al. (2004) concluded that discrepancies in the interpretation algorithms may influence the use of resistance testing over virological outcomes. De Luca et al. (2004) studied the application of 13 interpretation algorithms of drug

naive patients and concluded that there are discordances. Wang et al. (2009) also determined that there is a high level of discordance between the interpretation of NRTI resistance, and goes on to suggest that there should be a "standardization of unique interpretative rules". Vergne et al. (2006) also confirmed some discrepancies and attributed it to the application of the interpretation algorithms to drug-naive or experienced patients. Snoeck et al. (2006) confirmed that there are low discordances between the algorithms tested and suggested it may be due to subtypes. Vercauteren and Vandamme (2006) indicated in their study that it seems that the newer versions of interpretation algorithms are converging and provides the same interpretation. Poonpiriya et al. (2008) also indicated that there are little discrepancies in the seven interpretation algorithms they studied. Yebra et al. (2010) concluded that there is little discordance in interpretation of subtype B sequences, but there are variations in non-B subtype interpretations.

Some of the studies previously mentioned have limitations which may result in their findings not necessarily be valid today (Singh and Mars, 2012). Interpretation algorithms change as the opinion of the expert managing the interpretation algorithm changes. Therefore, the conclusions of previous studies may not be valid when using the newest versions of interpretation algorithm rules. The previous studies also mainly used accuracy to compare the interpretation algorithm, which may be limited in describing actual differences. Furthermore, these studies did not compare the algorithms to the phenotype gold standard.

Since the genome mutation interpretation rules of REGA, ANRS, and HIV-db have been continuously changing, the aim of this study was to determine the variability among the three interpretation algorithms and to statistically compare them with the phenotype gold standard.

METHODOLOGY

2926 protease inhibitor (PR) and 1981 reverse transcriptase (RT) subtype B sequences where obtained from the Stanford HIV-db genotype-phenotype correlation database. Each couple of data consisted of a sequence identifier, the HIV subtype of sequence, the phenotype method employed, a unique number identifying the isolate, the fold resistance of different drugs as compared to the wild type, the amino acid sequence, and summary of the mutations in the amino acid sequence.

The amino acid sequence comprised of 99 amino acids for the PR sequences and 199 amino acids for the RT sequences. Each amino acid in the sequences obtained from the HIV-db database were represented such that a '-' indicates consensus, '.' indicates no sequence; '#' indicates an insertion; '~' indicates a deletion; '*' indicates a stop codon and a letter indicates one letter amino acid substitution. The wild type HIV consensus B sequence was obtained and the amino acid mutations in each sequence were incorporated into the wild type sequence, forming a new representation of the mutated sequence. The amino acid sequence was then converted into nucleotide sequences.

These sequences where then fed into the HIValg program. This publically available web portal tool produces the outputs of the latest versions of the HIV-db, REGA and ANRS algorithms (http://sierra2.stanford.edu/sierra/servlet/JSierra?action=hivalgs). The outputs of the algorithms consisted of a single output for each algorithm stating if a particular sequence is susceptible, intermediate or resistant to different ARV drugs. The phenotype which was used as the gold standard, associated with each sequence was then converted into a susceptible, intermediate and resistant measure using appropriate cutoffs for each phenotypic method (http://hivdb.stanford.edu)

The accuracy, percent of sequences that were supposed to be classified as susceptible but were misclassified (called S-error), percent of sequences that were supposed to be classified as intermediate but were misclassified (I-error), percent of sequences that were supposed to be classified as resistant but was misclassified (R-error), percent of sequences that were supposed to be classified as susceptible but were misclassified as intermediate (and vice versa) or sequences that were supposed to be classified as intimidate but were misclassified and resistant (or vice versa, called OneDiff), and percent of sequences that were supposed to be classified as susceptible but was misclassified as resistant (or vice versa, called TwoDiff) were calculated. Sensitivity, specificity, positive predictive value and negative predictive value were also calculated for each of the susceptible, intermediate and resistant measures.

Statistical methods

Proportional Chi-squared tests for independence were used to determine if any difference exists in accuracy, S-error, I-error, R-error, OneDiff and TwoDiff between the REGA, ANRS, and HIV-db algorithms. The Chi-square test for independence examines whether two or more populations have the same proportion of observations with a common characteristic, that is, it test the null hypothesis that there is no association between one or more populations by determining the observed and expected values of the aspect been tested in the populations. The Chi-squared statistic may be calculated as shown in Equation 1. If the calculated Chi-squared statistic value is equal to or greater than the critical value associated with a particular degree of freedom, the null hypothesis is rejected.

$$X^2 = \Sigma \frac{(\text{Observed frequency} - \text{Expected frequency})^2}{\text{Expected frequency}} \quad (1)$$

Random block design was used to determine if there are difference between S-error, I-error and R-error for each of the REGA, ANRS and HIV-db algorithms. The same statistical test was used to determine the difference between OneDiff and TwoDiff for each of the REGA, ANRS and HIV-db algorithms. Random block design is a statistical theoretical framework that is used to analyze variance. It is similar to a two factor fixed-fixed design, but is applied to datasets where there is only a single value for each factor.

RESULTS

In order to determine if there is a difference in the results produced by ANRS, REGA and HIV-db, Chi-squared tests were performed. Three individual chi-squared tests were performed to determine if there is a difference between ANRS and HIV-db, ANRS and REGA, and HIV-db and REGA. The results of these tests for both protease and reverse transcriptase inhibitors are shown

Table 1. Chi-squared statistic calculated on the raw data between the interpretation algorithms.

Sequence data	Interaction	Chi-squared statistic
Protease inhibitor	ANRS- HIV-db	8110*
	ANRS-REGA	7832*
	HIV-db-REGA	1250559*
Reverse transcriptase	ANRS- HIV-db	8532*
	ANRS-REGA	8515*
	HIV-db-REGA	15169*

*p < 0.0001.

Table 2. Average percentage accuracy, S-error, I-error, R-error, OneDiff and TwoDiff.

Algorithm	Accuracy	S-error	I-error	R-error	OneDiff	TwoDiff
ANRS	59	41	24	51	93	7
HIV-db	59	41	39	34	87	13
REGA	61	39	29	43	82	18

Table 3. Chi squared statistic for comparing average percentage accuracy, S-error, I-error, R-error, OneDiff and TwoDiff between interpretation algorithms.

Test	Accuracy	S-error	I-error	R-error	OneDiff	TwoDiff
Chi-squared	27#	573*	635*	1057*	328*	243*

*p <0.0001, #p = 0.23.

Table 4. Random block design F-score to determine difference between S-error, I-error and R-error.

Algorithm	F
ANRS	28.4*
HIV-db	27.6*
REGA	29.6*

*p <0.0001.

in Table 1.

The accuracies of the different algorithms as well as the errors associated with predicting susceptible, intermediate and resistant measures are shown in Table 2. In order to determine if these differences are statically significant, Chi-squared tests were performed and the results are shown in Table 3.

Table 4 shows the results of a RBD test to determine if there is a difference between the errors in predicting susceptible, intermediate and resistant measures. Table 5 shows the sensitivity, specificity, positive predictive value and negative predictive values for each of the susceptible, intermediate and resistant measures.

DISCUSSION

The average accuracy for ANRS was 59%, HIV-db 59% and REGA 61%. An associated Chi-squared statistic of 26.5 suggests that there is no difference between the three interpretation algorithms in terms of the accuracy obtained. This is also confirmed by p-score of 0.23. This result confirms previous studies findings (Poonpiriya et al., 2008; Snoeck et al., 2006; Vercauteren and Vandamme, 2006; Yebra et al., 2010) that there is very little difference when comparing the accuracies of ANRS, REGA and HIV-db. The difference in the accuracies between the three algorithms has decreased as compared to previous studies. Liu et al. (2008) reported an average discrepancy of 3%, as compared to an average of 0.75% reported in this study. Thus in terms of accuracy, the algorithms seem to be converging and the understanding of HIV resistance increases. However, accuracy should not be the only method used to determine if there is a discrepancy between the three algorithms.

Chi squared tests on the raw output of each interpretation algorithm were performed in order to determine if a difference exists between ANRS and HIV-db, ANRS and REGA, and HIV-db and REGA. As shown in Table 1, these tests were performed separately for the

Table 5. Random block design F-score to determine difference between sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV).

Parameter	Sensitivity	Specificity	PPV	NPV
Susceptible				
ANRS	80.1	67.8	84.4	35.2
REGA	84.1	73.3	82.0	57.0
HIVDB	92.1	64.4	68.0	82.6
RBD -F test	<0.0001	<0.0001	<0.0001	<0.0001
Intermediate				
ANRS	41.6	88.8	49.5	74.7
REGA	49.2	61.4	52.7	79.5
HIVDB	56.7	48.1	46.2	78.2
RBD -F test	<0.0001	<0.0001	<0.0001	<0.0001
Resistant				
ANRS	89.5	53.5	89.5	92.7
REGA	67.7	73.5	67.7	82.7
HIVDB	58.8	75.7	58.8	90.1
RBD -F test	<0.0001	<0.0001	<0.0001	<0.0001

protease inhibitor and reverse transcriptase sequences. All six Chi squared tests produced a $p < 0.001$. This indicates that for both the protease inhibitor and reverse transcriptase there is a difference between ANRS and HIV-db, ANRS and REGA, and HIV-db and REGA. This difference is in contrast with previous studies where they reported differences only in some reverse transcriptase sequences or no difference at all (Poonpiriya et al., 2008; Snoeck et al., 2006; Vercauteren and Vandamme, 2006; Yebra et al., 2010).

The Chi-squared statistic shown in Table 3 for the associated S-error, I-error and R-error all indicate that there are differences in the three interpretation algorithms ($p < 0.0001$). REGA and ANRS are more accurate in predicting a susceptibility resistance measure than HIV-db. The three interpretation algorithms all perform differently in terms of I-error. HIV-db more accurately predicts intermediate resistance than REGA, which is more accurate than ANRS. HIV-db and ANRS have similar accuracies in terms of predicting resistance measures and are both more accurate than REGA.

Similar results were obtained for OneDiff and TwoDiff. The RBD analysis as shown in Table 4 indicates that for each interpretation algorithm, there is a difference between S-error, I-error and R-error. Post-hoc Kunkey and Bonfferoni statistics indicate that each interpretation algorithm predicts S-error-I-error, S-error-R-error and S-error-R-error differently.

Table 5 shows that there is a difference between the sensitivity, specificity, positive predictive value and negative predictive values of the three algorithms for each of the susceptible, intermediate and resistant measures. This indicates that there is a difference in the ability of ANRS, REGA and HIV-db to predict resistance measures.

These results indicate that although the accuracy of REGA, ANRS and HIV-db are similar, the interpretation algorithms are in fact different. This difference in interpretations may lead to inconsistent treatment for patients failing ARV therapy. There thus needs to be a mechanism of providing a single interpretation of a genome, formed by collating the strengths of each of the interpretation algorithms. The gold standard algorithms may be collated by weighted output or applying machine learning on gold standard outputs.

One limitation of the study is that it uses a limited dataset to perform the learning and testing. It will be of more value if the algorithms could have been tested on real-time data of patients currently being treated. This study however, has the benefit of describing the differences between the latest versions of REGA, ANRS and HIV-db in more depth than what has been done to date. Other papers only reported differences in accuracy. This paper goes on to discuss whether these differences are statistically significant and also discusses the distribution of the different types of errors.

Conflict of Interest

The author(s) have not declared any conflict of interests.

ACKNOWLEDGEMENT

This work was supported by National Institutes of Health

Fogarty International Centre [grant number 5D43TW007004-11].

REFERENCES

Bartlett JC, Gallant JE, Pham PA, Pharm D, Hadden DD, Peirce GA, Hadden C (2004). Medical Management of HIV infection: John Hopkins University School of Medicine.

Cunningham A, Donaghy H, Harman A, Kim M, Turville S (2010). Manipulation of dendritic cell function by viruses. Curr. Opin. Microbiol. 13(4):524-529.

De Luca A, Cozzi-Lepri A, Perno CF (2004). Variability in the interpretation of transmitted genotypic HIV-1 drug resistance and prediction of virological outcomes of the initial HAART by distinct systems. Antiretroviral Therapy 9(5):743-752.

Gilbert PB, McKeague IW, Eisen G, Mullins C, Guéye-Ndiaye A, Mboup S, Kanki PJ (2003). Comparison of HIV-1 and HIV-2 infectivity from a prospective cohort study in Senegal. Stat. Med. 22(4):573-593.

Jaideep R, Bradley JB, Francoise BV, Anne-Mieke V, Diane D, Kristel VL, et al. (2003). Rapid communication: HIV-1 protease and reverse transcriptase mutation patterns responsible for discordances between genotypic drug resistance interpretation algorithms. J. Acquir. Immune Defic. Syndr. 33:8-14.

Liu L, May S, Richman DD, Hecht FM, Markowitz M, Daar ES, Smith DM (2008). Comparison of algorithms that interpret genotypic HIV-1 drug resistance to determine the prevalence of transmitted drug resistance. AIDS 22(7):835-839.

McGovern SL, Caselli E, Grigorieff N, Shoichet BK (2002). A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. J. Med. Chem. 45 (8):1712-1722.

Migueles S, Connors M (2010). Long-term Nonprogressive Disease Among Untreated HIV-Infected Individuals: Clinical Implications of Understanding Immune Control of HIV. J. Am. Med. Assoc. 304(2):194-201.

Mitton J (2000). The sociological spread of HIV/AIDS in South Africa. J. Assoc. Nurses AIDS Care 11:17-26.

Pierret J (2007). An analysis over time of the experiences of living with HIV Soc. Sci. Med. 65:1595-1605.

Poonpiriya V, Sungkanuparph S, Leechanachai P (2008). A study of seven rule based algorithms for the interpretation of HIV-1 genotypic resistance data in Thailand. J. Virol. Methods 151:78-86.

Singh Y, Mars M (2012). Predicting a single HIV drug resistance measure from three international interpretation gold standards. Asian Pacific Journal of Tropical Medicine, 5(7):566-572.

Snoeck J, Kantor R, Shafer RW, et al. (2006). Discordances between interpretation algorithms for genotypic resistance to protease and reverse transcriptase inhibitors of human immunodeÞciency virus are subtype dependent. Antimicrob. Agents Chemother. 50(2):694-701.

Tang M, Shafer R (2012). HIV-1 antiretroviral resistance: scientific principles and clinical applications. Drugs 72(9):e1-25.

Tang M, Liu T, Shafer R (2012). The HIVdb system for HIV-1 genotypic resistance interpretation. Intervirology 55(2):98-101.

Vercauteren J, Vandamme A (2006). Algorithms for the interpretation of HIV-1 genotypic drug resistance information. Antivir. Res. 71:335-342.

Wang D, Larder B, Revell A, et al. (2009). A comparison of three computational modeling methods for the prediction of virological response to combination HIV therapy. Artif. Intell. Med. 47:63-74.

Yebra G, de Mulder M, de Romero J, Rodriguez C, Holguin A (2010). HIV-1 non-B subtypes: High transmitted NNRTI-resistance in Spain and impaired genotypic resistance interpretation due to variability. Antivir. Res. 85:409-417.